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.

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:	**F244-0001-0010**
Proposal Title:	**Dynamic, Interactive Knowledge Graph**

Agency Information

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

Firm Information

Firm Name:	**InspiRD, Inc**
Address:	**24261 Chrisanta Dr, Mission Viejo, CA 92691-4003**
Website:	**http://inspird.com**
UEI:	**D859W8HNHD86**
DUNS:	**054205491**
CAGE:	**6DC34**
SBA SBC Identification Number:	**001317907**

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)	**12**
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_001317907**

Supporting Documentation:

- SBC_001317907.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: [**X**] Socially and economically disadvantaged SBC [ ] Women-owned SBC [ ] HUBZone-owned SBC [ ] Veteran-owned SBC [ ] Service Disabled Veteran-owned SBC [ ] None Listed	
13. Race of the offeror: [ ] American Indian or Alaska Native [ ] Native Hawaiian or Other Pacific Islander [**X**] Asian [ ] White [ ] Black or African American [ ] 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
Sandeep Mehta	Sandeep R Mehta	CTO	InspiRD, Inc	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:

We will deliver a comprehensive toolset for dynamic and interactive updates to Knowledge Graphs (KGs). We will address very large KGs comprising of a variety of entities, including objects, events, situations, or concepts. Our solution features an intuitive user interface designed to allow users to review and update KGs without requiring specialized training. We will help the USAF maintain an up-to-date and robust KG, enabling comprehensive insights and data-driven decision-making. Our situational awareness experience with the AFNWC will drive usability and commercialization.

We will implement cutting-edge algorithms to dynamically update KGs for manual KG modifications or automatic anomaly detection. The KG correction algorithm will efficiently propagate changes across the KG, without disrupting its overall integrity. KG Completion will correct for missing entities or relationships to improve accuracy. Schema Reshaping will automatically adjust the ontology to accommodate data changes. Entity disambiguation algorithms will be implemented to correct ambiguous or duplicate nodes after changes. Our custom Technical Language Processing (TLP) toolkit will underpin the KG toolset and provide the necessary accuracy, reliability, and performance. A set of encoder LLMs will enable Graph Neural Networks (GNNs) and Graph Language Models (GLMs). We will also consider Generative AI (Decoder) assisted automation.

In Phase I, we will collaborate with AFRL to identify dynamic KG toolset use cases, define requirements, and generate associated workflows. We will build and deliver a Proof of Concept (POC) of the dynamic KG toolset. The POC will demonstrate key components, including automated KG correction, graph neural networks, graph language models, and vector database. We will build on our proven codebase over 5 years of LLM productization experience. Our AI models have been validated on COTS products and on DARPA, MDA, and USAF SBIRs. Our deep USAF operational and veteran experience will ensure usability, reliability, and accuracy. Our decade-long track record of technology transition and delivering exceptional value will ensure success.

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

Knowledge Graphs (KGs) enable a comprehensive approach that improves decision-making and enhances mission success in various defense and intelligence applications. Dynamic KGs are transformative tools that enhance situational awareness, pattern-of-life analysis, threat detection, targeting operations, and more. Hence, the proposed dynamic KG toolset has broad applicability across the Department of Defense (DoD), including the Army, Navy, Missile Defense Agency (MDA), Space Force, and others. Additionally, it will appeal to various government sectors, such as the Department of Homeland Security and the Department of Energy. Local and state governments, law enforcement agencies, and public safety organizations will also benefit from this system. In the commercial sector, dynamic KGs will revolutionize how businesses manage and use data. The proposed toolset will attract industries ranging from e-commerce to healthcare.

Phase I will deliver a POC that proves efficacy and drives stakeholder engagement. Phase II will mature technologies demonstrated in Phase I and integrate them into a TRL 6 prototype. We will define detailed test cases and perform comprehensive validation. We will obtain an Interim Authority to Test (IATT) from AFRL ITS and deploy the prototype in AFRL environment. Phase II will prove the dynamic KG toolset's efficacy using representative conditions and realistic data.

Phase II will be followed by commercialization in Phase III. We will establish partnerships and collaborations with relevant stakeholders. We will mature the dynamic KG toolset to TRL 8. Phase III work will include development of help files and how-to videos to ensure fast adoption of the dynamic KG toolset. We will work with the AFRL ITS to obtain an Interim Authorization to Operate (IATO). We will deploy the dynamic KG toolset in AFRL environment and test using real data.

We will leverage our robust TRL 9 codebase to accelerate development. We will jumpstart development in each phase from our latest code repositories. In Phase III, we will transition the dynamic KG toolset code to COTS development practices. We will support users and maintain the dynamic KG toolset using our mature COTS agile development processes.

We will support three configurations for delivering the dynamic KG toolset. A low-cost cloud-based SaaS will attract small businesses. To drive adoption, we will offer a free trial subscription to students and not-for-profit organizations. For enterprise and government clients, we will provide an on-premise installation with high security and guaranteed performance. We have established marketing and sales processes to aid commercialization, using both direct and indirect sales strategies. We participate in conferences and publish papers to reach a wide audience. Our COTS software will serve as a beachhead for sales. Our industry partners, existing customer base, and executive relationships will ensure successful commercialization and jumpstart sales.

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

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

Knowledge Graphs, LLM, GNN, Graph Language Models

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. Please note that some components will not accept any deviation from the Percentage of Work (POW) minimum requirements. Please check your component instructions regarding the POW requirements.	**YES**
Firm POW	**67.2%**
Subcontractor POW	**32.8%**
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	**YES**

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.

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.

14. Are teaming partners or subcontractors proposed?	**YES**
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?	**25%**
17. Is the principal investigator socially/economically disadvantaged?	**YES**
18. Does your firm allow for the release of its contact information to Economic Development Organizations?	**NO**

Partners:

Partner Name	Partner Type	Point of Contact
Vanderbilt Univesity	University	Tyler Derr

VOL I - Contact Information

Principal Investigator

Name:	**Dr. Sandeep Mehta**
Phone:	**(818) 694-3944**
Email:	**smehta@inspird.com**
Address:	**24261 Chrisanta Ave , Mission Viejo, CA 92691 - 1111**

Corporate Official

Name:	**Dr. Sandeep Mehta**
Phone:	**(818) 694-3944**
Email:	**smehta@inspird.com**
Address:	**24261 Chrisanta Ave , Mission Viejo, CA 92691 - 1111**

Authorized Contract Negotiator

Name:	**Dr. Sandeep Mehta**
Phone:	**(818) 694-3944**
Email:	**smehta@inspird.com**
Address:	**24261 Chrisanta Ave , Mission Viejo, CA 92691 - 1111**

1 Identification and Significance of Problem

In an era where data is abundant yet fragmented, the need for cohesive analytical frameworks has never been more critical. Knowledge Graphs (KGs) have emerged as a transformative solution for the Air Force, facilitating situational awareness, pattern of life analysis, threat detection, targeting operations, and more. By integrating diverse data sources into a unified schema, KGs allow for comprehensive analytics of complex scenarios. However, as operational landscapes evolve, so too must the KGs. Periodic updates — ranging from user corrections to addressing inconsistencies, identifying information gaps, and refining the underlying ontology —are essential to maintain KGs. Conventional approaches for KG maintenance are labor-intensive, computationally expensive, and need custom information extraction tools.

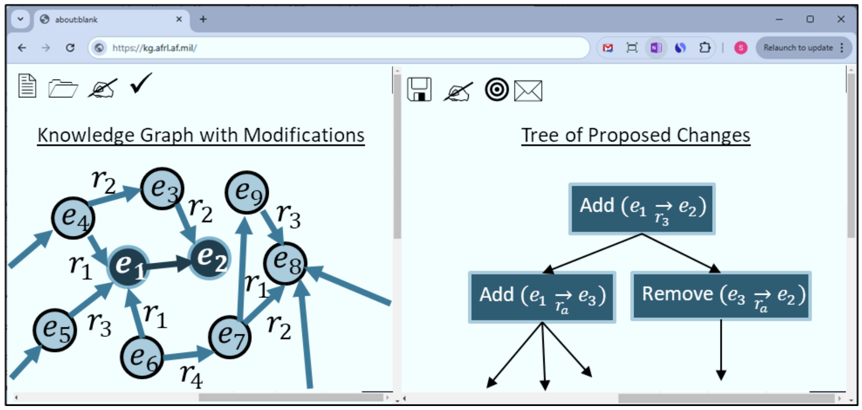

Figure 1: Illustrative User Interface for Dynamic, Interactive KG

The integration of Large Language Models (LLMs) has the potential to enable dynamic and adaptive KGs. Entities and relationships in KG tuples can be represented as text. LLMs (both encoders and decoders) excel at processing and interpreting vast amounts of text, allowing them to extract relevant entities and relationships that can be continuously incorporated into Knowledge Graphs. These dynamic knowledge graphs (DKGs) enable continuous updates, determine knowledge gaps, and update connections, or suggest corrections. Hence, DKGs improve accuracy and agility, ultimately leading to improved situational awareness and operational effectiveness. [blue text are hyperlinks to references]

However, despite their transformative potential, the adoption of LLMs has been slow, with only 5% of organizations integrating them into products. Key barriers to LLM adoption include inadequate security, accuracy, and reliability. Security is challenging because LLMs are known to have vulnerabilities such as sensitive data leakage or data poisoning. Accuracy is impacted by limitations in processing domain-specific jargon. Reliability is impacted by biases such as hallucinations (grammatically correct but factually false text). Additional tools are required to protect information or higher classification levels due to data aggregation. Approaches are needed to help LLMs access external sources of data, when needed.

Table 1: Summary of Proposed Innovations, Team Capabilities, and Commercialization Plan

Proposed Innovations	Inspird Differentiators
• World-record Technical Language Processing (jargons) • LLM Graph Language Models and Graph Neural Networks • LLM Encoders for secure KG processing and updates • AI Specialized Language Model for Gen AI KG processing • AI Agents for accessing dynamic data and computations	• 5+ years of LLM expertise & proven codebase • Direct AFRL/RI KG development experience • Direct AF situational awareness experience • 10+ years building complex enterprise apps • Successful DoD ITS code review and ATO
Commercialization	• Track record of high-SBIR value delivery and tech transition • Detailed commercialization plan and existing customer relationships • Mature commercial agile development processes and AI implementation pipeline

We will leverage our proven technologies and over 5 years of LLM productization experience to address these challenges. Prof. Tyler Derr is leading KG expert and has direct experience working on KGs at AFRL/RI. Mr. DePierre is a USAF veteran and former AF Nuclear Weapons Center operational officer. Our expertise and experience will ensure usability, reliability, accuracy, and performance (Table 1).

