# CUI

## ATTENTION

| | |
|---|---|
| **Controlled by:** OUSD(R&E) | |
| **Controlled by:** Sponsoring DoD Agency/Component SBIR/STTR Program Office | |
| **CUI Categories:** Source Selection | |
| **Distribution Statement:** F | |
| **POC:** osd.ncr.ousd-r-e.mbx.sbir-sttr@mail.mil | |

"Distribution Statement F. Further distribution only as directed by [Sponsoring DoD Agency/Component SBIR/STTR Program Office] [Date of Determination: BAA End Date for SBIR/STTR Phase I Proposals or SBIR/STTR PII Proposal Receipt Date (Proposal Released to the Agency/Component)] or higher DoD authority."

## ATTENTION

All individuals handling this information are required to protect it from unauthorized disclosure.

Handling, storage, reproduction, and disposition of the attached document(s) must be in accordance with 32 CFR Part 2002 and applicable agency policy.

Access to and dissemination of Controlled Unclassified Information shall be allowed as necessary and permissible to any individual(s), organization(s), or grouping(s) of users, provided such access or dissemination is consistent with or in furtherance of a Lawful Government Purpose and in a manner consistent with applicable law, regulations, and Government-wide policies.

Standard Form 901 (11-18)
Prescribed by GSA/ISOO | 32 CFR 2002

CUI

**Small Business Innovation Research(SBIR) Program - Proposal Cover Sheet**

## Disclaimer

**Knowingly and willfully making any false, fictitious, or fraudulent statements or representations may be a felony under the Federal Criminal False Statement Act (18 USC Sec 1001), punishable by a fine of up to $10,000, up to five years in prison, or both.**

## SBIR Phase I Proposal

| | |
|---|---|
| Proposal Number: | **F244-0001-0094** |
| Proposal Title: | **Faster, Easier, and More Accurate Automated Updating of Dynamic Knowledge Graphs for Intelligence and Operations Using Graph Attention Networks and Update Impact Prediction Networks** |

## Agency Information

| | |
|---|---|
| Agency Name: | **USAF** |
| Command: | **AFMC** |
| Topic Number: | **AF244-0001** |

## Firm Information

| | |
|---|---|
| Firm Name: | **Cenith Innovations, LLC** |
| Address: | **1861 9th Avenue, Sacramento, CA 95818-4111** |
| Website: | **http://www.cenithinnovations.com** |
| UEI: | **ML3NGPZG8PS3** |
| DUNS: | **116925606** |
| CAGE: | **88BH3** |
| SBA SBC Identification Number: | **001616419** |

# Firm Certificate

## OFFEROR CERTIFIES THAT:

| | |
|---|---|
| 1. It has no more than 500 employees, including the employees of its affiliates. | **YES** |
| 2. Number of employees including all affiliates (average for preceding 12 months) | **35** |
| 3. The business concern meets the ownership and control requirements set forth in 13 C.F.R. Section 121.702. | **YES** |
| 4. Verify that your firm has registered in the SBAS Company Registry at www.sbir.gov by providing the SBC Control ID# and uploading the registration confirmation PDF: | **SBC_001616419** |

**Supporting Documentation:**

- [SBC_001616419 (3).pdf]

| | |
|---|---|
| 5. It has more than 50% owned by a single Venture Capital Owned Company (VCOC), hedge fund, or private equity firm | **NO** |
| 6. It has more than 50% owned by multiple business concerns that are VOCs, hedge funds, or private equity firms? | **NO** |
| 7. The birth certificates, naturalization papers, or passports show that any individuals it relies upon to meet the eligibility requirements are U.S. citizens or permanent resident aliens in the United States. | **YES** |
| 8. Is 50% or more of your firm owned or managed by a corporate entity? | **NO** |
| 9. Is your firm affiliated as set forth in 13 CFR Section 121.103? | **NO** |
| 10. It has met the performance benchmarks as listed by the SBA on their website as eligible to participate | **YES** |
| 11. Firms PI, CO, or owner, a faculty member or student of an institution of higher education | **NO** |
| 12. The offeror qualifies as a:<br>[ ] Socially and economically disadvantaged SBC<br>[ ] Women-owned SBC<br>[ ] HUBZone-owned SBC<br>[**X**] Veteran-owned SBC<br>[**X**] Service Disabled Veteran-owned SBC<br>[ ] None Listed | |
| 13. Race of the offeror:<br>[ ] American Indian or Alaska Native<br>[ ] Native Hawaiian or Other Pacific Islander<br>[ ] Asian<br>[**X**] White<br>[ ] Black or African American<br>[ ] Do not wish to Provide | |
| 14. Ethnicity of the offeror: | **NON-HISPANIC** |
| 15. It is a corporation that has some unpaid Federal tax liability that has been assessed, for which all judicial and administrative remedies have not been exhausted or have not lapsed, and that is not being paid in a timely manner pursuant to an agreement with the authority responsible for collecting the tax liability: | **FALSE** |
| 16. Firm been convicted of a fraud-related crime involving SBIR and/or STTR funds or found civilly liable for a fraud-related violation involving federal funds: | **NO** |
| 17. Firms Principal Investigator (PI) or Corporate Official (CO), or owner been convicted of a fraud-related crime involving SBIR and/or STTR funds or found civilly liable for a fraud-related violation involving federal funds: | **NO** |

**Signature:**

| Printed Name | Signature | Title | Business Name | Date |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Kristopher Pruitt | Kristopher Pruit t | CEO | Cenith Innovations, LLC | 06/19/2020 |

# Audit Information

**Summary:**

Has your Firm ever had a DCAA review?**NO**

# VOL I - Proposal Summary

**Summary:**

| | |
|---|---|
| Proposed Base Duration (in months): | **6** |

**Technical Abstract:**

Department of Defense and Intelligence Community analysts and operators providing situational awareness, building patterns of life, performing threat detection, or conducting targeting operations have seen an overwhelming increase in data volume, velocity, and variety for decades. These analysts and operators face numerous priorities, including more complex and time-sensitive questions from commanders and decision-makers. Limited by time and technical tools to perform advanced data science on incoming data, they often must leave behind tranches of valuable unanalyzed data and delay updates to intelligence databases such as knowledge graphs. Knowledge graphs are powerful tools for understanding the world and supporting the types of analysis mentioned above, because they offer a structured, highly interconnected way of organizing and representing knowledge.  The knowledge graphs' capability to model complex relationships between entities allows them to aid analysts and other users in discovering insights and asking questions that other data structures often do not provide or allow. Knowledge graphs are the basis for modern internet searches and make Large Language Models and tools like ChatGPT more accurate and reliable. In a world with ever-increasing volume, velocity, and variety of data, knowledge graphs are pivotal in bringing the DoD and IC technology to a level where users can ask their intelligence systems the same questions they ask Google and ChatGPT in their everyday lives. A major part of delivering this capability to the DoD and IC will be creating more efficient, transparent, and accurate methods to update knowledge graphs. Currently, updating knowledge graphs and other intelligence databases often relies on time-consuming, heuristic-based, manual methods that can be arbitrary and oversimplified compared to reality. This can lead to knowledge graphs becoming brittle, unreliable, and inaccurate. Cenith Innovations proposes applying recent advances in Graph Neural Networks and novel attention mechanisms to build and test a more efficient, transparent, and accurate capability to make it easier for users to interpret, update, and organize knowledge graphs. This capability will allow users or automated data feeds to make changes and additions to the graph and have those changes recommend additional updates that should be made to the knowledge graph. We propose pursuing both a more

manual method by recommending changes to be verified by users, as well as a more automated and scalable method by building a transparent system to automatically update knowledge graph changes that users will trust.

## Anticipated Benefits/Potential Commercial Applications of the Research or Development:

By enhancing the scalability and accuracy of dynamic knowledge graphs, analysts can focus more on deriving insights rather than manual data management. We will test and prove the feasibility of using Graph Neural Networks, specifically Graph Attention Networks, and a novel attention mechanism we identify as an Update Impact Prediction Network to make updating the DOD and IC's dynamic knowledge graphs more scalable, accurate, adaptable, and easier. We will focus on ways to massively decrease the burden on users to upkeep knowledge graphs, as well as promote a transparent approach that helps build trust with users in an automated update system.

