"AI Foreign Policy Assistant" Powered by GraphRAG Techniques

Dev Mehra

Abstract

Retrieval-augmented-generation (RAG) allows for optimization of the output of a large-language model (LLM) by allowing it to use information outside of its training database as context for its responses. This allows the capabilities of the LLM to be applied to new datasets and tasks. However, the standard RAG method is limited in its ability to answer high-level questions about a dataset. In this paper, we discuss how LLMs can be supplemented with knowledge graphs, instead of traditional RAG, to assist in foreign policy decision-making. We developed a GraphRAG system using Neo4j and integrated it with an LLM to create an AI foreign policy assistant. Foreign policy is an area which requires a high-level understanding of multiple perspectives, global implications, and rich relationships between nations, which makes knowledge graphs better suited for it than traditional RAG.

Introduction

Foreign policy involves a state creating strategies to interact with other states. It plays a key role in international relations and global events by dictating how countries cooperate, compete, and resolve conflicts. The outcome of foreign policy decisions affect global peace, alliances, and trade, influencing the stability of the international order and domestic affairs within countries.

In this project, we harness Large Language Models (LLMs), supplemented by a knowledge graph "GraphRAG" system [1], to create an AI "foreign policy assistant." An LLM is an AI model trained on vast datasets to generate and understand human language [2]. It is provided context and relevant information to the task by a knowledge graph, a data structure that stores information as nodes— people, places, locations, events, countries, documents, etc.— and edges, relationships that connect the nodes. This approach leverages the language processing capabilities of an LLM, supported by structured data, unstructured data, and community summaries from the knowledge graph, to generate informed foreign policy drafts.

Researching this approach is significant because of the complexity of foreign policy analysis which requires consideration of vast amounts of interconnected information in the form of documents, books, speeches, news articles, etc. This information can be anything from the analysis of the history of tensions between two countries or the history of their economic and trade relations. Traditional

Retrieval-Augmented Generation (RAG) [3] methods are effective for situations in which an answer to the query can be found within regions of text that are retrieved [1].

We created an AI assistant that is able to ingest foreign policy data and store it in a large graph database using Neo4j. The AI assistant queries the graph, tracing connections relevant to the prompt. We evaluated its effectiveness as compared to traditional RAG methods.

Background

I. Politics

Today, there are 92 countries involved in international conflicts— the greatest number of countries since WWII [4]. According to the Global Peace Index 2024, there is a high likelihood of major conflicts occurring [5]. In 1947, US Senator Arthur Vandenberg asserted that the United States must stop "partisan politics at the water's edge," essentially saying that foreign policy must be made in a bipartisan manner. Today, however, political polarization is at an all-time high in several countries, such as the US, and can impact foreign policy decision-making [6]. Research by US foreign policy analysts has shown that it is this political polarization has made it more difficult for Congress to implement foreign policy [7].

Over the past few decades, the relationship between the US and China has become increasingly globally competitive and adversarial due to ideological differences, economic rivalry, and geopolitical strategies, which is evident from the rising polarization and extremism within both countries' foreign policy decisions. Recent studies have shown an increase in polarization between the two nations. This has led to decisions driven more by ideological agendas rather than data-driven policy evaluations.

Our AI assistant ingested data specifically pertaining to US-China relations, with the goal of creating an AI foreign policy assistant for the US government to inform its foreign policy actions with China.

An advanced foreign policy decision-making system can help countries navigate around this discourse to avoid violent conflict.

II. The Current State of Foreign Policy Assistance

Traditionally, foreign policy is informed by expert advice, intelligence assessment, and diplomatic channels. In the US, the Department of Defense, the State Department, and intelligence agencies work together to advise policymakers on foreign relations. However, they can focus primarily on short-term goals and solutions, rather than looking at the long-term, both in the past and in the future. Moreover, decision-making in international cooperation organizations like the United Nations can be politically motivated, rather than relying on data-driven reasoning.

III. Addressing Political Polarization with AI

The fluctuating, decades-long US-China relationship presents a valuable case study for the application of AI in foreign policy. There are vast amounts of data available, from trade agreements to military conflicts, that make this an ideal case to evaluate the effectiveness of an AI foreign policy assistant.

