Serapis Tasking Form

Tasking Form Part 1: *(to be completed by the Authority's Project Manager)*

To:	Lot 6 Frazer-Nash Consultancy Ltd	**From:**	The Authority

Any Task placed as a result of your quotation will be subject to the Terms and Conditions of Framework Agreement Number:

LOT 6 DSTL/AGR/SERAPIS/UND/01

VERSION CONTROL

0.3

REQUIREMENT

Proposal Required by:	22/07/2022	**Task ID Number:**	U104
The Authority Project Manager:	[REDACTED]	**The Authority Technical Point of Contact:**	[REDACTED]
Task Title:	MSSA Multi-Modality Data Validation		
Required Start Date:	15/08/2022	**Required End Date:**	15/03/2023
Requisition No:	RQ0000011397	**Budget Range**	250k inc. FNC fee.

TASK DESCRIPTION AND SPECIFICATION

Serapis Framework Lot	☐ Lot 1: Collect ☐ Lot 2: Space systems ☐ Lot 3: Decide ☐ Lot 4: Assured information infrastructure ☐ Lot 5: Synthetic environment and simulation ☒ Lot 6: Understand

Statement of Requirements (SOR)

The Machine Speed Strategic Analysis (MSSA) project is tasked with applying AI for Intelligence, Surveillance & Reconnaissance (ISR) of the sub-threshold information environment, focussing on military aspects. An important aspect of this environment is the spread of accidental misinformation, or deliberate disinformation, which can affect military operations or open-source intelligence gathering.

In order to counter misinformation and disinformation, MSSA has instantiated a task called 'Validation of Third-Party Data'. This task aims to develop AI techniques to help a human analyst assess the validity of third-party data. These data can be a variety of modalities, and may originate from open-source news articles, social media, forums, etc. As a sub-task, we wish to investigate multi-modality data validation - the ability for a tool to examine different modalities of data for evidence of synthesis or falsification.

In order to implement multi-modality data validation, a variety of tools and techniques should be developed. These tools should then be able to integrate with a lightweight central 'framework' to produce an easy-to-use proof-of-concept data validation suite. There are many data modalities and techniques of interest, which will be expanded upon below. Each modality (for example, imagery) could itself be the focus of multiple 'analytics'. Each analytic will represent a method focusing on a specific validation task. As an example, an analytic designed to detect JPEG compression, and another aiming to highlight spliced areas, could both be applied to a single image. Further, multiple analytics may be developed to perform the same kind of validation, if it is believed that they will approach the task in a sufficiently distinct manner.

When an analytic is applied, the output should contain a level of explanation appropriate to the specific technique used, and with an intelligence analyst user in mind. In some cases this may simply be the confidence level of a machine learning classifier, whereas for others image regions may be highlighted, or text explanations provided.

The system should be built in a modular fashion, likely making use of Dockerised services linked to a web-based API. Each analytic can be represented by a separate service, and the API can be used to provide data for analysis with these services. We do not expect every possible analytic of interest to be produced during the first phase of this project, so the tool should be built with future extensibility in mind. As such, the task can be split into the creation of a lightweight central framework, and the creation of individual data analytics. The framework creation should not detract from the investigation of the validation analytics any more than is necessary. As such, we anticipate that the budget allocated to the creation of the framework should be a maximum of £50k. Further, to safeguard future phases of work, the framework itself should be created under DEFCON 703. Analytics themselves can be produced under DEFCON 705, allowing suppliers to retain ownership of newly-developed capabilities.

The use of supplier-owned background IP will be considered for the analytics, provided several conditions are met. It should aid in the demonstration of a proof-of-concept toolkit (e.g., by providing access to more analytics than would otherwise be feasible); it should not prevent or otherwise hinder the standalone use of foreground IP developed during this task; and it should be integrated with the same API as foreground IP.

Data Modalities and Techniques of Interest

Data modalities of interest to this project include imagery, video, audio, and text (noting that, eventually, languages beyond English will be of interest). Under a previous Serapis contract (U53 - Trusted Datasets Initiative) we received the ability to produce datasets of images with different camera signatures, and GAN-generated imagery. The code for this can be provided as GFA for the purposes of this task, if the production of such evaluation data would be useful. These data could be used, for example, in the creation of a 'held back' validation dataset, for testing the data validator.

There are many ways in which media can be manipulated or falsified, so the data validator will ultimately need a substantial number of analytics. The following list should not be considered exhaustive, but illustrates a number of desirable techniques. The literature and open-source methods should be reviewed for each technique, which may spawn multiple relevant analytics. We wish to research a variety of novel techniques, so the task should not simply wrap existing code into an API.

