Proposal No. F244-0001-0024

Mapping Intel to Insight with AI-Augmented Dynamic Knowledge Graphs

Volume 2: Technical Volume

Mapping Intel to Insight with AI-Augmented Dynamic Knowledge Graphs

Volume 2: Technical Volume

This proposal includes data that must not be disclosed outside the Government and must not be duplicated, used, or disclosed – in whole or in part – for any purpose other than to evaluate this proposal. If, however, a contract is awarded to this offeror as a result of – or in connection with – the submission of this data, the Government has the right to duplicate, use, or disclose the data to the extent provided in the resulting contract. This restriction does not limit the Government's right to use information contained in this data if it is obtained from another source without restriction. The data subject to this restriction are contained in pages X-Y.

1. **Identification and Significance of the Problem or Opportunity**.

Analysts within the U.S. Air Force (USAF) utilize Knowledge Graphs, which captures relationships between entities in a graph-based format (e.g. technological, personal or physical interactions), to enhance situational awareness, support pattern of life analysis, and identify threats within a network. Knowledge Graphs are often derived via imperfect processes – e.g. Artificial Intelligence and Machine Learning (AI/ML) based methods – from heterogeneous data sources that may evolve over time. Accordingly, analysts often need to correct the contents and structure of a Knowledge Graph for a multitude of reasons, including:

a. aligning with updates to source information (i.e. information from which the knowledge graph was derived) as the situation on the ground and operational objectives evolve (e.g. new reports indicate that “Entity3 no longer reports to Entity2”);
b. adding entities or relationships that are not present in the Knowledge Graph (e.g. because an AI/ML-based method applied to source data did not capture the relationship that “Entity2 reports to Entity1”);
c. removing or modifying false relationships discovered within the knowledge graph (e.g. an AI/ML-based method may have earlier inferred that “Entity3 reports to Entity1” based on transitivity of the relationships “Entity3 reports to Entity2” and “Entity2 reports to Entity1”, the former of which is no longer true).

This concept of a Knowledge Graph that evolves based on user feedback and updates to data sources is known as a Dynamic Knowledge Graph (DKG). DKGs are powerful tools that provide analysts with interactive AI/ML-based capabilities such as information retrieval, automated reasoning and inference, and natural language querying. Even so, the USAF has found that maintaining the accuracy of a DKG involves a substantial amount of time and effort on the part of analysts as they manually correct and update the DKG, and this process involves substantial redundancy given that many user edits follow patterns or modes of inference that can be automatically learned and applied. *Virtualitics believes that augmenting DKGs with AI/ML-based methods for recommending graph modifications implied by user input can substantially decrease the time analysts spend correcting the results of up-stream AI/ML methods used to generate the DKG, while also extending the analyst’s capability to update and maintain DKGs so that they reflect evolving sources of heterogeneous data. In this way, Virtualitics can help analysts leverage DKGs with the speed and accuracy they need to quickly execute time-sensitive tasks (e.g. targeting operations).*

Virtualitics proposes to carry out a feasibility study to understand how AI/ML-based inference methods can automate the process of updating and correcting DKGs based on ongoing analysis of the manual modifications a user makes to the DKG. Specifically, Virtualitics proposes to: (i) develop an experimental prototype of a DKG based on synthetic or publicly available data that can be modified by simulated user inputs; (ii) evaluate and compare the performance of different AI/ML-based inference methods for predicting additional modifications deemed necessary based on analysis of the user’s input (modifications). Importantly, the prototype DKG will have an ontology and function that reflects the operational parameters of DKGs used by USAF analysts, and the edges within the DKG may be assigned weights to encode quantitative data expressing the strength or confidence of a particular relationship existing between two entities. Examples of candidates for AI/ML-based inference methods that Virtualitics can investigate include:

1. Dynamic recomputation of shortest-paths between entities (i.e. nodes) in a graph when a relationship between entities has been removed by the user – *this would expedite an analyst's ability to execute time-sensitive call-chain analysis*;
2. Adding or removing inferred relationships (i.e. edges) between entities based on correlated relationships they have with other entities in the graph (with the correlated relationships being modified by the user's inputs) – *this would help an analyst conduct pattern of life analysis by identifying a relationship between two persons who likely know one another based on the similarities in the places they visit at particular times of day and the other persons they meet at these places;*
3. Adding or removing inferred hierarchical relationships between entities based on group membership (with user edits resulting in dynamically changing group memberships) – *this would help an analyst flag a relationship between an individual and a commander by assessing the intersection (within the DKG) between the individual's set of acquaintances and the commander's set of (known) subordinates*.[1]

Virtualitics will seek to obtain baseline performance metrics for considered AI/ML-based inference methods and provide an assessment of which methods are most promising for further development in a full-scale prototype during a follow on Phase II effort.