Extremists in both the US and Chinese governments have advocated for more confrontational politics like trade embargoes that could further escalate tensions. An AI assistant could soften tensions by providing balanced, data-driven insights that take a wider, long-term context into account. The AI, which would not be swayed by political biases, could present alternative solutions to extreme policymaking and help show potential long-term consequences of impulsive, confrontational decisions.

IV.   How can LLMs assist in Foreign Policy?

LLMs, including models like OpenAI's GPT-4o, Meta's Llama 3.2, Anthropic's Claude 3.5 Sonnet, and Google's Gemini 1.5, show remarkable capabilities in processing and generating text, as well as analyzing and evaluating data.

We integrated LLMs with a knowledge graph to enhance the LLM's ability to generate context-aware foreign policy recommendations. The knowledge graph organizes foreign policy data into entities and relationships and allows the AI assistant to query the structured data. Moreover, we used a hybrid search approach, which means the AI assistant can retrieve both the structured data in the knowledge graph and unstructured data, similar to traditional RAG.

V.   Knowledge Graphs and GraphRAG

Knowledge graphs play an essential role in making this AI assistant effective by structuring the vast amount of data involved in foreign policy decision-making. We used a Neo4j graph database to store and manage the knowledge graph, primarily for its scalability and community creation features.

GraphRAG, improves on traditional RAG methods by traversing and extracting relationships between entities [1]. In a complex, interconnected field like foreign policy, this approach is essential to understanding the interplay between countries, events, and policies. By tracing these relationships, the AI assistant can generate responses accounting for broad patterns and long-term implications, which can be missed in traditional RAG systems that only retrieve pieces of text [1].

Data Collection

The success of our AI assistant is dependent on the quality and diversity of the data it has available. We built a comprehensive knowledge graph to feed into the LLM, enhancing its ability to generate informed, useful foreign policy recommendations. Data was collected from a variety of primary and secondary textual sources to ensure a wide coverage of relevant entities, relationships, and events. This section outlines how we gathered, cleaned, and ingested the data into the knowledge graph.

I.   Data Sources and Selection Criteria

To create a detailed, well-rounded dataset, we collected data from a variety of reliable textual sources found across the internet. These included official documents, treaties, diplomatic transcripts, Wikipedia pages, scholarly articles, books, speech transcripts, expert opinions, and news articles. We followed specific criteria to ensure good quality and relevant data.

- **Credibility:** Data was sourced from well-known and reliable platforms. These included official documents, government websites, scholarly books, and peer-reviewed articles.
- **Relevance to US-China Relations:** Every source was checked to ensure it was relevant to US-China relations with focuses on topics like military competition, diplomatic summits, trade wars, and political strategies.
- **Timeframe:** We primarily focused on data from the last 30 years, as that is where the most relevant geopolitical changes in US-China relations can be found. Some prior historical context was included.
- **Diversity of Perspectives:** Our data covered a range of perspectives from both US and Chinese sources, as well as insights from third-party experts. This helped the AI assistant consider a more holistic view of international relations.

II.   Data Cleaning, Ingestion, and Storage

Text was extracted from PDFs using LangChain and extracted from online sources using Jina.ai. These tools were essential in streamlining the process of data ingestion by allowing us to efficiently extract and clean up text from PDF and online sources. The data was split into chunks of 512 tokens with an overlap of 24 tokens.

- **Token:** a word or part of a word that the model processes.

Data cleaning was a critical step in maintaining a quality dataset:

- **Removal of Duplicates:** Duplicate nodes and connections were not added to the Neo4j knowledge graph.
- **Inconsistency:** With the large variety of sources being used, we ran the risk of inconsistencies in different accounts of events. We attempted to resolve this by also including unstructured data (raw text from the sources) as context for the LLM.

Scalability: While ingesting text-based data was straightforward with the tools available, we were unable to integrate other media types easily, such as video or audio, which could provide some additional helpful insights

III.   Knowledge Graph and Structured Data in Neo4j

Once the data was collected, it was transformed into a structured format for a knowledge graph using the Neo4j LLM Graph Builder [9], which transformed unstructured text into structured data by identifying key entities and determining relationships between them. Each chunk was analyzed to extract entities and relationships that are central to foreign policy, such as People, Organizations, Countries, Events, Laws, Institutions, and Agreements. These were stored as nodes and edges in a Neo4j graph database, serving as the foundation of our GraphRAG system.

