ACQUISITION INNOVATION RESEARCH CENTER

Collective Intelligence Decision Making via Dynamic Knowledge Graph with Trusted Cross-Organizational and Privacy Preserving Integration

EXECUTIVE SUMMARY AND REPORT

JULY 2024

PRINCIPAL INVESTIGATOR

Feng Liu, *Stevens Institute of Technology*

CO-PRINCIPAL INVESTIGATOR

Xinyue Ye, *Texas A&M University*

SPONSOR

Office of the Under Secretary of Defense for Acquisition and Sustainment

DISCLAIMER

The Acquisition Innovation Research Center (AIRC) is a multi-university partnership led and managed by the Stevens Institute of Technology and sponsored by the U.S. Department of Defense (DoD) through the Systems Engineering Research Center (SERC)—a DoD University-Affiliated Research Center (UARC).

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)) and the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) under Contract HQ0034-19-D-0003, TO#0285.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

RESEARCH TEAM

Name	Organization	Labor Category
Feng Liu	Stevens Institute of Technology	Principal Investigator (PI)
Xinyue Ye	Texas A&M University	Co-PI
Ruobing Liu	Stevens Institute of Technology	Graduate Student
Jiaxin Du	Texas A&M University	Graduate Student
Shoujia Li	Texas A&M University	Graduate Student
Suphanut Jamonnak	Texas A&M University	Research Scientist
Meng Jiao	Stevens Institute of Technology	Graduate Student
Shihao Yang	Stevens Institute of Technology	Graduate Student

ACKNOWLEDGEMENTS

We thank the sponsorship of the Acquisition Innovation Research Center (AIRC) and Department of Defense (DoD) to make work possible, and we are grateful for Drs. Philip Anton and Douglas Buettner for their guidance and constructive comments.

ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
AIRC	Acquisition Innovation Research Center
CATMS	Correspondence and Task Management System
DKG	Dynamic Knowledge Graph
DoD	Department of Defense
Fed-KG	Federated Learning-Based Knowledge Graph
FL	Federated Learning
KG	Knowledge Graph
LLM	Large Language Models
LoRA	Low Rank Adaptation
ML	Machine Learning
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
Q-A	Question-Answer

EXECUTIVE SUMMARY

Our objective is to develop a comprehensive decision-making framework that harnesses collective intelligence to effectively determine the optimal integration and collaboration mechanism for responsible agencies within the Department of Defense (DoD) when tackling complex tasks that entail cross-domain and cross-organizational operations. The framework will be based on privacy preserving and cross organizational federated learning to achieve the strategic goal of improved enterprise-wise interoperability. We aim to establish DoD domain specific knowledge graph (KG) aggregating data from multiple agencies of DoD without exchanging or sharing the data at each agency. To facilitate the interpretation of federated learning-based KG (Fed-KG), large language models (LLMs) under the federated learning paradigm are used and fine-tuned to understand the collaborative connections among all agencies within DoD.

Our team finished all the tasks for developing the DoD Fed-KG and further investigated the feasibility of fine-tuning LLMs under federated learning paradigm considering the data privacy constraint among different agencies within DoD. The data analysis, information retrieval, and visualization system design framework highlight the feasibility and effectiveness of federated learning approach on constructing enterprise level KG without sharing data among DoD agencies. The federated fine-tuning to enhance the domain specific understanding of LLMs while maintaining data privacy showcases the feasibility of building a domain specific privacy preserving LLM for DoD. The federated approach not only leverages diverse instructional data from multiple sources but also achieves competitive performance compared to established models like ChatGPT 4 and 3.5. The results underscore the potential of federated learning in overcoming the limitations of centralized data collection and paving the way for more secure and efficient training methodologies in the realm of natural language processing (NLP).

BACKGROUND

Significance: The Department of Defense (DoD) currently employs the Correspondence and Task Management System (CATMS) that allows users to create, delegate, assign, respond, and search the knowledge repository for tasks and correspondence (CATMS 2023). The CATMS primarily handles intra-organization task management, without an integration of the recent development of knowledge graph (KG) theory and privacy preserving mechanism. In this complex and volatile era, the acquisition, deployment, and long-term support of the DoD capabilities typically occur within the purview of the five military services and multiple agencies (Nguyen 2009). The current approach often prioritizes intra-service missions and requirements, potentially overshadowing the critical aspects of inter-service integration and interoperability (Starr 2005). The process of establishing, approving, and managing these tasks within the DoD is not only highly time-consuming, but it also tends to be fragmented into service-specific requirements. In addition, the exchange of information between different agencies can be challenging due to data sensitivity, format compatibility issues, and varied protocols. Designing a privacy-preserving federated learning-based knowledge graph (Fed-KG) system, leveraging recent advances from academia and industry that enables users to perform semantic searches at a global level, can facilitate enterprise-level decision-making and foster inter-organizational cooperation.

