Ministry of Defence

JSP 936 V1.1
Dependable Artificial Intelligence (AI) in Defence
Part 1: Directive

Dependable AI JSP 936 – Cover Note, November 2024

Events in 2023 highlighted the rapid advancements in AI and the potential of these technologies to transform all aspects of Defence. It is essential that we embrace these technologies at pace to realise productivity gains and to maintain our military edge within a competitive, volatile and challenging international security environment. While we modernise at increasing speed, it is equally important to assure our leadership, our staff, Parliament and the public that we are adopting AI technologies safely and responsibly.

This cover note sets out an overview of how Defence should seek to implement the JSP 936.

- This JSP (Part 1) is an important step to ensure that teams across Defence understand what is required from them when developing and using AI technologies. It provides clear direction on how to implement the MOD AI Ethical Principles to deliver robust, reliable, and effective AI-enabled services and capabilities.

- AI technologies are maturing at extraordinary pace and there is already a large number of related projects and programmes underway across Defence. At the same time, our understanding of related risks, safeguards and assurance standards continues to evolve. In many ways **this JSP represents the aiming point or ideal end state as many of the supporting tools and frameworks are necessarily still in development**. Where we cannot yet fully meet the JSP requirements, we should understand the risk and plan a route to compliance.

- Implementing the requirements set out herein will involve determining the right accountabilities and responsibilities in our respective organisations, as well as getting the technical assurance in place. Importantly, Defence organisations should not create unecessary duplicative processes, but update and adapt their existing, robust governance structures and assurance processes in a way so they can fully address AI-related risks and opportunities.

- **Implementing the JSP will be an ongoing process** - it is not expected that everything will be in place overnight. We must learn by doing and iterate and improve over time to ensure that we do not inadvertently handicap essential Research & Development and capability development efforts.

- As the department works to become ‘AI ready’, the Defence AI and Autonomy Unit (DAU) and the Defence AI Centre (DAIC) will be working with your teams to develop additional guidance and practical advice on the implementation of our safe and responsible approach.

We want to harness the innovation and creativity found across Defence and industry and intend that our approach will enable the ambitious adoption of AI-enabled solutions. Ethics can sometimes be thought of as a barrier to innovation. This could not be further from the truth; considering the ethical impacts of AI will help teams proactively anticipate and address potential barriers to success and deliver military advantage.

Responsible AI processes increase the assurance between Allies in NATO – with AI changing the global defence and security landscape we need to improve interoperability and coordination. It is vital that NATO and Allies share a common approach that is in accordance with our values, norms and international law.

Defence organisations should consider the next steps for implementing this JSP:

- Start conversations with the Responsible AI Senior Officer (RAISO) and the DAU/DAIC about how your organisation will implement the JSP. Assign roles and responsibilities where appropriate to support your organisation to do so.

- Identify in-use, under development and planned AI applications or components. This will help inform your JSP implementation priorities.

- Develop an implementation plan for the JSP, considering which elements should be priority tasks (for example, AI ethical risk assessments and skills and training) and determining where existing processes are either sufficient or need to be strengthened.

Part 2 guidance for the JSP will be developed over the course of 2024. Whilst that is happening, preliminary guidance via a range of minimum viable products (MVPs) have been made available internally; these include:

- Guidance on the role of the RAISO.
- A template model card and guidance.
- An AI ethics risk management framework.
- AI assurance question sets.
- A repository of good practice case studies.

As an emerging technology that is evolving rapidly, the MVP preliminary guidance will be iterated and tested through live AI projects before being integrated into a digitally enabled version of JSP Part 2.

Understanding whether or not AI-enabled systems or capabilities comply with MOD policies and ethical principles

- This JSP does not provide a list of AI-enabled systems or capabilities which are compliant or not compliant with our policies and ethics.

- Compliance with policy and ethical principles is not determined by what an AI-enabled system does; it is determined by how it does it. It follows 1) that any given concept could be delivered successfully or unsuccessfuly, depending on choices made potentially at any point during the system lifecycle; and 2) that it is not possible for MOD to generate a list of those that are acceptable and those that are not.

- We accept the possibility that a concept for a system or capability might be fundamentally and intrinsically incapable of being delivered in a safe, responsible and ethical manner – but we think it much more likely that a concept could ultimately be delivered safely, ethically and responsibly, provided the precepts and processes set out in this JSP are followed.

- Ultimately, our policy is grounded in the importance of meaningful human control (and therefore human responsibility and accountability) exercised through context-appropriate human involvement.

- Whatever the underlying concept, we expect that a system or capability could be developed or modified (either in itself, or – taking a systems engineering approach – in terms of the wider 'system of systems') to achieve and ensure meaningful human control.

- We recognise that this process could affect system or operational parameters in various ways. The most successful concepts and systems will be the ones which nevertheless deliver the greatest benefit while complying with our policies, legal obligations and ethical principles.

- Delivering ambitious, safe and responsible AI-enabled capability is therefore a shared endeavour between MOD and its suppliers.

Recognising the leading role of the private sector and academia, MOD's ability to adopt AI-enabled capabilities also depends on our relationship with these stakeholders and their trust that we will be responsible stewards of these technologies.

Likewise, MOD needs assurance that commercial vendor technologies are safe, reliable, and consistent with shared ethical principles and values.

Foreword

The integration of Artificial Intelligence (AI) in the military domain has the potential to revolutionise Defence capabilities, improve operational efficiency, and ultimately save lives. As examples, AI-enabled decision-support capabilities can accelerate the tempo and rigour of operational planning; we can deploy AI-powered systems with appropriate oversight in high-risk situations, such as reconnaissance missions or bomb disposal, to reduce the risk to human life and minimise casualties; and AI can streamline and optimize military logistics and supply chain operations, ensuring that our forces receive the necessary supplies and equipment in an efficient manner.

JSP 936 builds on our AI Ethical Principles, set out in the Ambitious, Safe and Responsible Policy Paper (ASR), as part of the Defence AI Strategy, which establish the ethical framework considerations of Human-Centricity, Responsibility, Understanding, Bias and Harm Mitigation and Reliability. At the heart of the ambitious, safe and responsible use of AI are our Defence People. By embedding our AI Ethical Principles, we will cultivate trust in AI technologies and their applications, realising the full potential of human-machine teaming, while mitigating the risks associated with its use, misuse or disuse and preventing unintended consequences.

The adoption and integration of novel technologies and capabilities is not a new challenge for Defence. We have established and effective risk management systems in place with clear lines of accountability and assurance and controls frameworks embedded throughout the lifecycle of any military capability. Defence has legal, safety and regulatory policies, processes and compliance regimes in place – this JSP is written in line with our existing framework and addresses the unique requirements owing to the nature or functionality of AI. By implementing AI assurance measures, including nominating Responsible AI Senior Officers, we aim to foster a culture of responsibility and accountability among AI developers, users, and policymakers, promoting the development of AI systems that are not only technically sound but also ethically aligned.

We recognise that AI ethics and assurance are dynamic fields that require continuous engagement, collaboration, and iteration. As such, we are committed to regularly reviewing and updating this policy to reflect the latest advancements in AI research, industry best practices, and societal expectations.

An AI/Human partnership was used to support the production of the cover page and this foreword – the rest of this JSP was solely written and reviewed by humans.

Alison Stevenson
Director General Delivery & Strategy
Ministry of Defence

Preface

How to use this JSP

1. JSP 936 mandates the application of ambitious, safe and responsible practices relating to all Defence projects that include Artificial Intelligence (AI). It is aligned to the Government's Technology Code of Practice and designed to be used by MOD staff responsible for developing and/or using systems which include AI, regardless of the application domain. This JSP contains the direction, high-level principles and guidance needed to achieve the policies outlined in the MOD Ambitious, Safe, Responsible [ASR] policy document. This JSP will be reviewed at least every two years.

2. The JSP is structured in two parts:

   a. Part 1 - Directive, which provides the direction that **must** be followed in accordance with statutes or policies mandated by Defence or on Defence by Central Government.

   b. Part 2 - Guidance, which provides the guidance and best practice that will assist the user to comply with the Directive(s) detailed in Part 1.

Must and *should*

3. Where this policy says **must**, this means that the action is a compulsory requirement to be completed by the actioner. Where this policy says *should*, this means that the action is not a compulsory requirement but is recommended good practice and therefore should be considered and applied as far as possible in the relevant context.

Coherence with other Policy and Guidance

4. Where this document contains references to policies, publications and other JSPs which are published by other Functions, these Functions have been consulted in the formulation of the policy and guidance detailed in this publication.

To support external publication, direct references to wider internal policy have been replaced by '[extant internal policy]'. Where this is seen, the internal publication includes direct references; it does not materially impact on the content of the JSP.

Related JSP	Title
JSP 200	Statistics
JSP 375	Management of Health and Safety in Defence
JSP 376	Defence Safety Acquisition Policy
JSP 441	Information, Knowledge, Digital and Data in Defence
JSP 536	Defence Research Involving Human Participants
JSP 604	Defence Manual for Information and Communications Technology
JSP 732	Research Integrity
JSP 815	Defence Safety Management System
JSP 816	Defence Environmental Management System
JSP 887	The Public Sector Equality Duty in Defence
JSP 892	Risk Management

JSP 912	Human Factors Integration for Defence Systems
JSP 939	Defence Policy for Modelling & Simulation
JSP 940	MOD Policy for Quality
JSP 945	MOD Policy for Configuration Management
JSP 985	Human Security in Defence
Technology Code of Practice	Government guidelines on design, build and buy technology for all software purchases
Data Ethics Framework	Guidance for public sector organisations on how to use data appropriately and responsibly when planning, implementing, and evaluating a new policy or service

Training

5. The Defence AI Centre (DAIC) is responsible for cohering training requirements for the AI-specific roles and responsibilities contained in this document, and for supporting the development of training material required.

6. AI is expected to form a key part of capability across all Top Level Budget (TLB) organisations (or equivalent[1]). To develop and deploy effective AI systems currently requires significant technical expertise and training needs to reflect this. In line with [extant internal policy], TLB nominated persons are responsible for the development and maintenance of technical capability.

7. The TLB nominated person is responsible for ensuring that all staff, including and beyond the roles identified in this JSP, working with AI are suitably qualified and experienced for the roles being fulfilled and the nature of the AI under development and use. The nominated person **must** liaise with the Responsible AI Senior Officer (RAISO) within the organisation to ensure that training meets the needs of the organisation.

8. To support MOD awareness of AI and its capabilities, some basic level training is available via the Defence Learning Environment. This *should* be taken alongside role specific AI competency training to work towards the goal of MOD as an 'AI-ready' organisation.

Further Advice and Feedback – Contacts

9. The owner of this JSP is the MOD Director General Delivery & Strategy and it is managed by the Defence AI and Autonomy Unit (DAU) and the Defence AI Centre (DAIC).