With the current data, the knowledge graph has 31,729 nodes and 104,356 relationships. However, it will continue to grow as more documents are added to the database.
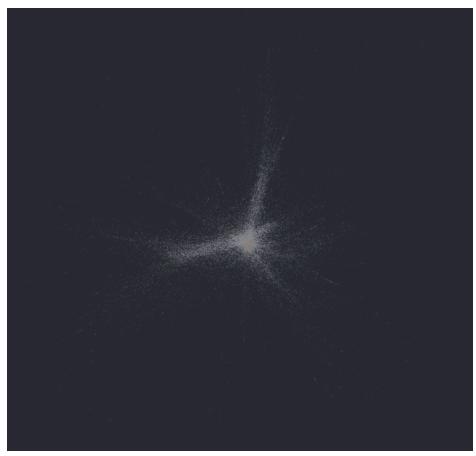
Figure 1: *A visualization of the full graph database, containing all nodes and relationships. Each of the dots represents a node— a person, country, organization, event, law, etc. The most densely connected nodes are located at the center and the connections get sparser further away.*
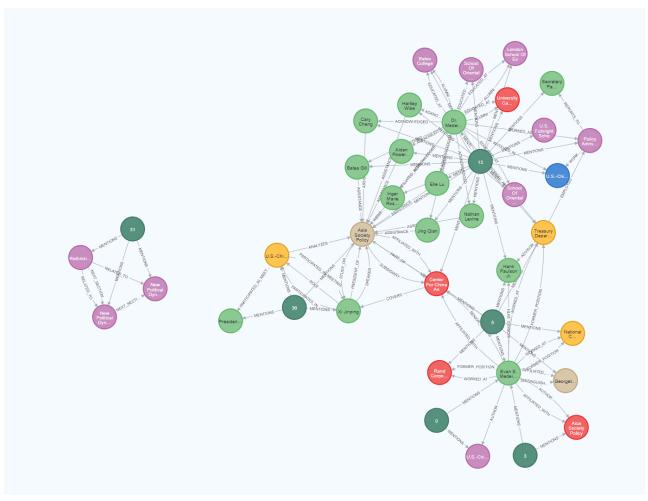
Figure 2: *A sample of 40 nodes from the data. The numbered nodes represent specific documents or pages from the ingested data.*
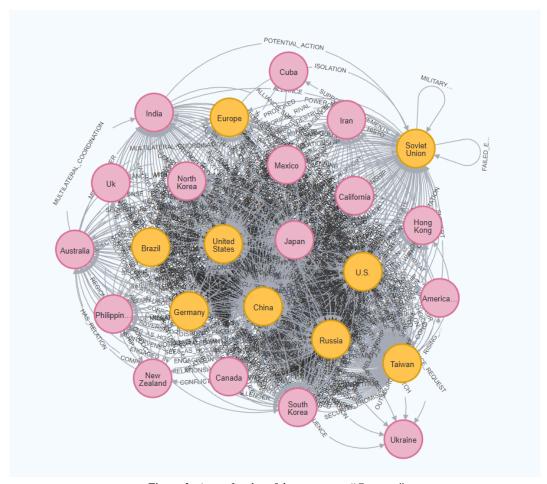
Figure 3: A set of nodes of the entity type "Country."

Methodology

I.   Overview

The objective of this research is to develop an AI-based foreign policy assistant using LLMs and knowledge graphs that can generate context-aware recommendations in complex geopolitical scenarios. We used US-China relations as a case study.

To achieve this, we integrated knowledge graphs, community detection algorithms, and traditional RAG techniques. A vast amount of structured data, unstructured data, and community summaries were stored in a graph database and retrieved when the LLM was queried. The LLM used this data to generate rich foreign policy insights based on the underlying connections in the data.

II.   Knowledge Graph Construction

The graph structure represents the core elements of foreign policy extracted from the text data provided in the forms of entities and relationships. Entities, which are stored as nodes, include people, countries, organizations, and events. Relationships, stored as edges, include conflicts, alliances, and diplomatic relations.

