



Enhancing Policy Generation with GraphRAG and YouTube Data: A Logistics Case Study

Naganawa, Hisatoshi

Hirata, Enna

(Citation)

Electronics, 14(7):1241

(Issue Date)

2025-04

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(URL)

<https://hdl.handle.net/20.500.14094/0100495613>



Article

Enhancing Policy Generation with GraphRAG and YouTube Data: A Logistics Case Study

Hisatoshi Naganawa ¹  and Enna Hirata ^{2,*} ¹ Faculty of Ocean Science and Technology, Kobe University, Kobe 658-0022, Japan² Graduate School of Maritime Sciences, Kobe University, Kobe 658-0022, Japan

* Correspondence: enna.hirata@platinum.kobe-u.ac.jp; Tel.: +81-78-431-6200

Abstract: Graph-based retrieval-augmented generation (GraphRAG) represents an innovative advancement in natural language processing, leveraging the power of large language models (LLMs) for complex tasks such as policy generation. This research presents a GraphRAG model trained on YouTube data containing keywords related to logistics issues to generate policy proposals addressing these challenges. The collected data include both video subtitles and user comments, which are used to fine-tune the GraphRAG model. To evaluate the effectiveness of this approach, the performance of the proposed model is compared to a standard generative pre-trained transformer (GPT) model. The results show that the GraphRAG model outperforms the GPT model in most prompts, highlighting its potential to generate more accurate and contextually relevant policy recommendations. This study not only contributes to the evolving field of LLM-based natural language processing (NLP) applications but also explores new methods for improving model efficiency and scalability in real-world domains like logistics policy making.

Keywords: NLP; LLM; GraphRAG; GPT; YouTube; policy generation; text generation



Academic Editors: Xiangjian He and Krzysztof Wolk

Received: 27 February 2025

Revised: 19 March 2025

Accepted: 20 March 2025

Published: 21 March 2025

Citation: Naganawa, H.; Hirata, E. Enhancing Policy Generation with GraphRAG and YouTube Data: A Logistics Case Study. *Electronics* **2025**, *14*, 1241. <https://doi.org/10.3390/electronics14071241>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent advancements in large language models (LLMs) have had a major impact on policy generation in various domains. Researchers are actively exploring innovative methods to improve the decision-making capabilities of LLMs without the need for extensive fine-tuning. For example, LLM-based policy generation has been proposed for intent-based management [1]. Despite the potential in improving decision making and task generalization, real-world applications remain limited.

To bridge this gap, we experiment with generating policy recommendation to address the logistics challenges using Japan as a case study, which is a pressing issue due to the need for sustainable and resilient solutions.

The logistics sector has moved from a cost deflationary phase to a cost inflationary era. In Japan, transportation capacity has exceeded logistics demand since the 1990s, largely due to deregulation. However, since around 2017, there has been growing concern about a looming logistics and courier crisis. Unlike past fluctuations caused by temporary supply and demand imbalances, the current logistics crisis stems from a sustained decline in transportation capacity, leading to delivery delays and operational challenges. As a critical component of one of the world's largest economies, Japan's trucking industry plays a vital role in maintaining its ultra-convenient consumer culture.

By 2021, Japan's domestic freight volume reached 4.3 billion tons and 405 billion ton-miles annually, with trucks handling about 90% of total freight by weight and 50% by

ton-miles. The rise of e-commerce and increasing demand from manufacturers seeking to minimize excess inventory are expected to drive commercial freight down to 3.2 billion tons in 2030. However, commercial freight capacity is projected to decline from 2.6 billion tons in 2020 to 2.0 billion tons in 2030, primarily due to a shrinking truck driver workforce. The number of truck drivers is expected to fall from 662,000 in 2020 to 519,000 in 2030, a decline of nearly 30% from 2015 levels. In April 2024, a new regulation capping truck drivers' annual overtime at 960 hours was implemented, exacerbating the existing driver shortage—an issue widely referred to as the “2024 problem”.

The driver shortage issue is not unique to Japan. A study by the International Road Transport Union [2] warns that if no action is taken to recruit more drivers, there could be a global shortage of more than 7 million drivers by 2028 in the 36 countries surveyed. The most commonly cited reasons include an aging workforce, long working hours, and low wages.

To address the issue, this study introduces a graph-based retrieval-augmented generation (GraphRAG) approach for policy generation aimed at addressing logistics challenges, such as the truck driver shortage, using Japan as a case study. The results show that the proposed model outperforms a GPT-only model for most of the prompts. In addition, this study leverages a large textual dataset by integrating both subtitle and comment data extracted from relevant YouTube videos, allowing for a more comprehensive analysis.

This study makes several important contributions. It is the first application of GraphRAG in logistics analysis, pioneering its use to analyze textual data derived from YouTube. The results highlight the potential of AI-driven methods in policy making. Using a large-scale natural language processing (NLP) approach, it captures public discourse and industry insights in real time. It also categorizes solutions by stakeholder (logistics companies, governments, drivers, and consumers) for a comprehensive approach to the labor shortage. Finally, it provides insights for optimizing GraphRAG's search strategies (local, global, and hybrid) for logistics applications.

The study is structured as follows: Section 2 reviews previous research, Section 3 introduces the dataset, Section 4 presents the proposed methodology, Section 5 details experiments and results to evaluate the RAG effectiveness of the approach and discusses the findings, and Section 6 presents the conclusion.

2. Literature Review

2.1. Studies on Driver Shortage

Research on truck driver shortages has been approached from a variety of perspectives and methodologies. Table 1 summarizes the major studies that have examined this issue. Wang et al. [3], Chandiran et al. [4], and Schuster et al. [5] have analyzed the societal impacts and contributing factors behind the truck driver shortage using specially developed models. Meanwhile, Suzuki et al [6], Sersland et al. [7], and Corell [8] have used interview surveys and economic indicators to predict truck driver turnover.