Objective: Our objective is to develop a comprehensive decision-making framework that harnesses collective intelligence to effectively determine the optimal integration and collaboration mechanism for responsible agencies within the DoD when tackling complex tasks that entail cross-domain and cross-organizational operations. The framework will be based on privacy preserving and cross organizational federated learning to achieve the strategic goal of improved enterprise-wise interoperability. We aim to establish DoD domain specific knowledge graph (KG) aggregating data from multiple DoD agencies without exchanging or sharing the data at each agency. In addition to the Fed-KG, to facilitate the interpretation of Fed-KG, large language models (LLMs) under the federated learning paradigm are used and fine-tuned to understand the collaborative connections among all agencies within DoD. The federated learning framework for LLM can extract information that can be used by Fed-KG and overcome the limitations of KG of incompleteness, lacking language understanding, and its inability to learn unseen facts.

Pertinent to DoD's Research Areas: The developed DoD specific Fed-KG with information embeddings extracted from the publicly available news release, websites, white papers, and other documents. The Fed-KG can predict the joint ownership of new tasks and provide collaboration suggestions between individual agencies and departments, so that enterprises-wise interoperation plans and subtasks ownership plans with underlying budgeting formation can be made. The Fed-KG system can provide suggestions and training to cross-functional teams that address the critical objectives and outputs outlined in the organizational strategy. This research project addresses the following research areas: 1) Acquiring and Integrating Interoperable Capabilities Across Organizational Boundaries and 2) Requirements Setting and Managing Requirements that Cut Across Organizations.

2. METHODOLOGY AND RESULT

2.1 FUNDAMENTAL CONCEPTS

2.1.1 KNOWLEDGE GRAPH

KGs are critical data structures to represent human knowledge and serve as resources for various real-world applications, which can illustrate relationships between different entities in a network. Key components of KGs include (1) entities (nodes), representing objects or concepts that the graph represents, (2) relationships (edges), representing how the nodes are related, and (3) attributes, where entities and relationships can have with various properties. The KG can be used for multiple applications and tasks, including search engines (e.g. Google), recommendation systems (e.g. Amazon recommendation), information query, etc.

2.1.2 DOD DOMAIN-SPECIFIC KG

Compared to encyclopedic KGs, domain specific KGs can be established from domains where the data is not accessible by the public and can characterize entities and relationships pertinent to an organization or a particular subject. The domain specific KG enables more precise and accurate information retrieval by focusing on the specific terminology, entities, and relationships within a particular domain. The DoD KG will enable the knowledge representation extracted from textual information and provide many down streaming tasks, such as information retrieval, semantic search, etc. In summary, the DoD domain-specific KGs provide a structured and detailed representation of information within the DoD and all related agencies, enhancing relevance, accuracy, and utility in various applications.

2.1.3 FEDERATED LEARNING (FL)

Federated learning is a machine learning (ML) technique where multiple decentralized devices or servers collaborate to train a shared model while keeping the data localized. This approach enhances privacy, reduces latency, and minimizes bandwidth usage. FL allows different clients to collaboratively learn a global model without sharing their local data (McMahan et al., 2017). In particular, the aim is to minimize: $\min_w f(w) = E_k[F_k(w)]$, where $F_k(w)$ is the local objective that measures the local empirical risk of k-th client.

2.1.4 LARGE LANGUAGE MODELS

LLMs are advanced artificial intelligence (AI) models designed to understand and generate human language. LLMs, including models and frameworks such as Llama and ChatGPT, leverage deep learning architectures and are trained on extensive datasets to achieve a high level of linguistic proficiency. LLMs have been used on various applications such as natural language understanding, text generation, question answering, code generation, etc.

2.2 FEDERATED LEARNING FOR DOD KNOWLEDGE GRAPH

The overall framework for a Fed-KG is illustrated below in Fig. 1. First, the agency level KG will be created and then aggregated using federated learning to construct the DoD level KG. To preserve privacy, the embeddings (low-dimension features) from agency level KG is passed to the DoD level KG.