## Attention:



## Addition:

Enter the page numbers separated by a space of the pages in the proposal that are considered proprietary:

**1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17**

List a maximum of 8 Key Words or phrases, separated by commas, that describe the Project:

**Knowledge Graph, AI/ML, GNNs, Update Impact Prediction Network, data management, knowledge management, GANs, LLM**

# VOL I - Proposal Certification

## Summary:

| | |
|---|---|
| 1. At a minimum, two thirds of the work in Phase I will be carried out by your small business as defined by 13 C.F.R Section 701-705. The numbers for this certification are derived from the budget template. To update these numbers, review and revise your budget data. If the minimum percentage of work numbers are not met, then a letter of explanation or written approval from the funding officer is required.<br>Please note that some components will not accept any deviation from the Percentage of Work (POW) minimum | **YES** |

| | |
|---|---|
| requirements. Please check your component instructions regarding the POW requirements. | |
| Firm POW | **100%** |
| Subcontractor POW | **0%** |
| 2. Is primary employment of the principal investigator with your firm as defined by 13 C.F.R Section 701-705? | **YES** |
| 3. During the performance of the contract, the research/research and development will be performed in the United States. | **YES** |
| 4. During the performance of the contract, the research/research and development will be performed at the offerors facilities by the offerors employees except as otherwise indicated in the technical proposal. | **YES** |
| 5. Do you plan to use Federal facilities, laboratories, or equipment? | **NO** |
| 6. The offeror understands and shall comply with export control regulations. | **YES** |
| 7. There will be ITAR/EAR data in this work and/or deliverables. | **YES** |
| 8. Has a proposal for essentially equivalent work been submitted to other US government agencies or DoD components? | **NO** |
| 9. Has a contract been awarded for any of the proposals listed above? | **NO** |
| 10. Firm will notify the Federal agency immediately if all or a portion of the work authorized and funded under this proposal is subsequently funded by another Federal agency. | **YES** |
| 11. Are you submitting assertions in accordance with DFARS 252.227-7017 Identification and assertions use, release, or disclosure restriction? | **NO** |
| 12. Are you proposing research that utilizes human/animal subjects or a recombinant DNA as described in DoDI 3216.01, 32 C.F.R. Section 219, and National Institutes of Health Guidelines for Research Involving Recombinant DNA of the solicitation: | **NO** |
| 13. In accordance with Federal Acquisition Regulation 4.2105, at the time of proposal submission, the required certification template, "Contractor Certification Regarding Provision of Prohibited Video Surveillance and Telecommunications Services and Equipment" will be completed, signed by an authorized company official, and included in Volume V: Supporting Documents of this proposal.<br><br>NOTE: Failure to complete and submit the required certifications as a part of the proposal submission process may be cause for rejection of the proposal submission without evaluation. | **YES** |
| 14. Are teaming partners or subcontractors proposed? | **NO** |
| 15. Are you proposing to use foreign nationals as defined in 22 CFR 120.16 for work under the proposed effort? | **NO** |
| 16. What percentage of the principal investigators total time will be on the project? | **51%** |
| 17. Is the principal investigator socially/economically disadvantaged? | **NO** |
| 18. Does your firm allow for the release of its contact information to Economic Development Organizations? | **YES** |

# VOL I - Contact Information

## Principal Investigator

Name: **Mr. Bryan O'Rourke**

| | |
|---|---|
| Phone: | **(661) 312-3063** |
| Email: | **bryan@cenithinnovations.com** |
| Address: | **61142 Ridge Falls Place, Bend, OR 97702 -** |

## Corporate Official

| | |
|---|---|
| Name: | **Kristopher Pruitt** |
| Phone: | **(916) 707-3178** |
| Email: | **kris@cenithinnovations.com** |
| Address: | **1861 9th Avenue, Sacramento, CA 95818 - 4111** |

## Authorized Contract Negotiator

| | |
|---|---|
| Name: | **Dane Kaiser** |
| Phone: | **(207) 441-5569** |
| Email: | **dane@cenithinnovations.com** |
| Address: | **1861 9th Avenue, Sacramento, CA 95818 - 4111** |

# Faster, Easier, and More Accurate Automated Updating of Dynamic Knowledge Graphs for Intelligence and Operations Using Graph Attention Networks and Update Impact Prediction Networks

## Volume 2: Technical Volume

### Graph Attention Networks and Novel Attention Mechanisms for Interpreting, Updating, and Organizing Knowledge Graphs

Department of Defense (DoD) and Intelligence Community (IC) analysts and operators providing situational awareness, building patterns of life, performing threat detection, or conducting targeting operations have seen an overwhelming increase in data volume, velocity, and variety for decades. These analysts and operators face numerous priorities, including more complex and time-sensitive questions from commanders and decision-makers. Limited by time and technical tools to perform advanced data science on incoming data, they often must leave behind tranches of valuable unanalyzed data and delay updates to intelligence databases, including knowledge graphs. Knowledge graphs are powerful tools for understanding the world and supporting the types of analysis mentioned above because they offer a structured, highly interconnected way of organizing and representing knowledge.[1 2] The knowledge graph's capability to model complex relationships between entities allows them to aid analysts and other users in discovering insights and asking questions that other data structures often do not provide or allow. Knowledge graphs are the basis for modern internet searches and make Large Language Models and tools like ChatGPT more accurate and reliable. In a world with ever-increasing volume, velocity, and variety of data, knowledge graphs are pivotal in bringing the DoD and IC technology to a level where users can ask their intelligence systems the same questions they ask Google and ChatGPT in their everyday lives.

A major part of delivering this capability to the DoD and IC will be creating more efficient, transparent, and accurate methods to update knowledge graphs. Currently, updating knowledge graphs and other intelligence databases often relies on time-consuming, heuristic-based, manual methods that can be arbitrary and oversimplified compared to reality. This can lead to knowledge graphs becoming brittle, unreliable, and inaccurate. Cenith Innovations proposes applying recent advances in Graph Neural Networks (GNNs) and novel attention mechanisms to build and test a more efficient, transparent, and accurate capability to make it easier for users to interpret, update, and organize knowledge graphs. This capability will allow users or automated data feeds to make changes and additions to the graph and have those changes recommend additional updates that should be made to the knowledge graph. We propose pursuing both a manual method that will recommend changes to be verified by users and a more automated and scalable method by building a transparent system that automatically updates knowledge graph changes that users will trust.

During this Phase I effort, Cenith Innovations will test and prove the feasibility of using GNNs, specifically Graph Attention Networks (GATs), and a novel attention mechanism we identify as an "Update Impact Prediction Network" (UIPN) to make updating the DOD and IC's dynamic knowledge

[1] Ji, Pan, Marttinen, Yu; "A Survey on Knowledge Graphs: Representation, Acquisition and Applications", IEEE Transactions on Neural Networks and Learning Systems, arxiv.org/abs/2002.00388, April 2021
[2] Peng, Xia, Naseriparsa, Osborne, "Knowledge Graphs: Opportunities and Challenges", arxiv.org/pdf/2303.13948, March 2023

graphs more scalable, accurate, adaptable, and easier. During this effort, we will focus on ways to massively decrease the burden on users to upkeep knowledge graphs, as well as promote a transparent approach that helps build trust with users in an automated update system.

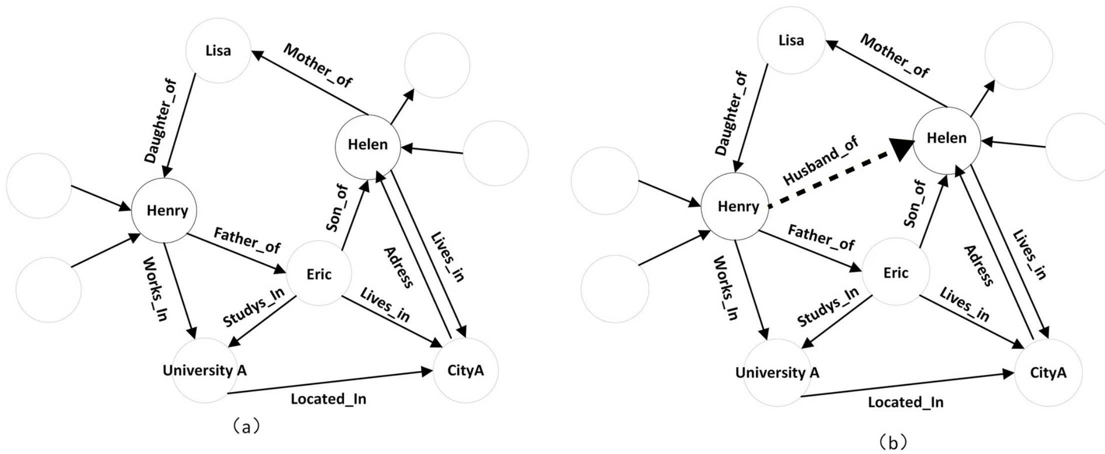


**Figure 1:** A common knowledge graph relationship node structure shown on the left (a) followed by a potential example of the employment of GATs unlocking previously unconnected node structures.

## Phase I Technical Objectives

### Introduction to Proposed Project:

Our proposed research and testing explores the feasibility of applying GATs with UIPNs, our novel attention mechanism,[3] to automate and increase the accuracy of updating and analyzing knowledge graph data to include flagging important updates for users to verify for approval. Delivering this capability will allow users to focus more on interpreting insights instead of manually managing data.

Additionally, since users may receive intelligence on numerous entities from numerous sources, "updating the graph" should not be limited to only entities already represented in the graph, so we have included work to use GATs and UIPNs to efficiently, intuitively, and accurately add, remove, and update nodes and edges from external, non-graph data sources.

Our proposed technical approach using GATs and UIPNs will make updating and managing dynamic knowledge graphs more scalable, accurate, adaptable, and easier. Our approach is differentiated from other approaches that only use Graph Convolutional Networks (GCNs) or GATs by themselves because of the addition of our UIPN, and the result will be:

- Scalable: Enables the benefits of GATs to efficiently handle large-scale knowledge graphs by limiting updates to affected nodes.[4] However, our proposed addition of the UIPN will enable higher-quality updating across a larger neighborhood of nodes compared to GATs or other types of GNNs alone.
- Accurate: Reality doesn't follow the heuristics leveraged by GATs alone to determine which knowledge graph elements should be updated (i.e., neighborhood size). Our UIPN is trained

[3] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin; "Attention is All You Need", 31st Conference on Neural Information Processing Systems, arxiv.org/pdf/1706.03762, June 2017
[4] Velickovic, Cucurull, Casanova, Romero, Lio, Bengio; "Graph Attention Networks", Published as a conference paper at International Conference on Learning Representations (ICLR) 2018, arxiv.org/pdf/1710.10903; April 2018

specifically to offer more accurate predictions about which graph elements should be updated and, therefore, improves the accuracy of automatically updated nodes and edges. This applies especially to long sequences of interconnected nodes across knowledge graphs.
- Adaptable: The GAT and UIPN learn and adapt to new patterns of dependencies as the underlying data evolves, keeping automated updates scalable and accurate, as described above.
- Easier and Faster: By simultaneously pursuing a manual verification of recommended changes and also building a capability to automatically make updates that will be more scalable in the future, our capability will greatly minimize the need for manual interventions and corrections.

