Intelligent Ship Phase 2
Summary Information to Support Phase 3 Proposals

DSTL/PUB148547

Cooke, C; Tate, A

DSTL/CR
Version 1.0

04/04/2023

Dstl
Platform Systems
Portsdown West
Portsdown Hill Road
Fareham
PO17 6AD

Release Conditions

Intellectual Property Department
Defence Science and Technology Laboratory
Porton Down, Salisbury, Wiltshire SP4 0JQ

Authorisation		
Role	**Name**	**Date**
Project Manager	Asha Sohail	04/04/2023
Technical Reviewer	Colin Cooke	04/04/2023
Author	Andrew Tate	04/04/2023

1 Introduction

1.1 Document Aim

This document summarises the second phase of the Intelligent Ship (Intelligent Ship) project funded by the Dstl Autonomy Programme with CSA research funding.

It aims to provide background information to potential suppliers for Phase 3 of the project

1.2 Project context & aims

Future military forces and platforms will operate within, and against, an increasingly complex, diverse and technology focused set of threats. This will increase the volumes and rate of delivery of the data and information that commanders and their supporting systems need to capture, process and respond to. This inevitably leads to a requirement for greater levels of automation and the wider use of machine intelligence. To fully optimise and exploit these approaches, effective alternative teaming relationships will be required between intelligent machine agents, automation, and human operators – a Human Autonomy Team (HAT).

The Intelligent Ship project was initiated with an aim of starting to understand, develop and evaluate concepts that address the challenge of enabling, integrating and managing complex HATs. The project was challenged to consider a clean sheet approach to avoid the constraints of current approaches and architectures.

Phase 2 of the Intelligent Ship project started to explore these future human-autonomy, and autonomy-autonomy relationships, the supporting architectures and its enablers. This was achieved through the development of component machine intelligent agents, human-machine interfaces, and through the development, evaluation and demonstration of those agents within a systems level architectural ‘sandpit’, called the “Intelligent Ship Artificial Intelligence Network” (ISAIN).

This research used a naval platform as a use case due to the decision making complexity and multiple systems and roles operating within a single platform. These are currently managed with large human command teams operating across 2-3 command spaces (operations room, bridge and machinery control spaces). The research, while focused on Naval C2, has clear and transferable applicability and relevance to any system or platform where humans and AI agents will work together.

1.3 Notes on language

The project quickly identified that the linkages between individual machine agents and operators were complex and highly dependent on the decisions to be made or influenced. The term Agents for Decision Making (ADeM) was used by the project in Phase 2 to define any group or combination of agents, or agents and operators that were required to make a collective decision in any given capability area.

The project also used the term Human Machine Teaming (HMT) in the first 2 phases of activity – this has now been replaced with the more universal term Human-Autonomy Teaming, reflecting the level of intelligence and complexity within the machine elements of a collaborative AI system.

1.4 Phase 1

Phase 1 was focused on developing ‘enablers’ for the project.

It was completed in November 2020 and included:

- The initial design and demonstration of the Intelligent Ship AI network (ISAIN), the project ‘sandpit’. (Delivered by CGI).
- A study looking at the Platform Design risks and Opportunities (PeDRO) of integrating AI and its enabling technologies into a naval platform. This looked at practicalities and wider DLOD impacts of integrating advanced machine intelligence across a range of naval Capabilities. (Delivered by BMT)
- The development of an underpinning Tactical navigation agent (TACNAV) with an aim of providing both Safety of Life at Sea (SOLAS) compliant navigation capability and novel capabilities able to manage more complex naval specific navigation functions such as the deployment and recovery of off board systems. (Delivered by CGI)
- A range of intelligent machine agents that provided a range of naval capabilities and provided enabling capability to a HAT. These were delivered via a Defence and Security Accelerator (DASA) themed competition and a range of different suppliers.

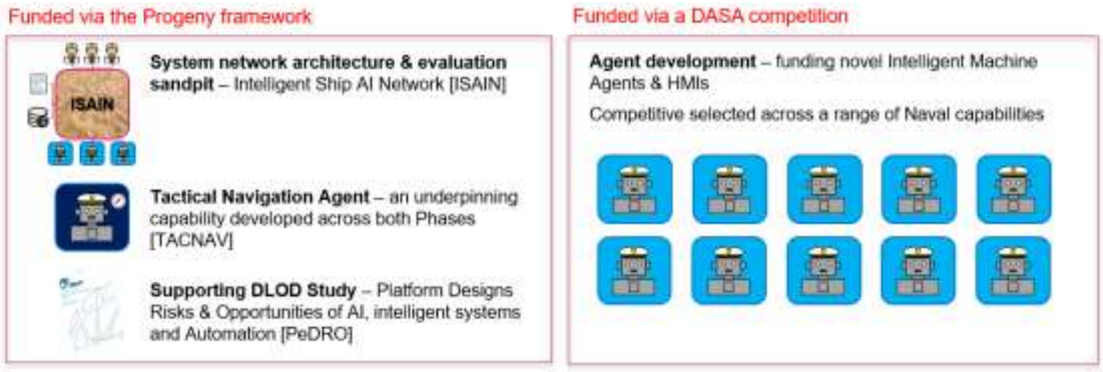


Figure 1 Summary of Intelligent Ship project’s Phase 1 components & tasks

1.5 Phase 2

Phase 2 focused on ‘Integration and evaluation’ of the Intelligent Ship HAT system.

It was delivered by another DASA competition, which was split into two challenges.

The successful challenge 1 supplier (CGI & DIEM) matured and further developed ISAIN. They then integrated the Machine Agents selected under challenge 2 of the call into ISAIN using the Dstl Command Lab facility at Portsdown West. Finally they developed approaches to, managed and delivered four evaluation events. Each consecutive event built up net capability, the number of agents integrated and the complexity of interconnectivity between them, as well as the number of operators involved. Operator usability was assessed against a developed representative naval scenario and component vignettes, using Dstl’s internal Military Advisors (MAs) as operators.

Under challenge 2, ten agents were developed or matured further. Again this represented a broad (but typically only 1 deep) mix of naval capability areas. Some were further developments of Phase 1 agents, but several new concepts were also developed. SYCOIEA, a Threat Evaluation and Weapon Assignment (TEWA) toolset developed by Dstl’s Above Water Systems programme was also included in the overall system to provide TEWA capability.

A summary of the agents developed and integrated into ISAIN is shown in Figure 2.

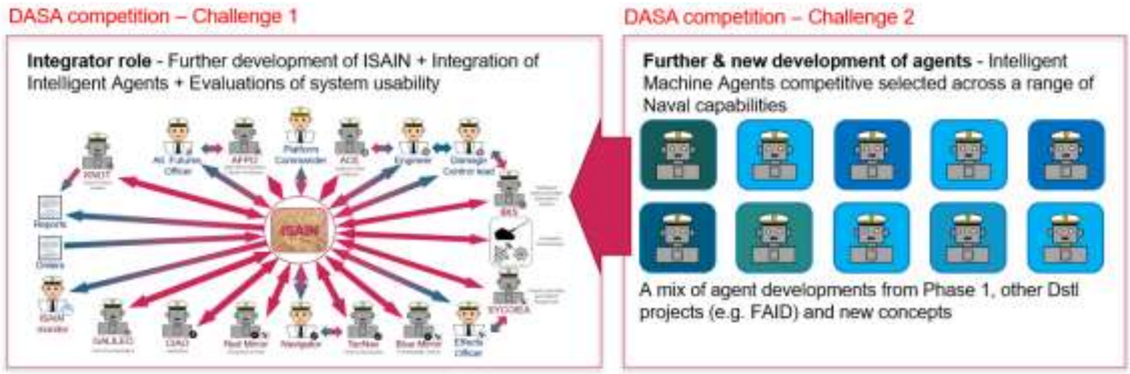


Figure 2 Summary of Intelligent Ship project’s Phase 2 components

1.6 Phase 2 technical approach

The technical approach and contributing activities within Intelligent Ship Phase 2 are described below. An agile approach was used so that lessons learnt from each evaluation event could be utilised in the follow-on events. This included modifying the agents used, connectivity levels achieved, the information provided to operators and the scope and details within the evaluation vignettes used.