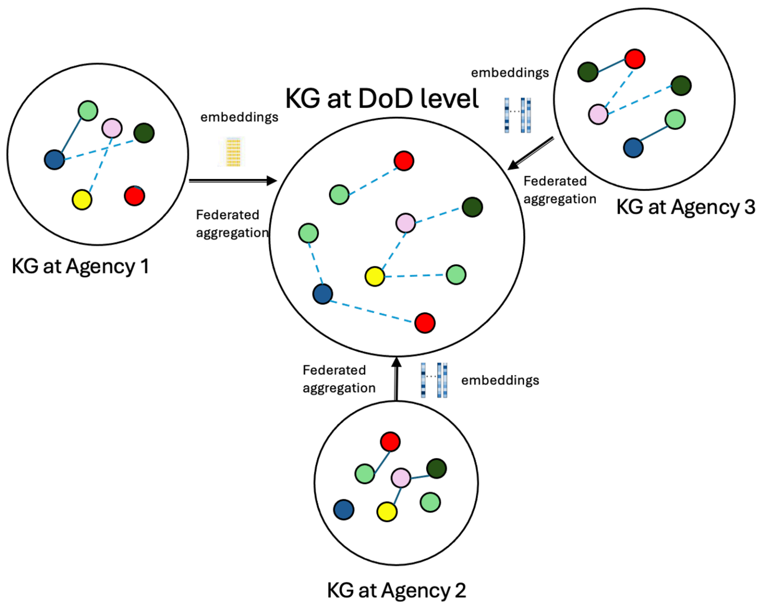


Figure 1: Fed-KG: framework of federated learning of DoD Knowledge Graph

2.2.1 DATA COLLECTION THROUGH WEB CRAWLING

We successfully curated a comprehensive cohort of dataset, utilizing Facebook posts, Wikipedia content, and X posts (tweets) from key DoD entities. According to the accessibility of official accounts of Facebook, the Selenium (the Selenium package is used to automate web browser interaction from Python) package was used for fetching Facebook posts for the Department of Homeland Security, National Aeronautics and Space Administration (NASA), and the Department of Defense. The number of posts for each department is around 30,000, with the year range of posts starting from 2017 and ending in 2024. Wikipedia content of United States Department of Homeland Security, United States Department of Defense, the Pentagon, United States Marine Corps, United States Space Force, and NASA were fetched by using the Python library Wikipedia. For Twitter/X platform, the developer API was used for fetching tweets of DoD (3222 tweets), DoD_DHA (1996 tweets), Homeland Security (3223 tweets), NASA (3216 tweets), SpaceForceDoD (3246 tweets), USAirForce (3247 tweets), USArmy (3241 tweets); the year range of tweets are from ~2021 to 2024. The detailed summary of postings is given in Table 1.

Table 1: Summary of Data Collection Across Platforms

Source of Data	Social media account	Number of postings	Time Period
Facebook	Department of Homeland Security	around 30,000 for each	2017-2024
	NASA		
	Department of Defense		
Wikipedia	United States Department of Homeland Security	N/A	
	United States Department of Defense		
	The Pentagon		
	United States Marine Corps		
	United States Space Force		
	NASA		
Twitter/X	DoD	3,222	2021-2024
	DoD_DHA	1,996	
	Homeland Security	3,223	
	NASA	3,216	
	SpaceForceDoD	3,246	
	USAirForce	3,247	
	USArmy	3,241	

2.2.2 LLM-POWERED KEYWORK EXTRACTION FOR KG

The KG construction mainly focused on data collected from Twitter/X and Wikipedia. For Twitter/X, our approach leveraged a LLM named *KeyBert* to extract entities embedded within social media text. We extract five key phrases from each tweet and the similarity of this extracted phrase with the original tweet is provided by this LLM. Table 2 shows some of these keywords and the related information including a unique ID, keyword extracted (with the similarity score between the keyword and the original tweets), the original tweet, group number and the group name (the department). The keywords were used for searching in the KG, which can help us quickly find the information and tweets we need. The LLM's capability to discern entities and their contexts significantly enhances our KG's depth, particularly in capturing contemporary and dynamic information.

The relationships between these tweets are represented by their co-occurrence within the data, with the sentiment of these relationships—ranging from positive to negative—quantified and used to weight the links between nodes. The relationships between them are extracted through the *sentence-bert* model. By grouping these tweets, the sentence relationships were analyzed using LLM and the relationality between two sentences was obtained (range from 0 to 1). A higher value means a stronger relationality between two sentences. The "similarity" represents the strength or importance of these connections (in an actual dataset, this could mirror the frequency of collaboration, command hierarchy, etc.). The relationship between two tweets is represented by an exclusive id. For example, [Source:"2b53573e-86f3-4d8d-8a18-7836fe308e26",Target:"8419888f-d53b-4b15-8d1f-13f152293323", Similarity: 0.58].

2.2.3 FEDERATED AGGREGATION FOR DOD KG

The federated aggregation is implemented using an existing method [Zhang et al. 2022]. The DoD KG maintains a complete table including entity embeddings and the corresponding entity IDs, and the DoD KG can identify if an entity exists in a client for entity alignment. This federated learning paradigm can address the data privacy issues among different agencies of DoD while learning the cross-organizational integration of data. In the subsequent visualization and retrieval, we can trace which data it is by the unique ID. At this stage, we have obtained the key information for constructing the knowledge graph, including "Nodes" and "Links". "Nodes" correspond to each tweet and also contain some information internally, such as IDs, keywords, and groups. These nodes are grouped according to their type or principal affiliation, and correspond to various military units, coalitions, operations, or notable entities such as universities or countries, like NASA, State Department, Homeland Security, USArmy, and so on. "Links" illustrate the interactions, collaborations, or conflicts among these entities, while the "Similarity" denotes the significance or intensity of these relationships.