[1] Hereafter all TLB equivalent organisations are grouped under the single 'TLB' banner.

Contents

1 Introduction

Policy

1. The Defence AI Strategy [1] and associated policy statement [2] (known as the Ambitious, Safe, Responsible, or ASR, document) set out how MOD will adopt and deploy AI in ways that are both effective and aligned to the UK's democratic values.

2. A holistic approach *should* be adopted when considering the AI Strategy; for example, it is closely linked to the MOD's strategies for data [3] and digital backbone [4].

3. In keeping with broad consensus, the MOD adopts the UK National AI Strategy [5] position that no single definition of AI is suitable across its range of applications. Therefore, a general characterisation is made for AI as follows:

'Machines that perform tasks normally requiring human intelligence, especially when the machines learn from data how to do those tasks.'

Note that Machine Learning (ML) is a subset of AI but has become so prevalent that AI is often referred to as AI/ML. ML is not further characterised here as it is encompassed in the characterisation above.

4. UK Government has also set out a legal definition in the National Security and Investments Act (see [6] for more information) which has greater clarity but is arguably less helpful for practical purposes in the context of this document:

'Technology designed to approximate cognitive abilities including reasoning, perception, communication, learning, planning, problem solving, abstract thinking or decision making.'

It does, however, serve to provide useful perspective with respect to the use of AI in decision-making and these perspectives *should* be borne in mind when considering the nature of AI in Defence applications.

5. For the purposes of this JSP, the MOD characterisation outlined in paragraph 3 will be adopted. Figure 1 provides an overview of indicative concepts, approaches and techniques that are found in AI. This is deliberately vague and incomplete as the field is complex and fast moving, any attempt at completeness is futile and risks excluding future developments. However, it does serve to characterise the types of technology that are often classed as AI and hence within scope of the JSP.

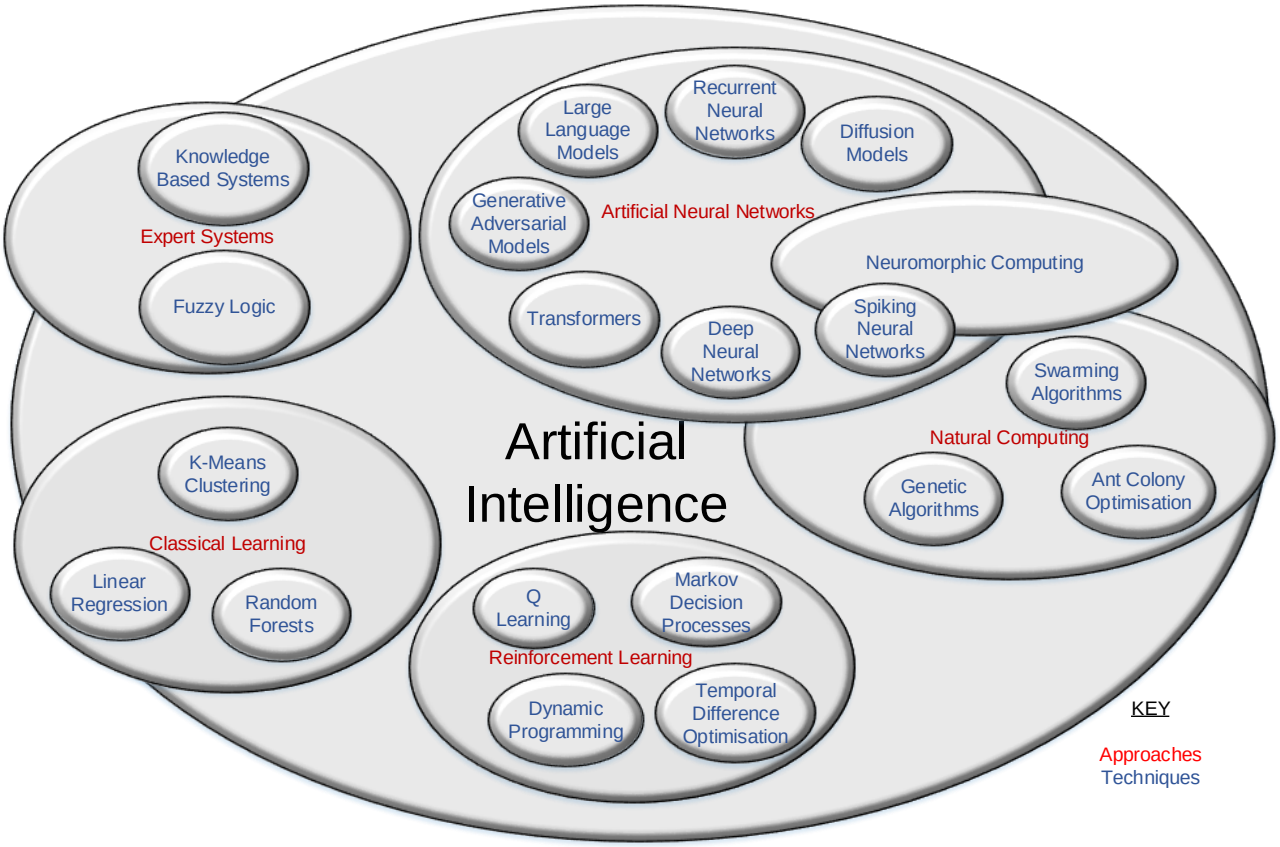

Figure 1: Indicative Approaches and Techniques for AI

6. The concept of dependability relates to how much the user can rely on the safe, secure and correct operation of the system within a given context (including mission or task, human involvement, the environment and system constraints). This JSP will not provide a dependability scale but the level of confidence developed in the system **must** be commensurate with the reliance placed upon it. Adequacy of evidence to support that confidence **must** be judged by risk owners throughout the lifecycle of the AI. In the Defence sphere, there might be times where there is considerable uncertainty (for example, due to a fast-evolving operating environment). Where this is the case owners of resultant risk **must** understand that uncertainty and communicate associated risks to all stakeholders involved in its operation. Whilst this JSP highlights new risk dimensions introduced by the unique nature of AI, the management of risk is not part of this JSP and **must** continue to be practiced as per JSP 892.

Scope

7. This JSP uses a broad definition of the term ‘software’ that includes program code, operating procedures, all relevant data, as well as associated documentation, such as the requirements, software specification, test plans, and user manuals etc. Its scope includes software that provides functionality for MOD equipment (as discussed in [extant internal policy]), Programmable Elements (PE) as defined by Defence Standard (Def Stan) 00-056, and software used in other digital information systems across Defence. The term AI is used to refer to the implementations in software of behaviours characterised in paragraph 3 but not restricted by the indicative examples shown in Figure 1. Software acquisition as a broader topic is addressed in JSP 441, JSP 604 and [extant internal policy]; the contents of this JSP are intended to be complementary to those JSPs and other JSPs referenced herein.

8. Similarly, with respect to Security, this JSP places requirements for special attention to be made regarding the unique security challenges associated with AI, it does not however replace [extant internal policy] and **must** be used in conjunction with it.

9. AI may take the form of algorithms and/or models. Where AI takes the form of a model or is used as a component within a conventional model, JSP 939, which provides Defence direction and guidance for the acquisition, development and usage of models and simulations across Defence, may apply. Where this is the case, it **must** be applied in combination with this JSP.

10. Due to the complexity of software development as a whole, it is neither feasible nor helpful to explicitly detail differences across the JSPs; however they *should* be addressed in conjunction with each other and any discrepancies resolved in a way that does not undermine confidence in the fulfilment of the policy requirement intent for dependable AI set out in this JSP.

11. Since AI has become increasingly prevalent across technology applications, it is important that a pragmatic approach is adopted in the interpretation of this JSP's scope. Given the breadth of technologies, it is impossible to be precise in stipulating the technology and application scope. In general terms, this *should* be tied to existing risk management and, in particular, where items are developed or purchased to meet a specific Defence need; some (non-definitive) examples are:

a. **In scope**.

(1) Object detection, recognition and identification software in Intelligence, Surveillance and Reconnaissance (ISR) capabilities.

(2) Reinforcement learning algorithms used in Command and Control (C2) application and autonomous systems.

(3) Large Language Models used in providing Human Resources (HR) applications.

(4) AI utilised within decision support toolsets or applications. For example: Course of Action (COA) development; task allocation (for example in a crewed/uncrewed force mix); mission tactics engines; and logistics planning.

b. **Out of scope**[2].

(1) Openly available, proprietary search engines.

(2) Predictive text facilities on commercial mobile messenging services (e.g. mobile phone).

(3) Common office productivity tools (e.g. MS Co-Pilot) where the risk associated with the resulting material is low.

12. Additionally, it should be noted that AI is increasingly being combined with non-AI approaches (such as, for example, propositional logic) to produce more effective capabilities. In some cases, the demarcation between the technologies may be so blurred

[2] Whilst out of scope for this JSP, responsible use (including maintaining security and checking the output) of such software remains the responsibility of the user.

that they appear to be a single entity. Some level of judgement is required when deciding which software components are within scope.

Applicability

13. The requirements set out in this JSP **must** be implemented from the outset of all Defence Capability development where MOD requirements lead to the use of AI[3]. For legacy AI that is still running or being repurposed, alignment with the requirements of the JSP *should* be achieved in so far as is possible. Where achievement is not possible, tailoring of the JSP **must** be applied as outlined in the Tailoring section.

14. Meeting the JSP requirements *should* be a shared endeavour between MOD and its suppliers. Through such an approach, it is anticipated that the challenges of responsible AI development and use can be met efficiently and effectively leading to better outcomes, in line with the MOD's policy set out in the ASR.

15. This JSP **must** be applied by all MOD staff in all phases of the system life cycle, from pre-concept through to equipment disposal / service termination, but especially the following capability stakeholders:

a. providers of Science and Technology (S&T) research at all Technology Readiness Levels.

b. customers (S&T research, capability planners, capability sponsors, programme Senior Responsible Owners (SROs) and requirements managers).

c. Delivery Agents (Project / Delivery Teams).

d. Defence Line of Development (DLOD) owners.

e. trials units/organisations.

f. Specialist Engineering Functions.

g. policy makers.

h. end users[4].

16. The application of this JSP *should* be proportionate to the maturity and nature of the AI application. Documentation *should* be produced and retained in line with extant policy for software in the application area.

Tailoring

17. It is recognised that the nature of AI is such that compliance with some of the requirements contained in this JSP may not be possible for some AI applications and uses. A demonstrably responsible approach to this challenge is needed.

18. Where tailoring is required, an assessment of risk **must** be undertaken and the additional risk managed as part of the risk management activities required by Section 3.