Notably, Virtualitics' proposed study is distinguished from more conventional approaches to developing interactive AI-augmented DKGs in two specific ways. First, Virtualitics's prototype DKG will make use of the (patented) state-of-the-art 3D data visualization capabilities built into the Virtualitics AI Platform (VAIP) – these interactive 3D visualizations of network data (e.g. see Fig 1. below), coupled with AI-based insights expressed using natural language, will provide the needed context in which analysts employ VAIP's built-in no-code AI routines to interrogate the graph's structure, identify anomalous graph elements, and highlight information gaps.[2] Second, Virtualitics will leverage its expertise in Network Graphs, applying VAIP's patented Network Extractor and Network Explainable AI (XAI) routines to provide analysis of the graph structure underlying the DKG and present the results in an intuitive, human interpretable form.[3,4] See Figures 2 and 3 below for additional examples of Virtualitics Explore's capabilities. Importantly, within the scope of the proposed feasibility study, utilization of VAIP to develop the experimental DKG prototype will enable Virtualitics to focus on: (i) creating a DKG based on publicly available or synthetic data (with an ontology relevant to end-user's operational criterion), (ii) generating simulated user inputs, and (iii) evaluating various AI-based inference methods. Moreover, Virtualitics has extensive experience deploying VAIP in operational environments (see related work), and so the development of the experimental DKG prototype using VAIP will help set the stage for continued deployment during a follow on Phase II effort.

[1] Note that this assessment can be made quantitative by computing the *Jaccard similarity coefficient* between the two sets, enabling a confidence score to be assigned.

[2] Donalek, C., Amori, M., Gantenberg, J., Sahu, S., Indurkhya, A., (2022) Systems and methods for high dimensional 3D data visualization (US11455759B2). U.S. Patent and Trademark Office.

[3] Indurkhya, A., Donalek, C., Amori, M., Sahu S., Anand, V., Gantenberg, J., (2021) Systems and methods for 3D data visualization and network extraction (US11481939B2). U.S. Patent and Trademark Office.

[4] Martinez, H.J.V., Indurkhya, S., Zanfardino, G., Indurkhya, A., Sahu, S., Donalek, C., Amori, M., (2022) Systems and Methods for Network Explainability (US20230004557A1). U.S. Patent and Trademark Office.

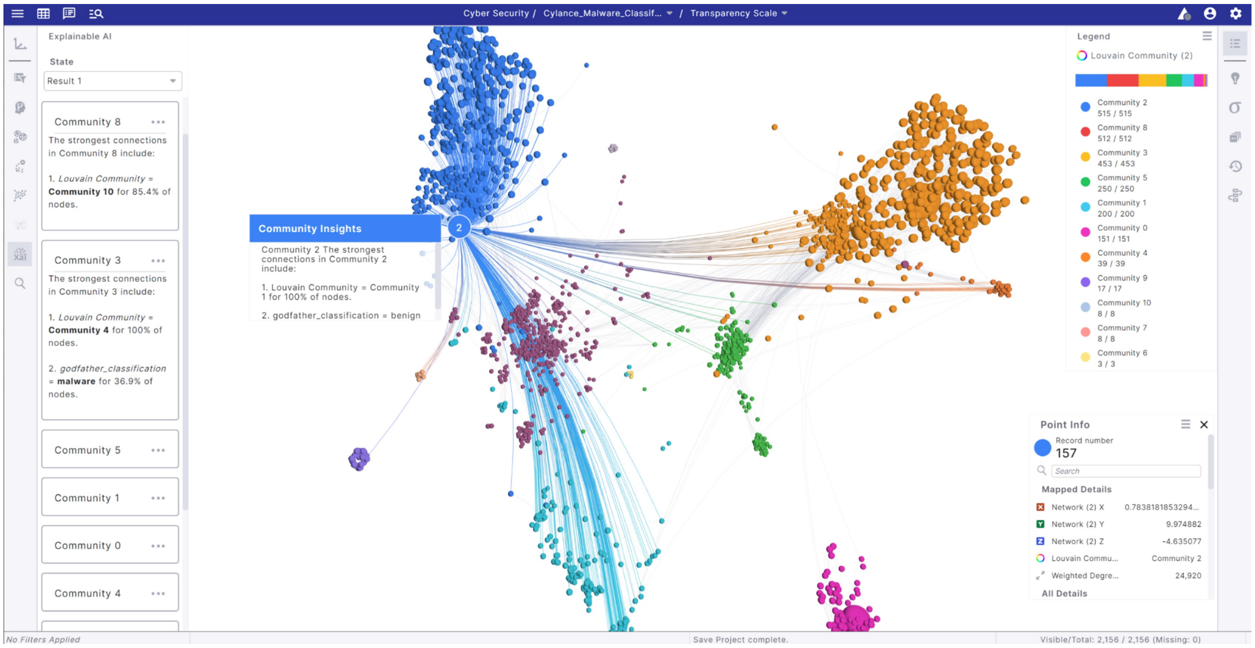

Fig 1. Virtualitics Explore displaying Network Graph with Explainable AI analysis.