Table 2: Samples of keywords and the related information extracted

ID	Original Tweet	Keywords	Group	Group Name
fcbf54d6-3c11-44e9-99b2-c5a4e4cc9aed	NEWS: U.S., Australian, British Land Chiefs Address Security Agreement's Progress https://t.co/QgRgR9nnEI	[('security agreement', 0.5118), ('british land', 0.4672), ('news australian', 0.4267), ('chiefs address', 0.3045), ('qgrgr9nnei', 0.1029)]	0	DoD
6c260e0e-7556-48f6-9c33-090d77304f76	NEWS: DoD Able to Send Additional Assistance to Ukraine Using Unexpected Army Savings https://t.co/UMqcICZ5c1	[('assistance ukraine', 0.7056), ('dod able', 0.413), ('send additional', 0.2675), ('savings https', 0.219), ('using unexpected', 0.1413)]	0	DoD
8af0b531-f383-4b1a-9c9f-e826c5e72453	NEWS: Leaders Discuss Security Priorities for Western Hemisphere https://t.co/aAvSEHzjr1	[('security priorities', 0.5961), ('news leaders', 0.4682), ('western hemisphere', 0.4475), ('discuss', 0.1283), ('aavsehzjr1', 0.0938)]	0	DoD
51f3a36d-c901-43a4-a5bc-9842546437a7	RT @DoD_Outreach: As part of the @ DeptofDefense Innovative Readiness Training program, the @MINationalGuard is joining forces with the @Mic...	[('rt dod_outreach', 0.6533), ('minationalguard', 0.5653), ('forces mic', 0.399), ('training', 0.3964), ('innovative readiness', 0.3709)]	0	DoD
...				
518985b3-04d1-479e-ac98-1f39bea16473	RT @USNATO: "History will see the accession of the Czech Republic, Hungary, and Poland as a key step towards a Europe of co-operation and i...	[('accession czech', 0.6721), ('europe operation', 0.5732), ('usnato history', 0.4023), ('rt', 0.2429), ('step', 0.1372)]	7	StateDept
361b68ed-b93c-47bd-99ba-c7a2c6d9c923	On the ongoing security crisis in Haiti: We welcome yesterday's announcement of a transitional governance structure in Haiti which paves the way for a peaceful transition of power, continuity of governance, and action plan for near term security. https://t.co/mAr3t6UTb5	[('crisis haiti', 0.6198), ('transitional governance', 0.5365), ('ongoing security', 0.3872), ('welcome yesterday', 0.1955), ('paves way', 0.146)]	7	StateDept
df6c92f7-669a-4753-a20f-a4ddfe51adb0	RT @StateDeptSpox: Watch the Daily Press Briefing from the State Department. https://t.co/8UatnbcxS1	[('press briefing', 0.4754), ('rt statedeptspox', 0.4148), ('state', 0.3344), ('https 8uatnbcxs1', 0.2217), ('watch daily', 0.2049)]	7	StateDept
2b53573e-86f3-4d8d-8a18-7836fe308e26	RT @SecBlinken: Haitians cannot wait any longer for a path to security, stability, and democracy. The United States and @CARICOMorg support..	[('haitians wait', 0.7074), ('stability democracy', 0.37), ('states caricomorg', 0.3483), ('secblinken', 0.2078), ('path security', 0.0872)]	7	StateDept

2.2.4 VISUALIZATION SYSTEM AND SEMANTIC QUERY OF DOD KG

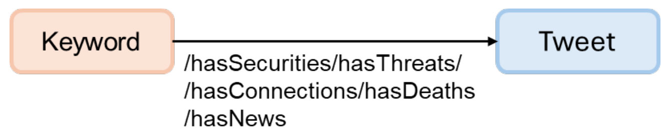
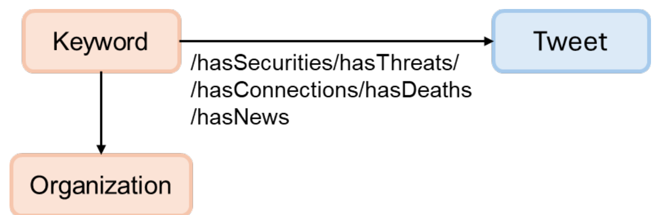
Semantic queries over knowledge graphs involve extracting a subgraph based on specific search criteria. This section discusses how knowledge graphs enhance information retrieval and analysis through these queries. However, constructing different semantic queries in structured query languages like SPARQL can be complex and time-consuming due to the knowledge graph's intricacy and diversity. To address this issue, we distill the core aspects of semantic queries and develop an algorithm that efficiently and scalably handles various semantic queries.

Some DoD knowledge graphs operate at the metadata level, supporting tasks like topic modeling or affiliation tracking that only need meta-information. However, more advanced tasks typically require human intervention to process text, such as reading, understanding, and summarizing. Therefore, we invest significant effort in constructing a KG by extracting, summarizing, and integrating information from both *textual content* and *metadata*. By combining these sources of information into a single knowledge graph, we can automate many real-world tasks that previously required human effort. Later, we discuss advanced semantic queries that leverage both semantic and metadata entities.

The DoD Fed-KG supports a wide range of queries by establishing rich connections among different entities through direct or multi-hop relations. We categorize these queries into two tasks: retrieval and summarization. For retrieval, connections from semantic entities to other metadata entities enhance information retrieval efficiency. For example, to identify potential threats in different agencies discussing "nuclear attacks," we can start from "nuclear attacks" and retrieve associated keywords and tweets by traversing the DoD KG. This approach provides factual data to support decision-making rather than relying on personal experience or domain knowledge. Conversely, connections from metadata entities to semantic entities are highly useful for information summarization. When dealing with multiple agencies, automatic summarization from various perspectives is valuable and can be achieved by utilizing relationships between agencies and keyword entities.

Semantic queries share the common characteristic of identifying a set of starting entities and specifying search criteria to navigate the knowledge graph and reach target entities. In Table 3, we summarize multiple queries based on the type of target entity. The ontology column displays the relevant entity and relation types utilized to support the query. The supported task column contains two example tasks related to retrieval and summarization, and we provide illustrative examples.

