

Statement of Requirement (SoR)

Purpose

[REDACTED]

Reference Number	[REDACTED]
Version Number	0.6.1
Date	31/05/2022

1.	Requirement
1.1	Title
	Explainability for Vulnerability Identification in AI systems
1.2	Summary
	The Research and Development submission to the Strategic Review (SR20) recognised the need to advance MOD's ability to adopt critical and game-changing technology, enabling autonomous systems on the battlefield and in the command space through the use of artificial intelligence. It proposed to do this by establishing a Defence AI Centre with the science and technology component delivered by a Defence AI Centre Experimentation hub (DAIC-X) led by Dstl. A key objective for DAIC-X is to understand and develop good practice in managing AI verification, validation, vulnerabilities as well as wider issues including trust and transparency and legal and ethical considerations. This task will research the potential to exploit artificial intelligence explainability (XAI) methodologies to identify and expose vulnerabilities in neural network-based machine vision algorithms.
1.3	Background

	AI-enabled military and dual-use systems are rapidly reaching a level of maturity where they may be considered for operational use. One of the key barriers to their adoption is assurance, and specifically security assurance and risk management. There exists a wealth of academic and online literature demonstrating the many risks attached to applying machine learning and broader artificial intelligence in real world scenarios. It is therefore essential that risk owners are equipped with the tools and data to understand the new and emerging threats posed these systems. Arguably the hardest challenge in this space is understanding how, and where vulnerabilities manifest at the algorithmic level. Explainable AI (XAI) has been a hot research topic for several years, aiming to provide justification to algorithmic inference. In this way the algorithm may provide an additional level of confidence to the developer or operator that the algorithm is making decisions for the correct reasons. Now, this area of research may offer an additional opportunity to tease out abstract vulnerabilities that could lie embedded within the algorithm.
1.4	Requirement

	The aim of this project is to explore the potential of repurposing artificial intelligence explainability methods to draw out and label vulnerable regions of the input surface of neural network-based machine vision algorithms. In this way, dynamic rules-based boundary conditions, or equivalent, may be crafted to protect against malign or accidental data causing a negative behaviour in the system. Here, the Authority is specifically concerned with camera systems, which will include stereocameras that have utility in many classes of autonomous platforms. The Authority is particularly interested in how explainability approaches may be used in real time, in real or representative systems. Accordingly, it is required that research will, at least in part, be conducted using physical sensors, and with at least one commercial-off-the-shelf (COTS) robotic system that supports a suitable sensor package (e.g. Husarion RosBot), even if data is processed offboard on another system (e.g. physical connection to a standalone). Costings should include purchase of one or more systems for this research. Representative machine vision perception algorithms may be selected from open-source or trained specifically for this task. The Authority requests that at least one object detection algorithm and at least one image segmentation algorithm is included as part of this research to maximise impact and utility to MOD. In all cases, the Authority requires that analyses be performed using high fidelity metrics, and include suitable controls and null hypotheses. In this work, the emphasis will be on applied solutions that show promise for near-term exploitation over theoretical proofs. Research will comprise open-source survey of explainability approaches that may aid vulnerability identification (15%) and applied research either using open-source or commercial tools or by development of novel approaches (85%). The Authority has an aspiration to promote security assurance and vulnerability research of AI systems in UK academia. It therefore requires that all research be published by the Supplier in Gold open-access peer-reviewed journals and conferences. Costs should reflect anticipated fees to meet this requirement.
1.5	Options or follow on work <i>(if none, write 'Not applicable')</i>
	Not applicable