2 Technical Objectives

We will deliver a comprehensive toolset (Figure 2) for dynamic and interactive updates to Knowledge Graphs (KGs). We will address very large KGs comprising of a variety of entities, including objects, events, situations, or concepts. Our solution features an intuitive user interface designed to allow users to review and update KGs without requiring specialized training. We will help the USAF maintain an up-to-date and robust KG, enabling comprehensive insights and data-driven decision-making. Our situational awareness experience with the AFNWC will drive usability and commercialization.

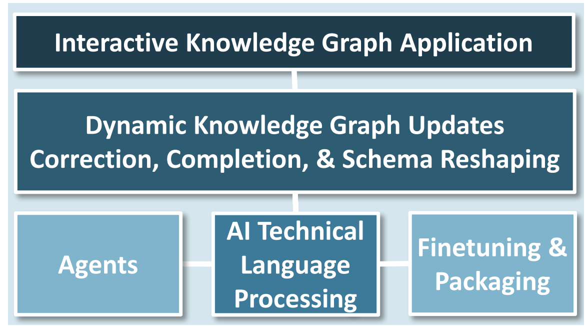

Figure 2: Dynamic Interactive KGs Toolset

Key functionalities will include automatic KG updates and change propagation for entity or relationship modification and periodic automated updates for inconsistencies and information gaps (Figure 2). Our algorithms will be exercised iteratively to dynamically propagate changes across the KG, keeping the ontology synchronized and consistent (Objective 1). Our custom Technical Language Processing (TLP) toolkit (Objective 2) will underpin the toolset and provide the necessary accuracy, reliability, and performance. A set of encoder LLMs will enable Graph Neural Networks (GNNs) and Graph Language Models (GLMs). We will also consider Generative AI (Decoder)-assisted automation.

TLP LLMs will integrate proven fine-tuning approaches to improve accuracy and minimize biases. We tailor LLMs to use USAF terminology and abbreviations. Innovative algorithms will satisfy security requirements (e.g., IL-5) and eliminate vulnerabilities such as data leakage or data poisoning that plague our competition. Advanced packaging algorithms will ensure performance and enable containerization.

In future phases, Agents will support access to dynamic data sources and facilitate extensibility. We will also optimize LLMs for temporal and geospatial information processing. We will explore Vision Language Models and Large Multimodal Models for improving the support for different types of entities. Our proven codebase and 5+ years of LLM productization experience will ensure success (Table 1).

We will deliver a state-of-the-art cloud-capable application (Objective 3). We will build a web portal with a GUI to facilitate intuitive user interactions. A robust server will provide access to the toolset and connect with other AF tools. Our state-of-the-art vector databases will provide reliable storage for the KG. Our extensive codebase and enterprise applications experience will ensure success (Table 1).

Subsections below provide details of technical objectives, our innovations, and their feasibility. In Phase I, we will collaborate with AFRL/RI to identify use cases, define requirements, and generate associated workflows. We will build and deliver a Proof of Concept (POC) of the DKG toolset demonstrating feasibility and efficacy. We will leverage the publicly available NBATransactions dataset for the POC. We will translate each capability demonstration to USAF utility.

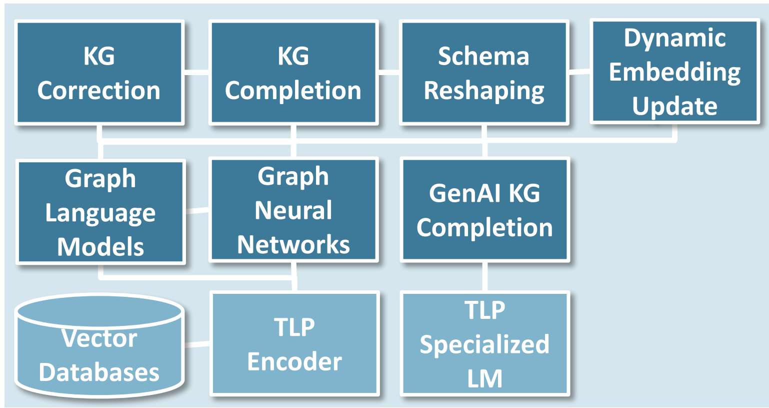

Figure 3: Dynamic Knowledge Graph Updates Innovations

Objective 1: Dynamic Knowledge Graph Updates

We will implement cutting-edge algorithms to dynamically update KGs for manual KG modifications or automatic anomaly detection (Figure 3). The KG correction algorithm will efficiently propagate changes across the KG (Figure 4), without disrupting its overall integrity. KG Completion will correct for missing entities or relationships to improve accuracy. Schema Reshaping will automatically adjust the ontology to accommodate data changes. Entity disambiguation algorithms will be implemented to correct ambiguous or duplicate nodes after changes.

Updating large-scale knowledge graphs dynamically poses significant computational hurdles. Traditional graph algorithms often require recomputing the entire graph structure when changes occur, which is computationally expensive for large graphs. Integrating new information from diverse sources while preserving existing knowledge is complex. Detecting and resolving inconsistencies or conflicts introduced by updates requires sophisticated mechanisms. Handling the temporal nature of dynamic updates adds another layer of complexity. Updating embeddings without full retraining is difficult, as changes can significantly affect the learned representations.

We will integrate KGs with LLMs to enable accurate, scalable, and reliable KG updates (Figure 3). We will represent the KG with embeddings - head, relation, and tail. We will use LLMs to initialize embeddings of textual nodes with inherent semantics. The KG embeddings will be generated using message passing to exchange and aggregate information from neighbors. We will utilize GNN-and GLM-based approaches to compute KG embeddings and underpin KG updates. LLMs will also be employed to improve disambiguation. Additional algorithms will be implemented for efficient embedding updates.

Subsections below describe our innovations, their feasibility, and Phase I deliverables.

Innovations

KG Correction

We will develop an innovative correction mechanism for KG updates that utilizes GNNs and GLMs. Both approaches capitalize on the powerful transformer self-attention mechanism to capture dependencies within graph data and effectively reason over graph structures while providing a systematic process for updating the KG. We will develop these approaches to specifically target situations where nodes or relationships are modified (Figure 4).

The workflow begins by accepting an analyst's modification, followed by extraction of a subgraph surrounding the affected node or relationship. Once the subgraph is extracted, it will be passed through

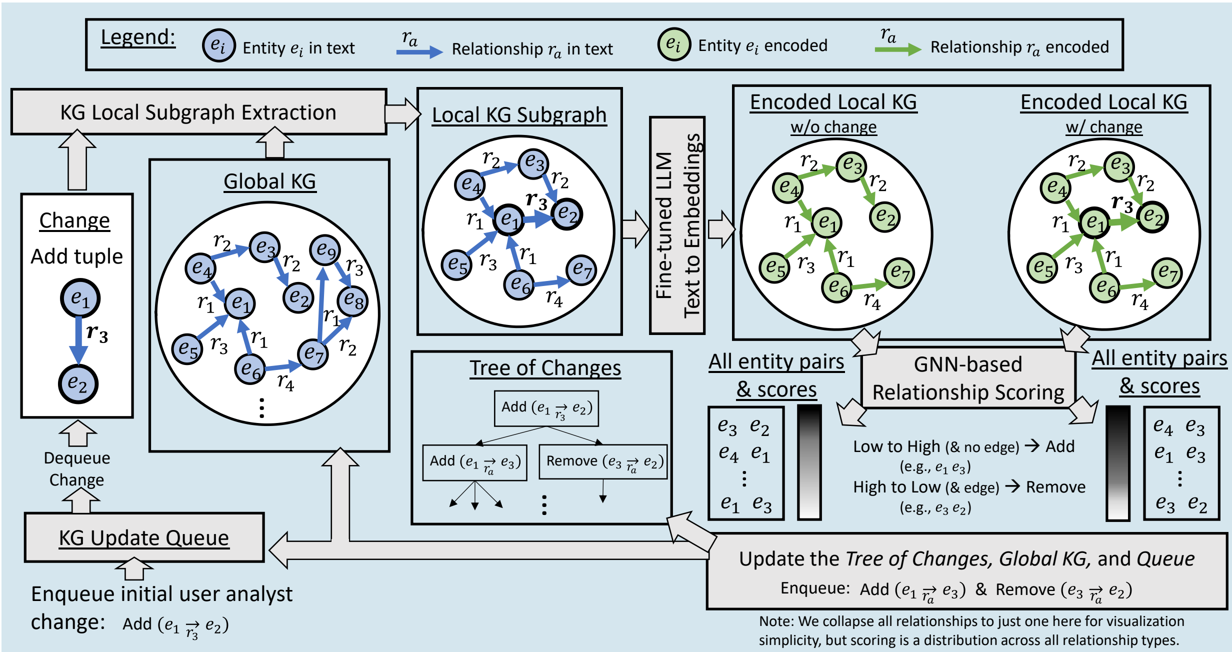

Figure 4: KG Correction Given an Analyst's Modification

our fine-tuned LLM encoder to convert the text representations of the entities and relationship names (and possibly additional context) into node and edge representations. Thereafter, two versions of the subgraph will be formed, one that includes the change and one without. Both will be passed through a GNN-based model to obtain relationship scores/probabilities for all pairs of nodes within the subgraphs.

These will then be used to determine inconsistencies between the subgraph with and without the change. As seen in Figure 4, if an edge had a high score before the change within the local subgraph and then had a low score, this suggests that the likelihood that this relationship within the KG is no longer valid (e.g., pair e3, e2). Similarly, if a previously non-linked pair of entities had a low score before the change and then has their score significantly increase (e.g., pair e1, e3), it suggests that this relationship is likely an addition to be made/suggested to the KG analyst. Once the set of modifications are identified, they are added to a queue that will investigate each potential change independently recursively following the workflow in Figure 4.