Table 3: Summarizing the utilization of our DKG in different real-world scenarios that involve semantic queries

Entity	Ontology	Supported Task
Keyword		• Tweet Retrieval • Web Crawling
Organization		• Organization Finder • Author Profiling
Similarity Score		• Reposted tweet across different organizations • Understand organization relationships

2.2.5 SYSTEM OVERVIEW OF DOD FED-KG EXPLORER

Based on the tasks, we have designed and developed a visual analytics and dataflow system called "DoD Fed-KG Explorer," as shown in Fig 2. The system includes a set of components with predefined inputs and outputs that users can select and connect to tweet metadata and information **(T3)**. These components fall into two categories: semantic queries **(T1)** and viewers for data visualization **(T2, T4)**. The operators generate and manipulate data, which can be passed to downstream components in the system, while the visualizers present data meaningfully and provide insights without transmitting the data to subsequent components.

We designed the DoD Fed-KG Explorer to facilitate visualization and semantic query. The DoD Fed-KG Explorer illustrated in Fig 3 consists of 5 visualizations components (1, 2) semantic queries include node and link query; (3) knowledge graph summary of entity in different organization; (4) relationship diagram representing more hierarchical view of relationship and its similarity scores; and (5) table summary shows a tweet information and meta-data.

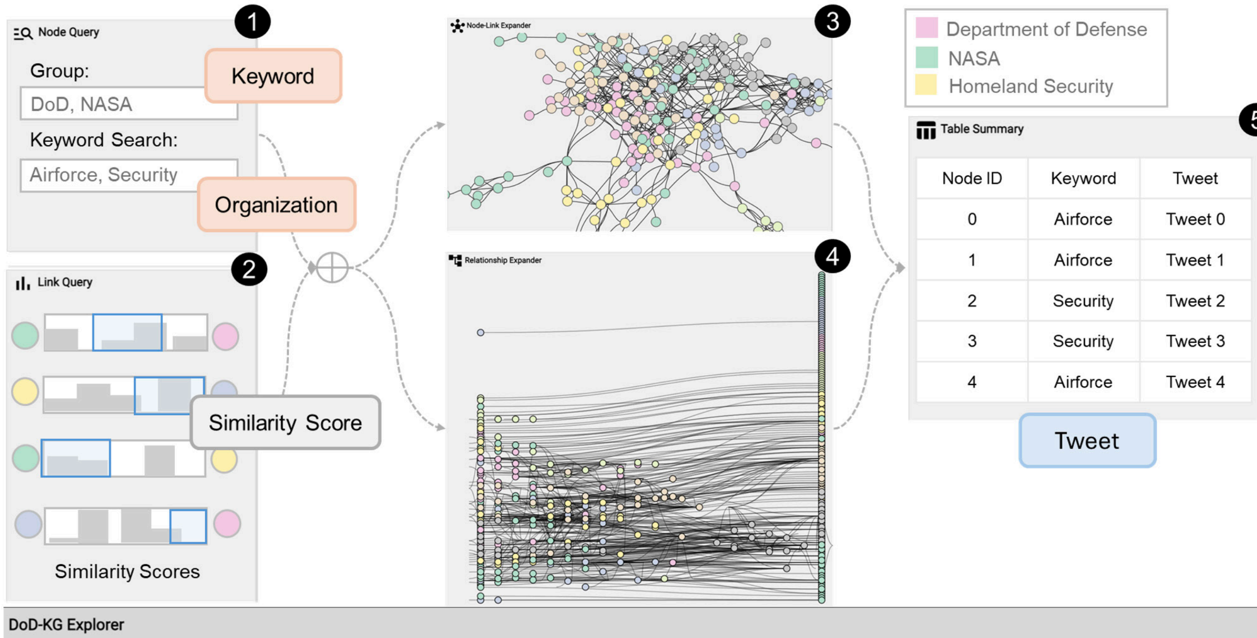


Figure 2: The DoD Fed-KG Explorer with the following components: (1, 2) semantic queries include node and link query; (3) knowledge graph summary of entity in different organization; (4) relationship diagram representing more hierarchical organization; and (5) table summary

2.2.6 SEMANTIC QUERY

To facilitate effective exploration of a Dynamic Knowledge Graph (DKG) **(T1)**, we have divided it into two stages: (1) Node Query and (2) Link Query.

Node Query: This stage allows users to specify “Keywords” as node entities and “Group” as organizational representations, which can encompass multiple entities. Different color codes represent different organizations and communities, making it easier for users to identify and distinguish between various groups and their associated entities. This visual differentiation aids in the intuitive understanding of complex relationships within the data.

Link Query: In this stage, instead of specifying the types of relationships, users are provided with a bar chart histogram that includes sliding windows. This interactive feature allows users to filter entities based on a range of similarity scores between groups of entities. The sliding windows enable users to adjust the similarity score range dynamically, filtering out entities that do not meet the specified criteria. This visual and interactive method simplifies the process of exploring relationships by providing an intuitive and immediate way to see how entities are connected based on their similarity scores.

Users can apply filters to refine their queries, focusing on specific attributes or conditions. This feature allows for more precise and targeted exploration, helping users to zero in on the most relevant data. The system provides an interactive query builder that guides users through the process of constructing complex queries without requiring deep technical knowledge. As users construct and refine their semantic queries, the system provides integrated visual feedback, displaying query results in real-time through various visualization tools. This immediate visual representation helps users to quickly interpret the data, validate their queries, and make informed decisions about subsequent steps.