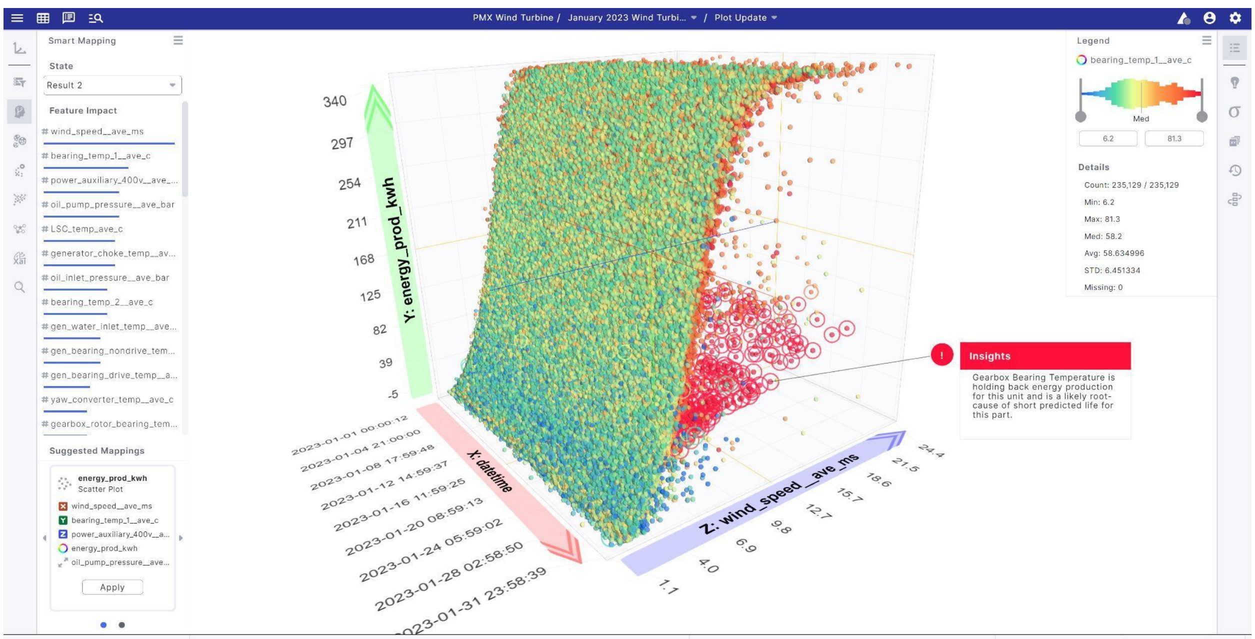

Fig 2. Virtualitics Explore displaying AI-generated 3D visualization annotated with AI-based insights.

2. <u>Phase I Technical Objectives</u>.

Virtualitics' proposed feasibility study has three objectives:

1. Virtualitics will develop a proof-of-concept prototype of a Dynamic Knowledge Graph (DKG) that is augmented with several AI/ML inference methods for predicting modifications to the graph structure based on analysis of the user input. In order to develop the concept, Virtualitics will work closely with government Subject Matter Experts (SMEs) to:

a. develop an ontology (i.e. schema) that is representative of Knowledge Graphs employed by USAF analysts and that will be used to generate a synthetic Knowledge Graph used for experimental testing of the DKG prototype – *this effort will result in a synthetic DKG used for evaluation in this feasibility study*;
b. identify several forms of DKG corrections, updates or other means of inference that USAF analysts are likely to carry out and that are suitable candidates for automation via AI/ML inference methods – *this scoping effort will identify which AI/ML inference methods are developed and evaluated during the feasibility study, as well as identify the characteristics of the simulated user input that will be used to evaluate the performance of the prototype*;
c. determine the Key Performance Metrics (KPMs) that can be used to quantitatively assess the feasibility of the overall concept – *specifically, these metrics will be used during the feasibility study to identify the most promising AI/ML inference methods and serve as a baseline to be improved upon in a follow on Phase II effort.*

Virtualitics will then evaluate the feasibility of the concept by creating a proof-of-concept prototype DKG that is augmented with each of the identified AI/ML inference methods, testing the AI/ML inference methods using the simulated user input (as appropriate for each type of inference method), and measuring the performance of the method using the identified KPMs. *Importantly, this evaluation will result in an assessment ranking the suitability of the AI/ML inference methods for continued development in a follow on Phase II.* Finally, Virtualitics will encapsulate the proof of concept in a Jupyter Notebook and an associated set of python modules.

2. Virtualitics will work with USAF and government Subject Matter Experts (SME) to determine specific use-cases for how the proposed software solution would be used. Based on this, Virtualitics will work closely with SMEs to identify specific ways in which Virtualitics Explore, a component of the Virtualitics AI Platform that supports 3D interactive visualizations and no-code AI-based analysis of knowledge graphs, can be used to interact with, understand and analyze the Dynamic Knowledge Graph. Concretely, Virtualitics will configure the Jupyter Notebook (encapsulating the proof-of-concept as detailed in Objective 1) to connect with Virtualitics Explore so that relevant 3D interactive visualizations (as identified after consulting SMEs) can be automatically populated into Explore. In this way, Virtualitics aims to demonstrate how the pairing of AI/ML-based no-code routines can further enhance the ways in which analysts can engage with the DKG and evaluate the modifications made by the AI/ML inference methods.
3. To prepare for a possible Phase 2, Virtualitics will formulate a plan for developing, testing and validating a full-scale prototype for adaptive, interactive dynamic knowledge graphs based on the proof of concept developed in the Phase 1 feasibility study. Specifically, the Phase 2 development would focus on further improving the AI/ML inference methods identified as promising during the Phase 1 study, with the goal of outperforming the baseline performance. Notably, the testing and validation would be in relation to a set of use cases identified in consultation with government SMEs. The plan, which would be executed during a subsequent Phase 2, will: (a) detail how the down-selected (during Phase 1) AI/ML inference methods could be improved; (b) outline how the Phase 1 prototype DKG can be extended to better represent DKGs used by USAF analysts, including use of a more sophisticated ontology; (c) outline how the software solution will operate on expected scales of data (as measured by the size of the graph structures underlying DKGs used by USAF analysts); (d) outline a set of specific tests (for each AI/ML inference method) that will be used to validate whether the software solution adequately meets the needs of the end user. (Where possible, the proof-of-concept to be developed during this feasibility study will help us identify the key performance metrics that can be used in a Phase 2.) Moreover, the plan will include a timeline that ensures multiple rounds of testing and validation so that identified deficits in the software solution can be improved and the software solution can be refined. Furthermore, the plan will detail resources needed to securely develop and deploy the software solution so that end-users can test the software and provide feedback. The plan will be included in the final report.