Furthermore, this research seeks to lay the groundwork for advanced analysis capabilities facilitated by GATs, such as pattern recognition and anomaly detection, which can help users uncover hidden connections and generate actionable intelligence, ultimately improving decision-making processes.

**Introduction to Knowledge Graphs and GNNs:**

Knowledge graphs are a data structure that represents information in a graph format, with real-world entities depicted as nodes and their relationships as edges. This structure allows for capturing complex interconnections between data points, which makes it easier to understand and analyze relationships within a given domain. Knowledge graphs can integrate diverse data sources, infer new knowledge, and provide a contextual understanding of information.[5] These uses are particularly valuable in applications like search engines, recommendation systems, and intelligence analysis.

GNNs are a class of neural networks designed to perform inference on data represented as graphs. Unlike traditional neural networks that operate on fixed-size input (like images or sequences), GNNs can handle the irregular structure of graphs, making them ideal for modeling relational data where entities (nodes) and their relationships (edges) are of primary interest.[6]

Graph Attention Networks are a type of GNN, but the defining characteristic of a GAT is its use of an attention mechanism to assign different weights to different neighboring nodes. This allows the GAT to focus more on the most relevant nodes in the graph for each node's representation and makes GATs particularly effective for tasks where not all neighboring nodes contribute equally to a node's feature.[7]

**Project Key Objectives:**

To prove the capability to intuitively and efficiently support users' interaction with and updating of knowledge graphs, our research will test and measure the feasibility of the following six objectives:

1. Research and test methods to convert non-graph, tabular data into graph topology suitable for GATs, allowing the construction of nodes, edges, and embeddings that integrate seamlessly with existing graph structures and ontologies.
2. Employ GATs and our novel UIPNs to determine related updates across the graph when making changes, ensuring consistency and coherence. Enable capabilities for users to verify recommended changes and to allow automated updates where appropriate.

[5] Chaudhri, Chittar, Chenesereth; "An Introduction to Knowledge Graphs", ai.stanford.edu/blog/introduction-to-knowledge-graphs/, May 2021
[6] Sanchez-Lengeling, Reif, Pearce, Wiltschko, "A Gentle Introduction to Graph Neural Networks", distill.pub/2021/gnn-intro/, September 2021
[7] Sanchez-Lengeling, Reif, Pearce, Wiltschko, "A Gentle Introduction to Graph Neural Networks", distill.pub/2021/gnn-intro/, September 2021

3. Identify methods and best practices to ensure consistency during the GAT's initial training and through user-provided feedback in operations. Include engagement with users to build best practices.
4. Use GATs and UIPNs to intelligently filter, batch, and prioritize human reviews of nodes and edges that were automatically updated by GATs and UIPNs. Include engagement with users to identify best practices.
5. Provide metrics at different levels of the graph (entire graph, subsets of a graph, embeddings across multiple nodes and edges, specific nodes and edges) to keep users aware of when and how the knowledge graph is changing over time based on automatic and manual updates. Include engagement with users to identify best practices.
6. Record snapshots of the GATs and UIPNs over time for use in subsequent graph analytics or other analytics.

We will use three data sets to test the feasibility of our approaches.

- The Panama Papers graph database created by the International Consortium of Investigative Journalists captures a tangled web of offshore accounts and is aligned with defense and intelligence use cases.
- The Armed Conflict Locator Events Database (ACLED) includes a graph of all battles, protests, and violent events in conflicts across the globe, including the identification of actors, and is aligned with defense and intelligence use cases.
- Open Graph Benchmark's ogbl-collab is an undirected graph representing a subset of the collaboration network between research paper authors indexed by MAG. Each node within the graph represents an author, and edges indicate the collaboration between authors. This dataset is commonly used to measure the performance of GNNs in the scientific community.

Additionally, we will use neo4j as our knowledge graph engine because of its integration with a well-known Cypher query language, large user community, and ability to transfer neo4j graph formats to other graph engine formats if necessary or desired.

**Objective #1: Research and test methods to convert non-graph, tabular data into graph topology suitable for GATs, allowing the construction of nodes, edges, and embeddings that integrate seamlessly with existing graph structures and ontologies.**

Intelligence analysts and other knowledge graph users rely on multiple data sources in varied formats to improve their analyses. This includes data from sources ***external*** to the graph that may be delivered in tabular, relational, or textual formats and that may be incorporated manually or through pre-built data pipelines. Users need help efficiently and accurately integrating these external data sources efficiently and effectively into their knowledge graphs and analyses.

We will achieve this objective by focusing first on tabular data. Textual data is another data type to focus on in future efforts. The first step for tabular data will be to preprocess the data into nodes and edges.

We will treat each row as a potential node. Then, a GAT will use each row's attributes, including spatial or other attribute data, such as a vehicle type, a person's name, or a military force name, to create an edge between a record in the tabular data and a node in the knowledge graph. For example, if a row recording a T-80 tank shares the same "Unit Name" attribute value with a node in the knowledge graph describing an Armored Tank Battalion, then the GAT will create a new "T-80 tank" node in the knowledge graph from the tabular data and create an edge between it and the Armored Tank Battalion.

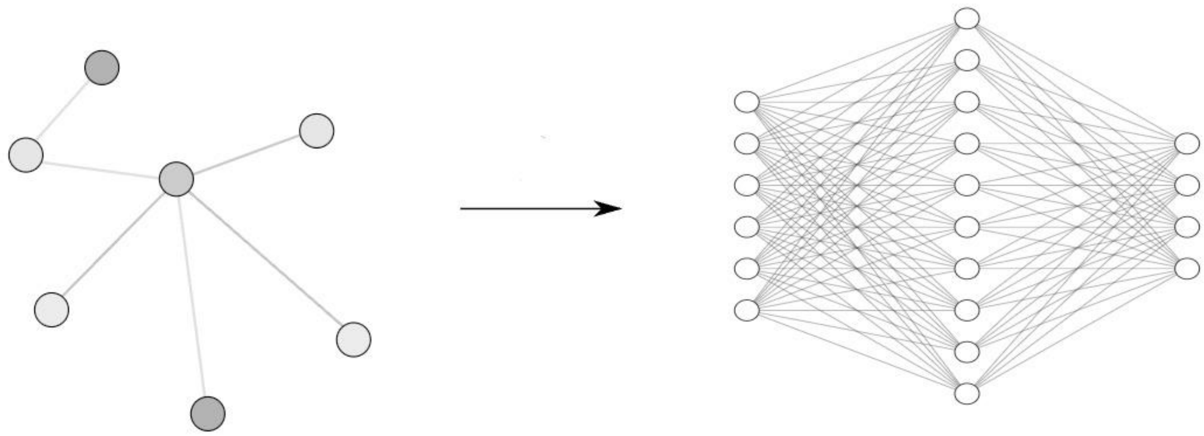

**Figure 2:** A notional example showing the translation of individual nodes from a knowledge graph to nodes within the GAT.

We will use GATs to improve the accuracy and scalability of the required entity matching, relationship definition, edge creation, and feature engineering steps that need to occur for tabular data[8 9 10].

- GATs will assist in automating entity matching through learned similarity measures, which will make matching more scalable and accurate than traditional methods of similarity analysis, such as string matching.
- GATs will assist with relationship definition by identifying patterns in the graph's structure and learned embeddings. This will add new nodes that are more consistent and rich for the user's context.
- GATs will suggest edge attributes for edge creation by analyzing existing nodes and their relationships. They will use attributes from nodes and tabular data to ensure the correct relationship is defined. This will often include spatial and other data contexts.
- GATs will assist in identifying the relevant features to extract from the tabular data and convert them into the graph for feature engineering. Additionally, they may convert attributes into the appropriate ingestible representations for the graph.

We expect these steps to deliver value because they will consider local and global knowledge graph contexts. This method will help determine where newly added nodes and edges best fit in the existing topology to keep the graph accurate and relevant.

**Objective #2: Employ GATs and our novel UIPNs to determine related updates across the graph when making changes, ensuring consistency and coherence. Enable capabilities for users to verify recommended changes and to allow automated updates where appropriate.**

Dynamic knowledge graphs allow users to externalize their knowledge in a digital format that can capture the complex, non-linear dependencies that traditional data structures and rule-based systems do not

[8] Zhu, Ma, Wang; "RAGA: Relation-aware Graph Attention Networks for Global Entity Alignment", arxiv.org/abs/2103.00791, March 2021
[9] Ji, Hui, Luo; "Graph Attention Networks With Local Structure Awareness for Knowledge Graph Completion", Institute of Electrical and Electronics Engineers, ieeexplore.ieee.org/document/9292922, December 2020
[10] Han, Liu, Zhang, Li; "Hierarchical Perceptual Graph Attention Network for Knowledge Graph Completion",

capture. Effective knowledge graphs must also be updated efficiently and accurately so that users don't spend all their time managing their knowledge but instead apply and deliver it to warfighters.