2.2.7 PLATFORM DEVELOPMENT FOR DATA VISUALIZATION

After users have constructed and refined their queries, DoD Fed-KG Explorer offers robust visualization tools to present the results comprehensively **(T2, T4)**. The system uses a KG presented as a node-link diagram, allowing users to visually explore the connections and relationships between different entities. This diagram helps users grasp the overall structure and key linkages within the data, providing a clear and immediate overview of the queried information. For more fine-grained details, the system employs a Sankey Diagram. This visualization technique is particularly effective for illustrating the flow and relationships between entities in a more detailed and granular manner. The Sankey Diagram enables users to see the volume of connections and the direction of relationships, offering deeper insights into the dynamics within the data. Together, these visual tools enhance the user's ability to analyze and understand the complex interrelations within the DKG, making the data more accessible and actionable.

In addition to visualizations, the DoD Fed-KG Explorer provides a summary of results in a tabular format. This table presents a concise overview of the queried data, allowing users to quickly review and interpret the findings. Users can also view the original tweets associated with the semantic queries directly from the system, providing context and additional detail for the analyzed data. Moreover, the system supports interactions from the visualization, enabling users to drill down into specific data points and explore the underlying information. For further analysis or reporting, users can export the results in .csv or .xlsx file formats. This export functionality ensures that users can easily share and utilize the data in other applications or workflows, facilitating broader collaboration and dissemination of insights.

2.2.8 FUTURE IMPROVEMENT FOR DOD FED-KG CONSTRUCTION

To achieve semantic queries interactively and efficiently, we have integrated a visual analytics system into our approach. Based on our finding, we have plans to implement the following tasks:

T1: Constructing Interactive Queries over DoD Fed-KG. Querying DoD Fed-KG with structured queries, such as SPARQL, can be challenging for users less experienced in programming. To ensure a smooth and efficient knowledge discovery process, our system will feature an intuitive interface that allows users to construct semantic queries efficiently and interactively.

T2: Visualizing Heterogeneous Data from KG. Due to the large volume and high diversity of DoD Fed-KG, data retrieved by semantic queries can be difficult to understand immediately. While visualization is a powerful tool for presenting patterns and generating insights, the system should offer clear and intuitive data visualizations for the queried data.

T3: Building Flexible Pipelines for DoD Fed-KG. Users with different backgrounds may have various goals for exploring DoD Fed-KG. The exploration process can involve multiple data queries, processing, and analysis to generate insights. Our system should allow users to customize analysis pipelines to meet their specific needs.

T4: Accessing Raw Data in DoD Fed-KG. While summarizing queried information is efficient for knowledge extraction, providing access to raw data is necessary to build user trust in the system and the data. Additionally, our system should link raw data to the web and offer an easy way for users to access it.

2.3 FEDERATED LEARNING OF DOD DOMAIN SPECIFIC LARGE LANGUAGE MODELS

Recently, LLMs have demonstrated their capabilities across a wide range of natural language processing (NLP) tasks. ChatGPT, a representative example, is trained on a diverse database of text from various domains and can emulate human experts with cross-disciplinary knowledge. ChatGPT exhibits an advanced proficiency in processing and interpreting extensive textual data, making it a valuable tool for information retrieval, decision support, and report generation. However, previous studies have raised concerns about the responses from LLM trained from the public data, especially its poor performance in highly specialized domains due to data privacy. To address this, fine-tuning LLM on locally collected and well-annotated data is a common approach that requires relatively low computational resources. Examples include ChatDoctor, which is a fine-tuned LLM based on Llama and trained on a dataset of 100,000 patient-doctor dialogues, providing informative, accurate, and professional advice. Yang et al. developed FinBERT-20, which is a finance domain specific BERT model, pre-trained on a financial communication corpus (4.9B tokens). FinBERT-20 also conducted a sentiment analysis task for fine-tuning experiments. The growing applications of LLM enriched by domain specific data records in a variety of applications inspired this study to develop a fine-tuned LLM for DoD and help facilitate the information retrieval and context understanding.

2.3.1 DATA PREPROCESSING AND EXPERIMENTAL SETUP

The data we extracted in Section 2.2.1 were organized into 5000 Question-Answer (Q-A) pairs (Fig. 3), for each DoD agency with 500 Q-A pairs. The training and testing ratio is 8:2. This corpus has been instrumental in training the Alpaca-Native model, enhanced by post-training quantization and advanced techniques like Grouped-Query Attention and Rotary Positional Embeddings to optimize performance. The federated learning steps for model training are shown in Fig. 4, while the pipeline of fine-tuning DoD domain specific LLMs using federated learning approach is given in Fig. 5. Components of Fig 5 include: client data representing the data from each agency of DoD, reflecting real-world non-independent and identically distributed (non-i.i.d) datasets; Simulated Local Training: Due to the high computational cost of simultaneous local training, the framework simulates this by sequentially training each client; and Client Participation Scheduling: Initially, a random selection approach is used for client participation. The fine-tuning process is done using Low Rank Adaptation (LoRA) approach [Hu et al. 2021].

```
    "response": "The US Marine Corps operates installations on land and aboard sea-going amphibious warfare ships around the world and has several tactical aviation squadrons embedded in Navy carrier air wi
    "category": "US Marine Corps"
},
{
    "instruction": "What historical action is commemorated by the Marine Band being dubbed the 'President's Own'?",
    "context": "",
    "response": "The Marine Band was dubbed the 'President's Own' by Thomas Jefferson and provides music for state functions at the White House.",
    "category": "US Marine Corps"
},
{
    "instruction": "What is the significance of the Battle of Princeton for the US Marine Corps?",
    "context": "",
    "response": "The Battle of Princeton was the first land combat engagement of the Marines, where they were personally rallied by George Washington.",
    "category": "US Marine Corps"
},
{
    "instruction": "What is the Marine Corps Martial Arts Program (MCMAP)?",
    "context": "",
    "response": "The Marine Corps Martial Arts Program (MCMAP) is a mix of different styles of martial arts that instills the 'Warrior Ethos' within marines, beginning training in boot camp.",
    "category": "US Marine Corps"
}
```