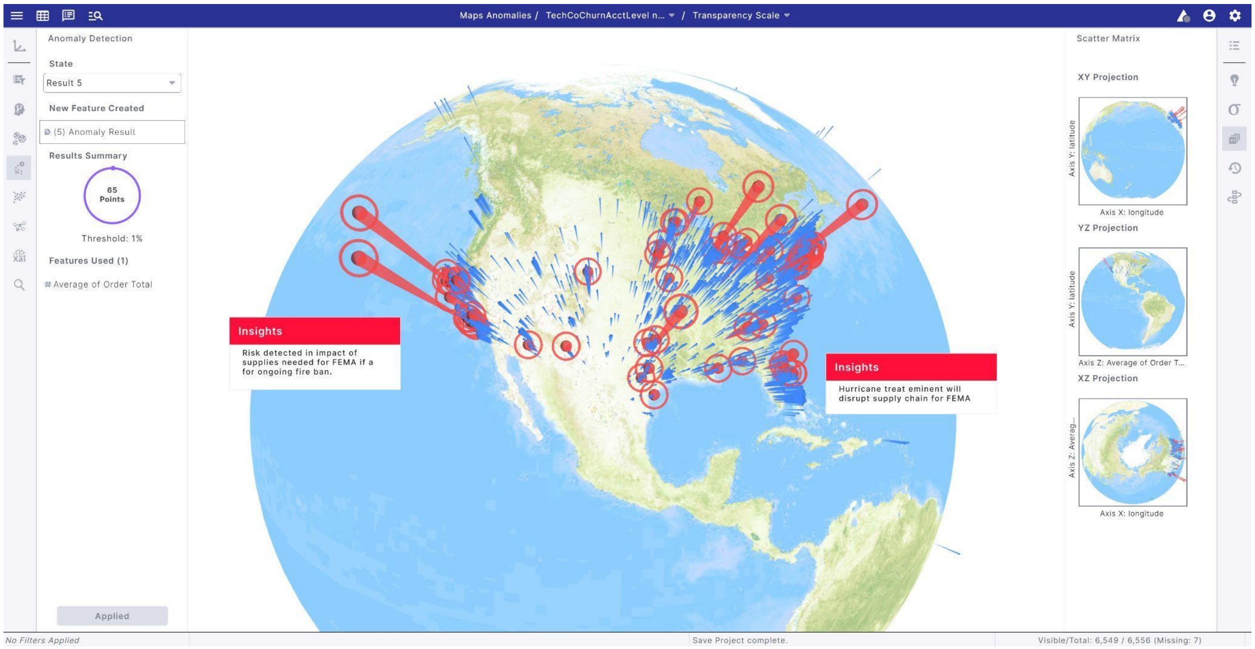

Fig 3. Virtualitics Explore displaying interactive 3D geospatial plot with AI-based anomaly detection.

3. <u>Phase I Statement of Work (include Subcontractors and/or Research Institutions).</u>

Task	**Duration**	**Description**	**Performer**
01 – Kickoff Briefing (DELIVERABLE)	Days 0-10	We will write and present an initial 10-slide report describing our research approach to accomplish this Phase I feasibility study.	Sagar Indurkhya, Andrew Green
02 – Customer Discovery	Days 10-45	Design relevant data schema (i.e. ontology) for Dynamic Knowledge Graph, determine graph modifications to model using AI/ML, and identify Key Performance Metrics. We will provide the TPOC with a progress report detailing outcomes of Task 2.	Andrew Green, Brandon Knight
03 – Develop Proof-of-Concept Dynamic Knowledge Graph (DKG)	Days 15-120	Develop routines for generating proof-of-concept DKG from identified data schema (i.e. ontology), generate simulated user input for DKG corrections, and integrate DKG with Virtualitics AI Platform. We will provide the TPOC with a progress report detailing outcomes of Task 3.	Sarthak Sahu, Brandon Knight

04 – Develop Proof of Concept AI/ML Inference Methods	Days 60-150	Develop proof-of-concept AI/ML-based Inference Methods that learn from simulated user input (for modifying DKG) and can automatically modify DKG.	Sarthak Sahu, Sagar Indurkhya
05 – Evaluate Technical Feasibility of Proof of Concept	Days 90-150	Apply AI/ML based Inference Methods to proof-of-concept DKG paired with simulated user input and evaluate technical feasibility using identified Key Performance Metrics. We will provide the TPOC with a progress report detailing outcomes of Tasks 4 & 5.	Sarthak Sahu, Sagar Indurkhya
06 – Review Feasibility Study Results and Write Final Report (DELIVERABLE)	Days 150-180	Virtualitics will review the initial test results of the developed proof-of-concept, identify limitations and opportunities for refinement, and incorporate information derived from tasks 02-05 to write the final report.	Sagar Indurkhya, Brandon Knight, Andrew Green

Task 01 - Kickoff Briefing

1. Kickoff Briefing with Technical Point of Contact (TPOC) in which Virtualitics will:
    a. Present TPOC with the capabilities of the Virtualitics AI Platform.
    b. Confirm with TPOC the outline and form of the feasibility study and the final report.