The suggested changes made to the KG analyst will be presented in the form of a Tree of Changes. This will provide the logical flow of changes starting from those that are one hop away and increasing iteratively throughout this process. While in Figure 4 we show a two-hop neighborhood subgraph extraction, we will also explore more advanced techniques. For example, once the subgraph is extracted, it will be converted into a tree, with the modified node or relationship placed as the root. The tree structure will be simplified by removing cycles, eliminating multi-parent connections, and flattening complex edges. This conversion transforms the graph into a hierarchical structure, making it easier to analyze the impact of the KG updates.

To correct anomalies, we will apply recursive message passing techniques (GNNs) and relative position matrices (GLMs), starting with one-hop suggested changes. For deeper relationships (e.g., a two-hop change), the model will recursively suggest updates further down in the tree. The recursive nature of this process allows for hierarchical suggestions, where the impact of one-hop changes informs subsequent updates. For instance, if a two-hop change is highly probable based on the one-hop update, it will be ranked higher than a less likely one-hop change. This ranking can be determined by multiplying edge weights along the path, where the overall probability of a change is the product of the weights from the root to the specific node.

This approach also allows for expert intervention. After the model suggests changes and ranks them, human experts can manually inspect the tree and prune branches containing unnecessary updates. This combination of automated GNN-driven suggestions and manual oversight ensures that only the most relevant updates are retained, enhancing both the accuracy and practicality of the system.

In Phase I, we will assume that updates are independent—meaning changes to different parts of the tree are treated separately. However, in future phases, models will be adapted to handle dependencies between updates, where the likelihood of one change affects others in the graph. This more advanced model will account for interrelated updates and improve the overall consistency of the KG.

In Phase I, we develop and evaluate KG correction tools. We will deliver at least one candidate tool for integration in to the POC and for full development in Phase II.

KG Completion

We will implement a robust method for KG completion using pre-trained encoder LLMs. Specifically, we will treat entities, relations, and triples as textual sequences, transforming the task of KG completion into a sequence classification problem. By representing these components as text, we will be able apply GLMs/GNNs to the problem. GLMs/GNNs will predict the plausibility of a given triple or relation. This approach is expected to deliver strong performance across various knowledge graph completion tasks, leveraging the power of language models to understand and predict the structure of the graph.

In Phase I, we research and evaluate KG completion methods for efficacy and integrability with our KG update stack. We will deliver a report outlining results and a list of methods for development in Phase II.

KG Schema Reshaping

We develop and evaluate an innovative schema reshaping algorithm that updates the predefined KG schema into a more data-oriented version, eliminating blank nodes while still preserving crucial elements from the original domain schema. In future phases, we will develop a query reshaping module that adapts queries written for one KG schema to function seamlessly with the reshaped KG schema. This ensures a uniform querying experience, regardless of schema variations. The resulting KG will be both more efficient and user-friendly due to their optimized structure. By eliminating blank nodes, our tools will reduce storage and processing overhead and accelerate querying.

In Phase I, we will research and evaluate schema reshaping algorithms for efficacy in our KG update tech stack. We will deliver a report outlining results and a list of methods for development in Phase II.

Dynamic Embedding Updates

This process of updating embeddings for KG changes is time-consuming and often necessitates retraining downstream models that rely on these embeddings. We explore methods for updating embeddings that eliminate the need for complete retraining. We will focus on strategies that position newly introduced nodes using local information. We design approaches for continuing the training of the existing embedding, while interspersing it with epochs dedicated solely to optimizing the newly added or removed nodes and edges. We will ensure that the new approach produces higher-quality embeddings faster than full retraining, enabling more efficient dynamic KG environments.

In Phase I, we will research dynamic embedding update algorithms for efficacy in our KG update stack. We will deliver a report outlining results and a list of methods for development in Phase II.

Graph Neural Networks

We will implement GNNs that incorporate encoder LLMs (transformers), often referred to as Graph Transformers (GTs), to under pin KG update methods. These models combine the strengths of GNNs in capturing local graph topology with the transformer's ability to model long-range dependencies. GTs typically extend the self-attention mechanism to operate on graph-structured inputs, allowing nodes to attend to their neighbors or even the entire graph. We will overcome limitations of traditional GNNs, such as over-smoothing and over-squashing.

In Phase I, we will leverage our TLP stack (Objective 2) to build a GNN optimized for the POC. Our 5+ years of LLM development experience will ensure success.

Graph Language Models (GLMs)

We will explore the application GLM architecture for KG updates. GLMs synergize the strengths of both LLMs and GNNs while overcoming their respective weaknesses. Published results show that GLM consistently outperforms traditional LM and GNN baselines, highlighting its versatility and effectiveness across various scenarios.

In Phase I, the GLM will be initialized with parameters from a pretrained TLP Encoder, allowing for a deeper comprehension of individual graph entities and relations. Furthermore, the GLM's architecture will be enhanced to integrate graph-specific biases, optimizing the representation and flow of knowledge throughout the graph. As a result, GLMs can seamlessly process graphs, textual data, and hybrid inputs consisting of both. We will integrate the GLM in Phase I POC and demonstrate efficacy.

Gen AI Knowledge Graph Completion

Depending on challenges faced by encoder-based algorithms we will explore innovative methods to use generative AI and decoder LLMs. We will integrate structural data into decoder LLMs, with the primary objective of enabling structure-aware reasoning. We will utilize LLM techniques such as in-context learning and instruction tuning and foundational approaches for injecting structural information into LLMs. We will incorporate a specialized structural pre-training phase, which will enable the model to grasp complex entities and relationships within KGs and encode them as structural embeddings. These structural embeddings will drive KG updates through schema reshaping, relationship predictions, and KG completions. We will leverage TLP SLMs for GenAI KG completion – improving accuracy, security, memory footprint, and performance.

In Phase I, we will research Generative AI-based KG update algorithms for efficacy in our KG update stack. We will deliver a report outlining results and a list of methods for development in Phase II.

Feasibility and Efficacy

Professor Derr is a leader in AI KG toolset development and has direct experience working on KG at AFRL/RI. We will leverage this expertise and experience to ensure success. We have identified several published peer-reviewed techniques for each KG update tool and also accessed associated permissive open-source code. We will leverage Inspird's proven in-house codebase and over 5 years of LLM productization experience to integrate and deliver these technologies. Additionally, we have a history of high SBIR value delivery.

Objective 2: Technical Language Processing (TLP) Toolkit

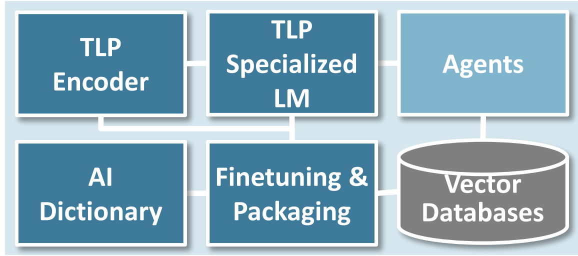

Figure 5: TLP Toolkit Underpins the DKG

Our custom TLP toolkit (Figure 5) will underpin the interactive, dynamic knowledge graph application. TLP provides a full stack of synergistic AI models that enable advanced KG updates and scalability. Our world-record TLP Encoder will enable GLMs and GNNs. Our state-of-the-art Specialized Language Models (SLMs) will provide an alternate KG update technology path. Comprehensive fine-tuning approaches will be integrated to optimize models for graph updates and package them for use in the DKG toolset. In future phases, agents will integrate external sources of data.

The TLP toolkit overcomes challenges faced by traditional LLM tools in meeting requirements of the dynamic, interactive KG toolset. Conventional LLMs need enormous compute resources and server farms to deploy, often requiring high-speed internet access, which raises information security concerns. The TLP toolkit is designed to support network isolation and secure operation.

Furthermore, conventional LLM vocabularies model only a small subset of domain-specific terms. In fact, these vocabularies contain only ~20,000 of the ~400,000 English words. While the LLM tokenizer's subword formulation allows high-level answers using non-vocabulary words, the quality of text processing declines significantly. Another challenge is the limited availability of data to train domain-specific and objective-specific models.

Subsections below describe our innovations, their feasibility, and Phase I deliverables.

Innovations

TLP Encoder

TLP Encoder will generate numeric representations of KG tuples and text semantics when available. These embedding vectors support KG update approaches such as GLM and GNN. TLP Encoder begins with a state-of-the-art Encoder Large Language Model, whose vocabulary is modified to integrate domain terms identified by the AI TLP dictionary. Unsupervised fine-tuning will help TLP Encoder learn domain context. Supervised fine-tuning approaches will be implemented to enhance KG processing and update accuracy. Novel semantic densification and adaptive masking algorithms address the training data scarcity. TLP Encoder is inherently secure against information leakage and robust against biases such as hallucinations or myopia. TLP Encoder achieved world record performance on benchmarks (CoLA, QQP, and MNLI).

In Phase I, we will deliver one TLP Encoder optimized for the dynamic KG POC. Our proven codebase and experience will ensure development success.

TLP Specialized Language Models (SLM)

Domain-optimized AI TLP SLMs will enable KG updates. The SLM will exceed the accuracy of very large LLMs (e.g., ChatGPT-4) for the target domain by eliminating unnecessary skills (e.g., multilingual support). More importantly, SLMs will integrate with the rest of the TLP toolkit to enhance security and reliability, while minimizing resource needs. We will build SLMs by updating their vocabulary with the TLP Dictionary.

Unsupervised and supervised fine-tuning will enhance accuracy for KG updates. In future phases, a suite of SLMs will be developed, each optimized for different aspects of KG updates.

In Phase I, we will demonstrate the ability to optimize SLMs for the dynamic KG application. Our proven codebase and 5+ years of LLM productization experience will ensure success.

LLM Fine-tuning and Packaging

Robust algorithms will be implemented to fine-tune and optimize LLMs used in the TLP stack for KG use. Fine-tuning LLMs is complex because it can lead to the model forgetting pretrained knowledge. TLP Encoder and SLMs will be built by updating the base LLM vocabulary with the TLP Dictionary discovered terms. Unsupervised fine-tuning with domain-specific repositories will accurately model semantics in the selected domain. Additional supervised fine-tuning will optimize TLP for KG updates.