Figure 3: Illustration of Q-A pairs organized

```
global_model: chavinlo/alpaca-native
data_path: ./data
output_dir: ./lora-shepherd-7b/
client_selection_strategy: random
client_selection_frac: 0.1
num_communication_rounds: 10
num_clients: 10
local_batch_size: 64
local_micro_batch_size: 8
local_num_epochs: 10
local_learning_rate: 0.0003
local_val_set_size: 0
local_save_steps: 3
cutoff_len: 512
lora_r: 16
lora_alpha: 16
lora_dropout: 0.05
lora_target_modules: ['q_proj']
train_on_inputs: True
group_by_length: True
resume_from_checkpoint: False
prompt template: alpaca
```

Figure 4: Training setting for Federated LLM

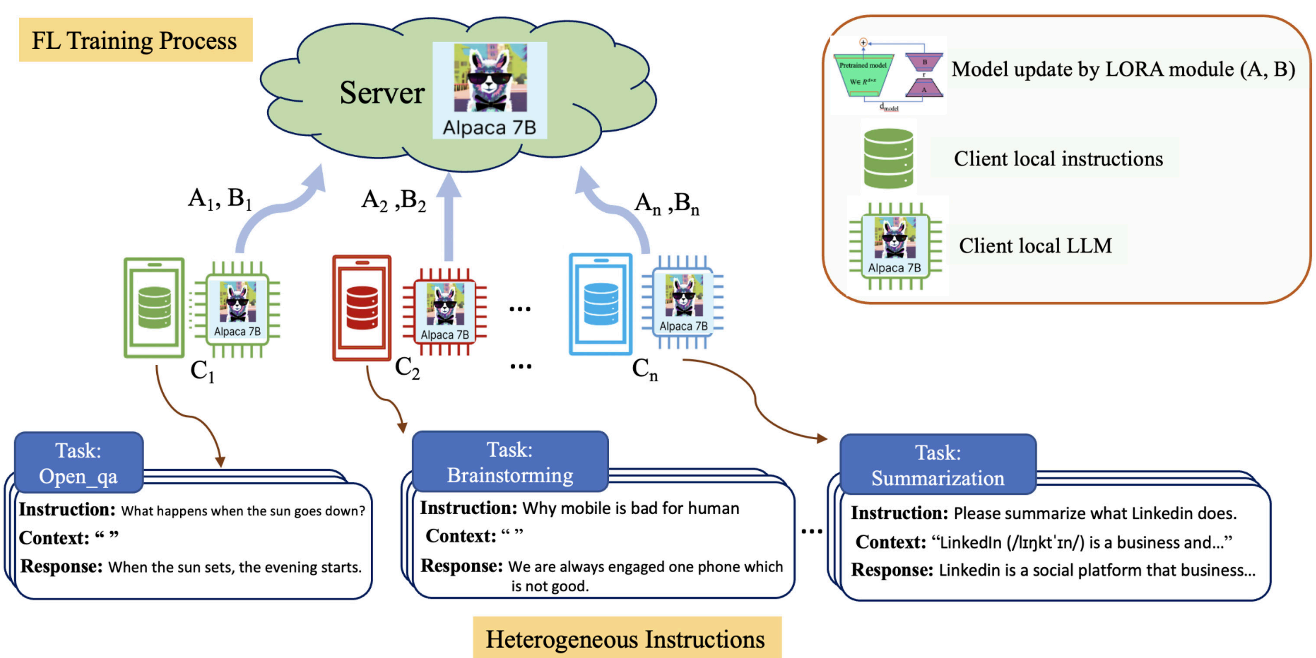


Figure 5: Training steps for Federated LLM-Client Data Allocation: The client data is partitioned to reflect real-world non-independent and identically distributed (non-iid) datasets. This can be done by varying class distribution and data volume across clients. Simulated Local Training: Due to the high computational cost of simultaneous local training, the framework simulates this by sequentially training each client. Client Participation Scheduling: Initially, a random selection approach is used for client participation.

2.3.2 FEDERATED LEARNING AND INSTRUCTION TUNING

In our study, we investigated the application of FL for the instruction tuning of LLMs. This approach addresses the challenges associated with acquiring high-quality instructional data, such as cost, accessibility, and privacy concerns. By leveraging FL, we enable the collaborative training of LLMs without the need for centralized data collection, thus preserving data privacy and security. The overall design of fine-tuning of the DoD-LLM is illustrated in Fig 5.