Task 02 – Customer Discovery

1. Virtualitics will design a data schema (i.e. ontology) for the proof-of-concept Dynamic Knowledge Graph that is relevant to expected use cases for USAF analysts (based on discussion with the government TPOC). Specifically, the data schema will specify the types of entities and the relationships (between various types of entities) that may appear within the Knowledge Graph.
2. Virtualitics will determine, on the basis of discussion with the government TPOC, specific forms of graph modifications that will be modeled using AI/ML-based inference methods during the feasibility study. Specifically, identified forms of graph modification should both align with the designed data schema and should also reflect the types of graph modifications USAF analysts make.
3. Virtualitics will identify (quantitative) Key Performance Metrics that will be used to evaluate the feasibility of utilizing AI/ML-based inference methods to automatically make graph modifications that correct for deficiencies or anomalies within a Dynamic Knowledge Graph. Specifically, these metrics will measure the accuracy of the AI/ML-based inference methods, and Virtualitics may determine that several different metrics are needed to measure the performance of different forms of inference methods. Examples of these metrics that may be considered include precision, recall and F1-score, and Area Under receiver operating characteristic Curve (AUC).

Task 03 - Develop Proof-of-Concept Dynamic Knowledge Graph (DKG)

1. Virtualitics will develop a proof-of-concept Dynamic Knowledge Graph (DKG) that is generated from the data schema designed in Task 2. Specifically, Virtualitics will develop routines (i.e.

computational procedures), implemented via the Python programming language and encapsulated within a Jupyter Notebook, that:

a. can be configured to generate, using the ontology developed in Task 02, DKGs of specified sizes (where the appropriate sizes are determined based on consulting with the government TPOC) to facilitate feasibility testing – note that the DKGs will be represented using the widely-used (open-source) *NetworkX* software library;
b. utilize random generation processes (e.g. sampling from a multinomial distribution that reflects the propensity of entity types appearing in the DKG) that take into account the structure of the generated DKG (e.g. by measuring network modularity as relationships are added to the graph) – importantly, the random generation process will be configurable by specification of a random-seed-value that enables deterministic reproduction of a generated DKG;
c. generate simulated user-inputs that modify the DKG per the identified forms determined in Task 2 -- these routines will be flexible and configurable to facilitate a variety of experiments during testing in Task 5 (e.g. the amount of user inputs, and how narrow an area within the graph the inputs impact).

2. Virtualitics will then integrate the Jupyter Notebook with Virtualitics Explore (a component of the Virtualitics AI Platform) so that relevant visualizations of the Knowledge Graph can be loaded into Explore to create interactive 3D visualizations of the Dynamic Knowledge Graph.
   a. This integration will leverage Virtualitics Python API, which enables Python code within Jupyter Notebooks to establish bidirectional communication with Virtualitics Explore.
   b. Virtualitics will work with the TPOC to identify specific 3D visualizations and interaction mechanisms that would be useful and provide enhanced situational awareness to end users (e.g. Virtualitics may augment nodes within the visualization of the network graph so that they can trigger actions in the Jupyter Notebook, such as removing relationships to other neighboring nodes).

Task 04 - Develop Proof of Concept AI/ML Inference Methods

1. Virtualitics will develop proof-of-concept AI/ML-based inference methods that are implemented using the Python programming language and encapsulated within the Jupyter Notebook (initially developed during Task 3) so that they may be considered to *augment* the proof-of-concept Dynamic Knowledge Graph. Examples of AI/ML-based inference methods that may be considered include: *Bayesian network classifiers, logistic regression, statistical difference in proportions, random forest, and rules-based expert systems*; final determination of the AI/ML-based inference methods to be implemented will be made after consulting with the government TPOC and will leverage open-source Python libraries for AI/ML such as *Scipy*, *Scikit-Learn*, *PyTorch* and *XGBoost*.
2. Virtualitics will demonstrate that the implemented AI/ML-based inference methods can first observe the (simulated) user-inputs modify the DKG and then be utilized to further modify the DKG.

Task 05 - Evaluate Technical Feasibility of Proof of Concept

1. Virtualitics will run a number of experiments that take the following form:
   a. a DKG is generated along with applicable synthetic user-inputs;
   b. a portion of the user-input is withheld (referred to as the "*test input data*"), with the remainder of the user-input constituting the "*training input data*";
   c. the AI/ML-based inference method observes the DKG modified via simulated application of the *training input data* and then outputs a set of additional graph modifications to be automatically applied;
   d. the AI/ML-based inference method is evaluated by computing applicable key performance metrics – *importantly, these key performance metrics correlate with how*

much the "test input data" overlaps with the output of the AI/ML-based inference method.

2. Virtualitics will consult with the TPOC to determine various operational parameters of the experiments carried out – e.g. what size DKGs should be evaluated, how much synthetic user input should be generated for evaluation, what subsets of the data schema should be used to generate the DKGs, etc.