Other studies, such as those by De Croon et al. [9], Prockl et al. [10], Swartz et al. [11], Wijngaards et al. [12], and Hage et al. [13], have examined the psychological aspects of truck drivers, particularly their well-being and its influence on turnover. In contrast, Beilock et al. [14] and de Winter et al. [15] have focused on job retention, examining the motivations and factors behind drivers' career decisions.

By using different research approaches, studies of the truck driver shortage aim to understand its social, economic, and psychological dimensions. Addressing this challenge requires aligning the goals of three key stakeholders: truck drivers, industry, and government [16].

In Japan, since the economic recession of the 1990s, companies have increasingly sought to reduce operating costs, including logistics costs. This trend has continued to the present day. Understanding public and industry perspectives on logistics challenges such as driver shortages is critical to shaping policy and practical responses, particularly in balancing supply and demand in the logistics sector [17].

Table 1. Summary of the literature on truck driver shortages.

Author	Year	Method	Contents
Beilock and Capelle [14]	1990	Questionnaire survey	Analysis of loyalty to occupation
de Croon et al. [9]	2004	Structural equation modeling	Analysis of psychological burden and turnover
Suzuki et al. [6]	2009	Econometrics method	Predicting truck driver turnover
Sersland and Nataraajan [7]	2015	Interview	Analysis of truck driver turnover
Prockl et al. [10]	2017	Statistical analysis	Analysis of well-being and safety environment
Swartz et al. [11]	2017	Structural equation modeling	Analysis of work attitude and safety environment
Belzer and Sedo [18]	2018	Statistical analysis	Analysis of long working hours
Burks and Monaco [19]	2018	Statistical analysis	Driver labor market analysis
Wijngaards et al. [12]	2019	Empirical sampling study	Determinants of well-being analysis
Hege et al. [13]	2019	Structural equation modeling	Survey of work–life conflicts
Lemke et al. [20]	2020	Data reviews	Impact analysis of COVID-19
Wang et al. [3]	2022	Structural equation modeling	Study of the impact of driver shortage
Chandiran et al. [4]	2023	System dynamic model	Study of the impact of driver shortage
Schuster et al. [5]	2023	Questionnaire survey	Factor analysis of driver shortage
de Winter et al. [15]	2024	Questionnaire survey	Survey of occupational image
Correll [8]	2024	Machine learning	Predicting truck driver turnover

2.2. Large Language Models in NLP

Artificial intelligence (AI) has received a great deal of attention in various fields and industries [21]. Over time, AI has gone through several stages of development and periods of stagnation known as AI winters. However, its global prominence skyrocketed after OpenAI released the chat generative pre-trained transformer (ChatGPT) in late 2022. The GPT family leverages massive datasets and LLMs trained with deep learning techniques to generate human-like text [22].

Built on a transformer architecture, GPT processes large amounts of text data to generate coherent responses based on user input [23]. While it excels at understanding context and handling diverse queries, its primary limitation is knowledge rigidity. It relies on pre-existing training data and cannot incorporate real-time updates. In addition, it struggles with factually accurate responses when confronted with unseen information, often producing misleading or incorrect content through inference. These limitations pose challenges for applications requiring high reliability [24,25].

To address these issues, retrieval-augmented generation (RAG) has been introduced. RAG extends the capabilities of GPT by retrieving external documents and databases to generate more factually grounded responses [26]. This approach is particularly useful for tasks involving real-time news updates and specialized knowledge domains. Miao et al. applied RAG in nephrology to improve its practical application. The authors created a specialized GPT model with RAG and demonstrated its potential in providing customized, accurate medical advice [27]. Hang et al. introduced an LLM and RAG integrated MCQGen framework for automated multiple-choice question generation. The authors suggest that MCQGen is effective at producing high-quality questions for different educational needs and learning styles, thus enhancing personalized learning [28]. However, RAG has its own challenges. It does not inherently consider document structure or interrelationships, which can lead to inefficient information integration. In addition, its effectiveness is highly

dependent on data quality and retrieval accuracy, and incorrect retrievals can still lead to misleading results [29,30].

GraphRAG was proposed to address the limitations of standard RAG; it extends traditional RAG by more effectively integrating both retrieval and generation processes [31]. It employs three different search strategies: (a) local search, (b) global search, and (c) hybrid search, each offering different trade-offs in search coverage, computational efficiency, and result completeness. Figure 1 illustrates the differences in generation methods between LLMs only, RAG, and GraphRAG.

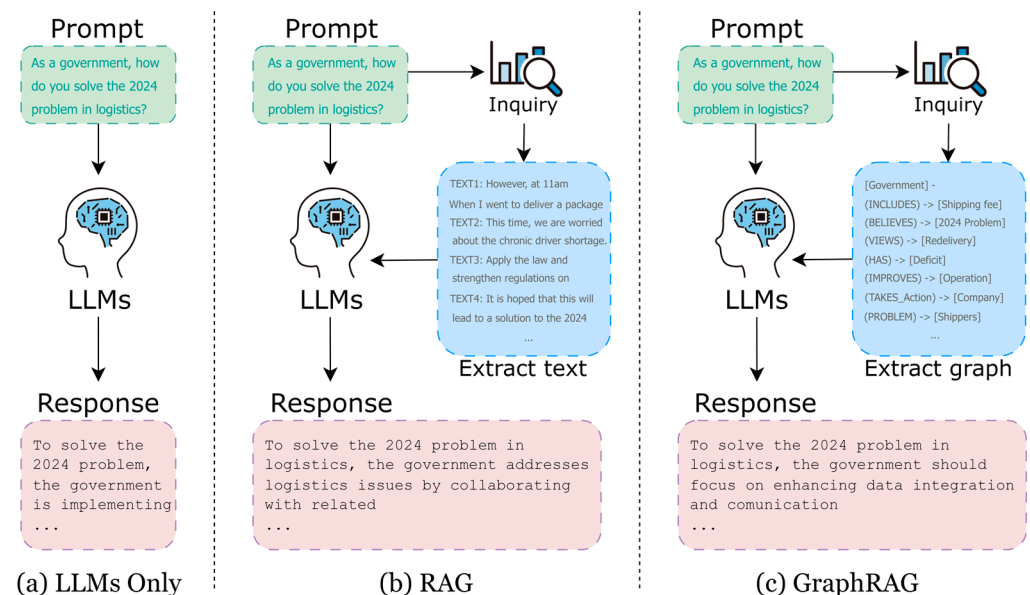


Figure 1. Differences in generation method.