Our proposed approach is designed to improve performance and deliver better technical and mission results. By offering efficient update propagation that avoids unnecessary computations, our method is scalable for the very large, dynamic graphs that the DoD and IC use for intelligence analysis and knowledge management. Incorporating the UIPN further boosts our efficiency and enables us to focus on the most impacted nodes, thereby improving prediction accuracy. The joint operation of GAT and UIPN, adjusting thresholds for node updates based on learned patterns, ensures flexibility in different contexts, further enhancing our efficiency.

We will accomplish this objective by testing the application of six steps to recommend nodes be updated and then verified by users, as well as establish the capability to automatically update (when appropriate) nodes, edges, and their attributes and embeddings across the graph. Our model training approach, outlined in Objective #3, is intertwined throughout these six steps, which are listed below in this objective. Enabling users to verify automatically recommended updates to the knowledge graph will aid in our model training process and help build a repeatable process that will enable users to understand how a knowledge graph can be automatically updated in the future.

We will also add the UIPN element to the architecture. The UIPN is a neural network built upon the GAT. It is trained on the connections between nodes and enhanced with additional attention mechanisms. The UIPN's input will be any updated node or edge embedding or contextual information from neighboring nodes; its output will be a probability distribution over other nodes that indicates the likelihood that each node will require an update.

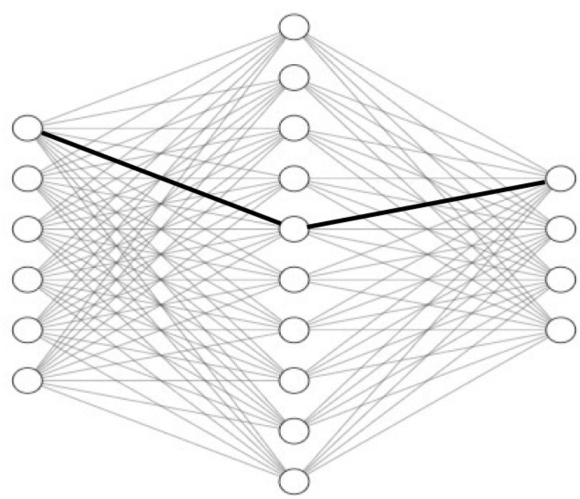

**Figure 3:** A notional example of the node structure affected by modified weights used by the described UIPN.

The additional attention mechanisms will focus the model's capacity on the most relevant parts of the knowledge graph by assigning attention weights to nodes based on their relevance to the updated node. This allows the UIPN to prioritize nodes that are more likely to be affected and to recognize nodes that are connected along long chains of connections. The six steps we will test are:

1. Initial Embedding Computation: Compute embeddings for all nodes and edges in the knowledge graph using their attributes and initial states.
2. Dependency Modeling with GAT: The GAT will process the graph to learn the dependencies between nodes. The node embeddings are updated by aggregating information from neighboring nodes through message passing.

3. Node Update Event: After a node or edge is updated, automatically or manually, its embedding is recalculated to reflect the new state. This updated embedding captures the change in the node's attributes and connections.
4. Update Impact Prediction: The updated node's embedding is fed into the UIPN, which uses the GAT's learned dependencies and attention mechanisms to assess the update's impact on other nodes. The UIPN outputs a probability score for each node that indicates the need for an update.
5. Propagation of Updates: Nodes with probability scores above a certain threshold determined by the GAT are selected for updating. These recommended updates can either be verified as correct or incorrect by a user or automatically propagated through portions of a knowledge graph or the entire knowledge graph. After updating, the nodes' and edges' embeddings are recalculated, and the process is iterated to capture multi-hop dependencies across multiple nodes.
6. Dynamic Adjustment: The UIPN and GAT integrated system continuously learns and adjusts the thresholds and attention weights based on user feedback, improving over time.

**Objective #3: Identify methods and best practices to ensure consistency during the GAT's initial training and through user-provided feedback in operations. Include engagement with users to build best practices.**

The GAT's initial training must be consistent with how ongoing training will occur in operations, so it is vital to judge tested approaches based on both model performance and the feasibility of ongoing model training via collecting user feedback when in operations.

Our approach for training GATs and UIPNs includes three main steps: **data collection, design of a loss function, and optimization.** We will **collect data** by simulating user-created, manual node and edge updates, additions, and removals, as well as creating new nodes and edges based on data meant to represent external, tabular data that a user would want to add to the graph. We will create this simulated data in an inductive environment to add new nodes or edges to a graph and in a transductive environment to update existing nodes or edges. Simulated data for existing node and edge updates will include randomized node and edge updates, removals, additions, and human, expert-directed updates. Simulated data for new node and edge additions will involve the removal of select nodes and edges in the knowledge graph and then re-adding them as training data points. During model operations, we will collect data on node and edge updates and the subsequent impact on other nodes and edges in the knowledge graph. This historical update data will be filtered, prioritized, and presented to users to identify as good or bad updates to neighboring nodes or edges.

We will **design a loss function** that penalizes incorrect predictions of node updates and balances precision and recall. Incorrect predictions will be identified by humans during our sampled, prioritized human assessment of automated changes, as described in Research Objective #4.

Lastly, for **optimization**, we will test using backpropagation to train the GATs and UIPNs jointly, which will ensure the model learns both the structural dependencies and dynamic propagation patterns.

We will conduct user-engagement sessions virtually or in person to verify assumptions and record best practices.

**Objective #4: Use GATs and UIPNs to intelligently filter, batch, and prioritize human reviews of nodes and edges that have been automatically updated by GATs and UIPNs. Include engagement with users to identify best practices.**

GATs are trained on a specific knowledge graph, so it will be necessary to leverage users' expertise to provide feedback to the model to improve its training and performance. Requiring users to provide training input to models in the middle of their analytic workflow can be annoying and disruptive because it can distract them from priorities, so we will identify ways to improve how users provide feedback to recommended knowledge graph updates that are identified below.

While the DoD has collected much information on the effectiveness of labeling data for Computer Vision use cases, the same level of lessons learned and information does not exist regarding labeled data for DoD knowledge graphs. We will seek to leverage best practices from Computer Vision labeling efforts and leverage our connections and experience in the DoD Computer Vision community so we can apply their expertise to GATs.

We will test the feasibility of effectively batching and prioritizing records for feedback and offer Active Learning options using the attention mechanisms and outputs of the UIPN to intuitively and easily collect feedback from users as they interact with the graph. For example, based on performance and user feedback, we can implement a strategy to batch updates of different categories or groups of nodes and cue users to complete these regularly. We can simultaneously provide Active Learning options for nodes or edges that we identify as more critical to the knowledge graph based on graph metrics.

We will ask users whether the GATs' and UIPNs' updates to a node or edge were good or bad, then use that feedback to continuously compute the graph embeddings and improve the UIPN and GAT through backpropagation as described in Objective #3. Much of the work in this objective will jointly accomplish training the model while also providing users the ability to verify recommended updates to the knowledge graph before propagating changes throughout it.

The users' collected feedback on adding new nodes or edges will match the inductive environment of the initial training for new nodes and edges, and feedback collected on the updating of existing nodes or edges will match the transductive environment of the initial training for existing nodes and edges.

We will conduct user-engagement sessions virtually or in person to verify assumptions and record best practices.

**Objective #5: Provide metrics at different levels of the graph (entire graph, subsets of a graph, embeddings across multiple nodes and edges, specific nodes and edges) to keep users aware of when and how the knowledge graph is changing over time based on automatic and manual updates. Include engagement with users to identify best practices.**

Automated methods for knowledge management and updating knowledge graphs in intelligence analysis must be measured objectively over time to build trust and rapport with analysts and other users. Users will quickly lose any time savings if they feel they must investigate the provenance of an automated update to a knowledge graph or if they can't easily understand why and how it was made. This is why, as part of our research and feasibility testing, we will research the effectiveness of various graph metrics at the node, subgraph, and graph level to allow users to understand how their knowledge graph is changing over the short and long term.

We will build graph metrics and survey users and other potential users to determine the most useful metrics for building trust and rapport with users. We will engage with potential users, either virtually or in person, to review several families of graph metrics and determine the best options to pursue:

- Graph Structure Metrics: Number of nodes and edges, degree distribution, connectivity metrics, graph components, clustering coefficient, graph centrality
- Graph Content Metrics: Attribute completeness, attribute changes, node and edge types, label changes
- Graph Quality Metrics: Consistency metrics, redundancy metrics, data completeness, accuracy metrics
- Dynamics Metrics: Change frequency, volatility metrics, temporal patterns
- Usage Metrics: Query volume and coverage, feedback and corrections, human review metrics

Cenith Innovations has extensive experience conducting user feedback sessions with analysts and operators. We will leverage our Design Thinking-based workshops to run interactive surveys using digital tools like iPads and survey tools such as Mentimeter to facilitate user sessions and record feedback in quantitative data that can be used in this Phase I project and beyond.

Ccnith Innovations proposes emphasizing this as part of a Phase I effort because even if a system is built that allows for faster, more automated updating of knowledge graphs, we will not achieve user adoption without a mechanism to explain the updates and characterize how they're affecting the overall graph over time.

We will conduct user-engagement sessions, virtually or in person, to verify assumptions and record best practices.