The suppliers in Phase 2 undertook the following activities:

1. Challenge 1 suppliers:
    - Further developed ISAIN as required;
    - Developed with Dstl the required scenarios and vignettes to support the evaluation events;
    - Developed an ISAIN Software Development Kit (SDK) to support off-site testing by Challenge 2 suppliers;
    - Planned the evaluation events working with challenge 2 suppliers through workshops and other interactive events to understand:
        - interface, hardware and software needs;
        - risks and opportunities for interactions between agents;
        - data, information and interconnectivity needs;
        - the maturity of each agent and hence plan what level of capability each agent will deliver.
    - Undertook 4 evaluation events, including:
        - agent integration and functionally testing at Command Lab;
        - training the MAs in their use;
        - whole system testing and familiarisation with the MAs;
        - undertake 2-3 days of evaluation against the a range of vignettes, including video and action logs to support analysis;
        - undertake post evaluation analysis.
2. Challenge 2 suppliers:
    - Developed their individual agents;
    - Supported a range of whole team integration events to identify data needs, gaps and opportunities for interactions with other suppliers;
    - Supported each evaluation event, providing on-site training to MA’s, support to integration of their agent and background support to evaluations (e.g. to inject faults or events).

1.7 Evaluation events & aims

A total of four events were conducted; 3 interim evaluations and a final evaluation and stakeholder demonstration.

An incremental approach was taken to the scope and focus of each of the events in order to iteratively de-risk and generate insight for the planning and running of the final demonstration and evaluation.

The aims of these events were to:

- Demonstrate and evaluate a number of different AI applications, working together with humans in a collaborative environment to deliver military effects.
- Evaluate the usability of the ADeM applications in different Human Autonomy Teaming (HAT) configurations to inform the design of the system (ISAIN and the ADeM applications) in future phases of the Intelligent Ship project.

1.8 Evaluation scenario development

The originally proposed scenario was developed by the Dstl Intelligent Ship team in conjunction with Dstl's MAs. The aim was to provide sufficient detail and diversity in sub-tasks (at an Official-level) to support the evaluation of the Intelligent Ship concept, the machine agents and ADeM applications and their usability.

It was not derived from any MOD accredited standard scenarios (to avoid classification issues), rather it focused on ensuring that all the selected agents and their respective interconnections are exercised whilst still being representative of a naval task.

The scenario assumed a time of tension/conflict around and over a waterway/choke point that is critical to maritime trade resulting in some traffic being harassed, boarded and turned away. For the purposes of the evaluations and due to ease of access to good data and models, the English Channel was used as this waterway. It assumed Red forces have committed a number of assets (sea and air) to establish and maintain a selective blockade on this vital shipping route, with land masses either side of the waterway sitting within Red territory. The UK carrier task group (CTG) was tasked to nullify the blockade, maintain freedom of navigation, DETER aggression against non-combatant shipping and DEFEND task group and non-combatant shipping. Prior to CTG transit into the disputed waters, the Intelligent Ship was deployed to act as a forward located asset tasked with information gathering.

The evaluation scenario was split into a number of phases that both mimicked a typical naval mission, but also allowed different aspects of the Intelligent Ship system to be tested in discrete vignettes. It was found after evaluation 2, that attempting to run a whole phase was experimentally inefficient. Hence, for the last 2 evaluation events, shorter (~30 min) vignettes were created under each phase allowing easier recovery from faults, more flexibility and the opportunity to re-run vignettes if required. The scenario phases and tasks are summarised below in Figure 3.

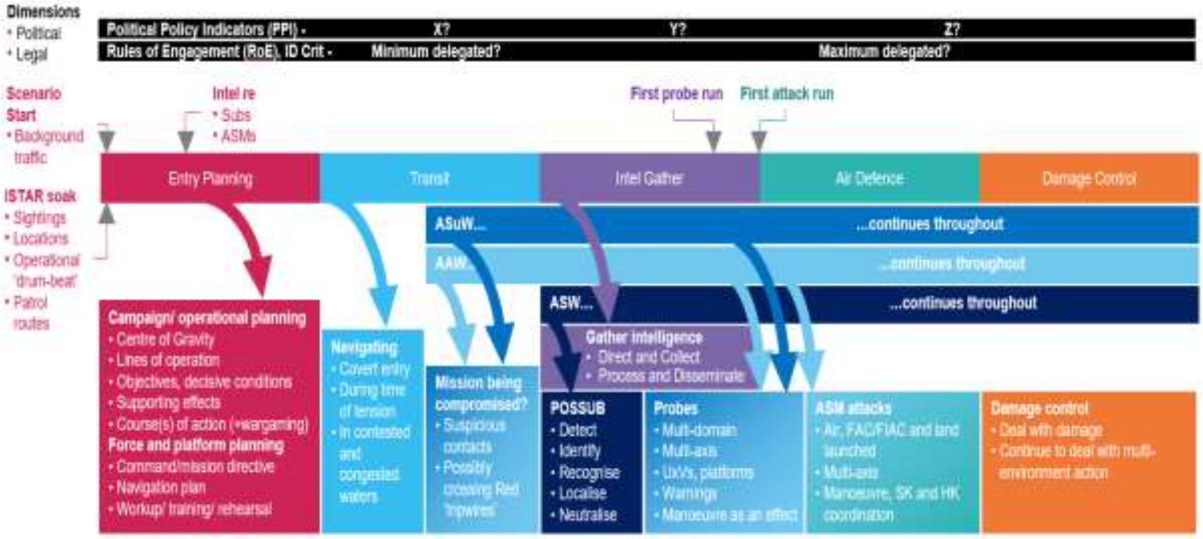


Figure 3 Evaluation scenario phases and tasks

2 Intelligent Ship System & components

This sections provides a high-level descriptions of each of the component of the final evaluated and demonstrated Intelligent Ship system.

2.1 ISAIN

The Intelligent Ship Artificial Intelligence Network (ISAIN) provided the Intelligent Ship project with a framework to support experimentation with Artificial Intelligence (AI) collaboration and HMT. This acts as a ‘playground’ or ‘sandpit’ for AIs’ to support inter-relationships between applications and human users. Developed by CGI in Phases 1 and 2, it is owned by MoD and utilises open systems approach and standard interfaces. It uses open commercially available or open source software packages such as Apache NIFI, MongoDB, Docker, Elastic Stack and Kibana derived diagnostics.

Additional functionality developed during Phase 2 included:

- Automatic configuration of the ISAIN data-flows between Agents/ADeMs through the use of a separation of concerns ISAIN Dynamic Dataflow Configuration (IDDC) extension;
- Improved fault diagnosis and post experiment analysis (Elastic Stack/Kibana dashboards);
- Providing a Data Query API that allows ADeM applications to treat ISAIN as a black box for querying data, publishing data or requesting a task be performed without requiring knowledge of the destination ADeM or data repository.

