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

EXECUTIVE SUMMARY
JULY 2024

PRINCIPAL INVESTIGATOR
Feng Liu, *Stevens Institute of Technology*

CO-PRINCIPAL INVESTIGATOR
Xinyue Ye, *Texas A&M University*

STEVENS
INSTITUTE OF TECHNOLOGY

TEXAS A&M
UNIVERSITY.

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).

DISCLAIMER