**Objective #6: Record snapshots of the GATs and UIPNs over time for use in subsequent graph analytics or other analytics.**

Analysts' and other users' knowledge is a living, breathing artifact because they must always adapt to new knowledge and circumstances. Representations of their knowledge in dynamic knowledge graphs should be just as responsive and adaptable. We expect that users from an operational and technical perspective will be interested in answering questions about how to analyze past versions of the graph that may have had differing numbers of nodes, connections between nodes, and other changed aspects.

For example, when knowledge graphs are updated automatically, it will be necessary for users to quickly comprehend what has changed in their graph since they last looked at a particular section of the graph in a specific context. We will research the application of Delta Graph Visualizations and Temporal Evolution Charts using graph snapshots to help users understand how and when their graphs and subsections are changing based on automated updates. This will also include the identification of structural anomalies and outliers that could indicate data quality issues or the addition of important new information. Similar to Objective #5, we will conduct engagement sessions with potential users, in person or virtually, to gather feedback on the best methods. We will leverage our extensive experience and Design Thinking approach to run interactive surveys using digital tools like iPads and survey tools such as Mentimeter to facilitate sessions with users and record feedback in quantitative data that can be used in this Phase I project and beyond.

**Plan to Accomplish Research Objectives:**

**Objective #1: Research and test methods to convert non-graph, tabular data into graph topology suitable for GATs, allowing the construction of nodes, edges, and embeddings that integrate seamlessly with existing graph structures and ontologies.**

Days 0-30: Initial design and testing by ingesting non-graph data into a dynamic knowledge graph

Days 30-60: Conduct user engagement to collect feedback on initial approaches

Days 60-90: Iterate on the design and methods based on user feedback

Days 90-120: Finalize design and methods in preparation for final testing

Days 120-150: Conduct final testing and record results of methods for inclusion in the final report

Days 150-180: Incorporate research objective findings into the final report and prototype design

**Objectives #2 & 3: Employ GATs and our novel UIPNs to determine related updates across the graph when making changes, ensuring consistency and coherence. Enable capabilities for users to verify recommended changes and to allow automated updates where appropriate. Identify methods and best practices to ensure consistency during the GAT's initial training and through analyst-provided feedback in operations. Include engagement with users to build best practices.**

Days 0-30: Initial setup of dynamic knowledge graphs, GATs, and UIPNs

Days 30-60: Initial training of GATs and UIPNs

Days 60-90: Initial testing of updating knowledge graph elements on the fly and using GATs and UIPNs to update appropriate nodes and edges

Days 90-120: Based on initial testing, iterate on overall GAT and UIPN training design and methods and implement changes in preparation for final testing

Days 120-150: Conduct final testing and record results of methods for inclusion in the final report

Days 150-180: Incorporate research objective findings into the final report and prototype design

**Objectives #3 & 4: Identify methods and best practices to ensure consistency during the GAT's initial training and through user-provided feedback in operations. Include engagement with users to build best practices. Use GATs and UIPNs to intelligently filter, batch, and prioritize human reviews of nodes and edges that have been automatically updated by GATs and UIPNs. Include engagement with users to identify best practices.**

Days 0-30: User engagement and initial designs related to updating a dynamic knowledge graph and recording feedback for continued GAT and UIPN training

Days 30-60: Iterate and design an initial testable implementation of capability for users to update graphs; the graphs will be automatically updated by the GAT and UIPN, which will be trained based on user feedback

Days 60-90: Initial testing of updating knowledge graph elements automatically, using GATs and UIPNs to update appropriate nodes, and collecting user feedback and training data to improve GATs and UIPN performance

Days 90-120: Based on initial testing, iterate on the design and methods for collecting user feedback and training data. Implement changes in preparation for final testing.

Days 120-150: Conduct final testing and record results of methods for inclusion in the final report

Days 150-180: Incorporate research objective findings into the final report and prototype design

**Objective #5: Provide metrics at different levels of the graph (entire graph, subsets of a graph, embeddings across multiple nodes and edges, specific nodes and edges) to keep users aware of when and how the knowledge graph is changing over time based on automatic and manual updates. Include engagement with users to identify best practices.**

Days 0-30: Set up initial metrics alongside dynamic knowledge graphs, GATs, and UIPNs

Days 30-60: Conduct engagement with end users to collect feedback on initial purposes and approaches to provide users with metrics. Separately, record metrics during the initial training of GATs and UIPNs. Implement any required changes to metrics.

Days 60-90: Record metrics during initial testing of automated updating knowledge graph elements using GATs and UIPNs

Days 90-120: Improve or change metrics based on initial testing to support final testing

Days 120-150: Conduct final testing, use developed metrics to measure performance, and include metrics in the final report

Days 150-180: Incorporate research objective findings into the final report and prototype design

**Objective #6: Record snapshots of the GATs and UIPNs over time for use in subsequent graph analytics or other analytics.**

Days 0-30: Setup ability to take snapshots of knowledge graphs, GATs, and UIPNs

Days 30-60: Record snapshots of knowledge graphs, GATs, and UIPNs

Days 60-90: Record snapshots of knowledge graphs, GATs, and UIPNs

Days 90-120: Build a plan for the best way to record snapshots during final testing

Days 120-150: Conduct final testing and record snapshots of knowledge graphs, GATs, and UIPNs throughout testing.

Days 150-180: Incorporate research objective findings into the final report and prototype design

## Phase I Statement of Work

This Statement of Work outlines the major tasks and deliverables for experimenting and testing the feasibility of GATs and UIPNs to allow for user modifications to a dynamic knowledge graph and predict additional necessary changes to the graph, obtain baseline performance metrics such as but not limited to accuracy and graph completeness, initial prototype design completion, and documentation of all work completed, all of which will aid in selecting the most promising approach for further research and development.

Additionally, Cenith Innovations will work with the Government to identify appropriate end-users to engage with in interactive feedback sessions during the period of performance.

**Scope:** During the six-month period of performance, Cenith Innovations will incrementally develop and deliver tests and test results that will demonstrate GATs’ and UIPNs’ performance in allowing user modifications to dynamic knowledge graphs and predicting additional changes, culminating in an initial prototype design and a final feasibility report by the end of the period of performance.

| Task | Task Name | Duration | Expected Milestone | Performer(s) |
|---|---|---|---|---|
| 01 | Initial knowledge graph, GAT, UIPN setup | Days 0-30 | **Deliverables**: Status report, virtual or in-person demo | Bryan O'Rourke |
| 02 | Initial training of GAT and UIPN | Days 30-60 | **Deliverables**: Status report, virtual or in-person demo | Colin Blackett |
| 03 | Initial testing and testing results | Days 60-90 | **Deliverables**: Status Report, virtual or in-person demo | Colin Blackett |
| 04 | Continued iterative development on overall GAT and UIPN training and data ingest | Days 90-120 | **Deliverables**: Status Report, virtual or in-person demo | Colin Blackett |
| 05 | Initial Prototype Design | Days 120-180 | **Deliverables:** Prototype design document | Ryan Tomlinson and Chris Lauber |
| 06 | Final Testing | Days 120-150 | **Deliverables:** Status Report, virtual or in-person demo. | Colin Blackett |
| 07 | Test Findings and Final Technical Report | Days 150-180 | **Deliverables:** Final Technical Report including test findings, baseline performance metrics, and documentation of work completed in this feasibility study, SF 298 | Bryan O'Rourke |

## Related Work

Cenith Innovations simplifies warfighter and analyst workflows and improves analytic products and outcomes by applying novel AI solutions and engaging closely with users. We exemplify this expertise across multiple projects, but most convincingly with PATH, an AI-driven product we developed for the US Army. PATH applies neural network-based Generative Adversarial Network (GAN) and Deep Reinforcement Learning (DRL) to aid soldiers in planning and navigating impossibly large combinatorial scenarios. Our work using these neural network-driven methods on PATH will directly benefit our proposed work when testing the feasibility of applying neural network-based GATs and UIPNs to provide automated, accurate updates to knowledge graphs. For example, our DRL, GAN, and GAT work are all grounded in deep learning principles, including neural network architectures, backpropagation, and gradient descent algorithms, and all involve complex architectures to model agents and environments and handle high-dimensional and unstructured data.

PATH's main customer is US Army SGM Corey Wilkens, corey.d.wilkens@army.mil, 417-650-1572. Cenith Innovations' work on PATH is currently ongoing.

PATH provides a robust human-in-the-loop Course of Action (COA) analysis powered by two primary AI capabilities that allow for human override and customization of the recommended outputs. This analysis ensures users are not beholden to a predicted COA and can still use their intuitive knowledge of the battlespace to work in concert with the technology and make optimal decisions. The first such AI capability uses a diffusion model, which is a type of generative AI model. Our diffusion model incorporates terrain features (e.g., surface, elevation, mobility factors, adversary locations, and doctrine) and creates a heatmap of undiscovered obstacles and threats. Our diffusion model uses a custom-built and trained Generative Adversarial Network (GAN) that learns how to "complete the picture" by filling in the missing information on the map. GANs are trained by building images and judging how good a generated image is by comparing it to a set of coded discriminators. Our discriminators are essentially map inputs, such as map layers, manual annotations, or even enemy doctrine. Our diffusion model can look at an incomplete threat picture and incorporate all layers and inputs to produce a probabilistic map overlay of the missing threats, drastically improving the accuracy of the threat pictures.