1.6	Deliverables & Intellectual Property Rights (IPR)						
Ref.	Title	Due by	Format	TRL *	Expected classification (subject to change)	What information is required in the deliverable	IPR DEFCON/ Condition <i>(Commercial to enter later)</i>
D1	Kick off workshop at Supplier's site	T0	Workshop	n/a	[REDACTED]	Supplier and Authority to cross-brief. Supplier to provide overview of recent research.	R-Cloud terms similar to DEFCON 705
D2	Progress and Technical Reviews either over MS Teams or in-person	T0+2/4/ 6/8 months	Presentation	n/a	[REDACTED]		R-Cloud terms similar to DEFCON 705
D3	Final report	Due by end of March 2023	Report	1-3	[REDACTED]	- Record of research completed; to include successful and unsuccessful approaches researched, their efficacy and opportunities for exploitation. - A record of planned, submitted or published articles.	R-Cloud terms similar to DEFCON 705

						• Opportunities for future collaboration and research.	
D4	Code to accompany report	Due by end of March 2023	Software	1-3	[REDACTED]	• For all software based tools used or developed: ○ Python or C/C++ code with worked examples – the Authority specifically does not require formal software management, tests or assurance to be applied, however the Authority requests code to be commented either directly inline or in the form of notebooks or markdown files. ○ (Where appropriate) Binary or executable files, beyond standard packages – as above, the Authority does not expect formal testing, proof-of-concept is sufficient. ○ Any additional software or data that will be generated	R-Cloud terms similar to DEFCON 705

						○ User manual, either as a standalone document or as markdown files alongside code ○ Pseudocode (e.g. skeleton code or documentation of the code structure) ○ Identification of licencing and intellectual property requirements	
--	--	--	--	--	--	--	--

***Technology Readiness Level required**

Notes- IPR should be inserted / checked by commercial staff before sharing with the supplier(s) to ensure accuracy.

1.7	Standard Deliverable Acceptance Criteria
	<i>This could be 'as per Framework T&C's' once an appropriate framework is later confirmed (links to section 13 of RCA). Consider the timeframe for our review of deliverable(s) (acceptance/rejection).</i>
1.8	Specific Deliverable Acceptance Criteria
	The deliverables must meet the criteria specified in the sections above. All deliverables shall be grammatically correct, be in a format agreed with the Dstl TP, and delivered via email to the Dstl TP and Project Manager. The criteria for accepting deliverables is as follows: - Documents are sufficiently detailed and technically coherent to enable a member of the DAIC-X technical team to understand the scientific progress made within the year and that the content meets the needs of the requirement. In general all deliverables shall ensure that: - The deliverables fully satisfy the requirements as stated in this statement of requirements. - All acronyms are fully defined; - The deliverable uses British English standard; - The deliverable uses simple language that a layman would understand; - All terms have clear meanings; - The structure is logical and easy to follow; - Diagrams, pictures, graphics, tables are used to help illustrate the requirement(s); - All references to existing work must be appropriately attributed; - All deliverables to have correct IP markings; - Delivery on time; - Delivery to cost; - Include a Report Documentation Page (RDP) – this can be supplied by the Dstl TP and should be included at the end of all reports. Failure to comply with the above may result in the Authority rejecting the deliverables and requesting re-work before final acceptance.

2.	Quality Control and Assurance
2.1	Quality Control and Quality Assurance processes and standards that must be met by the contractor

	☒ ISO9001 (Quality Management Systems) ☐ ISO14001 (Environment Management Systems) ☒ ISO12207 (Systems and software engineering — software life cycle) ☐ TickITPlus (Integrated approach to software and IT development) ☐ Other: (Please specify below)
2.2	Safety, Environmental, Social, Ethical, Regulatory or Legislative aspects of the requirement

3.	Security	
3.1	Highest security classification	
	Of the work	[REDACTED]
	Of the Deliverables/ Output	[REDACTED]
3.2	Security Aspects Letter (SAL)	
	Yes If yes, please see SAL reference- [REDACTED]	
3.3	Cyber Risk Level	
	[REDACTED]	
3.4	Cyber Risk Assessment (RA) Reference	
	[REDACTED] If stated, this must be completed by the contractor before a contract can be awarded. In accordance with the Supplier Cyber Protection Risk Assessment (RA) Workflow please complete the Cyber Risk Assessment available at https://www.gov.uk/guidance/supplier-cyber-protection-service	