The resulting top level architecture for ISAIN developed during IS Phase 2 is shown in Figure 4:

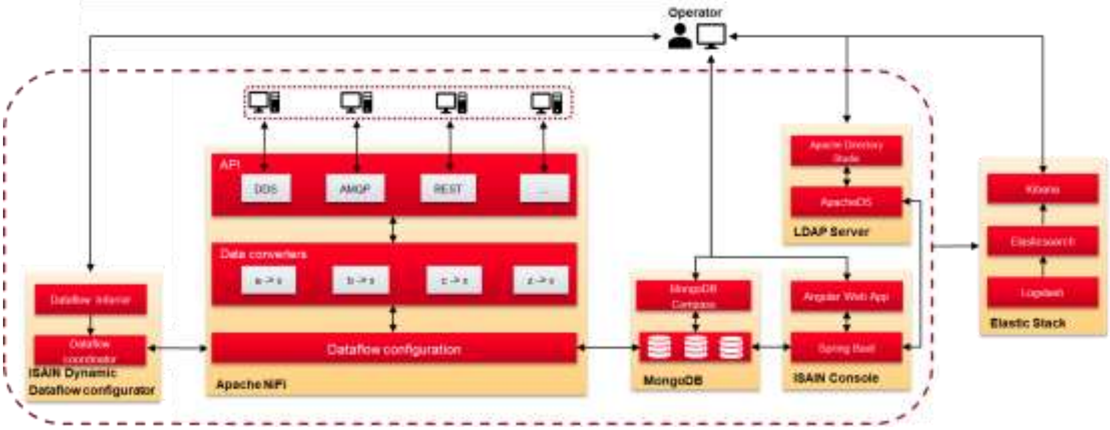


Figure 4 ISAIN Top Level Architecture

ISAIN uses a configurable technology, namely Apache NiFi[1] as the underlying framework to integrate standalone applications for the purpose of data routing, mediation and transformation. Apache NiFi is a domain agnostic technology that can support a range Application Programming Interfaces (APIs) and hence allow a range

---

[1] https://nifi.apache.org/

of applications developed in different languages and operating systems to interact within a common data system.

Each of the agents supplied to the project were delivered in a Docker[2] container to aid flexibility and integration. The Integration of a new AI or Agent was achieved through configuration of a Nifi flow that includes both input and output processors to connect it to the new system, and any additional appropriate data converters to connect to and from ISAIN's internal data model. In addition to the out-of-the-box Nifi input and output processors, DDS processors, built using Vortex OpenSplice Community edition, are provided.

The integration configuration of ISAIN and the ADeMs included for integration within the final evaluation event are shown in Figure 5.

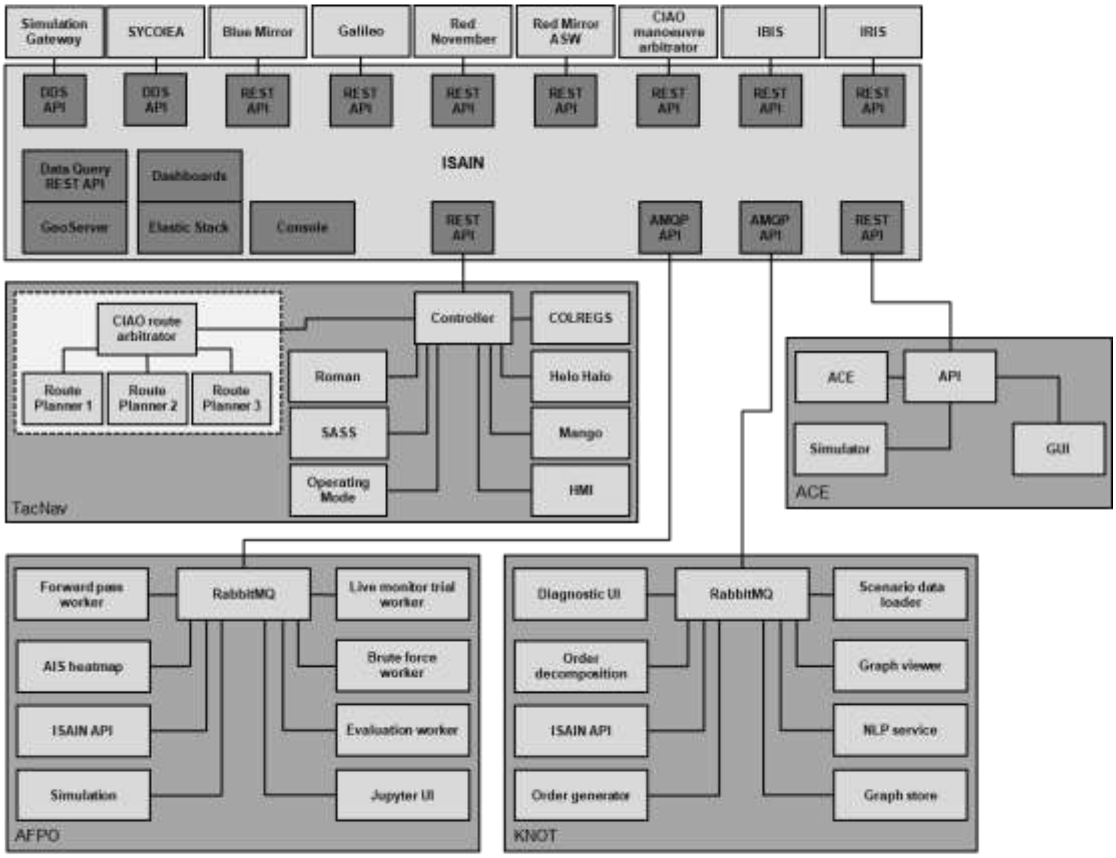


Figure 5 Final evaluation event – Integration configuration

[Note: REST or RESTFull, AMQP and DDS are all industry standard APIs]

Custom input and output processors can also be added for protocols not currently supported. This can be achieved though the addition of a Java based implementation of the AI/ Agent's data type and implementing a single Java interface to perform the conversion.

---

[2] Docker is a set of platform service products that use Operating System-level virtualization to deliver software in packages called containers.

2.2 Operators

A small group of Dstl Military Advisors (MAs) operated the ADeM applications in specific roles; i.e. commanding officer (CO), air-warfare officer (AWO), principal warfare officer (PWO), officer of the watch (OOTW) and marine engineering officer (MEO).

These roles largely reflected current service experience and roles, however there was very limited or no familiarity with the individual agents and their interfaces until pre evaluation event training. Scheduling pressures inevitably mean that the level of training and experience of the agents will be very limited when compared to experience with current fielded systems in the fleet.

2.3 Command Laboratory infrastructure & components

Dstl has developed a Command Lab facility at the Portsdown West site. It is seen as a key enabler in the modelling and simulation domain to exploit emerging technologies, experiment with new Operations Room layouts and conduct operator performance experiments in above water operations. It enables repeatable assessment of low Technology Readiness Level (TRL) concepts using Human in The Loop (HITL) experiments and provides a degree of validation prior to hardware integration evaluation using actual Combat System equipment and sea trials.

It reached initial operating condition early on in Phase 2 of the Intelligent Ship project and continued to mature in both stability and capability throughout this phase of work.

Note: The Information below reflects the capability as used in Phase 2; however the capability is subject to continuous development so may not reflect the available state at the beginning of Phase 3.

Command Lab is a configurable software and physical infrastructure. It consists of a synthetic environment including the required simulation components, Combat Management System (CMS) Emulator, and operator Graphical User Interface (GUI) outputs to provide an Initial Operating Capability (IOC). There is the ability to screen capture each operator position and it can provide timestamped audio recordings using a synchronised time source. Up to 12 Operator positions can be provided in different layout configurations with representative operator displays, including simulated Type 23 and Type 26 displays are available.

Figure 6 CLab image – MAs operating Intelligent Ship system during the 3rd Evaluation event

The Intelligent Ship project used a series of Virtual Machines within the high-powered servers to host ISAIN and the component agents. It also used the Simulation Gateway (under development) to access and use the VBS simulation engine to provide a visual simulation of a realistic environment and of other users and threats within it.

The Simulation Gateway also allows the use of other relevant toolsets to generate/ simulate, for example, the air picture in the test scenarios, or to support links to other data areas such as those sitting within Navy Digital.