[3] That is, at the earliest point at which the potential for adoption of AI is identified.
[4] *End Users* is an all-encompassing term to include all users of a capability, regardless of Armed Service (or if MOD civilians), rank or role. It includes operators, maintainers, trainers, support personnel, and so forth.

Delegation of Responsibilities

19. TLB Holders and Chief Executives are responsible for issuing appropriate direction within their area of responsibility. This includes adequate management arrangements to ensure their activities in relation to the through-life development and use of AI in Defence applications meet the requirements expressed in this policy.

20. TLB Holders and Chief Executives **must** ensure that Commanding Officers and managers to whom they may delegate authority (for example, RAISOs – see Section 4) are competent by virtue of suitable qualifications and experience and have adequate resources at their disposal.

21. Where further direction is required on Dependable AI policy, DAU *should* be consulted. The DAIC *should* be contacted for technical guidance and direction.

22. The delegation of responsibilities should note the Governance requirements set out in Section 4.

Associated Standards and Guidance

23. The primary standard for software is Def Stan 00-055 Requirements for Safety of Programmable Elements (PE) in Defence Systems [7]. Alongside non-AI software, Def Stan 00-055 currently only addresses ML and may therefore require additional considerations to be developed and applied for non-ML AI safety-related software. Additionally, since AI is abstract and will exist within some larger system to achieve effect, Def Stan 00-056 Safety Management Requirements for Defence Systems [8] is also relevant to many systems requiring dependable AI.

24. Other relevant standards and guidance are referenced throughout this JSP. Not all references were developed in the context of AI-based technology. It might, therefore, be necessary to view them through the lens of an AI-enabled context.

25. Noting that the field of AI is developing rapidly, organisations developing capabilities using AI **must** review the latest government and commercial standards and guidance for applicability, and consider their application even where they are optional.

26. In line with MOD policy, programmes *should* adopt civil standards where possible to achieve recognised good practice; however, these *should* be assessed for applicability. Where military standards exist, these *should* be consulted to address ‘military delta’ requirements.

2 AI in Defence Systems

Introduction

27. AI is considered a cross-cutting technology and as such will become increasingly prevalent across Defence, with applications ranging from back office corporate services to frontline military capabilities. A key strength is being able to perform in complex environments where the external input space is so large and complex that traditional, non-AI software is unable to perform effectively. AI therefore has the capacity to unlock important capabilities and make efficiencies that, when operated alongside human decision-makers, will maintain UK military advantage.

28. That said, the risks associated with AI, in particular its potential for unpredictable and opaque behaviour, means that a balance of risk judgement on the adoption of AI rather than existing technologies is needed. The rationale for that decision *should* be recorded and agreed at the appropriate level via the AI governance chain for all AI components that are assessed as having major and above risk.

29. Essentially, AI will always operate in the context of wider systems. These systems can be classified in several ways. For simplicity, we divide them into two main classes:

a. **Robotic and Autonomous Systems (RAS)**. These can be considered as physical platforms that typically move to achieve a desired outcome.

b. **Digital Systems**. These systems are designed to receive, process and store data and output information either to a human or another system.

Each of these are now further discussed. Before moving on, it should be noted that the lines between RAS and digital systems are blurred and some concepts are applicable across both.

Robotic and Autonomous Systems

30. The term 'robot' stems from the Czech word for forced labour ('robota'). In modern technology it has come to mean a system designed to undertake work that would normally be done by humans (or other living entities). In the context of this JSP, we consider only advanced robotics, that is those requiring the use of AI to perform complex tasks.

31. Similar to the problematic nature of defining AI, there is no overall consensus on what constitutes an Autonomous System (AS)[5].

32. Since this JSP addresses AI across all MOD AS, for the purposes of this JSP, we frame the type of system to which we are referring as follows:

'*An autonomous system is capable of acting on high-level goal-setting provided by human operators. From these set goals and its perception of its operating environment, such a system is able to take action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous system may be predictable, individual actions may not be. Additionally, autonomous systems may contain machine-learning capabilities which endow them with some abilities for changing their own actions without the intervention of a human. Autonomous Systems are contrasted to automated systems that can function with little human involvement, but only perform pre-programmed actions*[6].'

In all but the simplest environments, AI is highly likely to be a key component of an AS.

33. Having noted there is no universally accepted definition of AS, we should note that NATO has published the following [9]:

[5] AS may reside on crewed or uncrewed platform systems.

[6] We have endeavoured to include a comprehensive characterisation of all AS in order to apply the right frameworks to these systems – this will help teams to ensure that AS take the appropriate actions and are sufficiently explainable and predictable.

'Autonomy is the ability of a system to respond to uncertain situations by independently composing and selecting among different courses of action in order to accomplish goals based on knowledge and a contextual understanding of the world, itself, and the situation. Autonomy is characterised by degrees of self-directed behaviour (levels of autonomy) ranging from fully manual to fully autonomous.'

34. Clearly there is an overlap between robots and autonomous systems but advanced robots do not have to be autonomous and similarly autonomous systems do not need to be robotic. Generally though, they are grouped together due to their similarity.

35. Notably there is a continuum in RAS behavioural capabilities from manually operated, through automatic to fully autonomous. JSP 936 is concerned with dependable AI regardless of the overall system level of autonomy; however, higher levels of autonomy typically reduce the potential for human decision-making within the control loop and this **must** be considered when applying the requirements of this JSP[7].

36. Many RAS are considered to be or to contain 'dual-use' technology. Where a civil-sector RAS containing AI is procured for Defence use, the 'military delta' in the Operational Design Domain (ODD) **must** be identified, and the AI tested to ensure that any drop-offs in performance from the possibly benign, designed-for ODD are understood and additional risks identified. The difference in safety and security needs, introduced in the military domain by the increased potential for malign actions carried out by an adversary, should be of particular note.

Digital Systems

37. AI may be used in a range of digital system applications. These may be part of physical platforms such as crewed aircraft, road vehicles, submarines etc. but the essential difference between AI in digital systems and RAS is that digital systems are often more 'open loop'[8] in nature. Because the output of such systems is typically not physical, incorrect behaviours may be more subtle. For example, credibly incorrect outputs from generative AI (including Large Language Models, LLMs) may mislead human operators into poor decision-making.

38. The ODD for the AI *should* be identified and include information about its context of use and the digital systems in which the AI is designed to operate.

39. All digital systems containing AI that are within scope of this JSP **must** be clearly identified as such (i.e. as AI-based).

40. Products (such as documents, images etc.) that have had AI applied in their development **must** clearly state that AI has been used in their production. In products that have high-levels of risk, relevant information on the way in which AI has been developed, used and assured *should* be available to the risk owner of decisions being made on the basis of the product. This *should* include an appropriate capture of risk in the relevant risk register.

41. AI used in the production of official statistics *should* be clearly communicated in the associated Technical Annex (detailed in JSP 200).

[7] We are clear that there must always be context-appropriate human involvement throughout the AI-lifecycle which achieves meaningful human control over the operation and effects of the autonomous system.
[8] That is, the output is not fed directly back into the system providing the input to the AI.

42. Where the AI in a digital system provides 'advice' to an operator, alternate sources, and their provenance, of information *should* be identified. This will provide an understanding of the level of influence the AI may have on the operator's decision-making processes.

43. The level of influence and consequences of AI outputs on overall digital system performance *should* be identified and incorporated into overall system risk analysis.

3 Legal & Ethical Considerations of AI

Introduction

44. Whilst the ASR policy promotes the ambitious adoption of AI, it does so within the context of: good governance; the demonstration of safety; and legal compliance. Within these three perspectives sit the five ethical principles as shown in *Figure **2***. Each of the five principles are connected with all of the three perspectives.

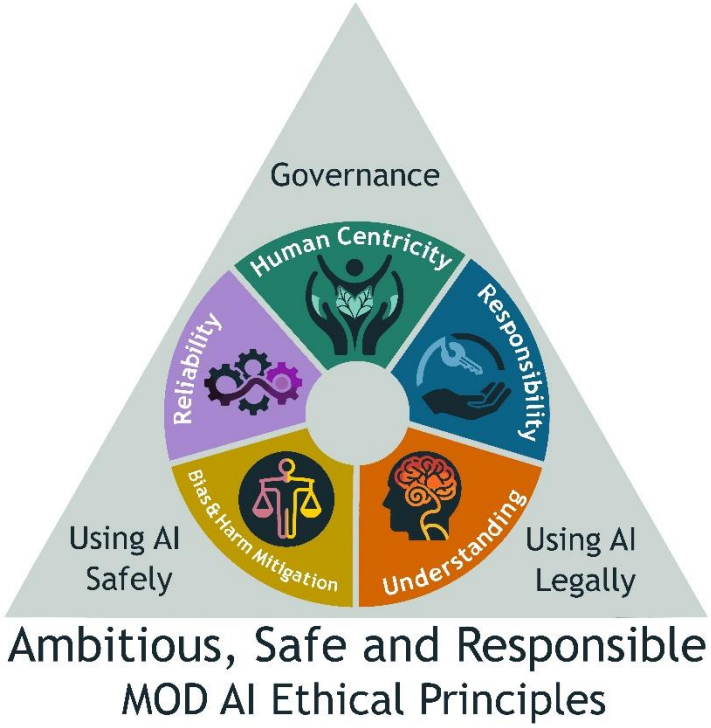

Figure 2: Conceptual approach of MOD AI policy, Ambitious, Safe and Responsible.

45. The remainder of this section focusses on legal considerations and the ethical principles. Governance is addressed specifically in Section 4, Ethics Governance and safety is more broadly addressed throughout the JSP.

Legal Considerations

46. As with all aspects of its activity, the MOD's development and use of AI is governed by national and international law. Defence always seeks to abide by its legal obligations across the full range of activities from employment law, to privacy and procurement, and the law of armed conflict, also known as International Humanitarian Law (IHL). It has robust practices and processes in place to ensure its activities and its people abide by the

law. These practices and processes are being – and will continue to be – applied to AI-enabled capabilities.

47. Appropriate legal advice **must** be sought through existing channels at the beginning of the AI project, and at any relevant stages throughout the programme, to ensure that all relevant national and international law relating to the development and use of AI, including associated data, are identified and complied with. Industry and developers will need to obtain their own legal advice as to the legal framework that applies to their development of the product, including in respect to the use of data and intellectual property. Within Defence, advice should be sought in accordance with established capability development and acquisition frameworks. Any wider legal queries should be directed to MOD Legal Advisers (MODLA).

48. Any systems of control to ensure the use of the AI enabled capability is in line with national and international legal obligations **must** be clearly communicated to relevant stakeholders. Such controls may apply at any point in the development and use lifecycle of the capability and may include, for example, controls on the use of personal data when operating internationally where legal obligations may be different from, or conflict with, the those in the UK.