- Deepfake detection

  There has been much well-publicised effort towards the goal of detecting deepfakes - videos in which a 'target' identity has been used to replace the original source persona (https://www.kaggle.com/c/deepfake-detection-challenge). It is also possible to manipulate facial expressions - even while keeping the same identity - to present a false impression (Mazaheri & Roy-Chowdhury, 2022 - https://arxiv.org/abs/2103.08134). The data validator should aim to characterise deepfakes produced using a variety of methods including first-order motion, encoder-decoder, and splicing.

- GAN detection

  Generative Adversarial Networks can be used to synthesise images of a wide variety of things, including realistic human faces (https://thispersondoesnotexist.com/). They can also be used to alter the style of an image that already exists. GAN-generated images contain data artefacts which can be detected (e.g.,

Frank et al., 2020 - https://arxiv.org/abs/2003.08685 - examined in U53). Artefacts may not be in frequency space. For example, GAN images of human faces were noted to have irregular pupil shapes by Guo et al. (2021 - https://arxiv.org/abs/2109.00162). U53 focused on StyleGAN detection, so further work should seek to expand to include other GAN implementations (e.g., CycleGAN, Zhu et al., 2017 - https://arxiv.org/abs/1703.10593).

- AI-generated text detection

  Natural language processing can be used to produce convincing synthetic text outputs. One well-known implementation is OpenAI's GPT-3 (Brown et al., 2020 - https://arxiv.org/abs/2005.14165). Synthesised text can be used to produce false postings and articles which nevertheless seem convincing, so the data validator should aim to detect such text.

- AI-generated audio detection

  Voices can be synthesised using text-to-speech systems (e.g., Nvidia's Tacotron 2 & Waveglow implementation - https://pytorch.org/hub/nvidia_deeplearningexamples_tacotron2/). These voices can be those of real people, learned by the system, or synthetic profiles. Further, voices within provided speech samples can be converted by separating the identity of the speaker from the speech content (Chou, Yeh and Lee, 2019 - https://arxiv.org/abs/1904.05742). Such voices could be used to create fake news, or otherwise misrepresent a speaker, so the data validator should aim to identify when speech has been synthesised in these ways.

- Image alteration / editing detection

  Machine learning techniques do not need to be used for an image to be manipulated. Regions of an image can be warped, spliced with another image, compressed, etc. While the use of such techniques does not necessarily indicate deceptive intent, they can be used to present false images or try to disguise them. The characterisation of manipulated regions of an image can be used to detect splicing, copying, and other simple image-altering techniques (Bappy et al., 2017 - https://ieeexplore.ieee.org/document/8237794). Mayer & Stamm (2021 - https://arxiv.org/abs/1912.02861) use forensic similarity graphs to localise image forgeries, and camera fingerprint inconsistencies can also highlight manipulations. The data validator should include ways to spot these kinds of altered images.

- Detection of mismatch between images and (embedded) captions

  A common way of creating mis-or-disinformation is to provide a false caption alongside an image, which misrepresents the image content. For example, an image of a warship of country A - in the waters of country A - could be captioned as belonging to country B, and therefore operating in a contentious manner. We are particularly interested in the use of optical character recognition (OCR) to extract text that has been embedded within an image or graphic. Dstl has developed a tool - Baleen - which could be used for this purpose (https://github.com/dstl/baleen3). This text should then be compared with other semantic information from the image. In the warship example, this could be a naval insignia, or the warship itself. Here, a mismatch would highlight potential misinformation. As this is a large problem, a proof-of-concept could be built around a specific example (to be discussed between the supplier and the Dstl technical partner) and included with the data validator.

- Detection of mismatch between image metadata and signatures

  The EXchangable Image File format (EXIF) stores metadata as part of an image file. If these metadata do not match information that can be gleaned from an image (such as camera model signature, also examined within U53) then the image, the metadata, or both may have been falsified. Therefore, the data validator should aim to characterise such inconsistencies.

To summarise, the output of this task should be a modular, proof-of-concept data validation tool. The tool should contain analytics which are designed to analyse input data modalities for inconsistencies which may indicate deceptive intent, and present this information in an easy-to-read and explainable fashion. A lightweight central framework should be used to provide a single point of interaction between data and analytics. The tool should be

easy to use, and extensible so that additional analytics may be added in future. Full source code should be provided, along with simple instructions for install and use. Mid-point reports should be produced to explain progress, with a detailed final technical report provided at the end of the project. This report should include a discussion of desired or suggested future avenues of investigation. Finally, a presentation should be provided to Dstl staff at the end of this phase of the project.*

In addition, we would like to retain the option of continuing this project beyond the first phase described in this SOR. A second phase of work would be able to focus on the creation of analytics not produced during phase 1, the improvement of analytics provided as phase 1 deliverables, and potential changes to the framework and API. These tasks would be able to be implemented by one or multiple suppliers, as each analytic could be the focus of separate sub-tasks. Each analytic would, as with phase 1, need to interface with the central framework.