Local search focuses on retrieving information from nodes closely related to the query, scanning only the surrounding nodes within a predefined depth. This method enables fast retrieval with minimal computational cost because it relies on local graph structures. However, its limited scope may result in missing globally optimal nodes, which limits the completeness of the results.

Global search, on the other hand, scans the entire graph to identify the most relevant nodes based on embedding similarity and other search techniques. This approach improves comprehensiveness by uncovering relationships and insights that local search may miss. However, it requires more processing power and longer processing times, making it less efficient for real-time applications.

Hybrid search combines the best of both approaches. It starts with a fast local search to gather immediately relevant information and then expands to a global search to retrieve additional insights that may have been missed. This method ensures highly accurate search results by maintaining a balance between local relevance and global context [32]. However, it is more complex to implement, and the computational cost can be significant due to the dual search process.

As described above, there is a tradeoff between search coverage and computational cost in the GraphRAG generation method. Overall, RAG-based techniques, especially GraphRAG, enhance factual grounding by integrating external document retrieval, making them highly suitable for real-time logistics analysis.

3. Method

3.1. Base Model and Fine-Tuning

We compared VectorRAG and GraphRAG and chose the latter for the following reasons. VectorRAG has a strong focus on semantic search, identifying documents that are conceptually related to the query, even without exact keyword matches. GraphRAG, on the other hand, is characterized by its ability to retrieve information by exploiting relationships and connections between entities. There are a few other contemporary RAG technologies, such as Hybrid RAG and Light RAG. Hybrid RAG has a primary focus on improving retrieval quality. While it applies both neural and keyword-based retrieval, it has increased computational cost, complexity in implementation and tuning, and risk of conflicting results. Light RAG has its primary focus on reducing computational cost; for this reason, it has disadvantages such as low retrieval accuracy, limited knowledge depth, and is less effective for large-scale data such as the data applied in this study.

To fine-tune GraphRAG for logistics policy generation, we first used a pre-trained version of GraphRAG as our base model. The model was adapted for policy generation by incorporating domain-specific logistics datasets obtained from YouTube. We made modifications to the input processing and output generation layers to handle domain-specific constraints and goals (Table 2). For GPT tuning, the temperature was set to 0 to ensure more definite and consistent answers. The number of nodes used to generate responses was limited to 100 to balance the thoroughness of search results with response speed. In addition, fuzzy matching was used as the search mechanism in GraphRAG. This technique allows searches based on partial matches and similarities between queries and node names. Cosine similarity was used to calculate the fuzzy match score, with a threshold set at 0.5. This approach effectively combines flexible natural language understanding with the relational structure of the graph, improving both the accuracy and practical utility of information retrieval and response generation.

Table 2. Tuning Metrics.

Metrics	Value
Temperature	0
Output Node Limit	100
Fuzzy Match Score	0.5

For fine-tuning, data obtained from YouTube videos are used. YouTube was selected for two main reasons. First, as Japan's largest online video platform, it has a usage rate of 87.9% across all generations [33]. Second, it provides multiple perspectives, capturing both video creators' viewpoints and audience comments, making it a valuable two-sided source of information.

The search keywords included "2024 problem in logistics" and "logistics crisis" in Japanese, which were retrieved using YouTube Data API v3 [34]. In total, 3037 videos and corresponding comment data were collected. Whisper v3 large was then applied for speech-to-text conversion. Finally, the data were translated to English and structured for NLP handling. The details of data collection are described in Section 4.

3.2. GraphRAG-Based NLP Analysis

GraphRAG was used to generate responses to four prompts employing four methods, namely (I) GPT-only, (II) local search, (III) global search, and (IV) hybrid search, to generate policy recommendations.

GraphRAG works by using a graph-structured knowledge base that stores domain-specific entities and their relationships. When a query is received, it is first mapped to

a relevant subgraph using a graph traversal technique. The retrieval process starts by identifying the nodes most relevant to the query through a search of the graph, which may involve traversing edges, analyzing node labels, and considering relationships within the graph. The goal of this process is to find the subgraph that best matches the semantic content of the query.

The information retrieval from the graph is guided by a scoring function based on cosine similarity, which measures the closeness between the query vector and the node representations. Let us denote the query vector as Q and the node vector as N_i . The similarity score between the query and each node N_i is computed as follows:

$$\text{Score}(Q, N_i) = \frac{Q \cdot N_i}{\|Q\| \|N_i\|} \quad (1)$$

where Q and N_i represents the dot product of the query and node vectors, and $\|Q\|$ and $\|N_i\|$ are the magnitudes of the respective vectors. Nodes with a similarity score above a defined threshold (e.g., 0.5) are retrieved as the most relevant for the query. The pseudocode for the Algorithm A1 is presented in Appendix A.

GraphRAG was created using neo4j desktop as the database and GPT 4o-mini as the LLM. The four prompts are designed from the perspective of four key stakeholders: logistics companies, government agencies, truck drivers, and consumers (Table 3). It is important to note that consumers are also a key stakeholder because they are the recipients of logistics services. The shortage of truck drivers directly affects delivery times, which in turn affects customer satisfaction. Therefore, the consumer perspective should be a key consideration in policy making.