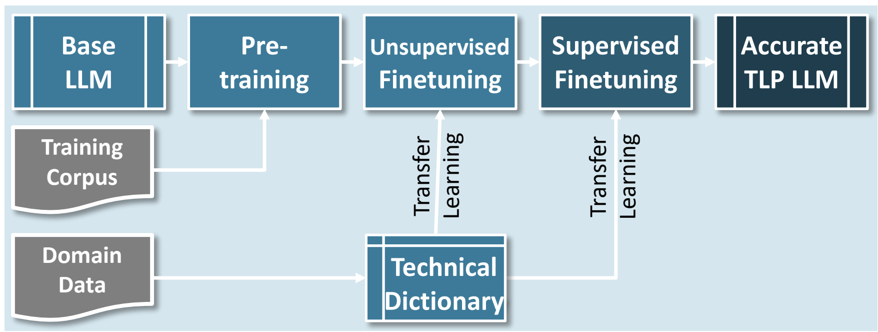

Figure 6: Proven TLP Fine-tuning Approach

Parameter efficient fine-tuning and Low Rank Adaptation (LoRA) will accelerate fine-tuning. Novel algorithms will ensure fine-tuning does not destroy foundational model learning. Approaches such as Direct Preference Optimization (DPO) will improve content quality and reduce user curation load. Synthetic data generation will maximize the benefit from any authoritative data or domain-specific text.

A variety of compression methods will be integrated to decrease model memory footprint and increase inference speed, while minimizing accuracy loss. These methods include distillation, quantization, and structured and unstructured pruning. Quantization will reduce model memory footprint and performance without greatly impacting accuracy. Unstructured and structured pruning will construct sparse models and reduce model size. Reduced model size will ensure adequate performance on portable devices.

In Phase I POC, we will leverage our mature codebase and demonstrate Encoder fine-tuning and quantization. In future phases, we will implement a full suite of compression and optimization tools.

AI Dictionary

Generating LLM vocabularies manually is cost prohibitive, especially with evolving terminologies. TLP will automatically generate a dictionary (Figure 7) of any domain's terminology and abbreviations. TLP will ingest available dictionaries and glossaries and automatically discover additional terms by scanning available repositories. Frequently used terms will be integrated into the LLM vocabularies to improve accuracy. The dictionary will allow different expansions for the same abbreviation. For example, depending on context, the term LOC may mean location, local, line of control, line of contact, etc.

In Phase I POC, we will leverage our mature codebase and demonstrate dictionary generation for representative USAF data. In future phases, we will optimize AI models for the dynamic KG application.

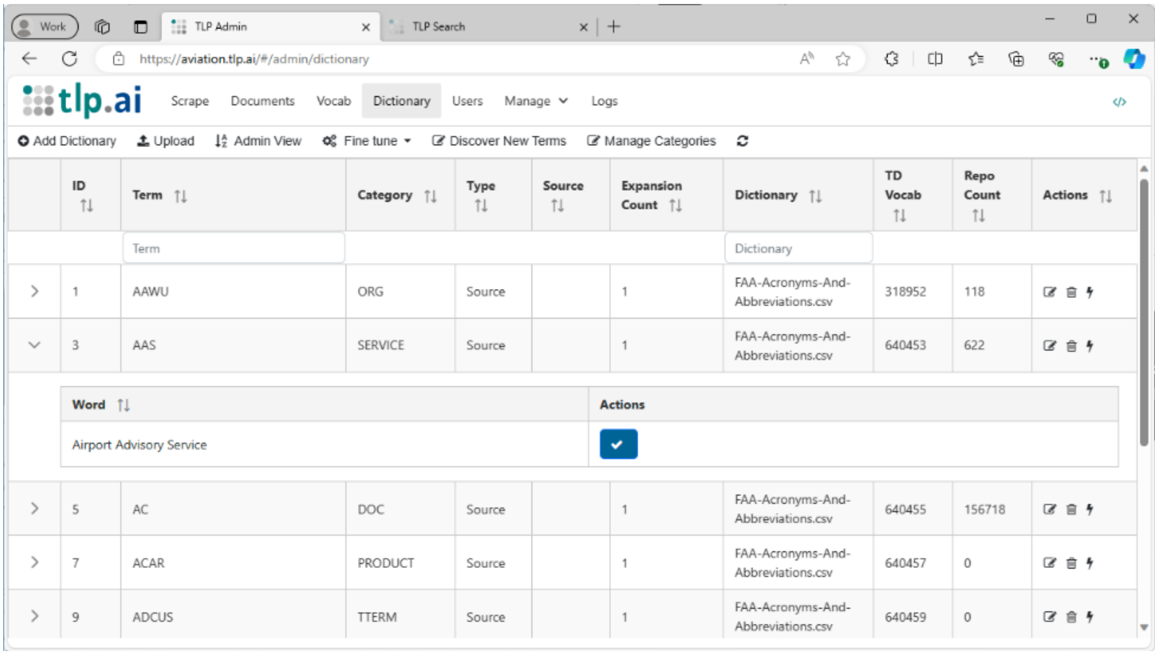

Figure 7: Inspird's Proven AI Dictionary

Agents

In future phases, LLM agents will be implemented to support tool usage. Tools will allow simple integration of external data sources and streaming data into the knowledge graph. In future phases, agent Reflection will be implemented to evaluate and improve update quality.

Feasibility and Efficacy

We will ensure successful development of KG update toolset using proven in-house codebase, robust open-source code, and over 5 years of LLM productization experience. We have a mature codebase for the TLP Encoder and vector database. TLP Encoder achieved world record performance on three benchmarks (CoLA, QQP and MNLI) – beating industry giants such as Google and Microsoft with models 10X larger. We have developed unique approaches for updating, fine-tuning and deploying generative AI SLMs. We have built a trusted in-house curated training corpus to train TLP. Our training corpus eliminates vulnerabilities such as data poisoning. Our commercial tools implement a full ML pipeline that includes data collection, LLM fine-tuning, testing, validation and integration.

We have COTS TLP products (Figure 10) for building an Authoritative Source of Truth for global aviation and medical device regulations. We have demonstrated the ability to semantically process tens of millions of lines in less than 1 second on desktop-class computers. We initially developed TLP under a USAF Digital Airworthiness SBIR. TLP was further matured under commercial funding. TLP was proven under DARPA Smart Model Repository and MDA Authoritative Source of Truth SBIRs. Our codebase, skills, and experience will ensure successful delivery of Objective 1.

Objective 3: Interactive Knowledge Graph Update Application

Our KG update solution will be delivered as a modern web-based and cloud-capable tool (Figure 8). A user web portal will allow intuitive access to artifacts. An AI search engine will be integrated to help users quickly navigate to the node of interest. An administration web portal will enable security, access control and maintenance of the toolset. A server will provide secure access to the data.

All data will be stored in an industry-standard database with a robust, extensible schema. The database will support compartmentalization and segregating data into distinct compartments based on the principle of least privilege. We will architect the database for multi-level security and ensure that the database does not contain data that exceed classification rulesets by compilation.

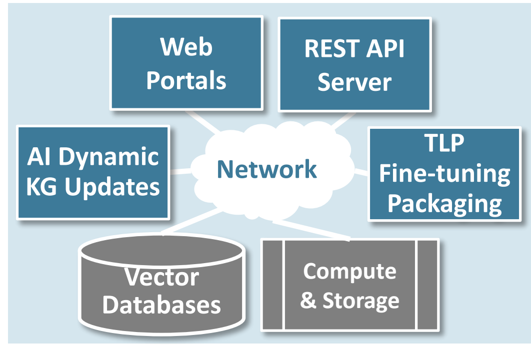

Figure 8: KG Update Toolset Architecture

In Phase I, we will design a web portal to demonstrate user interaction with the knowledge graph. We will design a state-of-the-art vector database to support the POC and store the open-source data. We will design the application for prototype development in Phase II. We leverage over a decade of experience developing enterprise applications, data curation, security, and access control, along with our proven codebase to ensure success.

3 Phase I Statement of Work

Phase I will define use cases and architectures, identify candidate AI models, and validate usability. We will build a POC, demonstrate feasibility, and engage AFRL/RI stakeholders in prototype design for Phase II development. Subsections below provide details:

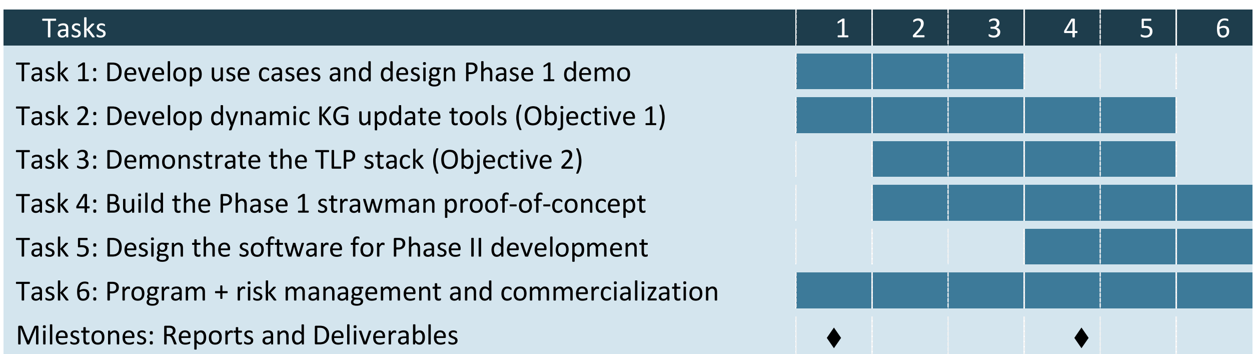

Figure 9: Phase I Schedule

Task 1: Develop use cases and design Phase 1 demo

We will define a set of use cases for the dynamic interactive KG toolset. We will build on our team member Kyle DePierre's experience as a veteran USAF officer to address user needs. We will perform a literature review to identify leading-edge frameworks and add to use cases. We will work Prof. Derr, a leader in the field of knowledge graph. We will work with the TPOC to refine use cases. If desirable for AFRL/RI, we will survey stakeholders to identify user needs. We will aggregate use cases and build related workflows. These use cases and associated workflows will drive Phase I tasks.