The second AI capability Path uses for first-level terrain analysis is the ability to produce a detailed set of COAs. To do this, we take the known key terrain features and fuse them with the probabilistic heat map from the diffusion model. The combined data feeds into an RL algorithm, combined with a pathfinding algorithm (A*- pronounced "A Stars") to produce an optimal set of possible COAs, balancing survivability with speed. We built this model using a Markov Decision Process (MDP) that evaluates all possible next and future moves and actions while assigning a multivariate score and assessing the highest scoring potential. These scores are determined using a reward function that incorporates the particular terrain, threats, location on the map, previous action, and resource availability. We train the reward function and RL agent in simulated terrain environments called an RL gym. In practice, there are too many possible combinations of moves and actions to practically compute the best path. This is where A* comes into play. A* is an algorithm used by applications like Google Maps to quickly find the optimal path while following an impossibly large decision tree.[11] We utilize A* to help the MDP explore the graph of possible moves and associated actions and to come up with a list of optimal COAs that maximize the reward function. The end result is an output list of COAs, where each COA consists of attributes like an action to take at each step, the expected time to take that action and movement, and the expected danger or attrition rate for personnel and equipment along the way.

## Relationship with Future Research or Research and Development

### Anticipated Results of a Successful Phase I:

Cenith Innovations records objective test results to inform and identify the most promising approach. We anticipate that using GATs and UIPNs for knowledge graph updating will yield new, beneficial results to include more accurate neighborhood sizing for attention mechanisms, new findings regarding best training approaches for knowledge graphs related to DoD and IC use cases, how to efficiently and effectively retrieve GAT and UIPN training feedback from users, and improved methods for measuring how automated updates change the knowledge graph.

### Phase I Provides Foundation for a Phase II Research and Development Effort:

With an objective foundation of test results and metrics from a successful Phase I, Cenith Innovations will build the momentum for delivering a functional prototype in a potential Phase II effort. By successfully accomplishing the six identified research objectives, Cenith Innovations will have all the required technical architecture and proven methods and will have built an initial relationship with end users.

**Objective 1:** Testing this capability requires us to build a basic prototype to identify and test how non-graph data can be ingested in the knowledge graph to update related nodes. This basic prototype will be built in an extensible manner following Modular Open Source guidelines to facilitate continued development during a Phase II effort.

**Objectives 2 & 3:** The required machine learning architecture and training environments to support a successful Phase II effort will be built during the Phase I effort. Additionally, our testing results lead to an

[11] Patel, Amit. *Introduction to A**, Amit Patel, 4 Nov. 2023, theory.stanford.edu/~amitp/GameProgramming/.

improved GAT and UIPN training process, which will help more quickly and accurately ingest and train new customer data sources provided by the Air Force.

**Objective 4:** Testing this capability will require a basic prototype for users to provide feedback on, which will be built in an extensible manner following Modular Open Source guidelines to facilitate continued development during a Phase II effort.

**Objective 5:** Cenith Innovations is a data-driven organization. Metrics developed during Phase I will be continuously improved across Phase I and Phase II to objectively measure our capability's performance.

**Objective 6:** We included this objective in our Phase I almost exclusively to be able to use this data during a potential Phase II effort. Time is a fundamental component of analysts' and operators' workflows; therefore, viewing and understanding how knowledge graphs change over time will be pivotal. We will use snapshots generated in Phase I training to improve our Research and Development of temporal analysis of knowledge graphs in a potential Phase II effort.

### Clearances and Approvals Required to Conduct Phase II Testing:

Cenith Innovations includes several assumptions in its plan to support the Air Force's Phase II goal of demonstrating a full-scale prototype via on-site testing in the customer's environment. If these assumptions are not correct, then additional conversations with the Government will need to be conducted before a Phase II proposal is submitted.

1. "Full-scale" prototype means a prototype fully demonstrating the agreed-upon capability identified through a Phase II proposal, which will be based on Phase I results.
2. "On-site testing in the customer's environment " means testing on a network or system that does not require any cyber authorizations, such as an Authorization to Operate (ATO).
3. The Air Force will provide data in support of the Phase II demonstration, which may be classified.

To address assumption 1, Cenith Innovations will continuously communicate with the Air Force Technical Point of Contact (TPOC) throughout Phase I to ensure a common understanding of Phase II goals, which will be included in a Phase II proposal.

To address assumption 2, Cenith Innovations will architect our solution so that it can run locally on a server or computer that Cenith Innovations or the Air Force will provide. Cenith Innovations will work with the Air Force to identify the exact technical details for the demonstration and include them in the Phase II proposal.

To address assumption 3, all Cenith Innovations team members for this project already maintain Secret-level clearances, with several team members maintaining TS clearance with access to SCI.

All Phase I work is unclassified and performed in our secure IL4 AWS GovCloud environment. Cenith Innovations will engage the Air Force during the Phase I period of performance to understand the Air Force's goals and data classification requirements for the Phase II demonstration. Based on the Air Force's goals and requirements, Cenith will determine how development and demonstrations will occur at what classification levels to support a Phase II demonstration and overall meeting of Air Force goals.

## Commercialization Strategy

Cenith Innovations is a profitable defense technology company with >$9M in annual revenue, specializing in rapidly building and deploying applied AI and advanced software at the speed and scale warfighters deserve. Cenith was founded in 2019 by Kristopher Pruitt and Bryan O'Rourke. Kristopher was the U-2S Dragon Lady Functional Area Manager on the Headquarters Air Force Staff, working for the Deputy Chief of Staff for ISR and cyber effects operations. Bryan was a Silicon Valley ML software engineer who began working with the DoD on Project Maven. Their unique combination of operational

and technical expertise has led to significant growth of the company and the commercialization of multiple product offerings. Cenith has supported several customers in the US Air Force, US Army, and the USSOCOM on efforts including Advanced Synthetic Aperture Radar System (ASARS) and Avionics Technology Refresh (ATR) for the U-2 Dragon Lady, Holistic Health and Fitness application development for the XVIII Airborne Corps, and LiDAR-generated Alternative Position and Navigation for the Family of Special Operations Vehicles. Our team includes PhD researchers, software and hardware engineers, remote sensing specialists, and geospatial analysts, each with decades of experience. Since our inception, we've increased our revenue by an average of 164% year-over-year by applying our expertise, service, and products to meet DoD needs while providing exceptional quality. Our customers include One Nation Innovation, Leidos, General Dynamics Information Technology, Modern Technology Solutions Incorporated, Huntington Ingalls Industries, Draper Laboratories, Collins Aerospace, the United States Army, US Special Operations Command, and the USAF, among others. We deliver bleeding-edge software development to those who need it the most.
Our GAT and UIPN technology is highly dual-use. Our commercialization strategy effectively executes market entrance, adoption, and growth in both commercial and DoD markets. Due to the specificity of the topic, required technology, and our go-to-market capabilities, the Defense market is our first target. We will align requirements, resources, contracting, and testing/validation to meet this goal.

LOE 1: We will initially target existing DoD and IC programs that leverage knowledge graphs for intelligence analysis by focusing on business-to-business relationships with the Primes for these large enterprise contracts. The particular contracts we're interested in include DIA's Machine-Assisted Analytic Rapid-Repository System (MARS), NGA's Cedalion capability, NGA's Structured Observation Management (SOM) program, and the Air Force Research Laboratory's Insight Integration program.

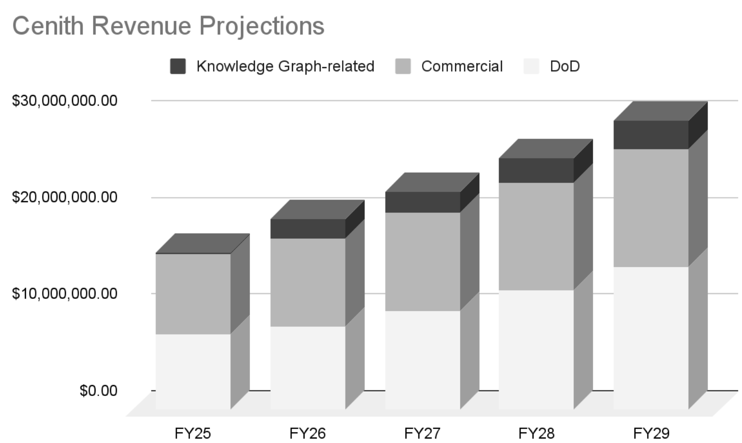


**Figure 4:** Cenith Innovations Revenue Projections through 2029.

LOE 2: We propose a pilot integration of third-party capabilities developed by small businesses, such as the ones developed through this SBIR, into the AFRL Insight Integration program. This work may be conducted under a Cooperative Research and Development Agreement (CRADA) or in the form of a SBIR Phase III. During this LOE, we will work with GSA to stand up a SBIR Phase III indefinite delivery, indefinite quantity (IDIQ) contract with assistance from the SBIR Technical Point of Contact. In addition to the GSA IDIQ we will maintain and mature our existing contract vehicles, which include multiple consortium OTAs (Consortium Management Group, Expeditionary Mission Consortium - Crane, and the Defense Industrial Base Consortium) along with large value and scope vehicles such as LOGIX (through FedSim), all of which can be leveraged by AFRL and other customers.

LOE 3: Cenith Innovations will take advantage of the large and growing markets that are already leveraging GNNs to automatically and accurately update enterprise dynamic knowledge graphs but may be limited by current implementations using Graph Convolutional Networks or GATs on their own. Our two target market segments are Knowledge Management & Customer Support and Real-time Fraud Detection in Financial Services. For Knowledge Management & Customer Support, we will solve the pain points of constant update requirements that are costly and time-consuming. In this segment, the GAT would infer the relevance of solutions, while the UIPN would predict the impact of the new support tickets or content updates on the existing knowledge graph. The Total Addressable Market (TAM) for this market is approximately $500 billion, the Serviceable Available Market (SAM) is approximately $125 million, and the SOM is valued at around $6.25 billion with a SOM growth rate of approximately 18% annually, according to market research firm MarketsandMarkets.