Task 06 - Review Feasibility Study Results and Write Final Report

1. Virtualitics will review with the government TPOC the results of the experiments (carried out in Task 5) and discuss overall technical feasibility of the proof-of-concept DKG and AI/ML-based inference studied during the Phase 1; moreover, Virtualitics will work to identify limitations of the proof of concept, and determine:
    a. which limitations can be addressed during a subsequent Phase II – e.g. if there are concerns about whether one or more of the inference methods can be trained on larger volumes of data than used in the Phase I study;
    b. whether one or more of the inference methods should be recommended (on the merits of the experiments carried out in Task 5) for further development in a subsequent Phase II;
    c. how the capabilities of the Virtualitics AI Platform can be applied to develop a scalable, deployable software solution that encapsulates and extends the proof of concept developed during the Phase I study.
2. Virtualitics will produce and deliver a final report that aligns with expectations set with the TPOC at the beginning of the Phase I study (Task 01). The final report will include a comprehensive, detailed presentation of the work carried out during the Phase 1 study, including technical details about the DKG implemented in Task 3, the AI/ML-based inference methods implemented in task 4, and the experiments carried out during Task 5. Furthermore, the final report will include a discussion about limitations of the proof-of-concept identified during the feasibility study, as well as a plan and timetable for how the proof of concept (of the software solution) that was developed in this Phase I study could be actualized in a subsequent Phase II.

4. Related Work.

Virtualitics has carried out prior work involving Knowledge Graphs:

1. In a prior SBIR Phase II R&D effort carried out by Virtualitics under the guidance of a Department of Defense entity from 2021-2022, Virtualitics developed the AI-enabled TIR Analysis Tool (ATAT), which can be used to help the Test and Evaluation (T&E) operators evaluate Test Incident Reports (TIRs) more efficiently, thereby reducing the total time required for evaluation. The ATAT system consists of two interoperating subsystems: (i) the Knowledge Graph tool, which consumes a corpus of TIRs and produces a Knowledge Graph of TIRs that can be visualized, queried, and analyzed within Virtualitics Explore (a component of the Virtualitics AI Platform); (ii) the Plot Generation tool, which takes as input a TIR and associated sensor data and produces relevant plots that can be visualized and analyzed within Virtualitics Explore. Together, the Knowledge Graph tool and the Plot Generation tool, along with Virtualitics Explore, form the user-facing component of the ATAT system, which in turn provides a T&E operator with a flexible platform that can be used to analyze TIRs both individually and at a corpus level.
2. Virtualitics has worked extensively with a Department of Defense entity to apply the use of knowledge graphs, primarily focused on technological interrelationships in large quantities of data. This enabled analysts to quickly identify key nodes within large sets of data and have a visual representation of the interrelationships.
3. Virtualitics has worked with a Department of Defense entity to integrate the use of ontologies into the organizations standard workflow. The adaptability of both doing standard network representation and network extraction from flat data sets allow a comprehensive capability for both traditional and non-traditional analytic techniques.

Virtualitics has also carried out prior work involving PCAP Analysis:

1. Virtualitics has worked extensively with a Department of Defense entity to apply machine learning technology to cybersecurity data. Virtualitics was able to demonstrate automated network mapping and predictive model modeling from raw PCAP data by developing our own state of the art network spatialization algorithms by extending force-atlas 2D to three dimensions. We used community detection clustering on the 3D network spatialization to automatically isolate normal network traffic and ran network algorithms to find IPs with anomalously high betweenness centrality and load traffic, as such nodes are usually victims of botnet attack.
   a. This work was performed for an organization in the US Department of Defense
   b. This work was completed in October 2020.
2. Virtualitics is currently supporting a Cyber Pilot for a Department of Defense entity. The intent of the pilot is to combine Vulnerability data, Threat Intelligence, and Key Terrain data to create a unified view of risk across the network. The adoption of AI into the project will encompass automated device fingerprinting, device classification, and network classification. These elements will all be displayed into dashboard views specific to network owners, cyber defense analysts, vulnerability analysts, and intel analysts.

Beyond this, Virtualitics has worked on other complex projects in related areas, demonstrating our dedication to innovation. For example, our work on Integrated Resource Optimization (IRO), a Phase III SBIR project with the U.S. Air Force, involved working with NSN data for parts and equipment as they relate to maintenance and logistical systems. Moreover, our IRO solution uses AI to make recommendations for how an RSP could be tailored for specific mission parameters. Overall, Virtualitics has extensive experience working with Air Force logistics subject matter experts, and this experience has been crucial to developing the IRO application.

5. Relationship with Future Research or Research and Development.

(a) If the proposed approach for the feasibility study is successful, it will lead to the development (in a subsequent SBIR Phase II effort) of a full-scale prototype AI/ML-augmented Dynamic Knowledge Graph that assists USAF analysts by automating the process of updating and correcting DKGs based on ongoing analysis of the manual modifications the analyst makes to the DKG. More broadly, the approach outlined in this feasibility study will set the ground for equipping USAF analysts with DKGs that are AI-accelerated, enabling analysts to spend less time manually curating and updating the DKG and more time employing it for operational purposes.

(b) Overall, the Phase I feasibility study will be instrumental in defining the parameters within which, during Phase II, Virtualitics can develop the full-scale prototype of an AI/ML-augmented Dynamic Knowledge Graph using the Virtualitics AI Platform. Moreover, the proof-of-concept prototype we develop during the Phase I feasibility study can be used as a starting point for the Phase II research and development. Specifically, the Phase I feasibility study will set up the Phase II R&D effort for success by documenting the following: (i) what specific USAF analyst operational use cases the solution would address; (ii) the operational parameters of each use case, which includes attributes of representative DKGs ontologies as well as estimates of the size and network complexity of real DKGs; (iii) what key performance metrics will be used to evaluate each of the AI/ML-based inference methods, and how these key performance metrics relate to the identified use cases; (iv) how integration with Virtualitics Explore and making use of its interactive 3D visualization and its built in no-code Explainable AI routines can provide end-users with additional network situational awareness in which to assess recommended graph-modifications yielded by the AI/ML-based inference methods; (v) a specific plan, including a timeline, that lays out how the concept developed in Phase I could be actualized into a software solution in Phase II.