Ethical Principles

49. The ASR policy [2] is clear that the MOD is committed to responsibly developing and deploying AI for purposes that are demonstrably beneficial whilst upholding human rights and democratic values. To support that it sets out five key principles[9]:

a. **Human-centricity**. The impact of AI-enabled systems on humans must be assessed and considered, for a full range of effects both positive and negative across the entire system lifecycle.

b. **Responsibility**. Human responsibility for AI-enabled systems must be clearly established, ensuring accountability for their outcomes, with clearly defined means by which human control is exercised throughout their lifecycles.

c. **Understanding**. AI-enabled systems, and their outputs, must be appropriately understood by relevant individuals, with mechanisms to enable this understanding made an explicit part of system design.

d. **Bias and harm mitigation**. Those responsible for AI-enabled systems must proactively mitigate the risk of unexpected or unintended biases or harms resulting from these systems, whether through their original rollout, or as they learn, change or are redeployed.

e. **Reliability**. AI-enabled systems must be demonstrably reliable, robust and secure.

Adoption of these principles is key to developing trust in our use of AI-based systems across the range of stakeholders, from the operator to system owners and our wider society. When speaking of trust, we do ***not*** suggest an abdication of all of our control to AI-based systems, rather we speak of building reliable systems operating under meaningful human control exercised through context-appropriate human involvement. Whilst it is tempting to treat the principles in isolation, they *should* be considered in the

[9] The text in these five principles are verbatim copies of that found in the ASR for ease of reference.

context of the wider ASR document (for example Legal and Governance aspects are considered separately in the ASR but are related to ethical values) as well as other relevant publications such as JSP 985 Human Security in Defence.

50. In many respects, but not all, the MOD's AI Ethical Principles can be considered as driving the safe[10] adoption of AI across the entire business, from back office functions to front line operations. Organisations **must** consider these principles as early as practicable, for ease of implementation. Teams will almost always need to undertake a balancing and judgement exercise between principles in order to adopt them – what good looks like in terms of meeting the principles will look different for each use case. Additionally, teams will need to consider military requirements and operational effectiveness, recognising that developing AI responsibly by implementing the MOD AI Ethical Principles will ultimately result in more robust, reliable, and effective AI-enabled capabilities, thereby advancing our military edge.

51. Whilst not directly related, [extant internal policy] may be relevant to AI ethics, particularly in relation to the reporting of ethical concerns. Where AI is in use and may impact human well-being, there **must** be clearly signposted avenues for redress as laid out in [extant internal policy].

Ethical Principles: Human Centricity

52. All humans (e.g. MOD personnel, civilians, targets of military action etc.) interacting with or affected by the development and/or use of an AI-enabled system **must** be clearly identified. An assessment **must** then be made of the impact the AI could have on each stakeholder group to ensure that effects are as positive as possible and justified as outweighing negative effects where these may arise.

53. Whilst conducting the assessment of any impact on humans, considerations *should* include, but not be limited to, the seven factors associated with Human Security (see JSP 985). Summarising JSP 985, these factors are that:

a. **Personal/Physical**: the potential for unnecessary physical harm *should* be minimised.

b. **Political**: the democratic values of the UK, where there is freedom from repression and the right for freedom of expression, *should* be upheld.

c. **Economic**: quality of life due to economic pressure *should* be maintained or enhanced where possible.

d. **Cultural/Community**: traditional relationships with cultural heritage *should* be maintained.

e. **Health**: illness *should* be prevented through maintenance of healthy lifestyle.

f. **Food**: physical and economic access to food that meets dietary needs *should* be maintained.

[10] Def Stan 00-056 [8] defines 'safe' as the 'freedom from unacceptable or intolerable levels of harm'.

g. **Environmental/Climate**[11]: the impact on the environment *should* be minimized whilst providing equitable access to natural resources or industrialization (for example, the energy consumption of LLMs is fairly high).

h. **Informational**: appropriate access to information that empowers the individual *should* be provided whilst not being manipulative or controlling.

54. Where the system has more than one mode of operation or 'level of autonomy' (see diagram on page 4 of the Defence AI Strategy [1]) the impact analysis **must** be conducted for all modes and 'autonomy levels'.

55. The concept of harmful effects is distinct from the intended military effects of certain capabilities. It is necessary to understand the factors set out in paragraph 53 in order to assess the military effectiveness of capability. Even when deploying a military effect it *should* be clearly demonstrated that the positive benefit of AI use outweighs any wider negative impacts factoring in the information available in the context of use.

Ethical Principles: Responsibility

56. In the ASR, it is made clear that, as unique moral agents, humans retain responsibility and accountability for the lawful and ethical use of AI in Defence. The behaviour of AI-enabled systems is not solely the responsibility of the operator or the duty holder. It is the responsibility of the governance chain to ensure that responsible persons have been clearly identified **at all the right levels** for the various contributing factors to the outcomes resulting from the entire AI lifecycle.

57. Responsibility is ensured through good governance (see Section 4). However, due to the nature of AI, and AI ethics, responsibilities are distributed throughout the AI design and use lifecycle. It is the responsibility of the governance chain to ensure that responsible persons have been clearly identified for the various contributing factors (e.g. the correctness of data used in machine learning) and their responsibilities clearly identified and agreed.

58. An overall articulation of where context-appropriate human involvement is being exercised and who is ultimately responsible and accountable for use of a system **must** be provided for all agreed contexts of use. Where contexts of use are changed, the manner of exertion of control **must** be re-examined as well as checks of continuing clear lines of responsibility and accountability.

Ethical Principles: Understanding

59. Understanding is driven by a combination of transparency and explainability.

60. The potential challenges around explainability of AI-enabled systems may vary depending on the AI system and its use context. Whilst designers and operators may have a deep understanding of how machine learning systems are designed and trained and how they make decisions based upon the weights and biases within the network, it is also true that it can be difficult or impossible to explain the way in which an individual decision has been reached in human-understandable terms. What matters is the effectiveness and performance of these systems. Therefore, Testing, Evaluation, Verification and Validation processes are critically important. While we may be unable to explain any individual

[11] JSP 816 and Def Stan 00-051 [12] provide MOD policy and requirements for environmental management in Defence systems.

operation, we can understand the overall level of risk associated with a system and its performance and can therefore make judgements on its effectiveness, safety and the risk-benefit of its use.

61. AI-enabled systems require a level of understanding that is sufficient for their responsible development and use. Stakeholder groups[12] that interact with the AI capability or its development **must** be identified, and their required interactions analysed to determine the level of understanding they require to engage with the system, or its development, responsibly.

62. Each stakeholder group will have a different perspective and not everyone will need to know every aspect of AI design or operation. Explanations *should*, therefore, be appropriate to the differing stakeholder needs.

63. Appropriate mechanisms *should* be put into place to permit an appropriate level of understanding of AI development and operation for each of the stakeholder groups throughout the AI development and use lifecycle. This *should* include awareness that systems contain AI components so that stakeholders can be alive to potential risks it may present.

64. Wider considerations such as security, privacy and intellectual property rights **must** be taken into account when providing insights to assist stakeholder understanding. Where full visibility cannot be provided for well-founded reasons, alternative strategies for achieving sufficient understanding *should* be implemented.

65. Where AI interacts with other systems (in particular where they also include AI); the AI behaviour *should* be understood in the system of systems context.

Ethical Principles: Bias and Harm Mitigation

66. Unintended bias is a common problem in AI and may result in unfair outcomes that can cause harms to groups or individuals, even where the AI output is used as the basis for human-based decision-making. Rather than being an intentionally malevolent act, discrimination often comes from subconscious biases being transferred from the designers, or skewed data sets, into the end product. This can be the case even when AI is designed to avoid human prejudices, and so an understanding of the unintended consequences of AI and how it may impact disparate groups of humans is particularly pertinent.

67. Defence organisations **must** set an open and inclusive culture so that multidisciplinary and diverse teams and people (regardless of rank/grade) feel safe to discuss potential issues with an AI system and take part in ethical risk assessment.

68. Harms include, but are not limited to, physical, psychological and discrimination against protected characteristics.

69. An analysis of data, AI learning algorithms and models **must** be made for unwanted bias that may lead to unintentional harms.

[12] Stakeholder groups may include, not exhaustively, designers, operators, end users, regulators, civilians, allied forces etc.

70. Where harms may arise, monitoring and mitigation strategies **must** be developed that include sufficient understanding of the AI behaviour.

71. Where bias is intentional, the ethical harms arising from it **must** still be considered and mitigated as appropriate.

Ethical Principles: Reliability

72. The ASR principle of reliability uses the term ‘reliability’ in its broadest sense.

73. There are three main focal areas for the principle of reliability:

a. **Reliable**. In this context, the ASR is referring to the correct operation of the AI. This does not mean that outputs are entirely predictable but that the behaviour meets the intended outcomes within acceptable performance criteria.

b. **Robust**. Robustness is the ability of the AI to handle inputs outside of its intended design and respond appropriately. It is unlikely that all possible inputs outside of those that are part of the intended design can be feasibly identified.

c. **Secure**. There are three key aspects to security: protection against loss of data due to non-adversarial activity; protection against adversarial action; and the way in which AI components interact with the broader system of systems having an impact on security. The traditional concepts of Confidentiality, Integrity and Availability apply across all aspects.

Alongside these, and related to reliability but not mentioned in the ASR, is the need for resilience and maintainability. Resilience is the quality of the AI to manage and recover from failure. Maintainability is the software’s amenity to error fixing and to updates in response to changes in the environment.

74. As with traditional software, AI cannot be expected to be reliable outside the operating context for which it is designed (its ODD) and furthermore, it cannot be guaranteed to produce error-free behaviour when inside its ODD. It is therefore essential for dependable AI that the risks associated with reliability are properly understood within the context of its ODD.

75. The operating context for the AI components **must** be clearly defined and communicated to relevant stakeholders such as risk-owners and operators. In addition to the risks being understood, mitigations or measures **must** be in place to constrain or bound the system’s behaviours such that the risks can be better quantified through the very use of such bounds or constraints. These constraints may be geospatial, temporal, or functional.

76. Performance targets - especially those relating to the concept of reliability - for AI components **must** be clearly defined for the operating context and demonstrated to a level of confidence that is commensurate with the risk associated with failure. Where AI is complementary to, or replacing, an existing approach, whether that is provided through software, human decision-making or a combination thereof, existing performance targets **must** be considered for continuing acceptability.

77. Where existing performance targets are deemed acceptable then the principle of demonstrating Globally At Least Equivalent (GALE) performance targets may be appropriate.

78. Appropriate response to reasonably expected inputs outside of the intended design **must** be defined and demonstrated.