The Command Lab space includes a range of larger screens and a track table to facilitate both C2 emulation and planning, but also to support visibility of actions occurring within the system and evaluation, as well as to provide a visualisation to operators and observers.

The project used the Clamshell toolset developed by Southampton University for the Command Team-working Experimental Test-bed (ComTET) project. This enabled the screen capture of events during the scenario run-throughs at evaluation events 3 & 4. Clamshell is a tool designed to support human factor experimentation. It is, for example, capable of supporting multiple human factors metrics that can appear on screen for the operators to fill in in sync with the experiment being undertaken. For the Intelligent Ship evaluations this capability allowed the team to start and stop screen capture to allow easier syncing and analysis after the evaluations. It is also capable of recording other video inputs such as webcam and capturing key strokes.

2.4 Component Agents

A high-level description of all the machine agents integrated into the Intelligent Ship system is provided below:

Agent/ ADeM Name & Description	Supplier
KNOT – Knowledge-based Naval Orders Toolset	Montvieux
KNOT decomposes human-written formal naval orders, in the NATO APP-11 format, into a knowledge graph that other AI agents can query. KNOT also works in the reverse direction generating new orders or status updates based on changes to its knowledge graph by other agents. KNOT uses Natural Language Processing (NLP) techniques to decompose the free text fields. Naval orders is not structured in the same way as the language, so standard NLP models do not work well, hence a bespoke model was developed and trained.	
GALILEO – GoAL based decomposition for InteLligent ship AI nEtwOrk	Seebyte
Breaks down mission aims into tasks and information and then allocates tasks to different ADeMs.	
AFPO – Alternative Futures Planning Officer	Montvieux
Alternative Futures Planning Officer (AFPO) toolkit supports mission planning through the testing of automatically & manually generated courses of action (CoA). AFPO can monitor unfolding live events against the mission intent (constraints), to determine if the CoA is in line with mission objectives (desired effect). If this is not the case, AFPO will raise an alert and auto-generate a new CoA should the live CoA be moving away from the required outcome. At the heart of AFPO is a neural net that has been trained with the spatial and temporal interactions/behaviours (e.g. patterns of life) required to predict the future locations of objects of interest.	
ACE – Artificial Chief Engineer	Rolls-Royce
ACE was developed to manage and optimise the use of a platform's power, propulsion and wider marine systems in a fully autonomous vessel. For Intelligent Ship ACE was adapted to provide decision support and matching interfaces for operators and to include various modes optimised to current command aims – e.g. optimise to minimise noise/emissions, or operate at maximum economy or availability.	
TACNAV – Tactical Navigation	CGI
Supporting safe navigation against international Collison Avoidance Regulations (COLREG); It also includes various optimisation agents that allow the toolset to manage more complex naval specific navigational requirements. Such as those required: to launch and recover off-board assets; to support environmental signature reduction or optimise firing arcs.	
RED MIRROR	DIEM Analytics
RED MIRROR (RM) seeks to build a rapid 'mirror' of Red's decision-making process, whether human or AI driven, to predict what Red will do next. Red Mirror has been applied to Anti-Submarine Warfare (ASW), Anti-Air Warfare (AAW) and Anti-Surface Warfare. The RM ASW application can highlight the likely area a submarine is operating in under certain conditions, based on the submarine's maximum speed and observed	

Agent/ ADeM Name & Description	Supplier
speed. The assumption is made that the submarine has detected the Intelligent Ship and is following one of four behaviours - transit, shadow, attack or retire.	
BLUE MIRROR	DIEM Analytics
Reverses RM's capabilities with an aim to assess the predictability of the Intelligent Ship system itself (BLUE) from RED, within the Intelligent Ship system.	
CIAO – Compounded Intelligent Agents for Optimisation	Decision Lab
Provides recommended advice by managing and arbitrating between conflicting advice from different agents or ADeMs.	
HADES & IRIS – Human-Agent Design & Evaluation	Fraser Nash/ Decision Lab
Developed a holistic design approach and matching tools to design HMTs with the human an integral part of the design. The approach was used to generate the IRIS display that was demonstrated by acting as a UI for IBIS and for the CIAO arbitration between SYCOIEA and TacNav AIs. The UI provides the operator with several windows displaying mission and system information.	
IBIS – Intelligent Battle Information System	Fraser Nash
A decision support system that uses reinforcement learning approaches to optimise recovery actions prioritisation against current command aims, and to estimate recovery times. This aims to support the currently human-centric and experience based Damage Control & Fire-Fighting (FF&DC). It supports both real time tasking but also what-if assessment of the damage impacts of threats. This built on Navy Develop and Dstl Funded activities by considering the use of AI to create a more generalised agent than achieved through existing rules based approaches.	
SYCOIEA – SYstem Coordinating Operational Interaction for Effects Assignment	CGI, HFE & Diem Analytics
SyCOIEA is a Threat Evaluation and Weapon Assignment (TEWA) research demonstrator developed within the Above Water Systems programme. Its objective is "to create and demonstrate a system that will support TEWA operators during times of high stress, high tempo warfare with the expectation that decision making is improved". SyCOIEA uses AI techniques to assess if tracks are threats and then generates recommended courses of action that the human operator can accept or veto depending on the automation mode. SyCOIEA has been developed to support Anti-Aircraft Warfare (AAW), but has been recently upgraded to simultaneously support Anti-Surface Warfare (ASuW) and Anti-Submarine Warfare (ASW).	

Table 1 Phase 2 Agents - high-level descriptions and Links to Appendices

3 Observations, results & conclusions

This section highlights some of the key observations, results and conclusions drawn from both the individual agent developments and the whole system evaluation events in Phase 2. These are discussed against a series of themes.

3.1 Overall system design

The project was primarily focused on proving the overall concept of a system using collaborative AI. The evaluations successfully showed the potential to design a system that could host multiple agents and that would allow them to interact and communicate, both machine to machine but also with its human operators. It also showed the ability to integrate agents within different capability areas, in different programming languages, with different levels of autonomy (and hence operator interaction) and levels of technical maturity.

The project provides further evidence for the need for greater levels of openness in ship-board information systems design, the benefits for early clarity on IP ownership of the overall host system and interfaces, and the benefits of using existing, often open source, tools, APIs and systems as building blocks.

While the overarching aim was met, it is important to note the limits of the current system make-up in considering collaborative AI and HAT design aspects going forward. The component agents were competitively selected with an aim to show a broad range of naval capabilities, and hence were not designed against a specific capability requirement. The result of this approach is:

- There was no way of comparing and contrasting, or arbitrating decisions of a number of agents in most capability areas, with the exception of navigation focused COAs (i.e. agents in other areas were typically 1 deep).
- Intelligent Ship operator's roles effectively mirrored current roles, with the only design decision being around the number of agents each operator should manage and some limited impact on matching UI design. As such no analysis or design work was undertaken to understand any potential changes to tasks, roles and responsibilities of the future crew of the Intelligent Ship and the balance of tasks between the human operators and machines.
- Each agent had varying technical maturity in both its AI and its UI. This resulted in a mix of UIs, styles and messaging approaches with the potential to impact assessments due to operator confusion, unfamiliarity or frustration.

The system design selected is also focused on being optimised to experimentation and control, and on being flexible enough to allow use in other tasks and projects. Earlier ISAIN design work suggested a range of overarching architecture design options and these will need to be re-visited if the concept is taken forward into real in-service applications. Equally security and system (and component agent) verification, validation and safety approval issues have not yet been considered in detail and will need to be addressed going forward.