The dataset used for training our model comprised 5000 Q-A pairs, with each of the ten participating institutions contributing 500 pairs. After shuffling, the data was split into a training set and a testing set in an 8:2 ratio. This corpus was pivotal in training the Alpaca-Native model, which was further optimized using post-training quantization, Grouped-Query Attention, and Rotary Positional Embeddings.

The training process was carried out using an NVIDIA 4090 setup, and the entire procedure took over one hour, highlighting the computational intensity of the task. We compared our federated model against ChatGPT 4, ChatGPT 3.5, and a non-federated version of our model to evaluate performance differences.

2.3.3 EVALUATION AND RESULT

The answer relevancy metric measures the quality of the LLM's output by evaluating how relevant the output of the LLMs is when compared to the provided input. To assess the quality and relevance of the generated answers, we employed DeepEval's answer relevance metric. This metric evaluates how relevant the output of the LLM is compared to the provided input, offering self-explaining feedback for its scores.

Table 4: Performance summary of fine tuned LLM vs the benchmark frameworks

Model	Relevancy Score	Explanation
Federated Model-Alpaca-Native Model	0.88	Responses are relevant but occasionally miss nuanced context.
ChatGPT 4	0.95	Demonstrates excellent performance with very high relevance in answers.
ChatGPT 3.5	0.90	Generates relevant answers but slightly less accurate than ChatGPT 4.
Non-Federated Model	0.85	Produces generally relevant answers but lacks the diversity and robustness of the federated model.

The DoD specific accuracy score is subjective, and our team will engage DoD stakeholders to ensure a good metrics to be established for quantifying the responses of all LLMs. Given our model was fine-tuned using DoD data, conversably, it can provide domain specific responses than other LLMs trained on public data.

CONCLUSIONS

Our team finished all the tasks listed in the Phase 1 proposal and further investigated the feasibility of fine-tuning LLMs under federated learning paradigm considering the data privacy constraint among different agencies within DoD. The data analysis, information retrieval, and visualization system design framework highlight the feasibility and effectiveness of federated learning approach on constructing enterprise level KG without sharing data among DoD agencies. The federated fine-tuning to enhance the domain specific understanding of LLMs while maintaining data privacy showcases the feasibility of building a domain specific privacy preserving LLM for DoD. The federated approach not only leverages diverse instructional data from multiple sources but also achieves competitive performance compared to established models like ChatGPT 4 and 3.5. The results underscore the potential of FL in overcoming the limitations of centralized data collection and paving the way for more secure and efficient training methodologies in the realm of NLP.

REFERENCE

1. Ali, Mehdi, et al. "PyKEEN 1.0: a python library for training and evaluating knowledge graph embeddings." The Journal of Machine Learning Research 22.1 (2021): 3723-3728.

2. Banjade, Rajendra. "DeepEval: An Integrated Framework for the Evaluation of Student Responses in Dialogue Based Intelligent Tutoring Systems." (2014).

3. CATMS: https://www.esd.whs.mil/CMD/CATMS/, last accessed, June 2023.

4. Dusmanu, Mihai, Johannes L. Schonberger, Sudipta N. Sinha, and Marc Pollefeys. "Privacy-preserving image features via adversarial affine subspace embeddings." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14267-14277. 2021.

5. Dwork, Cynthia. "Differential privacy: A survey of results." In Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings 5, pp. 1-19. Springer Berlin Heidelberg, 2008.

6. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

7. Ji, Shaoxiong, Shirui Pan, Erik Cambria, Pekka Marttinen, and S. Yu Philip. "A survey on knowledge graphs: Representation, acquisition, and applications." IEEE transactions on neural networks and learning systems 33, no. 2 (2021): 494-514.

8. Jiang, Meng, Taeho Jung, Ryan Karl, and Tong Zhao. "Federated dynamic graph neural networks with secure aggregation for video-based distributed surveillance." ACM Transactions on Intelligent Systems and Technology (TIST) 13, no. 4 (2022): 1-23.

9. McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial intelligence and statistics*. PMLR, 2017

10. Nguyen, Quy H. Achieving unity of Effort: leveraging interagency cooperation between the Department of Defense (DOD) and the United States Agency for International Development (USAID). Army Command and General Staff Coll Fort Leavenworth KS, 2009.

11. Singhal, Amit. "Introducing the knowledge graph: things, not strings." Official google blog 5, no. 16 (2012): 3.

12. Starr, Stuart H. The Challenges Associated with Achieving Interoperability in Support of Net-Centric Operations. BARCROFT RESEARCH INST FALLS CHURCH VA, 2005.

13. Wang, Qingyun, Xuan Wang, Manling Li, Heng Ji, and Jiawei Han. "Knowledge extraction to assist scientific discovery from corona virus literature." (2020).

14. Yang, Y., Uy, M. C. S., & Huang, A. (2020). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*

15. Zhang, K., Wang, Y., Wang, H., Huang, L., Yang, C., Chen, X., & Sun, L. (2022). Efficient federated learning on knowledge graphs via privacy-preserving relation embedding aggregation. *arXiv preprint arXiv:2203.0955*