III. Clustering and Community Detection

The use of community detection algorithms to group closely-related entities in the knowledge graph based on their interconnectedness was a key feature of this project. We tested two different clustering algorithms for this purpose.

- **Louvain** detects communities by maximizing a modularity score. However, it can also lead to badly connected or completely disconnected communities [10].
- **Leiden** improves on the Louvain algorithm by ensuring well-connected communities and hierarchical clustering.

To compare the efficacy of each algorithm on our data, we created communities in the graph using each of the algorithms and printed out the list of communities (community), the number of nodes per community (componentSize), and the list of all the nodes in the community (comp).

Results using Louvain clustering algorithm:

```
      community  componentSize                                    comp
0            46           4397  [Ukraine, Middle East, Taiwan, United States, ...
1          6178           3908  [Geopolitics, Covid, 2010S, Huawei, U.S. Milit...
2           833           2514  [Mitt Romney, Panama Canal, Germany, Thucydide...
3         10296           2425  [National Security Council, U.S. Policy, U.S.,...
4          4862           1704  [Aidan Powers-Riggs, U.S. Business Lobby, Zte,...
...         ...            ...                                             ...
1216      30220              1          [87782e7c62c9688a9e8d5316788c37ef]
1217      30232              1          [c4ec2571e5934b68c11b4af07c9cb24e]
1218      30249              1          [ff77ef503a1e82d6fc9e6bc730c7caf8]
1219      30290              1                          [Factories Abroad]
1220      30298              1                 [Central Party School Bookstore]

[1221 rows x 3 columns]

Single node communities: 954
Double node communities: 46
Triple node communities: 47
Quad node communities: 28
Communities between 4 and 25 nodes: 126
There will be 267 community summaries
```

Results using Leiden clustering algorithm:

```
Communities:
      community  componentSize                                    comp
0          1180           3920  [Ukraine, Middle East, Nuclear Weapons, United...
1          1042           3902  [Geopolitics, Covid, 2010S, Huawei, U.S. Milit...
2           674           2581  [8d5d0465e616da951aae15049d18cead, U.S.-China ...
3          1189           2405  [U.S., Mao, 19a24cd556a2a0524c17d85063bed7ca, ...
4           759           2021  [Panama Canal, Uk, Germany, Central Europe, Au...
...         ...            ...                                             ...
1208        311              1              [Intel'S Disruption Is Now Complete]
```

```
1209        312                 1  [Interviews By Arnold Thackray And David C. Br...
1210        1058                1                        [Andrew Goldstein]
1211        1111                1                          [Happy Meal Toys]
1212        883                 1      [From Classic Failures To Global Competitors]

[1213 rows x 3 columns]
Single node communities: 954
Double node communities: 46
Triple node communities: 48
Quad node communities: 27
Communities between 4 and 25 nodes: 121
There will be 259 community summaries
```

The Leiden algorithm produced fewer communities than the Louvain algorithm, but the results were largely the same. However the Leiden communities were better connected internally, so we ultimately chose to use the Leiden communities to generate community summaries. Single node communities were the result of isolated nodes being present in the graph. Since these would not be helpful in querying, as they were not connected to any other nodes and would not be reached, we did not generate community summaries for single node communities.

IV.   Community Summaries

Next, we created summaries of each community. The nodes in each cluster were stored in a list, which was included in a prompt fed to the LLM.

Community Summary Generation Prompt

```
SYSTEM:
You are an experienced data analyst who is assisting the US government in consolidating
foreign policy data. The data is stored in a list of components that are all related to each
other. You use natural language to summarize the data.

USER:
Using only the components provided below, create a comprehensive and coherent summary that
captures the essence of the data. Ensure your summary includes a title that is specific to
the content, focusing on distinct details such as names, countries, concepts, policies, or
significant events mentioned in the data.
=============================================================
        Here is the data:

        {components}
=============================================================
**Title:** [Insert a specific title related to the data]
 **Summary:**
In your summary, avoid generic terms like 'foreign policy' or 'global relations.' Instead,
delve into the particulars of the components, clearly articulating their significance and
interconnections. Your summary should be thorough, easy to understand, and devoid of bullet
points, ensuring that all important components are mentioned.
Please provide your detailed summary and title below:
```