79. Analysis for reasonable security threats and vulnerabilities **must** be carried out for the intended operating context. This *should* include the development environment, supply chains, data chains, and any other reasonable threat vectors.

80. Appropriate guards and mitigations **must** be put into place to provide an appropriate level of confidence that security is maintained throughout development and use of the AI capability. This *should* include all threat vectors identified in the security analysis.

81. The AI design, within the context of the system in which it operates, *should* minimise insofar as is reasonably practicable the adversarial attack surface. All remaining known vulnerabilities **must** be communicated to the risk owner for risk management action.

82. Analysis for potential effects of reasonable failure modes **must** be carried out and design mitigations put into place where possible. Where design mitigation is not possible, extant risks **must** be communicated to the risk owner. Additionally, as technology advances and AI becomes ubiquitous, AI-based Digital Systems and RAS will become increasingly interconnected. The potential for inter-system emergent effects and cascaded failures **must** be considered.

Research and Development Ethics

83. The AI development processes conducted across the wider systems engineering life cycles might involve: research trials; experiments; tests; surveys; or other forms of assessment or data collection involving human participants. In such cases, the research activities **must** comply with JSP 536 (Defence Research Involving Human Participants).

84. In line with the principles of the Concordat to Support Research Integrity (see JSP 732), where MOD funded research makes use of AI to achieve research outcomes (for example, the use of Large Language Models to conduct research analysis) these *should* be clearly identified. This *should* include an analysis and communication of the risk to scientific rigour.

85. The MOD Ethical Principles set out in the ASR and further elaborated on in this JSP

must be applied regardless of the AI technology readiness level. This will require early analysis and subsequent management of any potential AI-specific ethical risk.

AI Ethical Risk Assessment and Management

86. An AI ethical risk assessment that addresses the Ethical Principles and potential harms **must** be used at the outset of a project or programme and at any points where material changes to scope or outputs suggest changes to the overall risk profile. The assessment will determine the appropriate approvals pathway for AI-enabled projects or programmes. A guidance framework for the ethical risk assessment will be provided in Part 2 of this JSP.

87. Senior leaders responsible for overseeing development and operation of AI -enabled systems or capabilities (front-line or back office) **must** review AI ethical risks and maintain adequate and proportionate evidence of any related risk management decisions, supported by evidence and advice from suitably qualified and experienced personnel.

These steps will provide assurance to Defence Ministers – and thereby to Parliament – that any use of AI technologies within Defence is safe, responsible and policy compliant.

88. Each AI use case **must** be given an overall risk rating determining the level of approvals necessary as per Table 1 below. The overall risk rating should be calculated by assessing both the impact and likelihood of a risk, and the level of approval should be based on the *residual risk* (i.e. the severity of a risk if controls and mitigations to manage it are in place and working as intended) rather than inherent or target risk levels. Note, a programme may therefore need to be submitted to the Joint Requirements Oversight Committee / Investments Approvals Committee (JROC/IAC) or even Ministers, even where it would otherwise fall below the reporting and approvals thresholds set out in JSP 892 (financial risk, impact on outputs etc).

89. Additionally, extant MOD policy dictates that some applications of AI merit special attention. These are:

a. AI in kinetic effects.

b. AI in novel and contentious applications.

90. Special attention applications and any AI-enabled projects and programmes **must** be referred to the DAU - notifying the DAIC - where the level of ethical risk warrants top-level departmental ownership in line with Table 1 below. This includes particularly novel or contentious use cases that are identified using the framework that will be provided in Part 2 of this JSP.

AI Ethical Risk Rating (Impact and Likelihood criteria assessed together)	**Referral Level for Approval**
Critical	2PUS, or Ministers by exception.
Severe	
Major	Defence-Level Oversight: e.g. through the JROC/IAC
Moderate	TLB-Level Oversight, with delegation to business level processes as appropriate.
Minor	

Table 1: Ethical risk referral levels for approval

91. In cases where an AI system presents unacceptable negative ethical risks (i.e. where significant negative impacts are imminent, severe harms are actually occurring, or catastrophic risks are present) deployment or development **must** be halted in a safe manner until risks can be sufficiently managed. For example, if an operational system is found to be behaving in a way that is outside the acceptable bounds (including on ethical grounds), then it must be taken out of use until reviewed at the appropriate level.

92. Different actors may have different perspectives on ethical risks, potential impacts may not be easily foreseeable, and risks may change throughout the AI lifecycle as latent risks emerge or as AI systems adapt and evolve. Risk management must therefore be carried out throughout the lifecycle. This can be achieved through traditional risk management approaches whilst applying the AI ethical risk assessment overlay, including through effects review and re-assessment as necessary.

93. For Defence applications there are times where AI ethical risks may require trade-offs with operational effectiveness and military requirements. The Responsible AI Senior Officer (RAISO) must ensure that the risks of such trade-offs are identified and managed.

94. *To provide initial guidance on ethical risk assessment and management, the DAU has produced a range of Minimum Viable Products (MVPs). These include more detailed guidance on conducting AI Ethics Risk Reviews and mechanisms for tracking risk, as well as a set of assurance questions structured by AI lifecycle stage. These MVPs will be iterated and tested before they are formally integrated in the forthcoming part 2 of this JSP.*

Communication of AI Ethics

95. Critically, the ASR notes that Defence **must** not only behave ethically, it **must** also be seen to be ethical. This means that evidence supporting the ethical development and use of AI **must** be developed and communicated appropriately, including as much transparency as possible within the security constraints of Defence activity.

4 AI Ethics Governance

MOD Governance of AI

96. There are well-established roles and responsibilities for the governance of software products developed and used across the MOD (for example, but not limited to, as set out in JSP 939 and [extant internal policy]). These do not change simply because the software includes AI. However, AI as outlined in the ASR requires additional governance due to societal concerns associated with its responsible use in Defence applications; for example, systems that achieve kinetic effect or systems with direct implications for people (e.g. Human Resources software).

97. Several of the ethical principles in the ASR explicitly mention roles and responsibilities associated with ethics governance. The nature of these will vary depending on the legal and ethical risk associated with specific AI development and operational use.

98. Figure 3 provides an overview of the AI Ethics Governance structure within MOD.

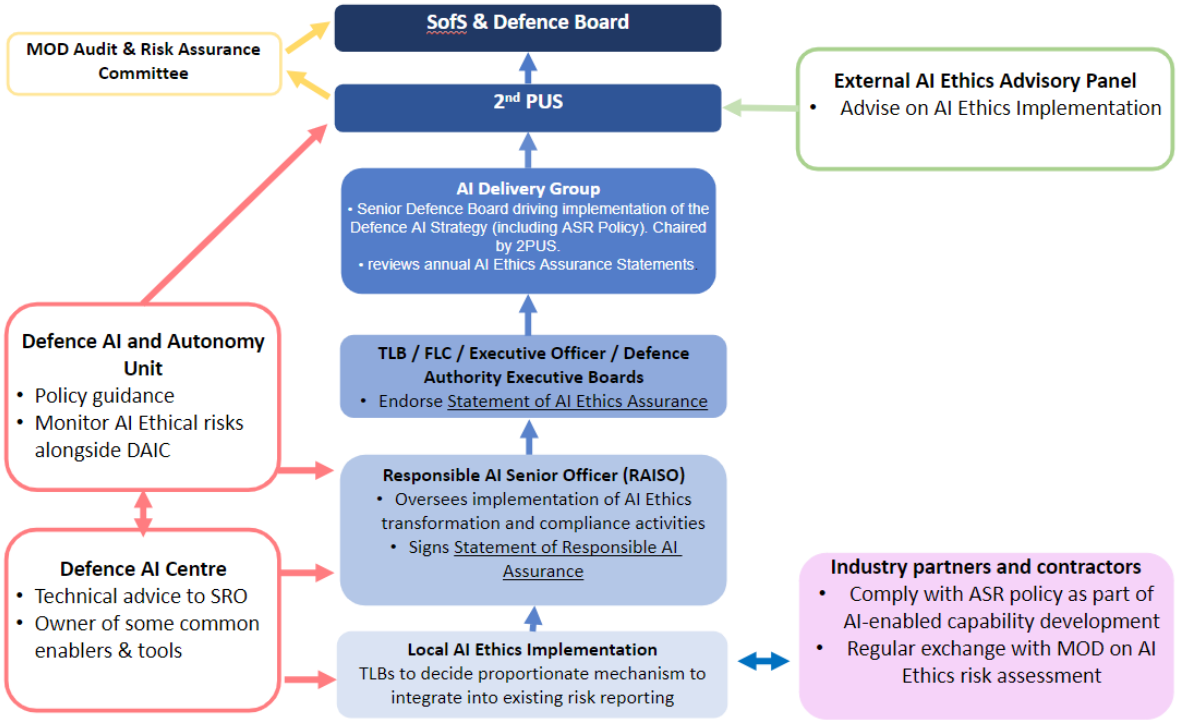

Figure 3: Diagram of MOD Governance Structure

99. TLB Executive Boards **must** provide Statements of AI Ethical Assurance to 2PUS on an annual basis. These **must** be underpinned by appropriate, auditable evidence.

100. Each TLB **must** appoint a RAISO. In practice, allocating this responsibility to the right person will look slightly different for each Defence organisation and will involve the head of the organisation (or senior duty holder) and the delivery/operating duty holders. More detail on the roles and responsibilities of the RAISO and the delegated responsibilities within the governance chain will be included in Part 2 of this JSP – please contact the DAU for further guidance if required.

101. The RAISO **must** ensure that appropriate assurance for the ethical and responsible development and use of AI is in place across all AI-based projects under the scope of their accountability – this entails setting the right culture within their organisation to enable everyone to develop/use AI responsibly.

102. Where AI is transferred, for example from development to operational use, the incumbent RAISO **must** ensure that accountability mechanisms in the receiving organisations are properly established. The new RAISO **must** understand and accept the risk of using the product.

103. Due to the variability of AI development, acquisition, and use etc. across the AI lifecycle in the various Defence organisations, it is not possible to be prescriptive on the roles and responsibilities required to meet the requirements of this JSP. However, the AI RAISO **must** put into place context approporiate roles with clearly defined responsibilities and competencies. This will require subject matter expert led analysis to ensure that organisations are ‘AI-ready’.

104. All projects developing or using AI **must** identify their organisation’s RAISO. Senior Responsible Owners of projects containing or making use of AI that are medium or high risk **must** inform the RAISO of their project or system. RAISOs **must** ensure their organisation has governance mechanisms in place which capture risks arising from other AI projects.

105. During implementation of large projects, and projects where ethics risk associated with the AI is significant or where teams exist to produce AI on a regular basis, there *should* be at least an Ethics Manager and Independent Ethics Assurance mechanism. These roles *should* also consider the wider aspects of the ASR such as checking that appropriate legal advice has been sought.