3.2 Observations & results from the final (4th) Evaluation event

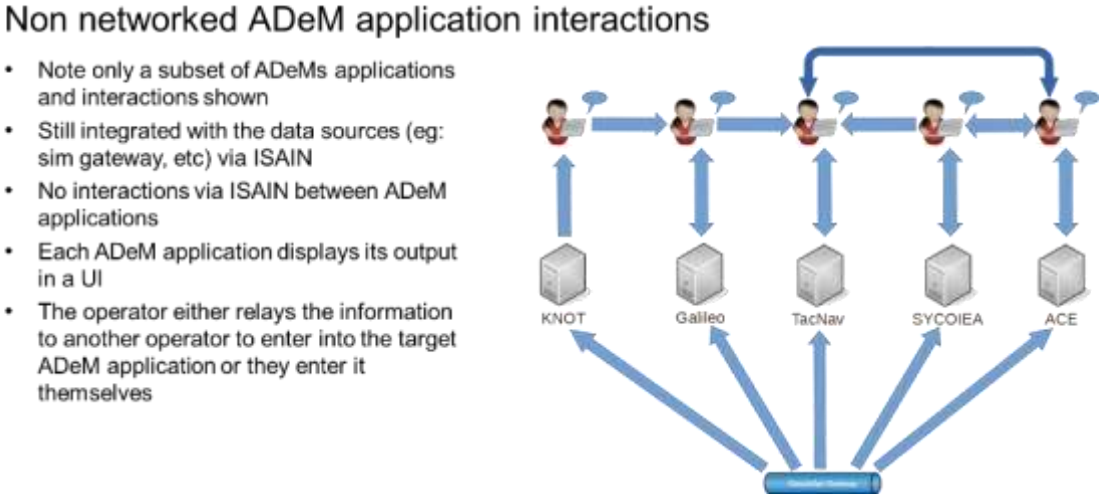
The evaluation events aimed to demonstrate the intelligent ship collaborative AI system concept as whole, but were primarily focused on assessing the usability of the system and some of its components. This was evaluated by running two series of vignettes from the agreed scenario, firstly with the Agents/ ADeMs working as stand-alone decision aides to the operators and then secondly by allowing them to be networked together within ISAIN.

Impacts were assessed through widely used usability metrics, comprising of questionnaires completed by the operators after each vignette, and through post processing of event logs, screen captures and real-time videos of the operators. These metrics covered workload, Situational Awareness (SA) and Trust in Automation (TiA) aspects. (See Annex A)

Five runs were conducted covering two sets of mission events from the scenario with the ADeM applications networked or not networked. As shown in Figure 3 these were based on:

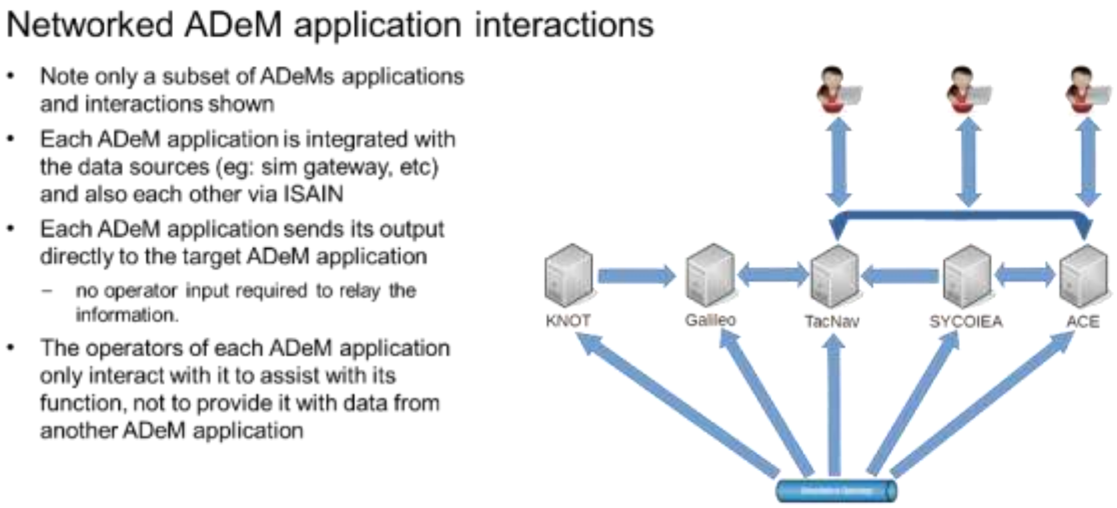
- Entry Planning;
- Transit and Probe;
- Air-Defence and;
- Damage Control.

The non networked and networked arrangements are illustrated below in Figure 7.



Non networked ADeM application interactions

- Note only a subset of ADeMs applications and interactions shown
- Still integrated with the data sources (eg: sim gateway, etc) via ISAIN
- No interactions via ISAIN between ADeM applications
- Each ADeM application displays its output in a UI
- The operator either relays the information to another operator to enter into the target ADeM application or they enter it themselves



Networked ADeM application interactions

- Note only a subset of ADeMs applications and interactions shown
- Each ADeM application is integrated with the data sources (eg: sim gateway, etc) and also each other via ISAIN
- Each ADeM application sends its output directly to the target ADeM application
  - no operator input required to relay the information.
- The operators of each ADeM application only interact with it to assist with its function, not to provide it with data from another ADeM application

Figure 7 Descriptions of Networked and Non-networked Intelligent Ship system operation

Analysis of variance (ANOVA) was conducted to identify the potential effects of networking the AI applications, and of different operator roles using the applications. Metrics that appeared to represent the main identified effects were then analysed further to determine whether they improve or degrade.

While these outputs can only be considered as ‘indicative’, they can be used to suggest future experimental needs and challenges to address. There are clear limits to this analysis due, in part, to the following factors:

- The limited number of operators available, and the limited time available to train them in the various agents and their matching UIs;
- The operators did not swop roles – i.e. a risk of bias due to existing experience and roles;
- Age & experience profile – MAs are all experienced officers and have good understanding of Science and Technology developments and approaches. They may not represent, for example, the experiences of perhaps more digitally literate, younger in-service operators.

3.2.1 Effect of networking Agents to work together

The analysis indicated that networking agents/ ADeMs has a potentially significant effect in reducing the temporal demand component of workload i.e. how hurried or rushed operators viewed the pace of the task. The analysis indicated that networking applications may also have a potentially significant effect on reducing the mental demand component of workload.

In the entry planning, transit and probe events, the analysis indicated that networking the AI applications has a potentially significant negative effect on the operators’ SA. It is not known whether this was due to the loss of SA that comes from delegating the process of working-up a plan (i.e. delegated to the Agent/ ADeM), the learning effect from running the networked cases first followed by the non-networked cases, or the specific details of the entry planning, transit and probe events.

Similarly, the analysis indicated that there may be a potentially significant effect of networking the applications on propensity to trust. Although, as with SA, it was found that the propensity to trust decreased with networking the AIs rather than increased.

3.2.2 Usability of individual Agents/ ADeMs

It was found that the individual agents/ ADeMs themselves had a potentially significant effect on the workload components of physical demand, performance and frustration, and all of the 6 components of the TiA metric used (See Annex A).

For all the operators, each using one to three agents/ ADeMs relevant to their existing role/ experience, the physical demand component of workload was reported as being relatively small. The performance component and frustration component stretched from low to very high, with some applications showing a wide range of values.

For the TiA assessment components, there was a general trend in the ADeM applications for understanding, familiarity, propensity and overall TiA.

These observations are likely to reflect:

- The widely varying maturity of agents and their connectivity;
- Equally wide variation in the maturity of UI;
- A wide variation in design languages – i.e. lack of commonality/ standardisation between agents when operators were interfacing with several at once. Some agent's developers found the MOD/ Navy digital style guides useful, others had pre-existing formats;
- The level of autonomy had an impact, where several agents were adding UIs to agents fundamentally designed for fully automated operation, while others were considering human needs from the outset of the agent design;
- More integrated UIs (e.g. IRIS) arrived relatively late in the project and hence there was limited opportunity to evaluate their functionality.