Sample Community Summary – Leiden Community ID 1180, the largest community

Title: The Complex Interplay of U.S.-China Relations and Global Geopolitical Dynamics

Summary: The data provides an intricate overview of the multifaceted relationship between the United States and China, highlighting the geopolitical, economic, and security dimensions that shape their interactions. Central to this discourse is the ongoing U.S.-China trade war, which has precipitated significant economic tensions and has influenced domestic policies in both nations. Key legislative measures such as the Uighur Forced Labor Prevention Act and the Chips and Science Act reflect the U.S. government's response to China's human rights practices and technological advancements, respectively. The U.S. public opinion, as gauged by Gallup in March 2022, indicates a rising wariness towards China, further complicating bilateral ties.

The data also underscores the role of various interest groups, including the U.S. Business Lobby and scholars, in shaping policies related to China. Figures like Robert Suettinger and Laura Silver emerge as pivotal voices in analyzing these dynamics. The Belt and Road Initiative, spearheaded by China, is presented as a strategic endeavor aimed at expanding its influence across Asia and beyond, with implications for U.S. interests in the region.

Military considerations are paramount, particularly in the context of Taiwan, where the Taiwan Relations Act of 1979 remains a cornerstone of U.S. policy. The military capabilities of both nations are scrutinized, with references to advanced weapons systems and missile technology, reflecting a landscape characterized by strategic competition. The ongoing Ukraine conflict and tensions in the Middle East further illustrate the interconnectedness of these geopolitical challenges, as they shape global security dynamics.

The data also touches on the implications of climate change and the U.S.'s efforts to engage in international climate agreements, such as the Paris Agreement, amidst rising global temperatures and environmental crises. The intersection of trade, technology, and climate policies illustrates the complexity of modern diplomacy, as nations navigate shared challenges while vying for influence.

Moreover, the rise of authoritarian capitalism in China poses a significant challenge to the liberal international order, prompting discussions about human rights and democratic values. The role of multilateral organizations, such as the World Trade Organization and the United Nations, is highlighted as crucial in mediating these tensions.

Overall, the data encapsulates the intricate web of interactions that define U.S.-China relations, emphasizing the strategic calculations, domestic pressures, and global contexts that inform policy decisions on both sides. The relationships among various stakeholders, including government actors, interest groups, and international institutions, underscore the complexity of navigating a world increasingly characterized by rivalry and interdependence.

V.   Prompting

Prompting was an essential part of this system, as it determined how the LLM interacted with the knowledge graph and generated policy recommendations.

```
SYSTEM:
You are an experienced advisor and international diplomat assisting the US government in
shaping foreign policy. Your role is to provide insightful and comprehensive answers to
inquiries by synthesizing structured data, unstructured data, and community summaries.
Please respond thoughtfully and thoroughly, ensuring that your answers reflect a deep
understanding of global issues and diplomatic nuances.

USER:
You are tasked with answering the question based solely on the provided context. Please
carefully consider the following sources:

1. **Structured Data**: This includes major entities and their relationships. Use this
information to understand the connections and significance of each entity.

2. **Unstructured Data**: This contains relevant text extracted from various documents.
Analyze this text for pertinent information that supports your response.

3. **Community Summaries**: Review the summaries that outline communities where closely
related entities exist. This context is crucial for a comprehensive understanding of the
relationships and influences among the entities.

Additionally, whenever applicable, cite historical precedents and events to strengthen your
answer. Ensure your response is well-supported by the information from the below sources.
{context}

Here is the question for you to answer: {question}
Use natural language and be detailed and thorough.
Answer:
```

The context data contained the Structured Data, Unstructured Data, and the Community Summaries in the following forms.

- **Structured Data**: presented in a node-relationship-neighbor format that demonstrates how entities are linked. The nodes include any entity mentioned in the user's prompt that is present in the knowledge graph. By querying the knowledge graph, we collect all the nodes directly connected to each node obtained from the question and the relationships between them.
  - Example: US -> Trade War - China
  - Example: President -> REQUEST -> Congress
- **Unstructured Data:** similar to a traditional RAG approach, where chunks of text from the original documents provide original context that may not be fully captured by the structured data.
- **Community Summaries:** high-level overviews of related entities.