(c) Virtualitics holds an active Facility Clearance Level (FCL). Key members of our teams, spanning data science to cybersecurity, hold individual Secret, TS, and TS/SCI clearances. In line with our commitment to utmost security standards, Virtualitics is also CMMC-ready and in compliance with NIST 800-171 guidelines. The clearances and certifications currently held by Virtualitics will enable our data science team to seamlessly embark on data-centric tasks as Phase II initiates. Additionally, our robust experience in deploying software on secure DoD computing resources, inclusive of platforms like Iron Bank and networks like SIPR, ensures a smooth integration process. Collaborating closely with relevant stakeholders, we are confident in our capability to efficiently deploy our solution within the DoD computing environment.

6. Commercialization Strategy.

Virtualitics provides a powerful AI platform, making advanced analytic capabilities accessible to users of all experience levels, allowing them to analyze data swiftly and without bias. Our robust market presence is evidenced by a clientele featuring industry stalwarts like Lockheed Martin, Wells Fargo, and Isoplexis, contributing to our 30+ paid customers. This commercial success has been reflected in an 11x growth in sales over the last year across sectors including financial services, education, and manufacturing. Additionally, we've successfully raised significant funding from respected entities such as Centricus, The VR Fund, and Smith Point Capital.

In the defense domain, our engagements span contracts from the Air Force, Navy, and Army, with a cumulative worth of $23.9M across 23 projects. The Air Force Network Integration Center (AFNIC) has certified our software for use on key networks like NIPR and SIPR, and we've secured a TOP SECRET level Facilities Clearance. We're targeting expansion within the Air Force Global Strike Command, particularly focusing on units within the 8th Air Force. Our approach is twofold: initially liaise with officers for endorsement, followed by obtaining direct feedback from end-users.

As more organizations are beginning to realize the power of unlocking the numerous amounts of data at their fingertips, the use of DKGs will become more prominent. Within the DoD, Virtualitics feels this is an area we have a competitive advantage on and belief there is a great opportunity for commercialization. We feel that DKGs can be utilized on their own as well as form the backbone of AI Decision Support tools enabling humans to understand the data behind AI tools.

A testament to our R&D efficacy and commercialization strategy is our successful automation software module developed for the B-1 and B-52 aircrafts, which has subsequently been integrated into our general AI platform. We anticipate that the SBIR project will not only integrate seamlessly within the DoD ecosystem but will also resonate with federal agencies and the broader private sector.

Within the DoD, DKGs can be utilized within weapon system program offices, depots, headquarters staffs and to analyze data in support of problem definition, engineering analysis, readiness and resource optimization, and better data correlation for decision making. Upon successful completion of this SBIR Phase 1 effort, we feel a commercialized product can be created for sale within 12 months (3QCY26). We feel our competitive advantage along with our suite of AI products can further the spread of DKGs within the market and increase revenue. In the first year, there is potential to bring up to $5M in revenue fro the sale of DKGs within the DoD. This is compounded if Virtualitics is able to penetrate non-DoD markets. The use of DKGs can support financial decision making to support return on investment decisions as well as better identification of company and investment risks. Future details on market need, size, and commercialization results will further solidify our growth trajectory.

7. Key Personnel.

PI Name: Dr. Sagar Indurkhya
PI Education:
- Massachusetts Institute of Technology, PhD in Computer Science, 2021
- Massachusetts Institute of Technology, M.Eng in Computer Science, 2015
- Massachusetts Institute of Technology, B.E. in Computer Science, 2012

Relevant Experience

Dr. Sagar Indurkhya heads the AI Innovation group at Virtualitics, Inc. He has over eight years of experience with Artificial Intelligence (AI) and publications in top journals and conferences in the fields of AI and computational linguistics, as well as experience consulting with a number of companies that contract for the DoD. His research work has focused on high-precision semantic parsing, the development of computational models of language acquisition grounded in linguistic theory, and black-box analysis of deep neural network-based NLP systems. Dr. Indurkhya has a strong track record of using AI & ML to infer optimized solutions to complex problems ranging from automated inference of natural language grammars (which was the central topic of his doctoral thesis) to automated *Munition Storage Planning* (as part of SBIR Phase 2 work carried out for US Air Force that has been awarded a SBIR Phase 3). Dr. Indurkhya has also served as the principal investigator for several DoD R&D projects, including a SBIR Phase II for developing the *AI-enabled Test Incident Report Analysis Tool*.

Relevant Awards or Patents

- Conference Best Paper Award: *Indurkhya, S., & Berwick, R. C. (2021) Evaluating the Cognitively-Related Productivity of a Universal Dependency Parser. In 2021 IEEE 20th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) (pp. 7-15). IEEE.*
- Indurkhya, S., Martinez, H.J.V., Zanfardino, G., Donalek, C., (2024). Systems and Methods for Natural Language Querying (US12086134B2). U.S. Patent and Trademark Office.
- Martínez, H.J.V., Indurkhya, S., Zanfardino, G., Indurkhya, A., Sahu, S., Donalek, C., Amori, M. (2024). Systems and Methods for Network Explainability (US11928123B2). U.S. Patent and Trademark Office.