Topic area: disinformation, fake news, deepfakes, GANs, validation, neural networks, artificial intelligence

Format: Software, demonstration, report.

D1: Source code for 'Data Validator' framework, along with documentation, and instructions for install and use. To be provided under DEFCON 703. T0+7M.

D2: Source code for 'Data Validator' analytics, along with documentation, and instructions for install and use with D1. To be provided under DEFCON 705. T0+7M.

D3: Mid-point progress report at T0+3.5M

D4: Final technical report describing the methodology and technologies used, along with a discussion of recommendations for next steps. T0+7M.

D5: Final presentation and demonstration of the 'Data Validator' system. T0+6.5M.

D6: OPTIONAL: Continuation of task beyond FY 22-23. Deliverables will be analogous to D2-D5 for each supplier.

***Please note, we expect Phase 1 of this task to be delivered by a variety of suppliers.**

Procurement Strategy

☒ Lot Lead to recommend ☐ Single Source / Direct Award

Pricing:

☒ Firm Pricing ☐ Ascertained Costs* ☐ Other*

Firm Pricing shall be in accordance with DEFCON 127 and DEFCON 643

Ascertained Costs shall be in accordance with DEFCON 653 or DEFCON 802.

*only at Authority's discretion

Task IP Conditions

Task IP Conditions (Follow the NIPPY guide to identify your information and IP requirements for each deliverable)	**Summary of the Authority's rights in foreground IP (IP generated by the supplier in performance of the contract)**
DEFCON 703 ☒ (Framework only)	Vests ownership with the Authority
DEFCON 705 Full Rights ☒ (All else)	Enables MOD to share in confidence as GFI or IRC under certain types of agreements. Can be shared in confidence within UK Government.

OTHER IP DEFCONS: 14* ☐, 15* ☐, 16* ☐, 90* ☐, 91* ☐, 126* ☐	Generally only suitable for deliverables at TRL 6 and above.
BESPOKE IP Clause ☐ *	Details to be added and agreed by IP Group
* Do not use without IPG advice and approval	
Please state in this text box if MOD or the customer has a requirement a) that one or more Other Government Departments is able to share confidentially with their own suppliers, b) to publish but you do not think there is a requirement to own or control the deliverable, or c) to share under a procurement Memorandum of Understanding (MOU).* *If any of these three issues applies, please contact IPG for advice before completing this form. *Listing research MOUs is not required, but can be a helpful courtesy to the supplier.*	

DELIVERABLES

Ref	Title	Due by	Format	TRL	Expected classification (subject to change)	Information required in deliverable	IPR DEFCON
D1	Source code for 'Data Validator' system framework, along with documentation, and instructions for install and use.	T0+7M.	Software	3	[REDACTED]		**703**
D2	Source code for 'Data Validator' analytics, along with documentation, and instructions for install and use with D1.	T0+7M.	Software	3	[REDACTED]		705
D3	Mid-point progress report.	T0+3.5M.	Technical reports		[REDACTED]		705
D4	Final technical report describing the methodology and technologies used, along with a discussion of recommendations for next steps.	T0+7M.	Technical report		[REDACTED]		705
D5	Final presentation and demonstration	T0+6.5M	Demonstration		[REDACTED]		705

	of the 'Data Validator' system						
D6`	OPTIONAL: Continuation of task beyond FY 22-23. Deliverables will be analogous to D2-D5 for each supplier.	TBC	Software, technical reports, demonstration.	3+	[REDACTED]		705 (703 if changes are made to framework).

DELIVERABLE: ACCEPTANCE / REJECTION CRITERIA

Unless otherwise stated below, Standard Deliverable Acceptance / Rejection applies. This is 30 business days, in accordance with DEFCON 524 Rejection, and DEFCON 525 Acceptance.

Standard Deliverable Acceptance / Rejection:-

Yes ☒ (DEFCON 524 Rejection, and DEFCON 525 Acceptance)

No ☐ (if no, please state details of applicable criteria below)

Deliverable Acceptance / Rejection Criteria:-

If there are any other specific acceptance/rejection criteria you would like to apply to any of the deliverables, please state them here.