Table 3. Overview of prompts.

No.	Prompts
1	As a logistics company, how do you solve the 2024 problem in logistics?
2	As a government, how do you solve the 2024 problem in logistics?
3	As a driver, how do you solve the 2024 problem in logistics?
4	As a consumer, how do you solve the 2024 problem in logistics?

3.3. Performance Metrics

The performance of each model is evaluated using five metrics. Paraphrase-MiniLM-L6-v2 was applied to embed sentences for the calculation of cosine similarity (Table 4).

Table 4. Performance evaluation metrics.

Metrics	Abbreviations	Descriptions
Word count	WC	Total number of words in a sentence.
Type Token Ratios	TTR	Measures vocabulary diversity; the higher the TTR, the more diverse the vocabulary.
Average Dependency Distance	ADD	Measures the distance of dependencies between words in a sentence; higher ADD indicates more complex sentence structure and higher reading load.
Flesch–Kincaid Grade Level	FKG	Indicator for evaluating the readability of a text.
Cosine Similarity	CS	An index measuring the degree of similarity between texts.

4. Data

4.1. Methods of Data Acquisition and Processing

Data collection process is outlined in Figure 2. It consists of the following steps:

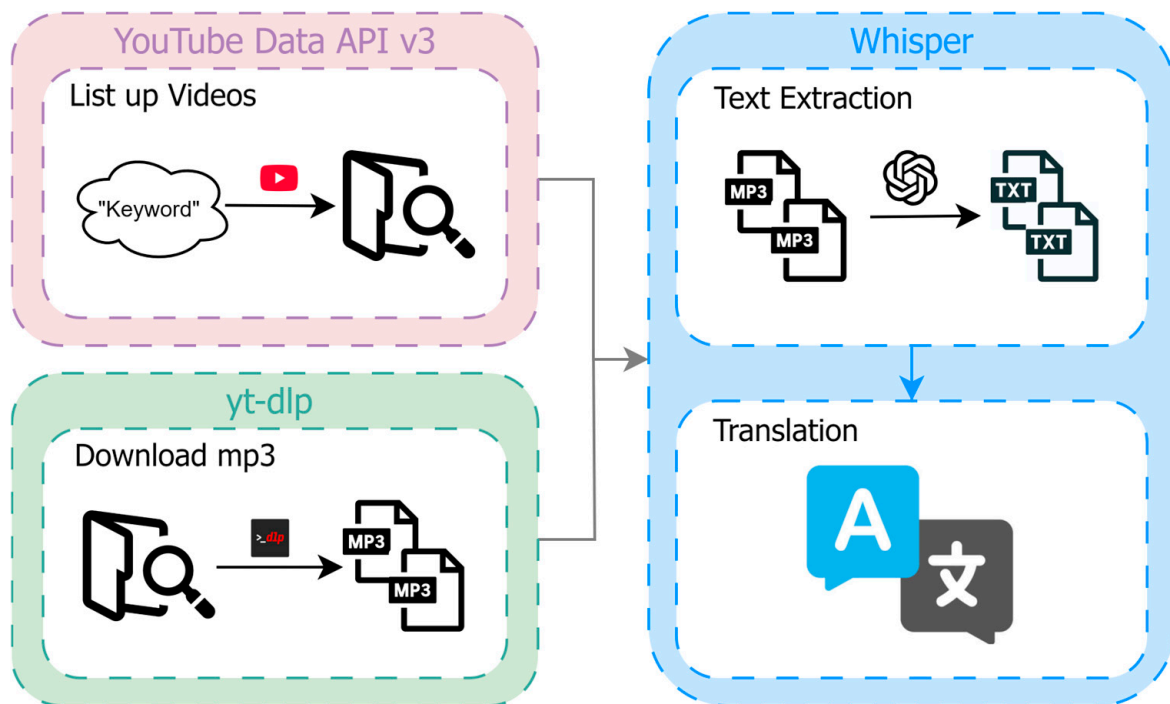


Figure 2. Data collection process.

(a) YouTube Data API v3 is used to create a list of videos related to the truck driver shortage. The search keywords include “2024 problem in logistics” and “logistics crisis” in Japanese.

(b) Duplicate or irrelevant videos were removed.

(c) Based on the video list, a comment dataset was created using YouTube Data API v3 [34]. The dataset included the video ID, date and time of posting, comment content, number of likes, and username.

(d) Based on the video list, the audio data of the videos were collected using yt-dlp [35]. The audio data were also converted into text data using whisper v3 large [36].

(e) All collected text data were converted to English text.

The Python programming language version 3.13.1 was used to convert these data into a structured format for further training and analysis. In contrast to the existing literature, which primarily analyzes data from video comments, our study includes both comment data and text data, which are converted from audio for analysis.

4.2. Overview of Data

A total of 3037 videos were listed as a result of the collection; the number of YouTube videos peaked between February and June 2021 and again between February and May 2024 (Figure 3). This is likely related to the increased attention surrounding the increase in the overtime premium rate from 25% to 50% for overtime exceeding 60 hours per month in small and medium enterprises beginning 1 April 2023. The second peak is likely related to the introduction of a cap on overtime work of 960 hours per year from 1 April 2024. The introduction of a driver hours and pay policy has led to an increased number of related videos being uploaded.