Selected Publications (See Full Bibliography: http://web.mit.edu/indurks/www/#publications)

1. Indurkhya, S. (2023) A Procedure for Inferring a Minimalist Lexicon from an SMT Model of a Language Acquisition Device. Proceedings of 16th edition of the International Conference on Grammatical Inference (ICGI), in Proceedings of Machine Learning Research pp. 217:35-58.
2. Indurkhya, S. (2022) Parsing as Deduction Revisited: Using an Automatic Theorem Prover to Solve an SMT Model of a Minimalist Parser. In Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL), pp. 157–175. Association for Computational Linguistics.
3. Indurkhya, S. (2020) Solving for Syntax. Doctoral Dissertation, Massachusetts Institute of Technology.
4. Indurkhya, S. (2020) Inferring minimalist grammars with an SMT-solver. Proceedings of the Society for Computation in Linguistics: Vol. 3, Article 56.
5. Indurkhya, S. Modeling Minimalist Grammars with SMT. Proceedings of the Seventh Annual
6. Conference on Cognitive Systems. (2019)
7. Indurkhya, S., (2019), Learning and Automata 2019. Automatic Inference of Minimalist Grammars using an SMT-Solver. LICS 2019.
8. Sprouse, J., Yankama, B., Indurkhya, S., Fong, S., & Berwick, R. C. Colorless green ideas do sleep furiously: gradient acceptability and the nature of the grammar. The Linguistic Review. (2018)

Sarthak Sahu (Head of Artificial Intelligence & Data Science): Sarthak holds a bachelor's degree in Computer Science from Caltech in 2017, focusing on machine learning and data science. He was hand selected by professors as one of the youngest teaching assistants for both the introductory machine learning course as well as the advanced data mining course during his time at Caltech. In addition, Sarthak also worked closely with Prof. Yisong Yue on both ongoing research as well as advanced independent research projects all tied to machine learning and data science. Before joining Virtualitics, Sarthak worked on designing robust, scalable software systems for classified government contracts at Raytheon and on creating advanced machine learning models for sports analytics at Second Spectrum, including a custom-built predictor which performed 2x better than any published model.

Andrew Green (VP of Customer Success): Andrew Green is the Vice President of Customer Success. He brings over a decade of experience in the Department of Defence to the Federal Team at Virtualitics. Andrew managed the Emerging Technologies Portfolio for Naval Special Warfare, securing over $40 million in funding for start-ups in just three years. As a Navy Chief Cryptologic Technician, he has led highly effective teams in the pursuit of National Mission and Intelligence Community requirements during Operation Enduring Freedom, Operation Inherent Resolve, and a multitude of tactical missions for the SOF community. Andrew is a founding member of the US Navy Artificial Intelligence Coalition and served as a trusted advisor to AI portfolio managers across the DOD. Andrew earned a Bachelors of Science in IT Management with Honors, graduated Summa Cum Laude with a Masters of Science in Integrated Design, Business, and Technology from the University of Southern California, and is a certified Project Management Professional.

Brandon Knight (Director of Customer Success - SOF/CYBER/IC): Brandon brings over 10 years of experience in Cyber and Intelligence Operations. During his active duty time he served as a Signals Intelligence Analyst with the following specialties: Target Digital Network Analyst, Defense Network Exploitation Analyst, Target Analyst Reporter, and Tactical Information Operations Operator directly under NSA, US Cyber Command, and Naval Special Warfare. After his time in service he worked within AT&T's Chief Security Office working to solve some of the most complex security issues at scale with the assistance of data scientists and software engineers. Brandon currently holds Security+ certification, a BS in Cybersecurity and Computer Networking, and a MS in Cybersecurity Operations and Leadership. His multi-domain experience in conjunction with our data scientists allow for the creation of tailor made and innovative cybersecurity solutions.

8. <u>Foreign Citizens</u>.

Virtualitics will not engage any personnel holding foreign or dual citizenship in the execution of work under this proposal. All team members assigned to this project are U.S. citizens.

9. <u>Facilities/Equipment</u>.

Virtualitics possesses all necessary facilities and equipment for the Phase I effort, negating the need for additional purchases from the USG. Our facilities comply with federal, state, and local environmental regulations, covering areas from emissions to waste disposal. We are self-sufficient and require no external resources for this project.

10. <u>Subcontractors/Consultants</u>.

Virtualitics will not be engaging with any subcontractors or consultants for the execution of this proposal. All work related to this project will be performed exclusively by our in-house team, ensuring full compliance with the stipulated requirements for small business concerns.

11. <u>Prior, Current or Pending Support of Similar Proposals or Awards</u>.

No prior, current or pending support has been provided for proposed work.

12. <u>Technical Data Rights</u>.

Virtualitics, Inc. acknowledges and agrees to the terms set forth by the Government concerning the rights in technical data and software developed under the resulting contract of this BAA. We hereby assert that the technical data and computer software termed as "Virtualitics AI Platform™" should have restricted rights in its use, release, or disclosure by the Government. Virtualitics AI Platform™, a commercial computer software platform, was developed exclusively at private expense by Virtualitics. It will serve as the primary software for the development, configuration, deployment of algorithms, analytics, workflow applications, and other capabilities proposed in this initiative. We emphasize that this platform falls under the "Commercial Technical Data License Rights," relating to commercial items developed at private expense and should be managed with Limited Rights. Furthermore, Virtualitics AI Platform™ will be licensed to the Government under terms customarily provided to the public, adhering to 48 C.F.R. 227.7202-1(a).