3.2.3 Predictability of Intelligent Ship actions

The project used the Red Mirror agent and applied it to the Intelligent Ship system itself, creating what the project described as Blue Mirror. The aim of this was to assess the predictability of the Intelligent Ship system's response. Due to several constraints this was limited to assess course predictability based purely on observations (i.e. what Red forces would visually see).

Depending on the precision of the prediction required e.g. 40 degree sectors down to 10 degrees sectors, Blue Mirror was able to achieve a high prediction accuracy, higher than standard prediction benchmarks e.g. assuming it will do what it has done most frequently previously.

This is a potentially useful issue to assess within the system, however in this context when focusing purely on navigation aspects, showing high levels of predictability is perhaps likely as it reflects safe and predictable navigation policy against COLREGs, which may be ignored or even actively avoided in warfare scenarios. A more nuanced assessment of predictability and what is ‘good’ and ‘bad’ is required in future experimentation.

3.2.4 Agent - Agent and Agent - Human interactions

Developing a whole system understanding of potential interactions between agents & operators helped the suppliers, and the Dstl team, to identify gaps and opportunities, e.g. where data generated by one agent had secondary utility by another. This was achieved via several workshops and to a significant degree within the margins of the build-up to the evaluation events (i.e. face-to-face interactions). This was probably the aspect most restricted by COVID, resulting in occasional identification of opportunities later in planning cycles than ideal.

An outline of the connections between agents and operators during each of the scenario phases is provided in Figure 8. The wider analysis of verbal communications in the team showed a general reduction in voice based communications when the system operated in networked configuration, although this varied from vignette to vignette (scenario phase to phase).

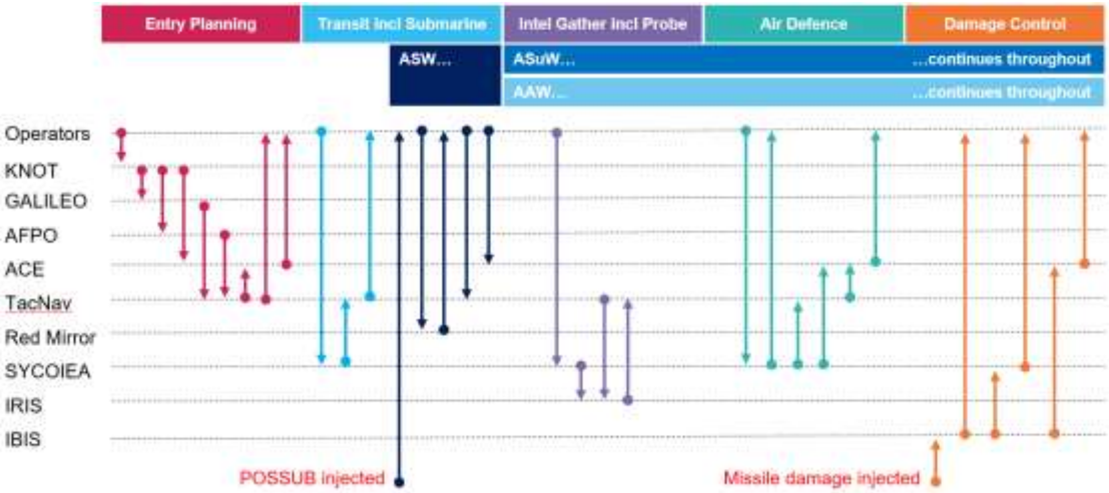


Figure 8 Outline of key Agent/ ADeM interactions during each phase of the scenario

Identified gaps in agent to agent communications or use included:

- The system had no Platform Management System[3] (PMS) equivalent to emulate both system state information and control responses in a platform’s systems. ACE delivered a limited emulation sufficient to allow ACE to work effectively, but several agents would have benefited from platform state information during the evaluations, and if hosted centrally this data layer could also capture historical, current and predicted states to inform prognostics or planning agents.

---

[3] PMS – the system on a naval platform that monitors and controls a Platform’s systems, including power, cooling and other services

- There was a concern from agent suppliers that there was no direct equivalent to a mission manager function. This resulted in no single source of information on mission (and hence capability) priorities, resulting in some agents translating non-functional requirements from the command aim or damage control state based rather than from a centralised source.
- Some agents had the capability to be used to answer ‘what-if’ questions; these capabilities were not able to be demonstrated under phase 2.

Figure 9 provides a high-level summary of the key interactions in the final evaluations[4].

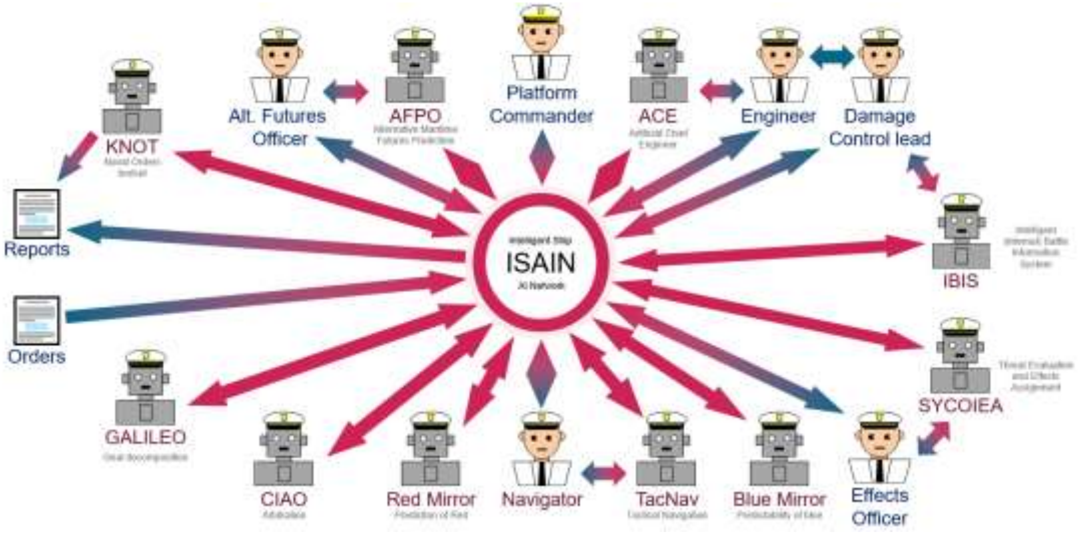


Figure 9 Final evaluation event – Agent/ ADeM interactions (high-level)

3.2.5 Internal system communications and process confirmations

A clear observation from the first 2 evaluation events, was that operators were using voice communication to assure themselves that the system was undertaking the required tasks. This reflected that many background and fully automated processes where offering limited confirmation messaging to operators to indicate that a task was being acted upon. This suggested a need for a system level dashboard to build confidence and trust with the operators and to show if systems have failed to pass on information or act as planned. This may only be needed during training, system failures or during a confidence building phase of operation.

This also highlights a future need to include within the fundamental design of an Intelligent Ship, a dynamic messaging system that can support development of trust, help to understand technical faults or agent serviceability, and that can manage messaging levels as tempo or overall trust develops or changes during a mission.

---

[4] Red arrows indicating digital information exchange; Blue, human interactions (verbal or via UIs)

To address this in Phase 2, CGI delivered an ADeM monitoring HMI that utilised Kibana[5] dashboards that could support two differing sets of users; both the operator and the system's engineer. These support both confirmation messaging to users but also the ability to track and analyse messaging with the system.