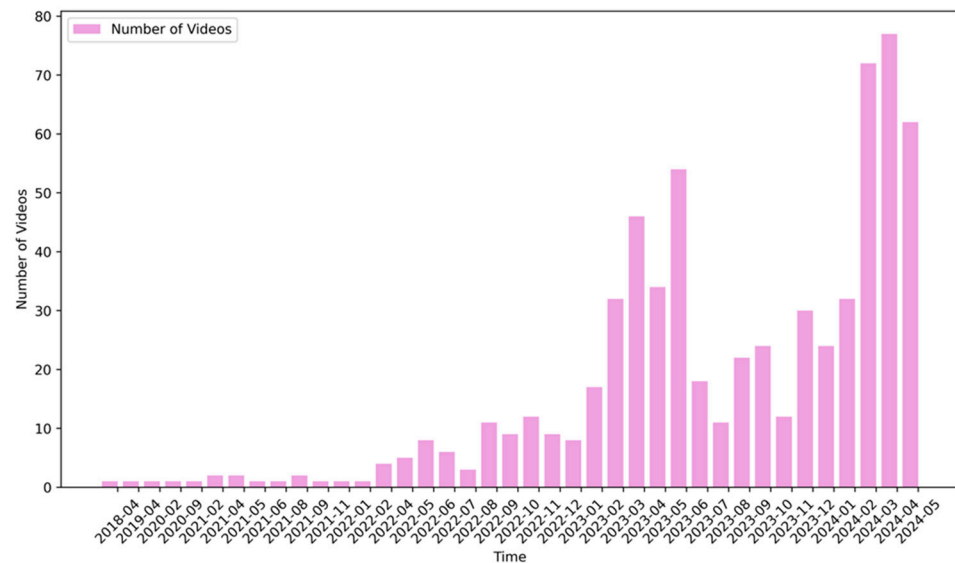


Figure 3. Numbers of videos.

5. Results and Discussion

5.1. Prompts and Answers

Following the definition explained in Section 3.2, the prompts are run in each of the four models, namely (I) GPT-only, (II) local search, (III) global search, and (IV) hybrid search, and the results are shown in Tables 5–8.

Table 5. Output of “As a logistics company, how do you solve the 2024 problem in logistics?”

Model	Output
I	The “2024 problem” in logistics isn’t a recognized term in the industry. If you’re referring to a specific issue expected to arise by 2024, please provide more details. However, general solutions for future logistics problems include investing in technology for better tracking and management, improving supply chain visibility, focusing on sustainability, and training staff to handle new challenges.
II	To address the 2024 problem in logistics, we focus on compliance measures and innovative solutions. This includes optimizing our operations through dynamic pricing strategies, enhancing delivery speed, and improving cost performance. We also leverage technology, such as robots, to streamline processes and ensure efficient order handling and delivery. Additionally, we actively participate in industry discussions, like the Supermarket Representatives Meeting, to stay aligned with market needs and collaborate with partners to develop effective strategies.
III	To solve the 2024 problem in logistics, we focus on enhancing our data management by integrating structured data systems for efficient tracking and inventory management, while also leveraging unstructured data analytics to gain insights from customer feedback and market trends. This dual approach allows us to optimize our supply chain, improve decision-making, and adapt to changing demands effectively.
IV	To address the 2024 problem in logistics, we focus on compliance measures and innovative solutions. This includes optimizing our operations through dynamic pricing strategies, enhancing delivery speed, and improving cost performance. We also leverage technology, such as robots, to streamline processes and ensure efficient order handling and delivery. Additionally, we participate in industry discussions, like the Supermarket Representatives Meeting, to collaborate on best practices and solutions.

Table 6. Output of “As a government, how do you solve the 2024 problem in logistics?”

Model	Output
I	To solve the 2024 problem in logistics, the government is implementing several measures, including providing guidance and support to shippers, improving transportation efficiency through logistics digital transformation (DX) and modal shifts, and addressing labor shortages by raising wages for truck drivers. Additionally, the government is introducing regulatory measures to ensure proper transactions and reduce redelivery rates by incentivizing consumers to choose flexible delivery options. Collaborative efforts among various ministries aim to enhance the overall logistics infrastructure and resolve the challenges faced by the industry.

Table 6. *Cont.*

Model	Output
II	To solve the 2024 problem in logistics, the government addresses logistics issues by collaborating with related ministries and agencies, providing a logistics policy package, and ensuring transportation capacity. They also take action by working with transport companies, offering support mechanisms, and raising wages for truck drivers to improve operational efficiency. Additionally, the government reviews shipping charges and implements regulations to enhance the logistics industry.
III	To solve the 2024 problem in logistics, the government should focus on enhancing data integration and communication between various stakeholders. This includes investing in advanced technologies for real-time tracking, optimizing supply chain processes, and ensuring compliance with regulations. Additionally, fostering collaboration between public and private sectors can improve resource allocation and efficiency. Implementing training programs for workforce development will also be crucial to adapt to evolving logistics demands.
IV	To solve the 2024 problem in logistics, the government addresses logistics issues by collaborating with related ministries and agencies, providing subsidies, and implementing a logistics policy package. They also ensure transportation capacity, improve operational efficiency, and raise wages for truck drivers to enhance the workforce. Additionally, the government interacts with the trucking association and requests input from shippers to develop effective solutions.

Table 7. Output of “As a driver, how do you solve the 2024 problem in logistics?”

Model	Output
I	To solve the 2024 problem, reduce waste and protect working hours by optimizing operation plans and utilizing digital technology. Promote shift work and appropriate freight rates, and achieve sustainable logistics through safe operations and efficiency.
II	To solve the 2024 problem in logistics as a driver, focus on optimizing routes to reduce fuel consumption and delivery times, utilize real-time data for traffic updates, and maintain effective communication with dispatch to adapt to any changes. Additionally, prioritize efficient loading and unloading processes to enhance overall productivity.
III	To solve the 2024 problem in logistics as a driver, focus on optimizing routes to reduce delivery times and costs, utilize technology for real-time tracking and communication, and ensure compliance with regulations. Additionally, prioritize efficient loading and unloading processes to enhance overall productivity.
IV	To solve the 2024 problem in logistics as a driver, focus on optimizing routes to reduce delivery times and costs, utilize real-time data for traffic and weather updates, and ensure efficient communication with dispatch to adapt to any changes. Additionally, prioritize vehicle maintenance to avoid breakdowns and enhance overall efficiency.