Evaluation

In order to determine the effectiveness of using a GraphRAG system, we compared its performance to a standard RAG system we created with a vector store that ingested identical data to ensure a fair comparison. Both the GraphRAG and standard RAG systems used GPT-4o as the underlying LLM.

I. Criteria

The responses were evaluated according to each of the following criteria:

- **Strategic Relevance:** Assessing whether the foreign policy recommendation addresses core issues and objectives, focusing on current global trends and challenges.
- **Clarity and Specificity:** Evaluating whether the recommendation clearly states its objectives, expected outcomes, and actionable steps.
- **Risk and Impact Assessment:** Examining potential risks and unintended consequences of the foreign policy recommendation, including its social, economic, and political impact.
- **Flexibility and Adaptability:** Determining if the foreign policy recommendation allows for adjustments to respond to domestic or international conditions changing in order to adapt to unforeseen challenges.
- **Ethical and Legal Integrity:** Determining if the recommendation adheres to international laws and ethical standards.
- **Sustainability and Long-term Viability:** Assessing the foreign policy recommendation's potential to provide sustainable solutions that will promote long-term international stability to reduce the risk of recurring issues.

II. Evaluating by Individual Criteria

The LLM was provided with the prompt, the criteria, and the answer from each RAG system and was instructed to state the "winner" for each criteria and a brief justification.

We used a set of different LLMs to evaluate the responses to the following prompt:

```
# Taiwan Crisis Response Policy Analysis

## Context & Requirements
You are a senior policy advisor tasked with developing a comprehensive U.S. response
strategy to a hypothetical Chinese naval blockade of Taiwan. Your analysis will be used to
brief Congress and must be thorough, well-researched, and politically viable.

## Instructions
Develop a detailed policy proposal addressing this crisis using the following structure:

### Executive Summary (250-300 words)

- Policy name and core objectives
- Key recommendations
- Estimated resource requirements
- Timeline for implementation

### Strategic Assessment
```

- Current situation analysis
- Historical precedents and relevant case studies
- Critical threats and opportunities
- Assessment of Chinese military, economic, and diplomatic capabilities

Response Strategy

- Detail specific recommendations across these domains:

Military response options
- Specify required resources and timeline
- Provide cost-benefit analysis
- Cite relevant historical precedents
- Address potential risks and mitigation strategies

Economic measures
- Specify required resources and timeline
- Provide cost-benefit analysis
- Cite relevant historical precedents
- Address potential risks and mitigation strategies

Diplomatic initiatives
- Specify required resources and timeline
- Provide cost-benefit analysis
- Cite relevant historical precedents
- Address potential risks and mitigation strategies

Cybersecurity considerations
- Specify required resources and timeline
- Provide cost-benefit analysis
- Cite relevant historical precedents
- Address potential risks and mitigation strategies

Economic Impact Analysis
- Effects on U.S. economy (short and long-term)
- Global economic implications
- Supply chain disruptions
- Market responses and mitigation strategies

Congressional Considerations
- Bipartisan support potential
- Key stakeholder concerns
- Required legislative actions
- Funding mechanisms

Risk Assessment
- Potential Chinese counter-responses
- Escalation scenarios
- Economic retaliation possibilities
- International alliance impacts
- Domestic political risks

```
### Implementation Framework

- Timeline for key actions
- Resource deployment schedule
- Coordination requirements
- Success metrics

## Requirements:
- All recommendations must include specific resource requirements (troops, funding, etc.)
- Each major assertion must be supported by historical precedent or credible analysis
- All cost estimates must be itemized and justified
- Include citations for all key data points and historical references
- Consider both immediate crisis response and long-term strategic implications
- Address potential opposition arguments
- Maintain tone appropriate for congressional briefing

## Format:

- Use clear section headings
- Include executive summary
- Number all recommendations
- Use bullet points for key details
- Include tables for budget and resource allocations
- Provide citations in footnotes
```

The LLM evaluations are included below:

GPT-4o as an evaluator (temperature 0.50):