106. The Ethics Manager role *should* work in coordination with roles that support ethical AI across the AI lifecycle such as Quality Assurance Managers and Safety Managers. In practice, provided appropriate competence and independence can be demonstrated, the Ethics Manager role may be filled by someone filling another role in the organisation; one of the aligned roles (e.g. safety or quality) may be most suited.

107. The DAU has responsibility for setting AI policy in relation to ethics. They **must** be consulted before engaging in external communications relating to ethically sensitive work.

108. Defence Legal Services (Navy, Army and RAF) and MOD Legal Advisers (MODLA) are responsible for the provision of legal advice. Where necessary, legal advisers **must** be consulted for guidance on the legal aspects of AI development and use (see paragraph 46).

Governance of Non-Sovereign AI Development and Use

109. With AI changing the global defence and security landscape we need to champion interoperability and coordination across allied nations. Ambitious, safe and responsible AI development is crucial to developing trust and assurance across all stakeholders, including these international partners. Processes that increase the trust in AI across allies by ensuring both an appropriate level of understanding of the technology and assurance that the technology has been developed in a manner that is reliable, safe, ethical and legal are vital. They help to eliminate uncertainty and hesitancy enabling decision-makers to leverage the AI tools at their disposal and move faster. An example is responsible AI processes increase the assurance between allies in NATO. Our UK MOD AI ethical principles are closely aligned with and complementary to the NATO Principles of Responsible Use for AI in Defence.

110. Details on how to ensure ethical and technical interoperability with our partners is a live discussion. The MOD is also engaging with a wide range of international bodies, partners, and stakeholders to promote our approach to responsible military AI and champion global norms and standards for the safe development and use of these technologies. For example, we actively support NATO's Data and Artificial Intelligence Review Board (DARB), which is developing a Responsible AI (RAI) Certification Standard and best practice risk management approaches. Outcomes of this and other fora will inform JSP Part 2 and future iterations of Part 1.

111. AI (either the AI component itself or the outputs from AI), including related data, may be shared amongst international partners. In such circumstances, the partner nations may have either developed their AI or use UK-developed AI in ways that may be incompatible with UK policy.

112. When UK-developed AI and/or related data is to be shared with international partners, whether this is as part of a coalition environment or Defence export, the RAISO *should* satisfy themselves as far as possible that its use would be in line with the ASR policy.

113. Where the UK is to make use of non-UK developed AI and/or related data (including where the AI will be subject to additional UK AI training activity) the RAISO (through their relevant delegated persons at the local implementation level) **must** satisfy themselves that the AI and/or data meets the UK ASR policy. This includes provision of a Statement of AI Ethics Assurance.

114. Defence will always apply the relevant UK legal and ethical framework to AI capability it uses abroad. In addition, as is appropriate to the operational context, AI and related data that is to be used in non-UK contexts *should,* so far as is reasonably practicable, be compatible with the legal and ethical considerations of the nation in which it is to be applied.

5 Human/AI Teams

Introduction

115. The MOD recognises the advantages that the teaming of humans with AI brings are central to the application of AI across Defence (see [1], [2], [9]). These advantages include enhancement of overall effectiveness, optimal use of resources, the practicalities of

integration and the ease with which we can address issues arising. Fully realising the benefits of AI depends on understanding the relative strengths of humans and machines, and how they best function in combination to achieve the desired outcomes within a particular context of use. This human involvement and teaming approach extends beyond the individual operator interacting with an AI-enabled system to include the wider team of people involved in supporting, training and maintaining the system.

116. As systems become increasingly interconnected, individual human/AI teams are likely to expand and become teams of human/AI teams extending across the Defence enterprise. The organisational consequences of ubiquitous AI will become increasingly important together with the critical role that humans play in supporting resilience, safety and maintaining human accountability.

117. The importance of addressing these human factors applies across all applications of AI from combat systems through to administrative and support systems.

118. Delivering effective human/AI teams is dependent on adopting a Human Centred Design (HCD) approach across the system lifecycle. HCD focusses on identifying user needs, involving users in the design and testing of a developing system solution and applying human factors best practice. A well-managed HCD approach will support projects in optimising system performance, reducing risk, enhancing cost-effectiveness and supporting user adoption of new systems.

Human Centred AI Design

119. A recognised HCD approach, appropriate to the system under development, **must** be applied across the system lifecycle, including introduction into service and in-service updates.

120. The HCD approach *should* be appropriately resourced, managed and integrated within the wider software development process.

121. For the development and acquisition of military capability the design approach **must** apply the Human Factors Integration (HFI) process mandated in JSP 912[13].

122. The through life HCD approach adopted **must** include the following activities:

a. identification of the users[14] of the system and understanding their characteristics.

b. identification of user needs and the development of Human Factors' related requirements with associated acceptance criteria for inclusion in project documentation.

c. involvement of users in the development and testing of the system solution; this includes software updates and testing of systems following training/retraining of AI systems.

[13] Note that JSP 912, associated Human Factors Integration Defence Standard 00-251 [11] and Technical Guides do not specifically address AI technologies.
[14] 'Users' includes: operators, maintainers, support personel and people that come into contact with or are affected by the system.

123. Alongside the involvement of users in the development and testing of the system's human factors, insights from human sciences *should* also be applied to the system solution. This *should* include:

a. the application of established Human Factors principles and accepted best practice to the design of the system.

b. the use of suitable methods, tools, techniques and data by projects to support design.

c. the application of principles of transparency, explainability and interpretability to the design of the system and the development of user interfaces that enable the user to understand system behaviour.

124. A key part of the design process of all systems, but one that is critical to those intended to be used within a human/AI teaming approach, is functional analysis and allocation of functions between human and machine.

125. An analysis of the allocation of functions between human and AI agents and AI behaviours across all modes of function and levels of autonomy **must** be conducted to:

a. provide evidence of compliance with the ethical principles.

b. ensure that there is no accountability gap, i.e. humans remain accountable for, and in control of, the effects of the system.

c. understand the roles of humans in ensuring system safety, performance, resilience, and preventing AI bias and how the design of the system supports this.

d. provide evidence to support broader legal and regulatory compliance arguments (e.g. safety cases and legal reviews).

People Implications of AI Technologies

126. AI technologies have the potential to change, in some cases significantly, the nature of tasks and roles currently undertaken by humans; augmenting rather than replacing them.

127. The implications of AI-based systems on the workforce *should* be considered from the outset of a project and continue to be re-evaluated as deeper understanding of the impact develops over time. These can include, but not be limited to, the Ethical Principle assessments detailed in Section 3.

128. Workforce implications such as: numbers and locations of personnel, positive or negative impact on safety, organisational structures and changes in the Knowledge, Skills, Experience and other characteristics, as identified on a context-by-context basis, required by personnel operating, supporting, maintaining and otherwise involved in the employment of a system *should* be identified and planned for.

Training Implications of AI Technologies

129. Appropriate training for users of systems utilising AI technologies is critical, not just to develop proficiency in system operation, but also to support users (including those responsible for making decisions regarding the employment of AI enabled systems) to

understand system behaviour, performance, limitations and in order to calibrate their trust in the system under different use cases and conditions. It is essential, therefore, that training addresses a wide variety of scenarios including edge cases that stress the human machine team and trigger reversionary ways of working.

130. Training Needs Analysis for users of AI-based systems *should* include consideration of the users' need to develop an understanding of system behaviour, performance and limitations and calibrate their trust in the system under different use cases and conditions.

131. To produce the most effective human/AI teams, collaborative training *should* also be considered where the humans and AI both learn from each other's behaviours. Where this occurs, assurance of the overall behaviour for each team **must** be provided on a case-by-case basis.

6 AI Lifecycles

Introduction

132. There are a number of AI development and use lifecycle concepts with the most common being aligned to the DevOps (a conjunction of Development and Operations) philosophy. DevOps seeks to speed up software releases through a continual highly automated process. The continuous nature of DevOps can make them problematic for dependable systems because technical debt[15] often arises due to planning and coding already being in progress whilst test and deployment monitoring are ongoing for earlier releases.

133. There are a number of variations on the DevOps, for example DevSecOps, which push security characteristics into the lifecycle. Of particular relevance to the development of AI (ML specifically) is the variant known as MLOps (Machine Learning + DevOps); see Figure 4. In the MLOps lifecycle, algorithms are created that then have data applied to them (training) to produce models that are then verified. All of the model elements are then packaged and released for deployment. Once deployed, the model will have input data applied and outputs inferred based on the functional approximation within the model. The performance of the model is monitored with new data often collected ready for the next cycle. This monitoring phase can be considered as continual validation of the ML in the context of use.

[15] Technical debt is where choices to develop technology at speed leads to gradual reductions in design quality. The longer the debt is accrued, the more difficult and expensive it is to fix and higher the likelihood of failure becomes.

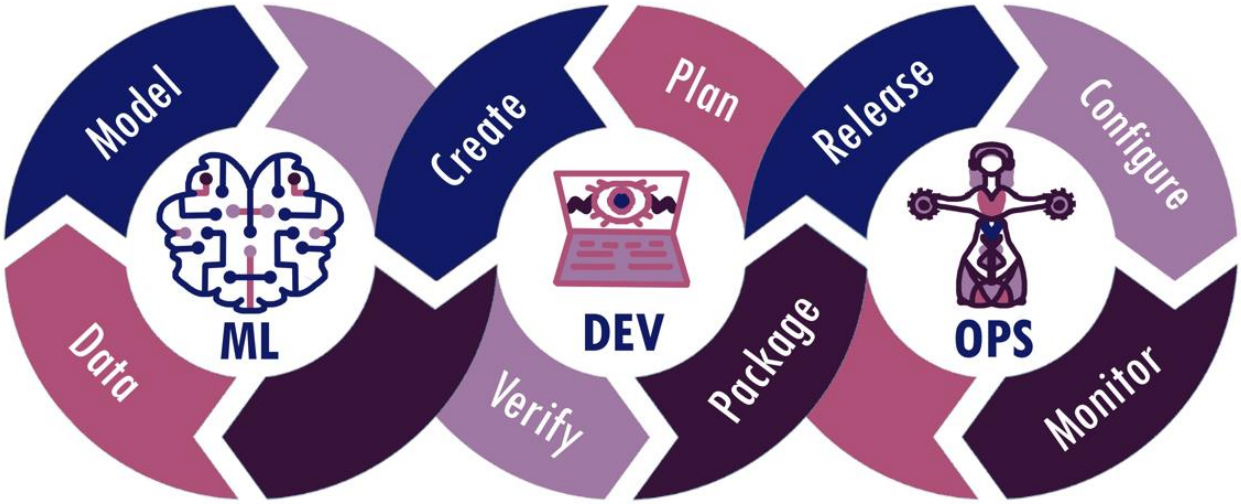

Figure 4: MLOps lifecycle

134. This JSP does not require, or advocate, any one lifecycle over any others that may exist. A lifecycle that is appropriate for a digital system may not be appropriate for RAS. The lifecycle adopted *should* be appropriate to the context that the AI will reside. Guidance on AI Lifecycles will be provided in Part 2. Lifecycle agnostic requirements are provided as follows.