4.	Government Furnished Assets (GFA)				
GFA to be Issued - No <i>If 'yes' – add details below. If 'supplier to specify' or 'no,' delete all cells below.</i>					
GFA No.	Unique Identifier/ Serial No	Description: <i>Classification, type of GFA (GFE for equipment for example), previous MOD Contracts and link to deliverables</i>	Available Date	Issued by	Return Date or Disposal Date (T0+) <i>Please specify which</i>

5.	Proposal Evaluation criteria
5.1	Technical Evaluation Criteria

The technical proposals will be assessed by the Authorities Technical Team using the following criteria and weightings.

The Technical Score will be calculated using the following formula:

Technical Score = Sum of (Score x Weighting x 10) for each of the Technical Criteria

Please note, proposals with a Technical Score of less than 65 will not be considered further. Furthermore, a score of 0 in any area will be categorised as inadequate and the Tender will not be considered further.

Technical Criteria		Score	Weighting
T1	The Tenderer's proposal must address all aspects of the requirement and display innovation, quality and accuracy. This must be supported by their history, record of accomplishment in the field, expertise and existing accreditations.	0, 3, 7, 10	40%
T2	The Tenderer's proposal must demonstrate strong evidence that there is the required expertise and knowledge to successfully complete the work.	0, 3, 7, 10	15%
T3	The Tenderer's proposal must demonstrate the availability of necessary facilities and assets to successfully undertake the work and that these will be available to the staff working on the project as required.	0, 3, 7, 10	10%
T4	The Tenderer's proposal must evidence a good track record of working on similar projects.	0, 3, 7, 10	10%
T5	The Tenderer's proposal must demonstrate confidence that the necessary outputs can be delivered within the required time-scales.	0, 3, 7, 10	5%
T6	The Tenderer must provide a clear, credible and appropriate project plan that lays out: - details of the technical approach; - schedule of events (Gantt chart); - basic organisation breakdown structure (OBS) of the Contractor's team, including partners and sub-Contractors (proposed split of effort and finance to be made clear in proposal); - assessment of the key project risks and mitigation actions; - list of dependencies, assumptions and exclusions;	0, 3, 7, 10	20%
Total			100%

The following scoring guide will be used to evaluate each of the Technical Criteria:

Unweighted Marking	Description
--------------------	-------------

5.	Proposal Evaluation criteria	
	0	Inadequate - the response does not address or explain how the requirement will be fulfilled and fails to demonstrate the ability to meet the requirement.
	3	Adequate - the response addresses the majority of elements of the requirement but is weak in some areas and does not fully detail or explain how the requirement will be fulfilled.
	7	Good - the response addresses all of the elements of the requirement and provides sufficient detail and explanation of how the requirement will be fulfilled.
	10	Excellent - the response addresses all elements of the requirement, and provides a comprehensive, unambiguous and thorough explanation of how the requirement will be fulfilled.
	Proposals will be scored according to their value for money, and awarded a total score. The total score will be calculated using the following formula: Total Score = Technical Score ÷ Proposal Price The contract will be awarded if the Tenderer obtains a minimum of 70 points from the Technical Criteria and passes the mandatory Commercial Requirements.	
5.2	Commercial Evaluation Criteria	
	The proposals will be reviewed against the below Commercial Requirements by a suitably qualified member of the Authorities Commercial Team. A proposal must pass the commercial mandatory requirements detailed below. Any proposal that does not meet the mandatory requirements will be deemed non-compliant and will not be evaluated further.	

5.	Proposal Evaluation criteria	
	Mandatory Commercial Requirements	Pass/Fail
	The Tenderer has provided a Firm price for the completion of this requirement.	
	The Tenderer shall clearly identify in the submission any background Intellectual Property that they intend to use in the execution of the contract and any limited rights terms (if any).	
	The Tenderer must accept the additional Terms and Conditions included within the Tasking Form.	