The Financial Fraud Detection industry uses knowledge graphs to track relationships between users, accounts, transactions, and devices. Knowledge graphs must constantly be updated with every new transaction so they can be analyzed and used to detect potentially fraudulent activity. In this example, the GAT would identify suspicious patterns based on existing relationships, while the UIPN would enable the network to adapt to new transactions. With updates being provided faster and more accurately by the UIPN, the system can also flag potential fraud more quickly and accurately. The TAM for Financial Fraud detection is approximately $40 billion, The Serviceable Available Market is approximately $24 billion, and the SOM for this market is valued at around $2.4 billion, according to MarketsandMarkets.

## Project Key Personnel

| **KEY PERSONNEL SUMMARY** | | | |
|---|---|---|---|
| **Name and Title** | **Employer** | **Qualifications** | ***Foreign National (Y/N)** |
| Bryan O'Rourke, *Principal Investigator* | Cenith Innovations | Bryan O’Rourke has over a decade of experience as a software engineer and architect at industry-leading companies in Silicon Valley and Defense Technology. His projects include several notable DoD-specific ML projects, such as Project Maven. Bryan was the architect for Cenith Innovation’s ML approach to build its successful PATH application for the US Army. His overlap of experience with DoD platforms, software engineering leadership in Silicon Valley, and experience delivering projects in secure government environments make him an ideal Principal Investigator. Bryan maintains a Secret security clearance. | N |
| Colin Blackett, *Principal Engineer* | Cenith Innovations | Colin Blackett has over 25 years of experience as a full-stack software engineer, software architect, and DevSecOps engineer in various roles across private technology companies and the DoD. He’s been an engineering team leader and software development manager and has broad expertise in producing rich and insightful AI/ML models and advanced mathematics. Colin maintains a Secret security clearance. | N |

| | | | |
|---|---|---|---|
| Ryan Tomlinson *Director of Product Management* | Cenith Innovations | Ryan Tomlinson has over 13 years of management experience with 8 years as a software engineer and 5 years as a product manager. Ryan is a product leader who turns complex user requirements into successful and beautiful products. He has extensive experience working with various tools to gain insight into user needs, workflows, and detailed technical integrations. Ryan maintains a Secret security clearance. | N |
| Chris Lauber, *Vice President of Defense and Intelligence Solutions* | Cenith Innovations | Chris Lauber has over 18 years of experience as Senior Intelligence Analyst and Product Manager in the Intelligence Community and DoD. He has extensive experience using Intelligence Community tools that leverage knowledge graphs, including leading the development of an analytic tool that extensively leveraged knowledge graphs while at the National Geospatial-Intelligence Agency. Chris maintains a Top Secret security clearance and is eligible for TS//SCI access. | N |

## Foreign Citizens

Cenith Innovations is a domestically owned company. No non-US citizens will participate in this effort. We have no foreign ownership, control, or influence and maintain a TOP SECRET facility security clearance under CAGE 88BH3. Currently, over 80% of our workforce has a SECRET clearance or above, and over 60% has a TOP SECRET, Sensitive Compartmented Information and/or Special Access Program clearance.

## Subcontractors/Consultants

None.

## Facilities/Equipment

All software development will be accomplished using Cenith Innovations' IL 4 AWS GovCloud environment. Software will be DFARS 252.204-7012, NIST 800-171, and NIST 800-53 compliant. All data is encrypted in transit and at rest with AES256 keys. All data in transit, including API calls, are encrypted with SSL/TLS1.2+. Authentication uses OAuth 2 protocol and SSO. Secure 2 Factor Authentication can be enabled, including for mobile. Authentication SIEM monitoring and alerting uses cutting-edge AWS GovCloud services.

As the work in Phase II may become classified, it's worth noting that Cenith Innovations is a domestically owned company. No non-US citizens will participate in this effort. We have no foreign ownership, control, or influence and maintain a TOP SECRET facility security clearance under CAGE 88BH3. Currently, over 80% of our workforce has a SECRET clearance or above, and over 60% has a TOP SECRET, Sensitive Compartmented Information, and/or Special Access Program clearance.

DSIP Defense SBIR/STTR Innovation Portal
Cost Volume

## SBIR Phase I Proposal

| | |
|---|---|
| **Proposal Number** | F244-0001-0094 |
| **Topic Number** | AF244-0001 |
| **Proposal Title** | Faster, Easier, and More Accurate Automated Updating of Dynamic Knowledge Graphs for Intelligence and Operations Using Graph Attention Networks and Update Impact Prediction Networks |
| **Date Submitted** | 11/06/2024 11:30:22 AM |

## Firm Information

| | |
|---|---|
| **Firm Name** | Cenith Innovations, LLC |
| **Mail Address** | 1861 9th Avenue, Sacramento, California, 95818 |
| **Website Address** | http://www.cenithinnovations.com |
| **UEI** | ML3NGPZG8PS3 |
| **Duns** | 116925606 |
| **Cage** | 88BH3 |

| | |
|---|---|
| **Total Dollar Amount for this Proposal** | $139,870.87 |
| Base Year | $139,870.87 |
| Year 2 | $0.00 |
| Technical and Business Assistance(TABA)- Base | $0.00 |
| TABA- Year 2 | $0.00 |

### Base Year Summary

| | |
|---|---|
| **Total Direct Labor (TDL)** | $93,838.84 |
| **Total Direct Material Costs (TDM)** | $4,038.00 |
| **Total Direct Supplies Costs (TDS)** | $0.00 |
| **Total Direct Equipment Costs (TDE)** | $0.00 |
| **Total Direct Travel Costs (TDT)** | $0.00 |
| **Total Other Direct Costs (TODC)** | $0.00 |
| **G&A (rate 35%) x Base (TDL+TOH)** | $32,843.60 |
| **Total Firm Costs** | $130,720.44 |
| **Subcontractor Costs** | |
| **Total Subcontractor Costs (TSC)** | $0.00 |
| **Cost Sharing** | -$0.00 |
| **Profit Rate (7%)** | $9,150.43 |
| **Total Estimated Cost** | $139,870.87 |
| **TABA** | $0.00 |

### Year 2 Summary

| | |
|---|---|
| **Total Direct Labor (TDL)** | $0.00 |

| | |
|---|---|
| **Total Direct Material Costs (TDM)** | $0.00 |
| **Total Direct Supplies Costs (TDS)** | $0.00 |
| **Total Direct Equipment Costs (TDE)** | $0.00 |
| **Total Direct Travel Costs (TDT)** | $0.00 |
| **Total Other Direct Costs (TODC)** | $0.00 |
| **G&A (rate 35%) x Base ()** | $0.00 |
| **Total Firm Costs** | $0.00 |
| **Subcontractor Costs** | |
| **Total Subcontractor Costs (TSC)** | $0.00 |
| **Cost Sharing** | -$0.00 |
| **Profit Rate (7%)** | $0.00 |
| **Total Estimated Cost** | $0.00 |
| **TABA** | $0.00 |

## Base Year

**Direct Labor Costs**

| **Category / Individual-TR** | **Rate/Hour** | **Estimated Hours** | **Fringe Rate (%)** | **Fringe Cost** | **Cost** |
|---|---|---|---|---|---|
| Software Developer/<br>Principal Investigator (Bryan O'Rourke) | $106.00 | 100 | 21.3 | $2257.80 | $12,857.80 |
| Software Developer/<br>Senior Software Engineer | $90.00 | 650 | 21.3 | $12460.50 | $70,960.50 |
| Life Scientists, All Other/<br>Research Lead and Reporting | $90.00 | 50 | 21.3 | $958.50 | $5,458.50 |

| | |
|---|---|
| Subtotal Direct Labor (DL) | $89,276.80 |
| Labor Overhead (rate 5.11%) x (DL) | $4,562.04 |
| **Total Direct Labor (TDL)** | **$93,838.84** |

## Direct Material Costs

| | |
|---|---|
| AWS GovCloud GPU Time | $4,038.00 |
| **Total Direct Material Costs (TDM)** | **$4,038.00** |

| | |
|---|---|
| **G&A (rate 35%) x Base (TDL+TOH)** | $32,843.60 |
| **Cost Sharing** | -$0.00 |
| **Profit Rate (7%)** | $9,150.43 |
| **Total Estimated Cost** | $139,870.87 |
| **TABA** | $0.00 |

## Year 2

**Direct Labor Costs**

| **Category / Individual-TR** | **Rate/Hour** | **Estimated** | **Fringe Rate** | **Fringe Cost** | **Cost** |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| | | **Hours** | **(%)** | | |
| Software Developer/<br>Principal Investigator (Bryan O'Rourke) | $0.00 | 0 | 0 | $0.00 | $0.00 |

| | |
|---|---|
| Subtotal Direct Labor (DL) | $0.00 |
| Labor Overhead (rate 0%) x (DL) | $0.00 |
| **Total Direct Labor (TDL)** | **$0.00** |

### Direct Material Costs

| | |
|---|---|
| None | $0.00 |
| **Total Direct Material Costs (TDM)** | **$0.00** |

| | |
|---|---|
| **G&A (rate 35%) x Base ()** | $0.00 |
| **Cost Sharing** | -$0.00 |
| **Profit Rate (7%)** | $0.00 |
| **Total Estimated Cost** | $0.00 |
| **TABA** | $0.00 |