Planning

135. The purpose of planning in the context of dependable AI is to support effective management of the AI lifecycle and provide confidence that the activities undertaken and artefacts produced will result in a product that meets its functional and non-functional requirements. It *should* also meet the delivery timescale needs of the AI operational context. Once it is confirmed that AI is the right solution for the problem, typical objectives for the planning stage include:

a. ensuring that the algorithm development, data (availability, relevance and manipulation), model management and other processes (e.g. quality assurance, configuration management etc) will meet the system-level requirements placed on the AI.

b. ensuring that transitions between lifecycle phases are properly controlled with timely feedback and resolution of errors.

c. clear definition of tools and development environments to be used and how confidence in their performance will be generated.

d. definition of recognised good practice standards to be applied to all areas of the lifecycle (including coding standards, data standards, test standards etc.).

e. the means of agreeing changes, deviations and waivers to plans that have approved for use.

f. controlled release and integration into wider systems. Where wider system requirements include the expectation that AI will be modified in-situ (e.g. through

online learning) then planning *should* include how this will be controlled and assured for continuing confidence.

g. how through-life monitoring of performance will be achieved and reported.

h. how obsolescence of hardware and software (including data) will be managed.

i. assessment and reduction of environmental impacts across the AI lifecycle.

136. Version controlled planning documents *should* be produced and agreed in a timely manner to minimise the risk of the AI being unable to meet its functional and non-functional requirements. The scope of the plans *should* include tools, models and simulations used to create algorithms, data and/or develop assurance evidence.

137. Planning documents **must** include steps throughout the development and use lifecycle to ensure that the ethical principles can be met.

Requirements

138. Traditionally, there is an expectation that requirements allocated to software are maintained with bi-directional traceability from the System-Level Requirements (SLR) down through the High-Level software Requirements (HLR), into Low-Level software Requirements (LLR) and onto the implementation and verification. In the broadest sense, the HLR can be regarded as outlining 'what the software is supposed to do to achieve the SLR apportioned to it' and the LLR details 'how the software will achieve the behaviour required by the system'.

139. In AI, particularly ML, the traceability in both directions is often lost at the LLR stage. Consequently, it is difficult to ensure that the system requirements passed to the software have been met. This, along with the inherently unpredictable outputs from AI components noted earlier, increases risk; reinforcing the need for recorded and agreed risk balance arguments about the use of AI versus other approaches in major and higher risk systems.

140. Performance requirements for the AI *should* be clearly stated alongside functional requirements.

141. A Hazard Analysis[16] **must** be undertaken to identify hazards introduced through the use of AI; this *should* include data failure modes where machine learning is adopted.

142. Any new hazards introduced through the adoption of AI **must** be passed back up to system-level safety processes for mitigation as derived requirements.

143. Where HLR are implemented through AI, requirements that are decomposable directly into LLR *should* be documented and traceable as for traditional software implementations.

144. Where HLR behaviours are to be implemented through AI and are not directly decomposable into LLR (i.e. those that are incorporated through a learning process), the combination of the training algorithm and data requirements **must** be demonstrated as meeting the intent of the HLR.

[16] Including safety and security hazards.

145. It is important that the operating context is fully addressed by the AI behaviours. The HLR **must** include clear articulations of the operating context including data distributions for training and test data.

146. Where analysis of the data requirements reveals potential for sparse or missing data, requirements and guidance for subsequent mitigation across the lifecycle *should* be provided.

147. Data requirements for machine learning approaches *should* include how set aside data will be curated for sufficiently independent verification confidence.

148. All requirements *should* include information on how they will be demonstrated through verification and validation activities. This information *should* also include meaningful measures of acceptable performance in the test and operational environment.

Architecture

149. The scope of the architecture in the context of this JSP relates to the AI algorithms, data and models.

150. The AI architecture **must** be traceable to requirements and be able to incorporate the intended behaviours whilst protecting against entry into failure modes identified during hazard analysis (e.g. increased potential for overfitting or underfitting data, handling sparse or missing data etc.). Consideration of this clause *should* include mitigations for AI failure put into the wider system integration and architecture.

Algorithm Design

151. Choices made during algorithm design (e.g. hyperparameter settings) *should* be justified and documented including tracing to functional and non-functional requirements.

152. Measures of performance that are used to optimise AI designs *should* be justified and documented.

153. It can be tempting to optimise AI design to maximise performance against training and test data. However, this risks limitations in data quality and availability driving operational performance. Evidence *should* be provided that algorithm design is optimised for the performance in the expected operational context; for example by demonstrating the avoidance of overfitting.

Algorithm Implementation

154. Traceability to requirements *should* be maintained in the algorithm implementation.

155. All frameworks and/or libraries that are utilised to develop algorithms **must** be justified and their output checked and demonstrated as instantiating the AI requirements.

Machine Learning Data Collection, Preparation and Control

156. Data required for ML can be classed as either:

a. **Training Data**. This is applied to the ML algorithms to produce an ML model;

b. **Test Data**. Used by the development team to test when the model has been trained sufficiently to achieve the intended performance; or

c. **Verification Data**. Used by a separate verification team to independently verify the model performance.

157. Incorrect data could impact on all of the MOD AI Ethical Principles; for example it could introduce biases (see the Bias and Harm Mitigation principle) if not selected and prepared carefully. Data *should* be demonstrated as correct through having the following properties[17]:

a. **Relevant**. Data *should* be validated as being relevant to the intended behaviour in the ODD.

b. **Complete**. Data *should* be drawn from across the potential input space taking into account real-world distributions and features.

c. **Balanced**. Consideration *should* be given to the relationship between the AI and the data to ensure that data is balanced from the perspective of the model it is being used to produce. For example, rare classes in the real-world might require over representation in the data to ensure that they are properly classified when encountered by the model during operational use.

d. **Accurate**. The data *should* be a sufficiently accurate reflection of the real-world application. There are several aspects to this - for example, that training data is drawn from sensors with the same characteristics of sensors on the target system. A further example is the accuracy with which class objects are bounded and labelled.

158. Separate training, test and verification data **must** be produced. These *should* be drawn from context relevant sources and maintained independently from each other.

159. When single datasets have been separated into training, test and verification data the resulting subsets *should* be verified as having the same characteristics as the original dataset.

160. Where possible, verification data *should* be collected separately from the training data. It *should* address the breadth of the ODD including edge cases that sit inside, on, and beyond the boundary conditions.

161. Data provenance **must** be assured and recorded from the point of collection to ingestion by the AI.

162. All data (including data metadata) that influences the output of the AI **must** be kept under configuration management controls that are commensurate with those applied to the AI. Special attention *should* be paid to the classification, complexity, volume and velocity of data as these may drive specialist data management solutions.

[17] Drawn from: Ashmore, R., Calinescu, R. and Paterson, C. 2021. Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges. ACM Comput. Surv., Vol. 54, No. 5, Article 111, Publication date: May 2021.

163. Where synthetic data is used to either supplement or even in place of real data, care *should* be adopted to validate it from the perspective of the AI and how the AI operates on it. This is necessary because AI can cue on different features of the data than a human.

Model Development

164. AI models are the combination of algorithms and data. As for all models, they represent an abstraction of the entity/phenomena they reflect.

165. All models developed for real-world applications **must** be accompanied by relevant information to permit their risk informed use, re-use and further development. This *should* include any assumptions, dependencies, known errors and weaknesses. As an example, information on the ODD for which they have been designed to operate will be critical to applying appropriate AI models within a given context of use.

166. Prior to re-use of existing models (this includes where models are re-trained), the risk owner **must** satisfy themselves that they have appropriate access to relevant information on the models such that they understand the risk associated with their usage. Again, this *should* include any assumptions, dependencies, known errors and weaknesses, information on the ODD etc.

167. All models *should* be designed for context appropriate interpretability. This means they *should* be transparent, include appropriate explanations of their output and provide measures of uncertainty that are understandable to the various stakeholders. This may mean providing different views to support differing stakeholder needs.

168. Where models are developed using security classified data, the resulting models **must** assume at least the same classification. Consideration **must** be given to the potential for increased classification due to the aggregation of data upon which the model is built.

AI Verification and Validation

169. Verification demonstrates that the AI meets its requirements; validation demonstrates that the AI meets the user's needs. This is sometimes referred to as 'did we build it right (verification) and did we build the right thing (validation)'.

170. Verification and validation of the AI **must** include demonstration across the ODD including boundary conditions and realistic edge cases.

171. Safe behaviours of the AI when exposed to inputs that fall outside of the ODD **must** be demonstrated.

172. Where AI is updated, for example through exposure to new training data, verification and validation activity **must** include testing for both existing and new behaviours (e.g. through new test cases and the application of regression testing). This is to provide assurance that the intended outcomes of the update have been achieved and unwanted emergent effects such as catastrophic forgetting and model drift have not occurred.

173. Evidence *should* be provided to support arguments for the validity of the claims being made for verification and validation sufficiency. This *should* include, for example, metrics on test coverage of the ODD, explicit and inferred requirements, internal algorithm/model structural coverage, performance metrics within the tested context etc.

AI Integration, Use and Modification

174. AI will always operate within some wider system or system of systems context, which may include other AI. Care *should* be taken to ensure that the AI will perform to its operational requirements in its context of use. This may require an assessment of the AI impact on broader operational governance requirements, such as operational authorities and Rules of Engagement.

175. Where performance is demonstrated in a system environment other than the final operational system, differences between the environments **must** be analysed, understood and mitigated. This includes where AI is transferred from one operational system to another with differing build standards (e.g. different computational hardware or system inputs).

176. When substantive modifications are made to an existing AI-hosting system continued satisfactory performance **must** be demonstrated. Substantive changes might include, but are not limited to, processors and operating systems etc.

177. The approved ODD, i.e. the environment in which the AI has been demonstrated as achieving acceptable performance, **must** be clearly defined in the operating context.

178. The management and use of AI may present particular operating risks. These *should* be considered in wider operating policies and procedures; for example, where necessary AI products **must** have specific and clearly written Security Operating Procedures (SyOPs) that are proportionate to the risk presented by their use.

179. The operational environment **must** be monitored at a contextually appropriate rate to ensure that the AI-based system continues to operate within the ODD of the AI.

180. AI performance **must** be monitored at an appropriate rate to provide assurance that it remains within an acceptable level with respect to the operational requirements placed on the system during use. In cases where AI is supporting important/risky decision-making, its output *should* have additional checks applied that are undertaken by a relevant subject matter expert.

