**STATEMENT OF REQUIREMENT (SOR): PROVISION OF AN ARTIFICIAL INTELLIGENCE (AI) CAPABILITY CONCEPT DEMONSTRATOR (CCD) 2 AT UP TO OFFICIAL-SENSITIVE**

|  |
| --- |
| **Introduction**The pressure to better manage and exploit our knowledge and information to enable us as an organisation to learn and adapt faster than our adversaries is currently a significant challenge for the Army. This is because our data repositories are large and disparate. The Army must look to adopt new tools, using cutting edge technology based on Machine Learning (ML) and Artificial Intelligence (AI), to support effective decision making; learning; our analytical tool-kits and broader knowledge exploitation.This SOR outlines the need for a proven software solution that provides an end-to-end intelligent platform to unlock the untapped potential of our disparate data sets and enhance exploitation of available data, information, and knowledge.  |
| **Background**Commander Field Army (CFA) has directed exploratory work be undertaken to identify state of the art functionality that can enhance the intellectual agility of the Force, to enable more effective exploitation of existing knowledge and to improve access to the most relevant information. In sum, we must better exploit our current and future data, seeing them as a strategic asset. This clear direction mutually supports Director Land Warfare (DLW) direction to develop an assured knowledge methodology, which seeks to improve evidence-based decision making through the harnessing of multiple cross-referenced feeds. From Sept 2020 – Dec 2020 a Capability Concept Demonstrator (CCD) was undertaken by Land Warfare Centre (LWC) in partnership with Adarga Ltd, to understand the art of the possible regarding an industry leading software solution. We now have a developing understanding of how a platform like Adarga’s can support assured knowledge and how it can improve both the quality, and tempo, of organisational decision making. But, having conducted a CCD with a limited scope, we seek to conduct a more in-depth examination of AI and ML capabilities, this time integrated into routine work to better determine its potential. The technology should be further assessed through the delivery of a longer-term deployment of a capability like Adarga’s, incorporating up to OFFICIAL SENSITIVE data, in order to maximise our understanding of its potential.  |
| **Objectives**The objectives of CCD 2 are to:* Increase productivity of LWC users who depend on understanding and corroborating AKX data and publicly available data
* Adopt AI and ML technologies that have been proven to enhance Information and Knowledge Exploitation (IKX)
* Surface hidden connections across publicly available, OFFICIAL classified data, and up to OFFICIAL-SENSITIVE classified data
* Enhance our trust in AI and ML
* Inform the future direction on how AI and ML may be adopted across the force
 |
| **Specific Requirements**The software will be required to support the following objectives:* Assist the Army in realising the full potential of available information to support better decision making
* Transform data into knowledge so users are empowered to make better decisions based on network discovery, corroboration, and knowledge bases
* Provide access to information more effectively and improve integration to reduce risk of knowledge gaps and duplication of effort
* Make use of AI and ML technologies to enhance data, information, and knowledge exploitation
* Allow users to identify, analyse and interpret existing and new information

adarga-engine™ and adarga\_bench™ will achieve the objectives through delivering the following:* Search and look across unstructured data (and agreed structured data) that will be made available for ingestion into adarga\_engine™
* Search and look across multiple formats of data. The Buyer data formats include .doc, .docx, .pdf, .mp4 and the Buyer understands that the quality of the results will be directly linked to the data quality and file format (e.g. .mp4 files with low quality audio, .pdf files that are not well structured and/or have undergone low quality OCR or other pre-processing techniques, etc., will not yield optimum results due to the initial data quality). Up to 10,000 documents a month shall be ingested. Each .doc/.docx/.pdf file shall be no more than 10MB. Each .mp4 file shall be no more than 250MB. Extracted text shall be no more than 1MB.
* Search and look across multiple domains of data, including reaching out to data repositories and agreed sources across the internet. The agreed data sources are listed in Annex 1 and further data sources may be agreed during CCD2
* Ability to be hosted and deployed as an appropriately accredited cloud-based solution that will allow for OFFICIAL and up to OFFICIAL-SENSITIVE material to be ingested and used, and for regular software updates/patches/releases to be deployed
* Demonstrate language translation from multiple different languages
* Transcribe and process appropriate video files that include audio
* Present any findings and analysis through adarga\_bench™ which provides visualisations as insights and dashboards
* Identify where repetition occurs in data and surface similar content through corroboration and related reading
* Support users’ workflows by enabling users to create projects and workspaces
* Locate event entities and make linkages in data
* Summarise documents, allowing users to choose between “Abstractive” and “Extractive” summarisation methods
* Provision of maps when analysing data
* Permit up to 60 users to get direct use of the software
* Facilitate agreed training to allow users to access the software and be able to properly validate the effectiveness of the system
* Be accessible from MODNET systems
* Process up to OFFICIAL data until the Buyer approves that adarga\_engine™ can process up to OFFICIAL-SENSITIVE data