Criteria	Winner	Justification
Strategic Relevance	GraphRAG	GraphRAG provides a nuanced strategic assessment, addressing China's military, economic, and diplomatic capabilities in detail, and emphasizes leveraging international norms and alliances. In contrast, while Standard RAG covers similar elements, it lacks the depth in analyzing critical threats and opportunities, which makes GraphRAG's response more strategically relevant.
Clarity and Specificity	Standard RAG	Standard RAG excels in clarity by providing precise timelines and resource allocations for each domain, making it

		straightforward for implementation. GraphRAG, while detailed, presents a more complex and less organized timeline, making Standard RAG's approach more accessible and actionable.
Risk and Impact Assessment	GraphRAG	GraphRAG offers a thorough risk assessment, including potential military escalation and domestic political risks, with comprehensive mitigation strategies. Standard RAG also addresses risks but focuses more on economic retaliation, making GraphRAG's analysis more holistic and impactful.
Flexibility and Adaptability	GraphRAG	GraphRAG's phased approach with immediate and long-term strategies demonstrates adaptability to evolving situations. While Standard RAG provides a clear action plan, it lacks the same level of flexibility in adjusting to unforeseen challenges, giving GraphRAG an edge in adaptability.
Ethical and Legal Integrity	Standard RAG	Standard RAG emphasizes legal and ethical considerations by proposing international diplomatic efforts, such as a UN Security Council session, aligning with global standards. GraphRAG, though comprehensive, does not highlight legal mechanisms as prominently, making Standard RAG more aligned with ethical and legal integrity.
Sustainability and Long-term Viability	GraphRAG	GraphRAG focuses on long-term stability by reinforcing alliances and international norms, suggesting a sustainable approach to regional security. Standard RAG, while addressing immediate actions, lacks the same depth in considering long-term implications, making GraphRAG's strategy more viable for enduring peace.

Claude 3.5 Sonnet (temperature 0.50):

Criteria	Winner	Justification
Strategic Relevance	GraphRAG	While both responses address the core issues, the GraphRAG response provides a more comprehensive analysis of the geopolitical landscape. It offers a deeper examination of Chinese capabilities across military, economic, and diplomatic domains, compared to the Standard RAG's more general assessment. The GraphRAG response also more explicitly links the crisis to broader U.S. commitments in the Indo-Pacific, demonstrating a keener awareness of the strategic implications beyond the immediate situation.
Clarity and Specificity	Standard RAG	Both responses present clear strategies, but the Standard RAG response excels in providing precise details. It offers specific troop numbers (15,000 personnel) and a clear 30-60-90 day timeline for various actions, which the GraphRAG response lacks. The Standard RAG also provides more detailed breakdowns of funding allocations for each aspect of the response, making it more actionable and easier to evaluate compared to the GraphRAG's more general resource estimates.
Risk and Impact Assessment	GraphRAG	The GraphRAG response offers a more thorough examination of potential risks and impacts. While both address economic consequences, the GraphRAG delves deeper into specific sectors like semiconductors and supply chains. It also provides a more nuanced analysis of market responses and mitigation strategies. Unlike the Standard RAG, the GraphRAG response explicitly addresses domestic political risks and public opinion, demonstrating a more comprehensive consideration of potential consequences.
Flexibility and Adaptability	GraphRAG	The GraphRAG response demonstrates greater flexibility by emphasizing phased military and economic measures and ongoing diplomatic efforts. It explicitly mentions the need for de-escalation channels, indicating a readiness to adjust strategies as the situation evolves. In contrast, while the Standard RAG response does outline a timeline, it doesn't emphasize the same level of adaptability to

		changing conditions, making the GraphRAG approach more suitable for a dynamic crisis scenario.
Ethical and Legal Integrity	Standard RAG	Both responses generally adhere to international norms, but the Standard RAG response shows a stronger commitment to ethical and legal integrity. It explicitly proposes convening a special session of the UN Security Council, demonstrating a clear intent to work within established international legal frameworks. The GraphRAG response, while not ignoring legal considerations, doesn't emphasize this aspect as strongly, giving the Standard RAG an edge in this criterion.
Sustainability and Long-term Viability	GraphRAG	The GraphRAG response places a stronger emphasis on long-term regional stability and security in the Indo-Pacific. It provides a more comprehensive approach to building and maintaining international coalitions compared to the Standard RAG. While both consider long-term implications, the GraphRAG response more explicitly addresses the need for sustainable economic solutions and market stabilization strategies, demonstrating a more forward-thinking approach to ensuring the policy's long-term viability.