181. Where AI is intended for modification, processes and procedures **must** be in place to ensure the continued assurance of acceptable behaviour. These *should* include steps to ensure that changes to risk are properly understood and managed appropriately.

MOD Staff Competencies

182. Competency is context specific and developed through appropriate training and experience. AI development and use is a complex undertaking and developing sufficient competency is a significant undertaking. The competence of all MOD personnel engaged in the AI lifecycle is essential to achieve its safe, responsible and effective application to Defence problems. At the time of this JSP publication there are no set competency profiles set out for any AI-specific roles in the MOD. Indeed, most roles associated with AI already exist for extant systems and as such those roles may need their competencies re-assessing for the introduction of AI.

183. It is the responsibility of leadership across all levels of Defence organisations in which AI is developed or used to ensure that staff maintain appropriate competencies.

184. All roles associated with AI development and use *should* be analysed for appropriate competency requirements.

185. Appropriate training and supervision **must** be provided to all staff until sufficient competence is acquired.

186. Where in-house competence is insufficient, support from external organisations **must** be incorporated with the aim of supervising MOD staff and developing MOD competence.

7 Quality, Safety and Security

Quality

187. The MOD policy for Quality (JSP 940) **must** be applied to AI product development and use.

188. Quality planning and subsequent activities **must** be commensurate with the risk and maturity of the AI. It *should* include cognisance of: the pivotal role of data; how users interact with the system; and the dynamic nature of AI, its ability to adapt to its environment and its rate of change.

189. Quality assurance activities **must** include broader aspects of AI such as compliance with configuration management, safety, security and ethical requirements.

Safety

190. JSP 815 provides the Defence Safety Management System (SMS) Framework of goals and guidance for SMS development and implementation. Evidence-based safety management is a fundamental principle of the JSP. For safety-related AI-based systems there **must** be sufficient evidence to support a meaningful SMS. Due to the size and complexity of many AI applications and their data sets such evidence may need to be collected and managed from the earliest stages of development.

191. AI may have unique safety risks associated with its development or behaviour. These **must** be analysed and included in the relevant wider safety cases and software and system risk assessments (see JSP 375 and JSP 376).

192. Safety-related risk assessments **must** be carried out early in the AI lifecycle. All risks identified **must** be managed in a coherent and proportionate way throughout the AI lifecycle. Where appropriate, this will include demonstration of compliance with Def Stan 00-056 [8] for system safety and Def Stan 00-055 [7] for software safety.

193. Risk analyses *should* include reasonably predictable behaviours of the entities with which the AI interacts in both benign and adversarial environments. This *should* include operating excursions beyond the ODD.

194. JSP 892 **must** be applied in the management of AI risk; the consequences of failure *should* reflect both direct impacts and wider societal concerns regarding the development and use of the technology. The activities undertaken to manage risk **must** be capable of keeping pace with the rate of change typically seen in AI-based system development and use.

Security

195. A Secure by Design approach is required for the definition, acquisition, development, maintenance, and disposal of information-based capabilities for the MOD as set out in [extant internal policy]; this includes AI. However, traditional approaches to security risk management may not be satisfactory, due to the rapidly evolving nature of AI. The focus for any organisation should be continuous assurance of system security, ensuring that their system is continually monitored, hardened and improved. It may no longer be enough to build a secure system: the system may need to be self-defending and self-monitoring, with minimal manual interference.

196. As for safety, AI may have unique security risks associated with its development or behaviour (for example, data poisoning). These **must** be analysed from the earliest practicable opportunity and included in the security aspects of relevant wider safety cases and software and system risk assessments (see JSP 375, JSP 376 and [extant internal policy]).

197. Security analyses of the AI and the systems in which it operates **must** include both conventional security risks and the potential for adversaries to interfere with the AI performance; e.g. loss of privacy through data breaches or through causing adversarial behaviour in which the AI is induced to provide an incorrect output due to a misleading input. This **must** include security of the data used to develop the AI since poisoning of data can lead to unsafe behaviour during operation.

8 Suppliers

198. As with all dependable systems technologies, AI that is acquired externally **must** attract the same level of confidence that the requirements of the JSP have been met as that developed within, or for, the MOD. With AI, this can be more challenging than is the case for traditional software and MOD teams may have to stand up additional assurance capabilities to address evidence shortfalls should they arise.

199. Many suppliers operating in the AI space are new to Defence. For suppliers delivering directly into Defence organisations, it is essential that they understand the Defence context; for example: information and operational security; MOD policies (including the ASR and Equality, Diversity and Inclusion, as set out in JSP 887) and standards; and operational contexts (where appropriate).

200. MOD teams contracting suppliers of AI **must** require them to demonstrate competence to deliver products to a level of confidence that is commensurate with their use. For example, if an ML model is to be used in a safety-related application then the supplier must be required to demonstrate they have the appropriate people, processes and experience to develop the product and evidence to support an appropriate level of confidence in the associated safety argument.

201. Due to the complexity of AI, most developers make use of commercially available (including open source) software applications, tools, libraries, frameworks and data. The MOD recognises that there are challenges with assurance to the same degree as bespoke software. Where suppliers make use of such third-party software they **must** be required to develop compelling arguments as to why confidence in the resulting AI has not been undermined.

202. The provenance for all data used in the training and testing of the AI, including legal and ethical compliance in line with MOD policy, **must** be assured and documented. This is to ensure that data is fit for purpose, does not contain adversarially introduced vulnerabilities and meets MOD legal and ethical obligations. This includes data supplied by MOD and third-party data.

203. When contracting for AI products, contracting conditions *should* consider through-life support including access to data and algorithms as well as the resulting models. This *should* include clear ownership of intellectual property for data, algorithms, models and tests. Contracting *should* also address the need for stakeholders to understand AI (see Section 3 - Legal & Ethical Considerations of AI).

204. Assessment of restrictions caused by foreign export controls *should* include data, as well as algorithms and models, to ensure that MOD has the necessary Freedom of Action to maintain, modify, upgrade and operate the AI.

9 AI Assurance

205. Assurance of AI shares many common features with its 'traditional software' counterpart. Traditional approaches *should* continue to be applied with this section providing an overlay of requirements that address the novel aspects of AI applications.

206. Given the breadth of technologies and development approaches, it is neither possible nor appropriate to be prescriptive on resolving aspects of 'traditional software' and AI assurance differences. Typically these differences may be as a result of imprecise requirements, reduced requirement traceability, excessive complexity, unpredictable specific behaviour and adaptive behaviour. This can be thought of as 'assurance risk', i.e. the risk associated with the assurance process itself. In some applications these issues will simply present too much risk and in such cases AI **must** not be adopted. Where 'assurance risks' can be tolerated, assurance activities still have to be undertaken that are commensurate with the level of risk posed by incorrect AI outputs.

207. AI assurance *should* be considered within the context of the system in which it operates. Some systems may be able to tolerate a level of incorrect specific outputs provided the overall intended outcome is acceptable (e.g. safe, secure, ethical etc.). This may be a key difference between traditional software assurance and that for AI. Without its accommodation, the potential gains of AI may be lost unnecessarily.

208. Where AI is replacing extant technology or human decision-making, AI assurance may be able to argue that the specific risk from the AI is no greater than the system being replaced. That is, for example, if AI is being used to plan vehicle trajectories (e.g. route planning) that were previously produced via human thought, then the risk of failure should be no greater than when the human planned them.

209. AI assurance **must** include assurance of behaviour where excursions from the AI ODD may reasonably be expected to occur. For many Defence applications this may include adversarial action causes. In such cases careful consideration **must** be made with respect to failure mode behaviour (e.g. options to 'fail safe' or 'fail operational'). The level of confidence sought in the assurance of the AI *should* be commensurate with the acceptable ethical, safety, security and mission risk associated with the function provided by the AI.

10 References

[1] Ministry of Defence, ‘Defence Artificial Intelligence Strategy’, 2022.

[2] Ministry of Defence, ‘Ambitious, Safe, Responsible: Our Approach to the Delivery of AI-enabled Capability in Defence’, 2022.

[3] Ministry of Defence, ‘Data Strategy for Defence. Delivering the Defence Data Framework and Exploiting the Power of Data’, 2021.

[4] Ministry of Defence, ‘Digital Strategy for Defence. Delivering the Digital Backbone and Unleashing the Power of Defence's Data’, 2021.

[5] UK Government, ‘National AI Strategy’, 2021.

[6] Department for Business, Energy & Industrial Strategy, ‘National Security and Investment: Sectors in Scope of Mandatory Regime’, 2020.

[7] Ministry of Defence, ‘Defence Standard 00-055: Requirements for Safety of Programmable Elements (PE) in Defence Systems’, 2021.

[8] Ministry of Defence, ‘Defence Standard 00-056: Safety Management Requirements for Defence Systems’, 2023.

[9] NATO, ‘Science and Technology Trends 2020-2040, Exploring the S&T Edge.’, NATO, 2020.

[10] Ministry of Defence, ‘Human-Machine Teaming’, 2018.

[11] Ministry of Defence, ‘Defence Standard 00-251, Human Factors Integration for Defence Systems, Version 2’, 2021.

[12] ISO, ‘Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems’, ISO, 2010.

[13] Ministry of Defence, ‘Defence Standard 00-051: Environmental Management Requirements for Defence Systems’, 2018.

11 Glossary

2PUS	2nd Permanent Under Secretary
AI	Artificial Intelligence
RAISO	Responsible AI Senior Officer
AS	Autonomous System
ASR	Ambitious, Safe, Responsible
C2	Command and Control
COA	Course of Action
DAU	Defence AI and Autonomy Unit
DCDC	Development, Concepts and Doctrine Centre
Def Stan	Defence Standard
DevOps	Development and Operations
DLOD	Defence Line Of Development
EMS	Environmental Management System
FLC	Front Line Command
GALE	Globally At Least Equivalent
HCD	Human Centred Design
HFI	Human Factors Integration
HLR	High-Level Requirements
HR	Human Resources
ISR	Intelligence, Surveillance, Reconnaissance
JROC/IAC	Joint Requirements Oversight Committee / Investments Approvals Committee
JSP	Joint Service Publication
LLM	Large Language Model
LLR	Low-Level Requirements
ML	Machine Learning
MLOPS	ML DevOps
MOD	Ministry Of Defence
NATO	North Atlantic Treaty Organization
ODD	Operational Design Domain
PE	Programmable Elements
RAS	Robotic and Autonomous Systems
RAF	Royal Air Force
S&T	Science and Technology
SLR	System-Level Requirements
SMS	Safety Management System
SRO	Senior Responsible Owner
TLB	Top Level Budget
UK	United Kingdom