Table 8. Output of “As a consumer, how do you solve the 2024 problem in logistics?”

Model	Output
I	As a consumer, you don’t directly solve the 2024 problem in logistics. This issue pertains to the logistics industry and professionals in that field. However, you can contribute to the solution by supporting companies that implement sustainable and efficient logistics practices, providing feedback to companies about their delivery services, and opting for delivery options that reduce environmental impact.
II	To solve the 2024 problem in logistics as a consumer, I would focus on leveraging technology to enhance efficiency. This includes using apps for real-time tracking of shipments, opting for services that utilize predictive analytics to anticipate delays, and choosing providers that prioritize sustainability and transparency in their supply chains. Additionally, I would advocate for better communication with logistics companies to ensure timely updates and resolutions to any issues that arise.
III	To solve the 2024 problem in logistics as a consumer, I would focus on leveraging technology to enhance efficiency. This includes using apps for real-time tracking of shipments, opting for services that utilize predictive analytics for better delivery estimates, and choosing companies that prioritize sustainability and transparency in their supply chains. Additionally, I would advocate for better communication with logistics providers to ensure timely updates and address any potential issues proactively.
IV	To solve the 2024 problem in logistics as a consumer, I would focus on leveraging technology to enhance efficiency. This includes using apps for real-time tracking of shipments, opting for services that utilize predictive analytics for better inventory management, and choosing providers that prioritize sustainability and transparency in their supply chains. Additionally, I would advocate for better communication with logistics companies to ensure timely updates and resolutions to any issues that arise.

The intuitive impression is that the GPT model (I) generates broader, more general proposals, while RAG-based models (II, III, and IV) provide more specific and targeted suggestions. Quantitative performance evaluation is discussed in Section 5.2.

5.2. Performance Evaluation

The performance of each model, as measured by the metrics described in Section 3.3, is shown in Table 9. The results indicate that overall model III (Global Search) outperforms the other models, especially for Prompts 1, 3, and 4. Interestingly, for Prompt 2 (government), model 1 (GPT only) outperforms GraphRAG models. This may be because GPT tends to excel at understanding broader, context-rich scenarios. On the other hand, RAG could be more useful when factual accuracy or transparent source traceability is the top priority. In a government policy recommendation context, the best model may combine both: using a retrieval-based system like GraphRAG for fact-based grounding and taking advantage of GPT’s capabilities for higher-level synthesis and creativity in policy design.

Table 9. Performance metrics (bold indicates best performance).

Prompt	1				2			
Model	I	II	III	IV	I	II	III	IV
WC	68	73	64	78	94	73	76	71
TTR	0.765	0.712	0.797	0.705	0.723	0.712	0.789	0.718
ADD	2.908	3.214	3.328	3.149	3.220	3.214	2.932	3.132
FKG	15.500	18.700	19.300	16.800	21.300	18.700	17.000	18.400
CS	-	0.709	0.660	0.721	-	0.901	0.810	0.907
Prompt	3				4			
Model	I	II	III	IV	I	II	III	IV
WC	39	55	49	56	66	78	78	79
TTR	0.846	0.818	0.816	0.804	0.712	0.756	0.762	0.747
ADD	3.162	3.509	3.571	3.464	3.302	3.195	3.312	3.244
FKG	14.800	18.700	18.800	17.800	15.500	17.300	18.400	18.600
CS	-	0.706	0.717	0.690	-	0.776	0.778	0.776

6. Conclusions

This study offers a new perspective on Japan’s truck driver shortage by using GraphRAG to analyze large amounts of text data from YouTube videos and comments. Going beyond traditional statistical analysis, it provides an alternative approach to understanding logistics challenges. Four models were used to generate responses to four prompts representing the perspectives of logistics companies, governments, drivers, and consumers. In three out of four cases, GraphRAG outperformed the GPT model, demonstrating higher vocabulary density, better readability, and greater similarity in responses.

This study contributes in the following ways. First, it is the first application of GraphRAG in logistics analysis. It pioneers the use of GraphRAG to analyze textual data retrieved from YouTube for logistics-related issues. The results demonstrate the potential of AI-driven methods in policy making. Second, the application of a large-scale NLP approach provides a rich and diverse dataset that captures public discourse and industry insights in real time. Third, the research categorizes solutions based on different logistics stakeholders (logistics companies, governments, drivers, and consumers), providing a holistic approach to addressing labor shortages. Fourth, by comparing different search strategies (local, global, and hybrid) within GraphRAG, the study provides insights into optimizing retrieval-augmented generation models for logistics applications.

This study has several limitations. First, the dataset is based on social media data. While YouTube provides a broad dataset, it may contain biases, misinformation, or exaggerated opinions that affect the validity of the conclusions. Second, the use of GraphRAG, particularly hybrid search, increases computational costs and may not be easily scalable for

real-time logistics applications. Third, the study does not compare the results with official logistics data or expert surveys, which limits the validation of its conclusions.

Future studies should combine GraphRAG with official industry reports, global positioning system (GPS) tracking, and transportation network datasets for more comprehensive insights. It would be worthwhile to develop improved methods for validating and interpreting AI-generated insights to increase industry confidence and adoption. To address potential biases or rumors in SNS data, future studies could incorporate LLM-based fact-checking tools such as TrumorGPT [37]. Finally, further research should assess how AI-driven analytics can shape government regulations and industry standards for sustainable logistics.