III.   Results

As shown by the tables above, when evaluated by both GPT-4o and Claude 3.5 Sonnet, the GraphRAG approach of our "AI foreign policy assistant" consistently outperformed Standard RAG across four criteria. However, Standard RAG performed better across the "Clarity and Specificity" and the "Ethical and Legal Integrity" criteria.

The evaluation demonstrates that the GraphRAG system offers substantial improvements over traditional RAG methods for foreign policy assistance. GraphRAG's performance in **Strategic Relevance**, **Risk and Impact Assessment**, and **Sustainability** suggests that this approach could be beneficial in high-stakes, long-term, dynamic geopolitical context. However, the Standard RAG system's strengths in clarity and adherence to ethical and legal standards indicate that both methods have complementary roles: GraphRAG providing strategic depth and Standard RAG offering procedural precision.

These insights reinforce the potential for GraphRAG to be used as an ineffective tool in foreign policy assistance, providing nuanced, adaptable, and sustainable recommendations that can assist governments in navigating complex international situations.

Conclusion

This paper presents an evaluation of a GraphRAG-based system to enhance foreign policy creation through the integration of a knowledge graph with an LLM. By comparing GraphRAG to traditional RAG, we explored how a structured knowledge graph can improve the depth, adaptability, and long-term viability of recommendations in high-stakes scenarios. Our findings show that the GraphRAG system significantly outperforms standard RAG in certain critical areas, including **Strategic Relevance**, **Risk and Impact Assessment**, **Flexibility and Adaptability,** and **Sustainability.** However, the Standard RAG system showed better performance in **Clarity and Specificity** and **Ethical and Legal Integrity**, indicating its utility in scenarios that prioritize procedural clarity and adherence to international norms.

The broader implications of this research suggest that GraphRAG systems can be transformative in dynamic policymaking. This could extend beyond foreign policy and into areas such as international trade, environmental policy, and crisis management. Future improvements in data collection could involve collecting a greater variety of types of data, such as images and video, and incorporating it into the knowledge graph.

Overall, our study suggests that GraphRAG is a powerful tool for nuanced policy assistance, representing a promising direction for AI-driven insights in foreign policy that has far-reaching potential to enhance decision-making in governments around the world.

References

[1] D. Edge *et al.*, "From Local to Global: A Graph RAG Approach to Query-Focused Summarization," Apr. 2024. Available: https://arxiv.org/pdf/2404.16130

[2] IBM, "What Are Large Language models? | IBM," www.ibm.com, 2023. https://www.ibm.com/topics/large-language-models

[3] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Apr. 2021. Available: https://arxiv.org/pdf/2005.11401

[4] Vision of Humanity, "Highest Number of Countries Engaged in Conflict since World War II," *Vision of Humanity*, Jun. 11, 2024. https://www.visionofhumanity.org/highest-number-of-countries-engaged-in-conflict-since-world-war-ii/

[5] Institute for Economics & Peace. Global Peace Index 2024: Measuring Peace in a Complex World, Sydney, June 2024. Available from: http://visionofhumanity.org/resources (accessed 30 Aug. 2024)

[6] G. M. Friedrichs and J. Tama, "Polarization and US foreign policy: key debates and new findings," *International Politics*, vol. 59, no. 5, pp. 767–785, Mar. 2022, doi: https://doi.org/10.1057/s41311-022-00381-0.

[7] L. L. Fowler, *Watchdogs on the Hill: The Decline of Congressional Oversight of U.S. Foreign Relations*. Princeton University Press, 2015.

[8] D. Murdick and O. J. Daniels, "The Perils of China's Great Information Wall," *TIME*, Jun. 26, 2023. https://time.com/6289567/china-restricts-open-source-data-research-backfire/

[9] M. Hunger, T. Bratanic, N. D. Jong, M. Senechal, and Persistent Team, "Neo4j LLM Knowledge Graph Builder - Extract Nodes and Relationships from Unstructured Text - Neo4j Labs," *Neo4j Graph Data Platform*. https://neo4j.com/labs/genai-ecosystem/llm-graph-builder/

[10] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Scientific Reports*, vol. 9, no. 1, Mar. 2019, doi: https://doi.org/10.1038/s41598-019-41695-z.