**Explanatory Material Relating to the Cost Volume**
**The Official From the Firm that is responsible for the cost breakdown**
Name: Aimee Kaiser
Phone: (207) 491-3865
Phone: aimee@cenithinnovations.com
Title: Proposal Owner

**If the Defence Contracting Audit Agency has performed a review of your projects within the past 12 months, please provide:** No
**Select the Type of Payment Desired:** Partial payments

## Cost Volume Details

### Direct Labor

**Base**

| Category | Description | Education | Yrs Experience | Hours | Rate | Fringe Rate | Total |
|---|---|---|---|---|---|---|---|
| Software Developer | Principal Investigator | Bachelor's Degree | 10 | 100 | $106.00 | 21.3 | $12,857.80 |
| Software Developer | Senior Software Engineer | Bachelor's Degree | 25 | 650 | $90.00 | 21.3 | $70,960.50 |
| Life Scientists, All Other | Research Lead and Reporting | PhD | 20 | 50 | $90.00 | 21.3 | $5,458.50 |

Are the labor rates detailed below fully loaded? **NO**

Provide any additional information and cost support data related to the nature of the direct labor detailed above.
**Cost of labor in line with state averages.**

Direct Labor Cost ($): $89,276.80

**Year2**

| Category | Description | Education | Yrs Experience | Hours | Rate | Fringe Rate | Total |
|---|---|---|---|---|---|---|---|
| Software Developer | Principal Investigator | Bachelor's Degree | 10 | 0 | $0.00 | 0 | $0.00 |

Are the labor rates detailed below fully loaded? **NO**

Provide any additional information and cost support data related to the nature of the direct labor detailed above.
**This is a Phase I effort and labor rates are not calculated for Phase II yet.**

Direct Labor Cost ($): $0.00

Sum of all Direct Labor Costs is($): $89,276.80

### Overhead

**Base**

Labor Cost Overhead Rate (%) **5.11**

| | |
|---|---|
| Apply Overhead to Direct Materials Cost? | **NO** |

Overhead Comments:
**This is a Phase I effort and labor rates are not calculated for Phase II yet.**

| | |
|---|---|
| Overhead Cost ($): | **$4,562.04** |

**Year2**

| | |
|---|---|
| Labor Cost Overhead Rate (%) | **0** |
| Apply Overhead to Direct Materials Cost? | **NO** |

Overhead Comments:
**This is a Phase I effort and rates are not calculated for Phase II yet.**

| | |
|---|---|
| Overhead Cost ($): | **$0.00** |
| Sum of all Overhead Costs is ($): | **$4,562.04** |

## General and Administration Cost

**Base**

| | |
|---|---|
| G&A Rate (%): | **35** |
| Apply G&A Rate to Overhead Costs? | **YES** |
| Apply G&A Rate to Direct Labor Costs? | **YES** |
| Apply G&A Rate to Direct Material Costs? | **NO** |

Please specify the different cost sources below from which your company's General and Administrative costs are calculated.
**This is a Phase I effort and rates are not calculated for Phase II yet.**

| | |
|---|---|
| G&A Cost ($): | **$32,843.60** |

**Year2**

| | |
|---|---|
| G&A Rate (%): | **35** |
| Apply G&A Rate to Overhead Costs? | **NO** |

| | |
|---|---|
| Apply G&A Rate to Direct Labor Costs? | **NO** |
| Apply G&A Rate to Direct Material Costs? | **NO** |

Please specify the different cost sources below from which your company's General and Administrative costs are calculated.
**This is a Phase I effort and rates are not calculated for Phase II yet.**

| | |
|---|---|
| G&A Cost ($): | **$0.00** |
| Sum of all G&A Costs is ($): | **$32,843.60** |

**ODC-Materials**
**Base**

Description: AWS GovCloud GPU Time

Vendor: Amazon

Quantity: 6

Total Cost($): $4,038.00

Consumable? no

Competitively Sourced? yes

Exclusive for this Contract? yes

Supporting Comments:

Supporting Documents:

- MAT 001 AWS GPU Time (1).pdf

**Year2**

Description: None

Vendor: N/A

Quantity: 0

Total Cost($): $0.00

Consumable? no

Competitively Sourced? no

Exclusive for this Contract? no

Supporting Comments:
**This is a Phase I effort and costs are not calculated for Phase II yet.**

**ODC-Summary**
**Base**

| | |
|---|---|
| Do you have any additional information to provide? | **NO** |

**Year2**

| | |
|---|---|
| Do you have any additional information to provide? | **NO** |

## Profit Rate/Cost Sharing

### Base

| | |
|---|---|
| Cost Sharing ($): | **-** |
| Cost Sharing Explanation: | |
| Profit Rate (%): | **7** |
| Profit Explanation: | |
| Total Profit Cost ($): | **$9,150.43** |

### Year2

| | |
|---|---|
| Cost Sharing ($): | **-** |
| Cost Sharing Explanation: | |
| Profit Rate (%): | **7** |
| Profit Explanation: | |
| Total Profit Cost ($): | **$9,150.43** |
| Total Proposed Amount ($): | **$139,870.87** |

# CENITH INNOVATIONS LLC

**DISCLAIMER:** Information provided herein is privileged and confidential, and not subject to disclosure, pursuant to 15 U.S.C. 638 (k)(4) and 5 U.S.C. 552. This information shall only be used or disclosed for evaluation purposes.

SBIR Company Commercialization Report

| Total Investments: | Total Sales: | Total Patents: | Government Designated Phase III Funding: |
|---|---|---|---|
| $0.00 | $0.00 | 0 | $0.00 |

## Company Information

**Address:**

1861 9TH AVE
SACRAMENTO, CA 95818-4111
United States

**SBC Control ID:** SBC_001616419

**Company Url:** http://www.cenithinnovations.com

## Company POC

**Title:** CEO
**Full Name:** Kristopher Pruitt
**Phone:** 9167073178
**Email:** kris@cenithinnovations.com

## Commercialization POC

**Title:** CEO
**Full Name:** Kristopher Pruitt
**Phone:** (916) 707-3178
**Email:** kris@cenithinnovations.com

## Additional Company Information

| % Revenue for last fiscal year from SBIR/STTR funding: | Total revenue for last fiscal year: |
|---|---|
| 0.0% | $5,000,000 - $19,999,999 |
| **Year Founded:** | **# Employees Currently:** |
| 2019 | 32 |
| **Year first Phase I award received:** | **# SBIR/STTR Phase I Awards:** |
| 2021 | 4 |
| **Year first Phase II award received:** | **# SBIR/STTR Phase II Awards:** |
| 2022 | 3 |
| **# Employees at first Phase II award:** | **Mergers and Acquisition within past 2 years:** |
| 19 | No |
| **Spin-offs resulting from SBIR/STTR:** | **IPO resulting from SBIR/STTR \| Year of IPO:** |
| No | No \| N/A |
| **Patents resulting from SBIR/STTR \| #Patents:** | **List of Patents:** |
| No \| N/A | |
| **Woman-Owned:** | **Socially and Economically Disadvantaged:** |
| No | No |
| **HUBZone-Certified:** | **SBC majority-owned by multiple VCOC, HF, PE firms \| By what percent (%):** |
| No | No \| N/A |

## Additional Investment From

| | Last Submitted Version (01-21-2022 06:34 PM) | Current Version |
|---|---|---|
| DoD contracts/DoD subcontracts | $0.00 | $0.00 |
| Angel Investors | $0.00 | $0.00 |
| Venture Capital | $0.00 | $0.00 |
| Self Funded | $0.00 | $0.00 |
| Private Sector | $0.00 | $0.00 |
| Other Federal Contracts/Grants | $0.00 | $0.00 |
| Other Sources | $0.00 | $0.00 |
| Additional Investment | $0.00 | $0.00 |
| Total Investment | $0.00 | $0.00 |

## Phase III Sales To

| | Last Submitted Version (01-21-2022 06:34 PM) | Current Version |
|---|---|---|
| DoD or DoD prime contractors | $0.00 | $0.00 |
| Private Sector | $0.00 | $0.00 |
| Export Markets | $0.00 | $0.00 |
| Other Federal Agencies | $0.00 | $0.00 |
| Additional commercialization by 3rd Party Revenue | $0.00 | $0.00 |
| Other Customers | $0.00 | $0.00 |
| Additional Sales | $0.00 | $0.00 |
| Total Sales | $0.00 | $0.00 |

## Government Phase III Contracts

| | Last Submitted Version (01-21-2022 06:34 PM) | Current Version |
|---|---|---|
| Funding Obligated | $0.00 | $0.00 |

## Commercialization Narrative

## Commercialized Awards

- Listed below are the sales revenue and investment details resulting from the technology developed under these SBIR/STTR awards.

# You have not entered any Commercialized Award Data.

# CERTIFICATE OF COMPLETION

THIS CERTIFICATE IS PRESENTED TO

Aimee Kaiser, Cenith Innovations, LLC

FOR SUCCESSFULLY COMPLETING FRAUD, WASTE AND ABUSE TRAINING AND MEETING ALL REQUIREMENTS SET FORTH BY THE OFFICE OF SMALL BUSINESS PROGRAMS

Nov 06, 2024

COMPLETION DATE

Nov 06, 2025

EXPIRATION DATE