Government Furnished Assets (GFA)

ISSUE OF EQUIPMENT/RESOURCES/INFORMATION/FACILITIES (*if not applicable, delete table and insert "None" in this text box)*

Unique Identifier/ Serial No	Description	Classification	Type	Available Date	Issued by	Return or Disposal Date	Any restrictions?
U53 code, report	Delivered outputs for the U53 Trusted Datasets Initiative task.	OFFICIAL	Software, report	T+0M	Dstl	T+9M	This IP is the property of [REDACTED], but may (if desired) be used for the provision of data and analytics for the Multi-modality Data Validation task, provided no foreground IP depends on their presence to function.

QUALITY STANDARDS

☒ **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 in free text below)

SECURITY CLASSIFICATION OF THE WORK

The highest classification of this SOR

OFFICIAL ☐ OFFICIAL-SENSITIVE ☐ SECRET ☐ TOP SECRET ☐ STRAP ☐ SAP ☐

The highest expected classification of the work carried out by the contractor

OFFICIAL ☐ OFFICIAL-SENSITIVE ☐ SECRET ☐ TOP SECRET ☐ STRAP ☐ SAP ☐

The highest expected classification of Deliverables/Output

OFFICIAL ☐ OFFICIAL-SENSITIVE ☐ SECRET ☐ TOP SECRET ☐ STRAP ☐ SAP ☐

Is a Security Aspects Letter (SAL) required? *(A Security Aspects Letter (SAL) will be required for each Task above Official-Sensitive and above)*

Yes ☐ No ☐

TASK CYBER RISK ASSESSMENT. *(In accordance with DEF STAN 05-138 and the* Risk Assessment Workflow)

Cyber Risk Level	[REDACTED]
Risk Assessment Reference	[REDACTED]

ADDITIONAL TERMS AND CONDITIONS APPLICABLE TO THIS CONTRACT

Please ensure all completed forms are copied to DSTLSERAPIS@dstl.gov.uk when sending to the Lot Lead.

Tasking Form Part 2: *(To be completed by the Lot Lead)*

To: The Authority	From: The Lot Lead

Proposal Reference **017318-97654L U104 MSSA Multi-Modality Data Validation - Frazer-Nash Proposal V3** **(attached)**

Delivery of the requirement:

The proposal shall include, but not be limited to:

- A full technical proposal that meets the individual activities that are detailed in Statement of Requirements (Part 1 to Tasking Form).
- Breakdown of individual Deliverables, with corresponding Intellectual Property rights applied.
- Breakdown of Interim Milestone Payments, with corresponding due dates.
- A work breakdown structure/project plan with key dates and deliverables identified.
- A list of required Government Furnished Assets from the Authority, including required delivery dates.
- A clear identification of Dependencies, Assumptions, Risks and Exclusions which underpin your Technical Proposal.
- Sub-Contractors Personnel Particulars Research Worker Form and security clearances (if applicable)

PRICE BREAKDOWN

You are to use the costs detailed in Item 2 Table I in the Schedule of Requirement and at Annex E Table 2 of the Serapis Framework Agreement. Please also provide a price breakdown which should include, but is not limited to: Lot Lead Rates, Sub-contractors costs and rates, travel and subsistence. In support of your Proposal you are requested to provide clear details of all Dependencies, Assumptions, Risks and Exclusions that underpin your price.

Offer of Contract: *(to be completed and signed by the Contractor's Commercial or Contract Manager)*

Total Proposal Price in £	**£249,629.80**		(ex VAT)
Start Date:	12/09/2022	**End Date:**	**15/03/2023**

Lot Leads Representative	Name	[REDACTED]
	Tel	[REDACTED]
	Email	[REDACTED]
	Date	27/07/2022
Position in Company	Serapis Lot 6 Project Manager	
Signature	[REDACTED]	

[REDACTED]
[REDACTED]
[REDACTED]

Core Work – Milestone breakdown costs

Proposed Milestones Payments

Your TMS bid costs shall be included in milestone 1.

The final Milestone must reflect the actual cost of the deliverable, and be greater than 20% of the Task value, unless otherwise agreed with your Commercial POC

Please duplicate the template per milestone table format below as necessary, and rename milestone number accordingly.

[REDACTED]

Total Cost (All Milestones)	**£249,629.80**

Tasking Form Part 3:

To be completed by the Authority's Commercial Officer and copied to the Authority's Project Manager.

1. **Acceptance of Contract:**		
Authority's Commercial Officer	Name	[REDACTED]
	Tel	[REDACTED]
	Email	[REDACTED]
	Date	17/08/2022
Requisition Number		RQ0000011397
Contractor's Proposal Number		FNC 017318-121828V
Purchase Order Number		DSTL0000006784
Signature		[REDACTED]
Please Note: Task authorisation to be issued by the Authority's Commercial Officer or Contract Manager. Any work carried out prior to authorisation is at the Contractor's own risk.		