Although the dashboards were installed in Command Lab in preparation for the final evaluation event, they were unable to be run simultaneously with the other ADeM applications due to server capacity. This was communicated to the MA's prior to the start of the evaluation. The dashboards were subsequently demonstrated in Command Lab post the final evaluation event, but their usability and utility still needs to be fully demonstrated and assessed.

3.3 Hosting infrastructure & Intelligent Ship system set-up

The following observations and conclusion were made from the use of Command Lab, and from the design of the Intelligent Ship system itself:

- When multiple AIs and humans are brought together in one system, the whole system design and management of this complex structure needs to be considered from the outset.
- The needs of the human operators should be considered early in design and continuously throughout development. Simple issues could add to operator frustration, potentially detracting from the core evaluation aims.
- With respect to Command Lab:
    - The Command Lab was under development in parallel with the development of the Phase 2 Intelligent Ship system – this presented a range of risks, which while managed, meant plans often needed to be changed and some capability could not ultimately be delivered or evaluated by phase end.
    - There is benefit in good access to Dstl owned engineering support for Command Lab and the Simulation Gateway – the lead contractors couldn't be expected to understand all the nuances and configuration of the host facility as well as their own systems.
    - It would be useful to have separate operation and observation spaces during evaluations. This would allow meaningful conversation with stakeholders while not disturbing the operators undertaking evaluations.
- The ability to duplicate/ mirror some of the Command Lab's environment and simulation engines at a contractors, or another site would have also reduced the de-bugging time needed at Command Lab and allow for 'soft' testing of agents remotely. AI/ Agent suppliers had meaningful interactions during those sessions but also had extended periods of inactivity while the integration team was focused on issues with another agent.
- While there were efforts made to ensure standards, frameworks and hardware requirements were declared and agreed (and any underlying assumptions

---

[5] Kibana is a free and open user interface that lets you visualize data and, for example to understanding the way requests flow through an application or system.

checked) as early as possible, there were gaps that required interim fixes to make system elements work as planned. Issues arose around the lack of commonality in the use of key data items including time, course headings and position were also experienced.

3.4 HAT aspects & evaluations

The project quickly understood during its development that enabling collaboration between multiple AIs also drove a clear need to understand HATs and the systems that are enabling them. This is an acceptance that any complex C2 system that contains high levels of AI, autonomy and automation is likely to retain humans (and their skills) particularly in areas of analysis and decision making.

While the project has a clear aspiration therefore of assessing HAT aspects, the ability to do this in Phase 2 were limited. A key part of this was the overall system design approaches used to develop Phase 2, as already highlighted. These meant that the evaluations and overall system design were not based on the premise of considering the requirements of the human operators from the start, and that roles and numbers were dictated by the delivered agents and their current skills rather than best practice with respect to role and task design.

Another restriction was the realisation of the limitations of trying to set useful HAT focused evaluation metrics. This is due to the impracticality of comparing a new, novel system with existing approaches as shown in the table below:

	Intelligent Ship	Existing Naval C2
Concept maturity	Low	High
Host system maturity	Low	High
Data used	Simulated	Real
World model	Simulated	Real
Communications	Mix of verbal and machine messaging	Largely verbal and visual
UIs	A wide mix of UIs (Potential to combine into a single customisable UI standard aiming to be intuitive, but this has yet to be addressed)	Separate to specific capability but mature; often legacy with limited commonality in style
Operator training & familiarity	Limited	Extensive
Team size	Small	Large
Operator experience	Very limited	Extensive

Table 2 HMT assessment metrics - Comparison between Intelligent Ship and current approaches

As suggested by Table 2 it is perhaps impractical to show a new approach is 'better' during a single phase of integration research, especially when the underpinning AI technology is also still in its infancy.

The agents used were a mix of those originally designed to be fully automated that had overlaid UIs to meet the needs of the project, and agents that were designed specifically for Intelligent Ship with greater attention paid to HF aspects.

3.5 Evaluation delivery

- The timings of each evaluation event could, without COVID limitations, have been spaced better. The need to finish integration of the component agents into the Intelligent Ship system, before debugging, training the operators and then carrying out the evaluations within a two week period was challenging. Creating buffer periods between those actions would allow greater focus on the current task and allow management of unexpected issues.
- Combining the usability metrics with operator debriefs and recordings of the interactions, provided useful support information particularly for the Agent developers. Tools such as Clamshell will be essential going forward.
- The evaluations balanced data gathering with demonstration and de-bugging. For future larger scale evaluations a strict experimental protocol should be implemented to ensure the needs of one set of stakeholders e.g. the AI/ agent developers, do not interfere with the running and analysis of the evaluation.

3.6 Project delivery, skills & team structure

In order to deliver this type of project you need a diverse range of skills and capabilities across multiple organisations, both government and industry, and across programmes and tasks. Some of these skills (e.g. human factors and software engineers) are in limite supply both internal to Dstl and in industry presenting a risk to projects such as Intelligent Ship. Specific skills in HAT design need development across UK enterprise reflecting its emergence as a new focus area.

To be successful close collaboration is needed with and within the industrial delivery team. This is essential during a research driven phase of work such as Intelligent Ship, where there has been a need to be adaptable and flexible in approach and delivery.

Considerable effort was put in by the Dstl team to encourage team and inter-team working across the industry suppliers and to break down potential barriers, due to commercial sensitivities, to deliver the cross teaming approach required. The need for this was further amplified by the constraints imposed on the project by COVID.

3.7 Intelligent Ship enablers – gaps & needs

The Intelligent Ship concept assumes future access to a suitable quantity and quality of data that is likely to be different, broader or more detailed than routinely collected currently. This need for new or different data needs to be considered in future systems design and architectures.

The Intelligent Ship AIs/ Agents have, to date, been tested and assessed using virtualised and hence ‘clean’ data. Future testing will need to assess the impact of real, incomplete or damaged data sets both to support Validation and Verification, but also to support trust and an understanding of the practical limits of certain agents.

Future developments will need to identify data collection opportunities or develop specific plans to collect the required data to support research in this area. This will highlight current gaps in data collection, as the concept is likely to drive for new data that may not be possible to collect on current platforms due to lack of sensors, data capture or issues over data quality. Hence future projects and evaluations will need to be able to generate representative data in these areas, and will need some form of data generator and simulation capability that can mix together real and simulated data.

3.8 Impacts of concept on wider DLODs

The PeDRO study highlighted that a balance of investment (e.g. software vs. hardware) is needed when assessing the future use of AI and automation. Just because it can done by AI or automation doesn't necessarily mean it should – e.g. the cost or size, weight and power needs of automation to support an AI function may be prohibitive. AI and automation will mature and become cost acceptable at different rates in different capability areas. (There will be a dependency on issues such as the level of investment in component technologies across wider commercial applications.)

Consideration needs to be given to the impact on other lines of development areas such as training, support or doctrine when assessing applied intelligence. For example, the training of such a system may be very different than that required for a more conventional approach. If the system is complex, and potentially making decisions that the human operator would usually make, then the training needs to address:

- How to provide the information the operator needs to develop trust in the system;
- How to provide sufficient information to maintain situational awareness, and
- How information needs to change dynamically with changes in tempo, threat or operator workload.

4 Recommendations

Intelligent Ship has developed an environment to explore system design options that enable collaborative AI based HAT.

It has become apparent that comparing, and contrasting, internal performance and human experience of a new concept (such as the intelligent ship system) against an existing established, highly trained for, system and processes, is challenging. It is recommended that the future aims should include:

- To use more output focused metrics to understand how a system's performance, gaps and issues are addressed and mitigated.
- The project has shown that designing systems with AI's and humans requires consideration of humans (and their matching UIs) early and throughout design.
- Confirming the significance of the effects of networking Agents/ ADeM applications on mental and temporal demand related workload, SA and system trust.
- Investigation of system robustness and the decision arbitration needs of a collaborative AI and the corresponding impacts on operators.
- Developing a best practise guide to the design of systems that support collaborative AI and HMTs.