We will design a demonstration POC around the use cases and workflows that prove dynamic KG update feasibility and efficacy in Phase I. We will build training and demonstration data repository to support the POC (NBATransactions). We will work to gather additional representative open-source data, if necessary. We will generate synthetic data to supplement open-source data, if required. We will incorporate any AFRL/RI info, if available.

We will develop realistic evaluation sets to measure and prove KG update effectiveness in actual use. These evaluation sets will prove accuracy, reliability, bias mitigation, and safety. Development of evaluation sets is a critical, hard, and frequently overlooked task in LLM tool development. We will strategically incorporate human evaluation to improve efficacy. We will develop comprehensive sets of metrics to measure accuracy, reliability, and quality. Performance metrics will include latency, throughput, time per output token, etc.

The deliverables for this task will include use cases, workflows, training data, evaluation sets, and metrics. We will also deliver a report outlining progress and lessons learned.

Task 2: Develop dynamic KG update tools (Objective 1)

We will collaborate with AFRL/RI to refine dynamic KG update requirements, if needed. We will develop cutting-edge algorithms to dynamically update KGs for manual KG modifications or automatic anomaly detection. We will leverage Vanderbilt's R&D and peer-reviewed work. We will implement algorithms for subgraph extraction based on user modification. We will use GNN/GLM to identify inconsistencies in the modified KG and identify suggested changes using recursive message passing (GNNs) and relationship matrices (GLMs). We will demonstrate subgraph to tree conversion to visually present proposed changes to the user. The deliverables will include a set of candidate models for the POC (Task 4).

We will select and optimize GNNs and GLMs for KG correction. We will utilize our proven TLP stack (Task 3). The TLP encoder will underpin Graph Transformers for both the GNN and GLM-based approaches. We will update, fine-tune and optimize the TLP encoder for each approach (Task 3). The deliverables will include

In Phase I, we research and evaluate several algorithms for efficacy and integrability with our KG update stack. These algorithms will include KG Completion, KG Schema Reshaping, Dynamic Embedding Updates, and Gen AI KG Completion. The deliverables for this research will be a set of candidate models for development in Phase II (Task 5). We will also deliver a report outlining progress, lessons learned, and requirements for Phase II prototype development.

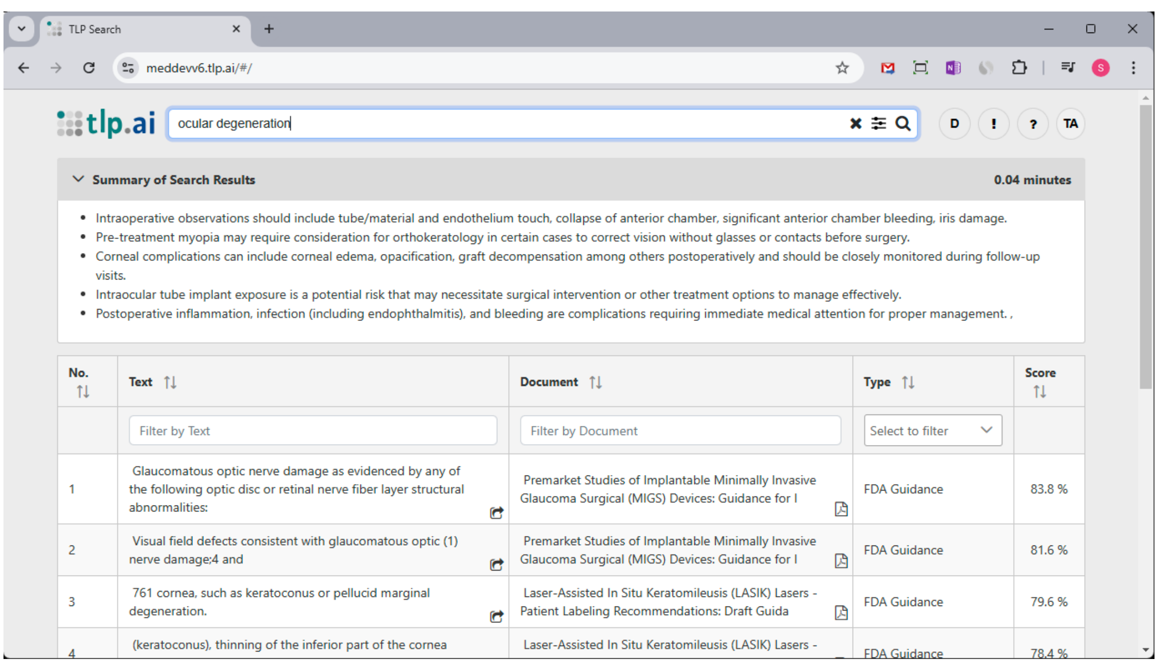

Figure 10: TLP Toolkit Enables Inspird's COTS Authoritative Source of Truth (ASoT)

Task 3: Demonstrate the TLP stack (Objective 2)

We will collaborate with AFRL/RI to refine Dynamic KG LLM requirements, if needed. We will select candidate AI models that best satisfy user needs defined in Task 1 and technical requirements in Task 2. We will leverage our proven codebase and use well-supported open-source models with permissive licenses as a backup. We will develop and evaluate USAF domain specialization approaches. We will demonstrate dictionary generation and vocabulary formulation. We will update selected TLP Encoder with the vocabulary. We will implement approaches to rapidly fine-tune TLP Encoder. We will perform supervised fine-tuning. We will optimize TLP Encoder to produce 90% precision and recall for retrieval.

We will optimize SLM for text quality metrics defined in Task 1. We will evaluate and select approaches for packaging AI models for use in the dynamic KG toolset. We will identify approaches to build an ML pipeline that can automate AI model fine-tuning and packaging in Phase II. Our proven codebase and 5+ years of LLM productization experience will ensure success.

The deliverables for this task will be a set of fine-tuned models for the POC (Task 4). We will also deliver a report outlining progress, lessons learned, and requirements for Phase II prototype development.

Task 4: Build the Phase 1 strawman proof-of-concept

Under this task, we will integrate various elements of the dynamic KG toolset to demonstrate the efficacy of technologies developed in Phase I. We will leverage available KG visualization components for the Phase I POC. We will deploy the POC on our servers and show the ability to satisfy user needs defined in Task 1. We will perform extensive testing and generate metrics. We will demonstrate effectiveness against evaluation sets. We will document lessons learned and plan further development in Phase II.

We will leverage Mr. Kyle DePierre's direct USAF work experience to translate POC results into situational awareness benefit. We will review results with AFRL/RI stakeholders and industry partners and incorporate their feedback in Phase II development plan (Task 5). The deliverables will include the POC and a report.

Task 5: Design the software for Phase II development

We will design the dynamic KG toolset for prototype development in Phase II. We will ensure that the dynamic KG toolset design satisfies all use cases identified in Task 1. We will design a mobile client and an administration portal. We will leverage frameworks such as D3.js to design best-in-class GUIs (Figure 11). We will design a state-of-the-art vector database to support the POC. We will leverage our extensive data curation, knowledge base development, security, and access control, along with our extensive codebase to ensure success.

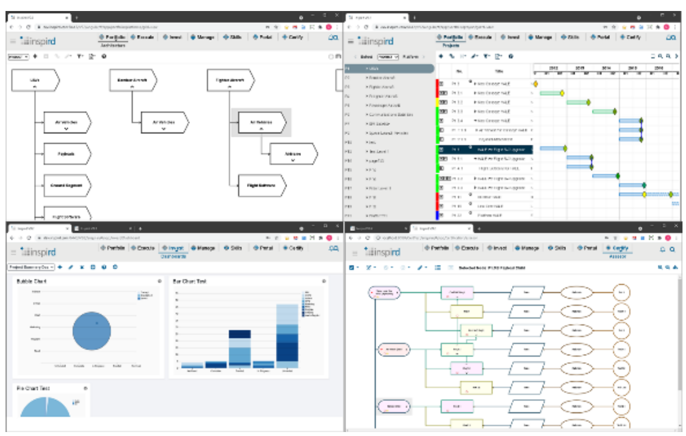

Figure 11: Inspird's Cutting-edge COTS GUI

We will design a cloud-friendly secure server. The server will be designed to be extensible, scalable, and with cybersecurity/IA necessary to run in the DoD intranet (DEVSECOPS 2.0). Inspird's proven client and server code will provide an effective foundation for the dynamic KG toolset development. We will leverage the Phase I POC to gain stakeholder feedback and incorporate it into the prototype design.

We will integrate ITS into the development plan with gradual progression from Interim Authority to Test (IATT) to Interim Authority to Operation (IATO) and full Authority to Operate (ATO). Inspird has successfully gone through a full DoD IT security source code review for our COTS software and received an ATO. The lessons learned will ensure successful software delivery and commercialization.

The deliverables for this task will include Phase II prototype requirements and metrics to demonstrate the dynamic KG toolset value. We will also deliver a report outlining progress and lessons learned.

Task 6: Program + risk management and commercialization

We will undertake tasks of technical project management. We will ensure adherence to the project schedule and fulfillment of contract requirements. The program manager will have intimate knowledge of all aspects of the dynamic KG toolset development. We do not anticipate extraordinary risks in Phase I. We have identified alternate technology development paths and published research that we can leverage in Phase I. We will use agile development methods to enhance development and minimize risks. We have an active DD-2345 from the DLA and processes in place to control CUI and ITAR information.

We will work with AFRL/RI stakeholders for commercialization. We will gain participation from additional DoD agencies in Phase II. We will engage with our industry partners on the dynamic KG toolset commercial needs. We will define broader use cases in large defense, automotive, and public safety industries (See the Commercialization section for details).

4 Related Work

Inspird has been building state-of-the-art applications that integrate AI and Engineering for over a decade. Our ongoing R&D and competencies (Figure 12) are directly applicable to the proposed effort.