The Buyer and Supplier shall agree the delivery measures and shall assess these through Quarterly Business Reviews and further ad-hoc meetings as agreed.The Supplier shall provide an appropriate and agreed technical support and maintenance arrangement as referenced in the Service Level Agreement.The Supplier may add to the features and capabilities during the CCD2.Definitions:* **Unstructured data:** From the perspective of the Adarga capability, the term unstructured data is generally defined as text organised into clean and coherent sentences
* **Structured data:** From the perspective of the Adarga capability, structured data is defined as the text (usually in complete sentence form) organised in a structured way, for example into tabular form inside of a document or spreadsheet file
* **Abstractive summarisation:** Abstractive summarisation is much the same as extractive in that the aim is to select the key pieces of information from an article such that a user can understand the main points of the article being summarised. However, an abstract is also an interpretation of the document, rather than an excerpt, providing the possibility of even more concise and useful information, albeit with the risk of errors of interpretation
* **Extractive summarisation:** Extractive summarisation is the identification of the key sentences within a document and providing these to a user unaltered. The number of sentences provided can be controlled though is typically chosen such that a user can understand the main point of the article being summarised
* **Entities:** Labels applied to text according to Adarga’s ontology, such as ‘person’, ‘place’, ‘organisation’, ‘equipment’
 |
| **Success Criteria** The software will be considered a success once it is able to demonstrate enhanced data, information and knowledge exploitation across publicly available, OFFICIAL, and up to OFFICIAL-SENSITIVE data. Success may be assessed through both tangible and intangible results when using a Knowledge Platform and may include the following:Tangible measures:* Improve productivity when engaging with AKX that is up to OFFICIAL SENSITIVE and publicly available data
* Increase volume of data considered when generating reports

Intangible measures:* Increase confidence in the data and evidence for decision making
* Support DLW direction to develop an assured knowledge methodology and the role of AI and ML within this methodology
* Enhance data, information, and knowledge exploitation
* Understanding of and trust in AI and ML technologies

To achieve the success criteria, the platform will be fully configured and integrated, and all users will have to be trained to operate it. At the end of CCD 2, an assessment will be made as to whether this technology is something which should be considered worthy of additional investment by the Army and/or wider Defence.  |
| **Impact Statement**The future impact will build on the results from the CCD1 users against the broader data sets and day-to-day working. CCD1 demonstrated: * **Increased productivity**.The platform was seen by **95%** of users as being faster at finding relevant information, and generating insights, as compared to current working practices.

 * **Improved decision making**, deliveredthrough the ability to rapidly access and synthesise a wider range of information. **81%** of users agreed that the results Adarga generated were more in depth when searching for information.
* **Improved exploitation of data, information and knowledge. 71%** of users agreed that the Adarga platform’s capability offered an improvement upon current working practices.
* **Increased accuracy and confidence in information. 83%** of users were more confident in the results Adarga produced in comparison to results produced using current working practices.

The undertaking of CCD2 will also reveal the key challenges that the Army must overcome when adopting AI, ML, and other cloud-based cutting-edge technologies. Identifying, understanding, and overcoming these challenges will be critical if the Army is to maintain and/or expand its technological competitive edge in relation to potential current and future adversaries.  |

Annex 1: Data Sources

Internal Buyer data:

* AKX OFFICIAL and up to OFFICIAL-SENSITIVE
* Army Research Reports Repository – MODNET

Supplier provided data:

* Global and threat intelligence news feed from November 2020 to current.