The final Phase 2 Intelligent Ship system was an amalgamation of the various agents offered and selected in the Phase 2 DASA call, and hence was not designed, or optimised, against a specific capability need or the needs of human operators. It is recommended that follow-on phases should also consider:

- Continuing to focus on collaborative AI and HMT.
- Focus on a more limited capability area and hence a more limited range of agents to allow a more optimised approach to system design. This should allow the system, the human operator roles (and numbers) and UIs to be designed and optimised against the capability.
- Testing system performance and robustness through a combination of more complex vignettes (e.g. large numbers of threats), forced failures in key components and/or the use of more realistic data.

A range of enablers, and/or constraints to further successful experimentation needs addressing. These include the following recommendations:

- A greater focus on standardisation covering: messaging; language; UI styles and system variables.
- Agree, test and implement an approach to understand and measure system performance and interactions during evaluations.
- Develop outline exploitation plans for porting the intelligent ship system (or a sub-set of it) to other evaluation or demonstration opportunities.
- Agree, test and implement appropriate approaches to messaging and providing system status and feedback between AI's-&-AI's and AI's-&-

humans. Aiming to understand the information requirements to provide situation awareness and build operator trust. And how this should dynamically change with operational tempo, trust level, training and other needs.

- Future assessment of predictability, and what is ‘good’ and ‘bad’, is required in future experimentation.
- Start addressing verification, validation and security considerations of the intelligent ship system approach.

List of abbreviations

AAW/ASW/ASuW	Anti-Air Warfare / Anti-Submarine Warfare /Anti-Surface Warfare
ACE	Artificial Chief Engineer
ADeM	Agents for Decision Making
AFPO	Alternative Futures Planning Officer
AI	Artificial Intelligence
ANOVA	Analysis of Variance
API	Application Programming Interfaces
ASuW	Anti-Surface Warfare
ASW	Anti-Submarine Warfare
C2	Command and Control
CIAO	Compounded Intelligent Agents for Optimisation
CMS	Combat Management System
CO	Commanding Officer (RN)
CoA	Course(s) of Action
COLREG	IMO’s Collision Avoidance Regulations
COVID (19)	Coronavirus
CSA	Chief Scientific Advisor
CTG	Carrier Task Group
DASA	Defence and Security Accelerator
DEW	Directed Energy Weapon
DLOD	Defence Lines of Development
FF&DC	Fire-Fighting & Damage Control
GALILEO	GoAL based decomposition for InteLligent ship ai nEtwOrk
HADES	Human-Agent Design and EvaluationS
HF	Human Factors (technical discipline)
HITL	Human-In-The-Loop
HMI	Human-Machine Interface
HAT/ HMT	Human-Autonomy Teaming/ Human-Machine Teaming
IDDC	ISAIN Dynamic Data-flow Configuration
IBIS	Internal Battle Intelligence Software
IOC	Initial Operating Capability
IP	Intellectual Property
IRIS	Information and Response for the Intelligent Ship
ISAIN	Intelligent Ship Artificial Intelligence Network
KNOT	Knowledge-based Naval Orders Toolset
MA	(Dstl) Military Advisor
MEO	Marine Engineering Officer (RN)
ML	Machine Learning
MOD	(UK) Ministry of Defence
NBO	Non-Binding Opinion (MODRec)
NDP	Naval Design Partnership
NLP	Natural Language Processing
OOTW	Officer Of The Watch (RN)
PeDRO	Platform Design Risks & Opportunities
PMS	Platform Management System
PWO	Principal Warfare Officer (RN)
SA	Situational Awareness/
SADM	Ship Air Defence Model
SDK	Software Development Kit
SOLAS	IMO’s Safety of Lives At Sea Regulations
SyCOIEA	SYstem Coordinating Operational Interaction for Effects Assignment
TacNav	Tactical Navigation
TEWA	Threat Evaluation and Weapon Assignment
TiA	Trust in Automation
TRL	Technology Readiness Level
UI	User Interface
USV	Uncrewed Surface Vessel
VBS	Virtual Battlespace

A.1 Trust in Automation metrics

Trust In automation (TiA) metrics used during the evaluations are a underpinned by a theoretical model[6] and consists of six subscales each of which is driven by the responses (from ‘strongly disagree’ to ‘strongly agree’) to a number of different statements, asked in a mixed order as follows (where the numbers indicate the order asked):

- Reliability/Competence:
  - (1) The system is capable of interpreting situations correctly;
  - (6) The system works reliably;
  - (10) A system malfunction is likely;
  - (13) The system is capable of taking over complicated tasks;
  - (15) The system might make sporadic errors; and
  - (19) I am confident about the system’s capabilities.
- Understanding/Predictability:
  - (2) The system state was always clear to me;
  - (7) The system reacts unpredictably;
  - (11) I was able to understand why things happened; and
  - (16) It is difficult to identify what the system will do next.
- Familiarity:
  - (3) I already know similar systems; and
  - (17) I have already used similar systems.
- Intention of Developers:
  - (4) The developers are trustworthy; and
  - (8) The developers take my well-being seriously.
- Propensity to Trust:
  - (5) One should be careful with unfamiliar automated systems;
  - (12) I rather trust a system than I mistrust it; and
  - (18) Automated systems generally work well.
- Trust in Automation:
  - (9) I trust the system; and
  - (14) I can rely on the system.

---

[6] M. Körber, “Theoretical considerations and development of a questionnaire to measure trust in automation,” Congress of the International Ergonomics Association, pp. 13-30, 2018.

Report documentation page v5.0

* Denotes a mandatory field

1a.	**Report number: ***	DSTL/PUB148547	**1b.**	**Version number:**	1.0
2	**Date of publication: ***	04/04/2023	**3.**	**Number of pages:**	27
4a.	**Report UK protective marking: ***	UK OFFICIAL			
4b.	**Report national caveats: ***	NONE			
4c.	**Report descriptor: ***	NONE			
5a.	**Title: *** Intelligent Ship Phase 2 - Summary Information to Support Phase 3 Proposals				
5b.	**Title UK protective marking: ***	UK OFFICIAL			
5c.	**Title national caveats: ***	NONE			
5d.	**Title descriptor: ***	NONE			
6.	**Authors: *** Tate, A.; Cooke, C.; A Tate				
7a.	**Abstract: *** This summary provides an overview of the second phase on Intelligent Ship (Intelligent Ship) project funded by Dstl Autonomy Programme with CSA research funding. It considers an outline of the activities within the phase, as well as key conclusion and recommendations. It’s an abbreviated version of CR140397 for sharing as part of the DASA competition for Phase 3 of the project to provide context to potential suppliers.				
7b.	**Abstract UK protective marking: ***	UK OFFICIAL			
7c.	**Abstract national caveats: ***	NONE			
7d.	**Abstract descriptor: ***	NONE			
8.	**Keywords:** Collaborative AI; Human-Autonomy Teaming; C2; Intelligent Ship				
9.	**Name and address of publisher: ***	Dstl Platform Systems Portsdown West Portsdown Hill Road	**10.**	**Name and address of funding source:**	CSA funding via Autonomy Programme
11.	**Funding source contract:**	CSA			
12.	**Dstl project number:**	PJ100338			
13.	**Programme:**	Autonomy			
14.	**Other report numbers:**	n-a			
15a.	**Contract start date:**	Nov 2020	**15b.**	**Contract end date:**	April 2022
16.	**IP conditions for report: ***	MOD Funded Dstl authored report			
17a.	**Patents:**	NO			
17b.	**Application number:**	n-a			
18.	**Release authority role:**	n-a			

THIS PAGE INTENTIONALLY LEFT BLANK

Ministry
of Defence