Technical Language Processing

For over 5 years, we have been developing technologies for accurate semantic processing of text with jargon and abbreviations. We have launched COTS TLP products to build an authoritative source of truth for global airworthiness regulations (Figure 10). We have developed tools to automatically build a tech terminology dictionary. Our COTS products use AI to parse PDFs and extract document structure. We have developed small, robust, and secure SLMs for use in technical text generation.

We are developing TLP products to answer questions from repositories such as papers, publications, and patents. We are developing AI models to build, extend, and enhance ontologies and knowledge graphs. We have built accurate AI classifiers for categorizing technical text, AI summarizers for large technical documents and simulations, and AI meta-analysis tools to summarize numerical data. Our TLP tools have been proven on USAF, MDA, Navy, and DARPA SBIRs.

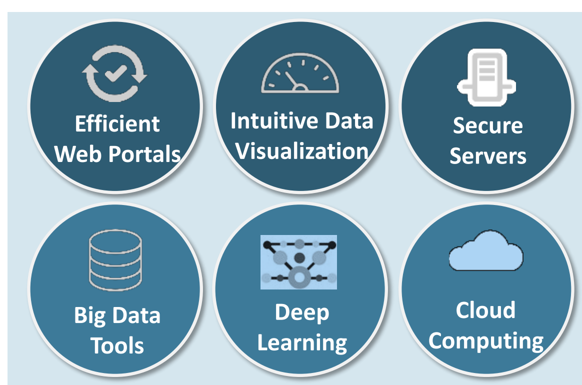

Figure 12: Inspird Core Competencies

Vision Language Models (VLMs) and Large Multimodal Models (LMMs)

We have been researching VLMs combine visual (images, videos) and textual data to enable tasks like visual question answering and semantic image retrieval. Integration of VLMs into the dynamic KG toolset will improve its utility, decision-making, and contextual understanding.

Digital Engineering R&D

We have been developing COTS tools to manage Digital Engineering Ecosystems for over a decade. Our R&D on cross-enterprise collaboration is directly applicable to the proposed effort. We have extensive experience with managing multidimensional taxonomies and have helped many large organizations develop portfolio architectures. We will leverage our practical experience to accelerate development.

Another active area of R&D for us is the use of AI to accelerate Digital Engineering. We are developing AI methods to enhance model reuse and accelerate analytics under DARPA Semantic Model Reasoning Technologies and MDA Innovative Automated Analytics SBIRs.

Software Development Competencies

We have over a decade of experience developing enterprise software. Our COTS products are supported by broad software development competencies including big data, cloud computing, and secure servers. Our software is used by diverse users including DoD primes such as Northrop Grumman. We are actively developing innovative web portals. We continue to make significant investments in data science for intuitively presenting complex information in a decision-ready format.

We have successfully obtained a DoD ITS ATO and deployed our COTS tools on the DoD intranet. We will use our COTS agile development and ensure successful development and long-term supportability.

Vanderbilt University Network and Data Science Lab

The Network and Data Science (NDS) lab is a leader in knowledge graphs (KGs) and question answering. NDS also developed an LLM-based KG traversal agent to intelligently retrieve question-relevant contexts by navigating the constructed KG. The agent was fine-tuned to adaptively traverse the most promising neighbors within the surrounding subgraph, based on the context of already collected information and what it deemed necessary to complete the question answering task. This work earned the Best Paper Award at NeurIPS 2023. In their acclaimed work, “Knowledge Graph Prompting for Multi-Document Question Answering,” NDS introduced methods for constructing knowledge graphs from sets of documents. NDS’s research on “Augmenting Textual Generation via Topology Aware Retrieval” offered a new perspective on RAG frameworks, moving beyond local proximity-based retrieval. NDS equipped retrievers with topological awareness and structural role-based similarity.

Additionally, the NDS lab has conducted numerous works in collaboration with industry partners, including Adobe Research, Snap Inc., The Home Depot, and Visa Research, especially in recommender systems. In “Knowledge Graph-based Session Recommendation with Session-Adaptive Propagation,” which constructs a KG according to historical user behavioral data paired with e-commerce domain knowledge to develop a heterogeneous graph transformer-based architecture. This work will be applicable to the proposed effort in providing improved context to an adaptive retriever.

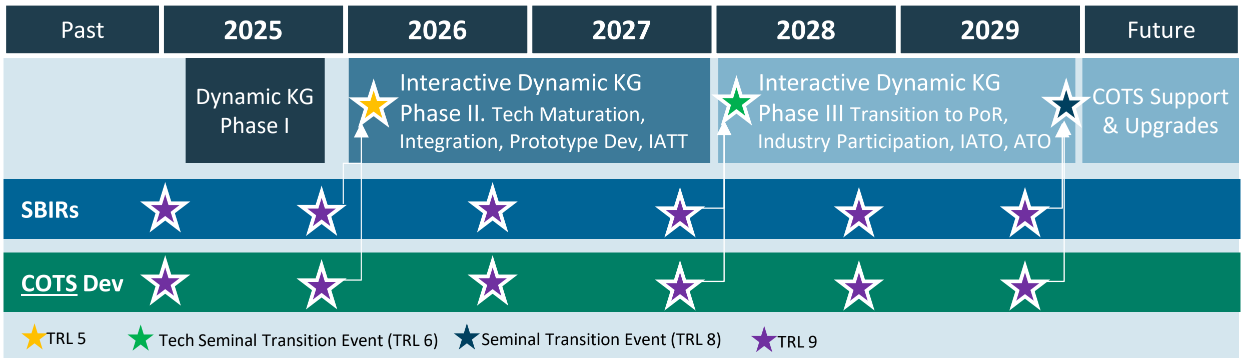

Figure 13: Interactive Dynamic Knowledge Graph Toolset Roadmap and Commercialization Plan

5 Relationship with Future Research or Research & Development

Figure 13 shows the roadmap and transition plan for the dynamic KG toolset. Phase I POC will be followed by prototype development in Phase II, where we will mature technologies and integrate them into a TRL 6 prototype. We will refine the use cases defined in Phase I to cover a broader range of scenarios and questions. We will refine and mature the dynamic KG toolkit.

Phase II will prove the dynamic KG toolset usability with more realistic KGs and more comprehensive AI models. We will adapt KG correction to handle dependencies between updates, where the likelihood of one change affects others in the graph. This more advanced model will account for interrelated updates and improve the overall consistency of the KG. We will implement a robust method for KG completion using pre-trained encoder LLMs. We implement innovative schema and query reshaping algorithm, that updates the predefined KG schema into a more data-oriented version. We will implement a full suite of LLM compression and optimization tools. We will design agents to support access to dynamic data sources and facilitate extensibility. We will also optimize LLMs for temporal and geospatial information processing. We will explore Vision Language Models (VLMs) and Large Multimodal Models (LMMs) for improving the support for different types of entities. We will optimize the AI dictionary for dynamic KG application.

We will implement robust servers, databases, and administration portals to support the dynamic KG toolset in Phase II. We will incorporate DEVSECOPS into the prototype development. We will obtain an Interim Authorization to Test (IATT) to test the dynamic KG toolset prototype in the AFRL/RI IT environment. We will perform extensive testing and validation to ensure the prototype's reliability and effectiveness. We will refine the user interface and experience based on iterative feedback from users.

Phase II will be followed by commercialization in Phase III. We will establish partnerships and collaborations with relevant stakeholders. We will mature the dynamic KG toolset to TRL 8. Phase III work will include development of help files and how-to videos to ensure fast adoption of the dynamic KG toolset. We will work with the AFRL/RI ITS to obtain an Interim Authorization to Operate (IATO). We will deploy the dynamic KG toolset in AFRL/RI environment and test using real data.

We will leverage our robust TRL 9 codebase to accelerate development. We will jumpstart development in each phase from our latest code repositories. In Phase III, we will transition the dynamic KG toolset code to COTS development practices. We will support users and maintain the dynamic KG toolset using our mature COTS agile development processes.

6 Commercialization

Military and Commercial Application

KGs enable a holistic data approach that not only improves decision-making but also enhances mission success and safety across a range of defense and intelligence applications. KGs are transformative tools for enhancing situational awareness, pattern of life analysis, threat detection, targeting operations, and more. By integrating diverse data sources—such as sensor outputs, communication logs, geospatial data, and human intelligence—into a semantically rich and interconnected structure, KGs enable a deeper understanding of complex and dynamic environments. For situational awareness, they provide real-time contextualization by linking entities, events, and relationships to offer a comprehensive operational picture. In pattern of life analysis, KGs reveal behavioral patterns and deviations, helping analysts identify unusual activities and assess potential threats. They support threat detection by cross-referencing multiple data points to flag suspicious entities or actions. In targeting operations, KGs facilitate precision by mapping out key assets, affiliations, and locations, helping decision-makers prioritize and assess risks effectively. Hence, the proposed dynamic KG toolset has broad applicability across the Department of Defense (DoD), including the Army, Navy, Missile Defense Agency (MDA), Space Force, and others. Additionally, it will appeal to various government sectors, such as the Department of Homeland Security and the Department of Energy. Local and state governments, law enforcement agencies, and public safety organizations will also benefit from this system.

KGs have found widespread commercial applications across various industries, revolutionizing the way businesses handle and utilize data. KGs have become an integral part of LLM-based Retrieval Augmented Generation (RAG) tools. In e-commerce, companies like eBay and Amazon have implemented KGs to enhance product recommendations, improve search functionality, and create more personalized shopping experiences. Search engines like Google, Yahoo, and Microsoft have integrated KGs into their platforms to provide more accurate and contextually relevant search results. In the cybersecurity sector, KGs are employed to map system endpoints, identify vulnerable areas, and fill security gaps, thereby strengthening overall network protection. The healthcare industry has also benefited from KGs, using them to establish connections between healthcare professionals, publications, clinical trials, and events, leading to improved patient care and research outcomes. Additionally, KGs are being utilized in customer relationship management to gain deeper insights into consumer behavior, product relationships, and marketing strategies. Our dynamic, interactive KG tool will have a broad applicability to these segments.