Author Contributions: Conceptualization, H.N. and E.H.; methodology, H.N. and E.H.; software, H.N.; validation, H.N. and E.H.; data curation, H.N.; writing—original draft preparation, H.N. and E.H.; writing—review and editing, E.H.; visualization, H.N.; supervision, E.H.; funding acquisition, E.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant Numbers JP 23K04076.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank the editor and referees for helping to improve this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LLMs	Large language models
GraphRAG	Graph-based retrieval-augmented generation
NLP	Natural language processing
GPT	Generative pre-trained transformer
WC	Word count
TTR	Type token ratios
ADD	Average dependency distance
FKG	Flesch–Kincaid grade level
CS	Cosine similarity
AI	Artificial intelligence
DX	Logistics digital transformation
GPS	Global positioning system

Appendix A. Pseudocode of the GraphRAG Algorithm

Algorithm A1. GraphRAG
Structured Retriever: (question)
- Input: Question
- Processing:
- Extract entities from the question (entity_chain.invoke(question))
- Search structured data in Neo4j using each entity (graph.query(entity))
- Output: structured data (result)
Retriever:
- Input: question, mode
- Processing:
- If mode is “gpt_only”, pass question directly to LLM (llm.invoke(question))
- If mode is “local” or “hybrid”, retrieve structured data (structured_retriever(question))
- If mode is “global” or “hybrid”, search for unstructured data (vector_index.search(question))
- Output: result of combining structured and unstructured data

Algorithm A1. Cont.**QA Chain:**

- Input: question, mode
- Processing:
 - Call retriever to get data (context) based on the question
 - Pass the data to LLM to generate the final answer (llm.invoke(context))
- Output: final answer

References

1. Dzeperaska, K.; Lin, J.; Tizghadam, A.; Leon-Garcia, A. LLM-Based Policy Generation for Intent-Based Management of Applications. In Proceedings of the 2023 19th International Conference on Network and Service Management (CNSM), Niagara Falls, ON, Canada, 30 October–2 November 2023; pp. 1–7.
2. IRU Global Truck Driver Shortage Report 2023 | IRU | World Road Transport Organisation. Available online: <https://www.iru.org/resources/iru-library/global-truck-driver-shortage-report-2023> (accessed on 15 January 2025).
3. Wang, M.; Wood, L.C.; Wang, B. Transportation Capacity Shortage Influence on Logistics Performance: Evidence from the Driver Shortage. *Heliyon* **2022**, *8*, e09423. [\[CrossRef\]](#)
4. Chandiran, P.; Ramasubramaniam, M.; Venkatesh, V.G.; Mani, V.; Shi, Y. Can Driver Supply Disruption Alleviate Driver Shortages? A Systems Approach. *Transp. Policy* **2023**, *130*, 116–129. [\[CrossRef\]](#)
5. Schuster, A.M.; Agrawal, S.; Britt, N.; Sperry, D.; Van Fossen, J.A.; Wang, S.; Mack, E.A.; Liberman, J.; Cotten, S.R. Will Automated Vehicles Solve the Truck Driver Shortages? Perspectives from the Trucking Industry. *Technol. Soc.* **2023**, *74*, 102313. [\[CrossRef\]](#)
6. Suzuki, Y.; Crum, M.R.; Pautsch, G.R. Predicting Truck Driver Turnover. *Transp. Res. Part E Logist. Transp. Rev.* **2009**, *45*, 538–550. [\[CrossRef\]](#)
7. Sersland, D.; Natarajan, R. Driver Turnover Research: Exploring the Missing Angle with a Global Perspective. *J. Serv. Manag.* **2015**, *26*, 648–661. [\[CrossRef\]](#)
8. Correll, D.H. Predicting and Understanding Long-Haul Truck Driver Turnover Using Driver-Level Operational Data and Supervised Machine Learning Classifiers. *Expert Syst. Appl.* **2024**, *242*, 122782. [\[CrossRef\]](#)
9. de Croon, E.M.; Sluiter, J.K.; Blonk, R.W.B.; Broersen, J.P.J.; Frings-Dresen, M.H.W. Stressful Work, Psychological Job Strain, and Turnover: A 2-Year Prospective Cohort Study of Truck Drivers. *J. Appl. Psychol.* **2004**, *89*, 442–454. [\[CrossRef\]](#)
10. Prockl, G.; Teller, C.; Kotzab, H.; Angell, R. Antecedents of Truck Drivers' Job Satisfaction and Retention Proneness. *J. Bus. Logist.* **2017**, *38*, 184–196. [\[CrossRef\]](#)
11. Swartz, S.M.; Douglas, M.A.; Roberts, M.D.; Overstreet, R.E. Leavin' on My Mind: Influence of Safety Climate on Truck Drivers' Job Attitudes and Intentions to Leave. *Transp. J.* **2017**, *56*, 184–209. [\[CrossRef\]](#)
12. Wijngaards, I.; Hendriks, M.; Burger, M.J. Steering towards Happiness: An Experience Sampling Study on the Determinants of Happiness of Truck Drivers. *Transp. Res. Part A Policy Pract.* **2019**, *128*, 131–148. [\[CrossRef\]](#)
13. Hege, A.; Lemke, M.K.; Apostolopoulos, Y.; Whitaker, B.; Sönmez, S. Work-Life Conflict among U.S. Long-Haul Truck Drivers: Influences of Work Organization, Perceived Job Stress, Sleep, and Organizational Support. *Int. J. Environ. Res. Public Health* **2019**, *16*, 984. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Beilock, R.; Capelle, R.B. Occupational Loyalties Among Truck Drivers. *Transp. J.* **1990**, *29*, 20–28.
15. de Winter, J.; Driessen, T.; Dodou, D.; Cannoo, A. Exploring the Challenges Faced by Dutch Truck Drivers in the Era of Technological Advancement. *Front. Public Health* **2024**, *12*, 1352979. [\[CrossRef\]](#)
16. Mittal, N.; Udayakumar, P.D.; Raghuram, G.; Bajaj, N. The Endemic Issue of Truck Driver Shortage—A Comparative Study between India and the United States. *Res. Transp. Econ.* **2018**, *71*, 76–84. [\[CrossRef\]](#)
17. Hirata, E.; Matsuda, T. Examining Logistics Developments in Post-Pandemic Japan through Sentiment Analysis of Twitter Data. *Asian Transp. Stud.* **2023**, *9*, 100110. [\[CrossRef\]](#)
18. Belzer, M.H.; Sedo, S.A. Why Do Long Distance Truck Drivers Work Extremely Long Hours? *Econ. Labour Relat. Rev.* **2018**, *29*, 59–79. [\[CrossRef\]](#)
19. Burks, S.V.; Monaco, K. *Is the U.S. Labor Market for Truck Drivers Broken? An Empirical Analysis Using Nationally Representative Data*; Institute of Labor Economics (IZA): Bonn, Germany, 2018; pp. 1–47.
20. Lemke, M.K.; Apostolopoulos, Y.; Sönmez, S. A Novel COVID-19 Based Truck Driver Syndemic? Implications for Public Health, Safety, and Vital Supply Chains. *Am. J. Ind. Med.* **2020**, *63*, 659–662. [\[CrossRef\]](#)
21. Hyder, Z.; Siau, K.; Nah, F. Artificial Intelligence, Machine Learning, and Autonomous Technologies in Mining Industry. *J. Database Manag.* **2019**, *30*, 67–79. [\[CrossRef\]](#)
22. Cascella, M.; Semeraro, F.; Montomoli, J.; Bellini, V.; Piazza, O.; Bignami, E. The Breakthrough of Large Language Models Release for Medical Applications: 1-Year Timeline and Perspectives. *J. Med. Syst.* **2024**, *48*, 22. [\[CrossRef\]](#)