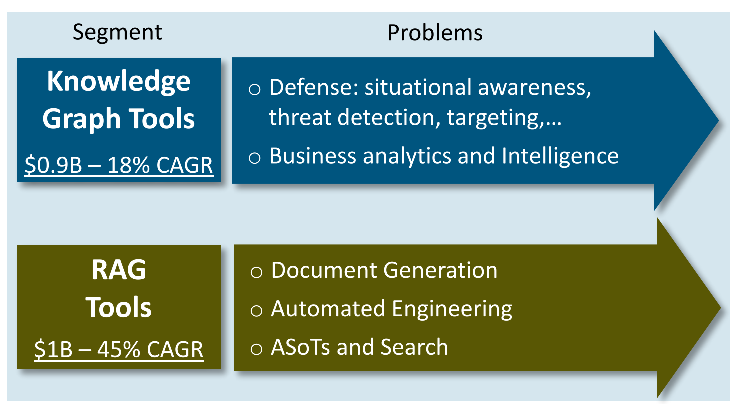

Figure 14: Market Segments Addressed

Market Size and Marketing Plan

This SBIR is developing technologies that target the market segments for KG tools (Figure 14). KGs are also applicable to the fast-growing RAG market segment. In 2023, the total market for KG was $2B and is expected to grow to $13B by 2030. A key driver for this growth is the availability of LLM tools. We will target high complexity and accuracy niche within this market. We estimate that the current size of this niche is $50M and growing fast.

We will support three configurations for delivering the dynamic KG toolset software. A low-cost ($100 per month per user) cloud-based SaaS will attract small businesses and emerging economies. To drive adoption, we will offer a free trial subscription to students and not-for-profit organizations. For enterprise and government clients, we will provide an on-premise installation with high security and guaranteed performance. Pricing will depend on the specific configuration but is expected to range between $100,000 and $400,000.

We have established marketing and sales processes to aid commercialization, using both direct and indirect sales strategies. We participate in conferences and publish papers to reach a wide audience. Our COTS software will serve as a beachhead for sales, and our existing customer base will help optimize transition and jumpstart sales.

Business Plan and Funding Requirements

Figure 13 illustrates the roadmap and commercialization plan for the dynamic KG toolset. The section on Relationship with Future Research or Research & Development provides a detailed description of the roadmap. The proposed dynamic KG toolset presents a compelling business case, especially with its synergies with our RAG product portfolio (Figure 15). We anticipate needing $1-2M in Phase III funding and expect a timeline of 18-24 months. Over the following 5-year period, our goal is to generate over $25M in revenue from government and industry sectors. We plan to supplement these revenues with venture funding, which will help us scale quickly, attract talent, build market presence, and enhance our marketing and sales efforts. Our AI subsidiary, tlp.ai, will be instrumental in securing venture funding and driving transition.

Figure 15: Inspird's Synergistic LLM/RAG Product Portfolio

Company Background and Competition

We develop and supply software tools to government and industry alike. Our AI RAG tools (Figure 15) are synergistic and complementary to the dynamic KG toolset. We have a proven track record of commercialization, high value delivery, and have generated commercial revenues since the company's inception in 2012. The DoD has designated Inspird a critical capability provider to secure the Defense Industrial Base and included us in the Trusted Capital Marketplace. Inspird was recognized by Gartner, a leading IT analysis firm, as a cool vendor in R&D.

Our customers include military organizations such as the Assistant Secretary of Defense for Research & Engineering and DoD primes such as Northrop Grumman. We have successfully obtained an ATO from DoD and deployed our tools on the DoD intranet. We have extensive experience packaging software and delivery. We have mature processes for agile development and for Cybersecurity Maturity Model Certification (CMMC) / DEVSECOPS 2.0 compliance. We have secure methods for delivering and updating software. We have experience building help files, developing user forums, and deploying support channels. We also have expertise developing training programs and providing implementation services.

Many organizations, such as Anthropic, Cohere, Google, Meta, and OpenAI are focused on general-purpose AI. Others are addressing smaller market niches. For example, Contextual is developing general-purpose RAG and Aitomatic focuses on domain knowledge capture. Inspird is unique in its integration of accurate, secure, and trusted technical language processing in specialized domains. We are differentiated by our world-record AI models, deep know-how, and patent protection. To drive commercialization, we have a strategic alliance with Chromologic, a company that has generated $3 in non-SBIR revenue for each $1 in SBIR funding. Chromologic also has an extensive track record of delivering exceptional value to AFRL/RI.

7 Key Personnel

Dr. Sandeep Mehta will lead the development of the dynamic KG toolset. Mr. Kyle DePierre, a USAF veteran, will provide user perspective. Overview of their expertise is below and CVs are attached. Data scientists and software developers from our team will support this effort as necessary.

Dr. Sandeep Mehta, Principal Investigator

Dr. Mehta brings over 35 years of software development experience, including 5+ years of leading-edge LLM and 10+ years of AI/ML development. He led the development of the TLP stack that holds world records on three benchmarks. He has also developed innovative new AI tools such as ontologically-assisted analytics and hierarchical classification. His doctoral research focused on advanced methods including Bayesian statistics and probabilistic optimization that are directly applicable to AI. Dr. Mehta has been the PI on ten AI and DE SBIR efforts funded by Army, DARPA, MDA, Navy, and USAF.

Dr. Mehta has held a wide variety of leadership positions in large organizations including Northrop Grumman and Boeing. At Price Waterhouse Consulting, he helped many customers with diverse challenges from strategic planning to supply-chain management. Over his career, he has cultivated a deep professional network of technologists and executives that can help drive tech transition and commercialization. He has extensive experience with best practice business process implementation and drives agile software development at Inspird.

Dr. Mehta has received several prestigious awards, such as *Outstanding Contribution to the Engineering Profession* from the American Society of Mechanical Engineers. He received an M.B.A. from the Anderson School of Management at UCLA and an M.S. and Ph.D. in engineering from Vanderbilt University. Dr. Mehta has published more than 15 technology and management articles. He is an inventor on 5 patents.

Prof. Tyler Derr, Knowledge Graph-assisted RAG Expert, Vanderbilt University

Dr. Derr is an Assistant Professor in the Department of Computer Science, Teaching and Affiliate Faculty in the Data Science Institute, and Faculty Fellow in the Frist Center for Autism and Innovation at Vanderbilt University. Tyler directs the Network and Data Science (NDS) lab, which focuses on data mining and machine learning. He has been prolific in the area of graph neural networks (GNN), a newer class of deep learning methods that leverage graph-structured data to improve predictions. His work focuses on developing robust methods to overcome real-world data quality challenges, such as imbalanced, biased, and noisy data. In one notable project, Prof. Derr introduced a GNN-based approach using zero-shot learning and a knowledge graph to enhance classifiers.

Prof. Derr has published multiple works at the intersection of graph and text mining, including "Knowledge Graph Prompting for Multi-Document Question Answering," which won the Best Paper Award at the Graph Learning Frontiers. He has led the development of tools that construct a knowledge graph from a set of PDF documents at various levels of granularity (e.g., sentence, page, and also from complex table data). He has leveraged a novel LLM-based knowledge graph traversal agent to retrieve the relevant context from the KG and improve performance on multi-document question answering.

Prof. Derr has received numerous awards including the NSF Career Award. He was part of the Visiting Faculty Research Program (VFRP) of The Air Force Research Laboratory's Information Directorate (AFRL/RI). He received his PhD from Michigan State University and MS from the Pennsylvania State University. He has published over 50 papers in highly ranked journals and conference proceedings.

Kyle DePierre, AFNWC Operational Expertise and System Architect

Kyle DePierre has over 18 years of Aerospace and Defense experience, including 8 years of System Engineering experience for DOD and Intelligence programs. He has held positions as chief architect for National Space Defense Center and Joint Space Operations Center ground software programs, ICBM, Aircraft and Army ground vehicles. He has held various senior systems engineering roles at Northrop Grumman and Booz Allen Hamilton. He has active USAF/USSF Top Secret clearance with SCI eligibility and counter intelligence polygraph.

Mr. DePierre is an Air Force veteran and former acquisition officer at the Space and Missile Systems Center. He has relevant experience in architecting multi-level security architectures, to include architecting databases that do not contain data that exceed classification rulesets by compilation. He received a Bachelor of Science in Systems Engineering from the United States Air Force Academy and his M.B.A from the Wharton School at University of Pennsylvania.

Rich Cormier, Principal Engineer

Richard Cormier is Inspird's Vice President for Engineering and Support. Mr. Cormier brings over 30 years of experience in all aspects of Information Technology, including 13 years as a Chief Information Officer (CIO) at a Fortune 500 company. Mr. Cormier has an outstanding track record for planning, analyzing, and managing enterprise systems implementation initiatives across global organizations such as Chevron, Monsanto, and Edwards Lifesciences.

Mr. Cormier is a seasoned executive who has blended disciplined program management with broad business acumen to deliver technology-enabled value creation. At Edwards, Mr. Cormier delivered a custom R&D Portal with data mining capabilities that provided easy access to a wide variety of engineering information, including requirements, specs, CAD drawings, etc., for both engineers and management.

8 Foreign Citizens

None.

9 Facilities/Equipment

No significant purchases will be required to support this contract. Inspird will utilize existing internal compute resources to develop and demonstrate the dynamic KG toolset in Phase I.

10 Subcontractors/Consultants

Vanderbilt University (Prof. Tyler Derr) is a subcontractor. Vanderbilt will provide expertise and codebase for Knowledge Graphs.

11 Prior, Current, or Pending Support of Similar Proposals or Awards

None.