23. Yenduri, G.; Srivastava, G.; Maddikunta, P.K.; Jhaveri, R.H.; Wang, W.; Vasilakos, A.V.; Gadekallu, T.R. Generative Pre-Trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *arXiv* **2023**, arXiv:2305.10435. [[CrossRef](#)]
24. Rawte, V.; Sheth, A.; Das, A. A Survey of Hallucination in Large Foundation Models. *arXiv* **2023**, arXiv:2309.05922.
25. McIntosh, T.R.; Susnjak, T.; Liu, T.; Watters, P.; Halgamuge, M.N. The Inadequacy of Reinforcement Learning From Human Feedback—Radicalizing Large Language Models via Semantic Vulnerabilities. *IEEE Trans. Cogn. Dev. Syst.* **2024**, *16*, 1561–1574. [[CrossRef](#)]
26. Zhao, P.; Zhang, H.; Yu, Q.; Wang, Z.; Geng, Y.; Fu, F.; Yang, L.; Zhang, W.; Jiang, J.; Cui, B. Retrieval-Augmented Generation for AI-Generated Content: A Survey. *arXiv* **2024**, arXiv:2402.19473.
27. Miao, J.; Thongprayoon, C.; Suppadungsuk, S.; Garcia Valencia, O.A.; Cheungpasitporn, W. Integrating Retrieval-Augmented Generation with Large Language Models in Nephrology: Advancing Practical Applications. *Medicina* **2024**, *60*, 445. [[CrossRef](#)] [[PubMed](#)]
28. Hang, C.N.; Wei Tan, C.; Yu, P.-D. MCQGen: A Large Language Model-Driven MCQ Generator for Personalized Learning. *IEEE Access* **2024**, *12*, 102261–102273. [[CrossRef](#)]
29. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T. Retrieval-Augmented Generation for Knowledge-Intensive Nlp Tasks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9459–9474.
30. Barnett, S.; Kurniawan, S.; Thudumu, S.; Brannelly, Z.; Abdelrazek, M. Seven Failure Points When Engineering a Retrieval Augmented Generation System. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering—Software Engineering for AI, Lisbon, Portugal, 14–15 April 2024; ACM: New York, NY, USA, 2024; pp. 194–199.
31. Edge, D.; Trinh, H.; Cheng, N.; Bradley, J.; Chao, A.; Mody, A.; Truitt, S.; Larson, J. From Local to Global: A Graph Rag Approach to Query-Focused Summarization. *arXiv* **2024**, arXiv:2404.16130.
32. Han, H.; Shomer, H.; Wang, Y.; Lei, Y.; Guo, K.; Hua, Z.; Long, B.; Liu, H.; Tang, J. RAG vs. GraphRAG: A Systematic Evaluation and Key Insights. *arXiv* **2025**, arXiv:2502.11371.
33. Hottolink 2025 Ranking of SNS Utilization in Japan and Worldwide. Available online: https://www.hottolink.co.jp/column/20250106_114872/ (accessed on 14 February 2025).
34. Google YouTube Data API. Available online: <https://developers.google.com/youtube/v3> (accessed on 15 January 2025).
35. pukkandan; Shirt-dev; Sepro; Sawicki, S.; coletdjnz; ashonly Yt-Dlp/Yt-Dlp 2025. Available online: <https://github.com/yt-dlp/yt-dlp/releases> (accessed on 14 February 2025).
36. Radford, A.; Kim, J.W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. Robust Speech Recognition via Large-Scale Weak Supervision. *arXiv* **2022**, arXiv:2212.04356.
37. Hang, C.N.; Yu, P.-D.; Tan, C.W. TrumorGPT: Query Optimization and Semantic Reasoning over Networks for Automated Fact-Checking. In Proceedings of the 2024 58th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 13–15 March 2024; pp. 1–6.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.