DSIP Defense SBIR/STTR Innovation Portal

Cost Volume

SBIR Phase I Proposal

Proposal Number	F244-0001-0010
Topic Number	AF244-0001
Proposal Title	Dynamic, Interactive Knowledge Graph
Date Submitted	10/28/2024 04:03:17 AM

Firm Information

Firm Name	InspiRD, Inc
Mail Address	24261 Chrisanta Dr, Mission Viejo, California, 92691
Website Address	http://inspird.com
UEI	D859W8HNHD86
Duns	054205491
Cage	6DC34

Total Dollar Amount for this Proposal	$139,937.76
Base Year	$139,937.76
Year 2	$0.00
Technical and Business Assistance(TABA)- Base	$0.00
TABA- Year 2	$0.00

Base Year Summary

Total Direct Labor (TDL)	$72,560.00
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 20%) x Base (TDL+TOH)	$14,512.00
Total Firm Costs	$87,072.00
Subcontractor Costs	
Total Subcontractor Costs (TSC) 1	$42,500.00
Total Subcontractor Costs (TSC)	$42,500.00
Cost Sharing	-$0.00
Profit Rate (8%)	$10,365.76
Total Estimated Cost	$139,937.76
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 20%) x Base (TDL+TOH)	$0.00
Total Firm Costs	$0.00
Subcontractor Costs	
Total Subcontractor Costs (TSC) 1	$0.00
Total Subcontractor Costs (TSC)	$0.00
Cost Sharing	-$0.00
Profit Rate (8%)	$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
Computer and Information Research Scientist/ Principal Investigator (Sandeep Mehta)	$125.00	240	25	$7500.00	$37,500.00
Engineers, All Other/ USAF Operations Expert (Kyle DePierre)	$100.00	64	25	$1600.00	$8,000.00
Computer and Information Research Scientist/ Senior Data Scientist	$75.00	98	25	$1837.50	$9,187.50
Software Developer/ Full Stack Developer	$50.00	32	25	$400.00	$2,000.00

Subtotal Direct Labor (DL)	$56,687.50
Labor Overhead (rate 28%) x (DL)	$15,872.50
Total Direct Labor (TDL)	**$72,560.00**

Subcontractor Costs

Subcontractor- Vanderbilt University

Vanderbilt University	$42,500.00
Total Subcontractor Costs (TSC) 1	**$42,500.00**
Total Subcontractor Costs (TSC1)	**$42,500.00**

G&A (rate 20%) x Base (TDL+TOH)	$14,512.00
Cost Sharing	-$0.00
Profit Rate (8%)	$10,365.76

Total Estimated Cost	$139,937.76
TABA	$0.00

Year 2

Direct Labor Costs

Category / Individual-TR	Rate/Hour	Estimated Hours	Fringe Rate (%)	Fringe Cost	Cost
Computer and Information Research Scientist/ Principal Investigator (Sandeep Mehta)	$125.00	0	25	$0.00	$0.00

Subtotal Direct Labor (DL)	$0.00
Labor Overhead (rate 28%) x (DL)	$0.00
Total Direct Labor (TDL)	**$0.00**

Subcontractor Costs

Subcontractor- Vanderbilt University

Vanderbilt University	$0.00
Total Subcontractor Costs (TSC) 1	**$0.00**
Total Subcontractor Costs (TSC1)	**$0.00**

G&A (rate 20%) x Base (TDL+TOH)	$0.00
Cost Sharing	-$0.00
Profit Rate (8%)	$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: Sandeep Mehta

Phone: (818) 694-3944

Phone: smehta@inspird.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
Computer and Information Research Scientist	Principal Investigator	PhD	35	240	$125.00	25	$37,500.00
Engineers, All Other	USAF Operations Expert	Master's Degree	18	64	$100.00	25	$8,000.00
Computer and Information Research Scientist	Senior Data Scientist	PhD	4	98	$75.00	25	$9,187.50
Software Developer	Full Stack Developer	Bachelor's Degree	8	32	$50.00	25	$2,000.00

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.
Standard rates

Direct Labor Cost ($): $56,687.50

Year2

Category	Description	Education	Yrs Experience	Hours	Rate	Fringe Rate	Total
Computer and Information Research Scientist	Principal Investigator	PhD	35	0	$125.00	25	$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.
Phase I is six months only. All work will be performed in base year.

Direct Labor Cost ($): $0.00

Sum of all Direct Labor Costs is($):	$56,687.50

Overhead

Base

Labor Cost Overhead Rate (%)	**28**
Overhead Comments:	
Overhead Cost ($):	**$15,872.50**

Year2

Labor Cost Overhead Rate (%)	**28**
Overhead Comments:	
Overhead Cost ($):	**$0.00**
Sum of all Overhead Costs is ($):	**$15,872.50**

General and Administration Cost

Base

G&A Rate (%):	**20**
Apply G&A Rate to Overhead Costs?	**YES**
Apply G&A Rate to Direct Labor Costs?	**YES**
Apply G&A Rate to Subcontractor Costs?	**NO**
Please specify the different cost sources below from which your company's General and Administrative costs are calculated.	
G&A Cost ($):	**$14,512.00**

Year2

G&A Rate (%):	**20**
Apply G&A Rate to Overhead Costs?	**YES**
Apply G&A Rate to Direct Labor Costs?	**YES**

Apply G&A Rate to Subcontractor Costs? **NO**

Please specify the different cost sources below from which your company's General and Administrative costs are calculated.

G&A Cost ($): **$0.00**

Sum of all G&A Costs is ($): **$14,512.00**

Subcontractor/Consultants

Base

Subcontractor/Consultant:
Vanderbilt University

Budget Contact Name	Budget Contact Title	Budget Contact Phone	Budget Contact Email
Dr. Tyler Derr	Assistant Professor	(717) 644-7601	tyler.derr@vanderbilt.edu

Do you have a letter of commitment from the subcontractor/consultant? **YES**

Document uploaded for the letter of commitment:

- Vanderbilt_LOA__Signed.pdf

Are you able to provide detailed budget information for this subcontractor/consultant? **NO**

Total Cost($) : **$42,500.00**

Do you provide the authority to the Government to contact this Budget Contact? **YES**

Year2

Subcontractor/Consultant:
Vanderbilt University

Budget Contact Name	Budget Contact Title	Budget Contact Phone	Budget Contact Email
Dr. Tyler Derr	Assistant Professor	(717) 644-7601	tyler.derr@vanderbilt.edu

Do you have a letter of commitment from the subcontractor/consultant? **YES**

Document uploaded for the letter of commitment:

- Vanderbilt_LOA__Signed.pdf

Are you able to provide detailed budget information for this subcontractor/consultant?	**NO**
Total Cost($) :	**$0.00**
Do you provide the authority to the Government to contact this Budget Contact?	**YES**

Total Subcontractors/Consultants Cost ($):	**$42,500.00**

Profit Rate/Cost Sharing

Base

Cost Sharing ($):	**-$0.00**
Cost Sharing Explanation:	
Profit Rate (%):	**8**
Profit Explanation:	
Total Profit Cost ($):	**$10,365.76**

Year2

Cost Sharing ($):	**-**
Cost Sharing Explanation:	
Profit Rate (%):	**8**
Profit Explanation:	
Total Profit Cost ($):	**$10,365.76**
Total Proposed Amount ($):	**$139,937.76**

Inspird Inc.

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:
$500,000.00	$0.00	0	$0.00

Company Information

Address:

24261 Chrisanta Ave
Mission Viejo, CA 92691-4003
United States

SBC Control ID: SBC_001317907

Company Url: http://inspird.com

Company POC

Title: Dr.
Full Name: Sandeep Mehta
Phone: 8186943944
Email: smehta@inspird.com

Commercialization POC

Title: Dr.
Full Name: Sandeep Mehta
Phone: (818) 694-3944
Email: smehta@inspird.com

Additional Company Information

% Revenue for last fiscal year from SBIR/STTR funding:	Total revenue for last fiscal year:
75.0%	$500,000 - $999,999
Year Founded:	**# Employees Currently:**
2011	12
Year first Phase I award received:	**# SBIR/STTR Phase I Awards:**
2018	6
Year first Phase II award received:	**# SBIR/STTR Phase II Awards:**
2020	1
# Employees at first Phase II award:	**Mergers and Acquisition within past 2 years:**
5	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	Yes
HUBZone-Certified:	**SBC majority-owned by multiple VCOC, HF, PE firms | By what percent (%):**
No	No | N/A

Additional Investment From

	Last Submitted Version (-)	Current Version
DoD contracts/DoD subcontracts	-	$0.00
Angel Investors	-	$0.00
Venture Capital	-	$0.00
Self Funded	-	$500,000.00
Private Sector	-	$0.00
Other Federal Contracts/Grants	-	$0.00
Other Sources	-	$0.00
Additional Investment	-	$0.00
Total Investment	-	$500,000.00

Phase III Sales To

	Last Submitted Version (-)	Current Version
DoD or DoD prime contractors	-	$0.00
Private Sector	-	$0.00
Export Markets	-	$0.00
Other Federal Agencies	-	$0.00
Additional commercialization by 3rd Party Revenue	-	$0.00
Other Customers	-	$0.00
Additional Sales	-	$0.00
Total Sales	-	$0.00

Government Phase III Contracts

	Last Submitted Version (-)	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.

Accelerated Modifications using Physics-substantiated Safety Arguments (AMPS) 1 of 1

Agency/Branch:	Department of Defense/Air Force	**Manufacturing related**	No | N/A
Program/Phase/Year:	SBIR/Phase I/2018	**Subsidiaries**	tlp.ai
Topic #:	AF181-025	**Other contributing SBIR/STTR awards**	Awarded Phase II (Contract FA864920C0022) and led to the commercialization.
Contract/Grant #:	FA8650-18-P-2121		
Achieved a cost saving or cost avoidance?:	No	**Used in Federal or acquisitions program?**	No

Additional Investment From		**Phase III Sales To**	
DoD contract/subcontract:	$0.00	**Dod or DoD prime contractors:**	$0.00
Other Federal contract/grants:	$0.00	**Other Federal Agencies:**	$0.00
Angel Investors:	$0.00	**Private Sector:**	$0.00
Venture Capital:	$0.00	**Export Market:**	$0.00
Self-Funded:	$500,000.00	**3rd Party Revenue:**	$0.00
Private Sector:	$0.00	**Other Customers:**	$0.00
Other Sources:	$0.00		
Investment Total:	**$500,000.00**	**Sales Total:**	**$0.00**

CERTIFICATE OF COMPLETION

THIS CERTIFICATE IS PRESENTED TO

Sandeep Mehta, InspiRD, Inc

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

Oct 28, 2024	Oct 28, 2025
COMPLETION DATE	EXPIRATION DATE