7th European Conference on Industrial Engineering and Operations Management
Augsburg, Germany, July 16-18, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/EU07.20240271
Published: July 16, 2024

GAI-based Intelligent Cognitive Assistant for Urban Rideshare Safety Recommendation Services

Chenxi Tao, Roosan Liyons, Roger J. Jiao and Seung-Kyum Choi
School of Mechanical Engineering
Georgia Institute of Technology
Atlanta, GA, USA
ctao40@gatech.edu

Abstract

Urban rideshares have become a cornerstone of urban transportation, offering unparalleled convenience and presenting a significant market opportunity. However, safety concerns are escalating as their usage proliferates. By harnessing advancements in computer vision and mobile technologies, alongside the widespread adoption of shared vehicles, urban rideshare systems can now deliver real-time road information and safety advisories. A critical gap remains in predictive capabilities and the translation of image data into human-understandable language or actionable safety recommendations. This paper investigates the potential of creating a real-time generative AI-based intelligent cognitive assistant for updating road information, specifically designed for urban rideshare safety recommendation services. Utilizing computer vision to collect road data during rides, the platform employs a generative AI database to comprehensively store and analyze this information. Users receive immediate safety updates through intuitive interfaces using a retrieval-augmented generation system. To demonstrate the ICA's viability, a test case of shared e-scooters is studied, showcasing the system's effectiveness and potential impact on urban scooter rideshare safety. This test case highlights how the GAI-based ICA can enhance the safety and reliability of urban rideshare services by providing timely and accurate safety information and recommendations.

Keywords

Generative Artificial Intelligence, Large Language Models, Retrieval-Augmented Generation, Knowledge Graphs, Urban Rideshare Safety

1. Introduction

The rideshare system is widely used in urban environments, providing convenience for daily public life and representing a significant market opportunity. Service providers such as Citi Bike, Lime, and Bird quickly achieved a strong service-market fit, reaching billion-dollar valuations shortly after their inception (Ma et al. 2021). Ensuring the safety of riders is paramount for both the users and the rideshare companies. Much research has focused on this area, examining data analysis aspects (Zakhem and Colin 2024) and policy regulation analysis (Riggs et al. 2021). However, practical methods for real-life applications to monitor road conditions and provide safety suggestions are still limited.

With recent advancements in computer vision technology, numerous methods have been developed to monitor road conditions. For instance, convolutional neural networks (CNNs) can be trained on pavement images to create deep learning models that provide detailed information on road conditions (Zhang et al. 2016). Mobile device cameras, being highly accessible and convenient, can be combined with computer vision technologies to become powerful and easy-to-use tools for daily applications. Modern computer vision technology has leveraged data from simple sensors in cell phones, employing methods such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVMs) for model training (Ashqar et al. 2020). Therefore, with the help of computer vision technologies and modern mobile device sensors, road safety conditions and traffic information can be effectively gathered and analyzed. However,

there is a lack of a comprehensive assistant that can monitor and store data for users, automatically analyze it, and provide actionable suggestions for data analysts.

In the field of vision research, large language models (LLMs) are receiving significant attention. The GPT-Vision model creates a link between image data and human language input, enabling the generation of related images. Several applications have emerged in this research area. Large vision models can be used for contextual emotion recognition through fine-tuned models, which can help enhance emotionally sensitive decision-making in robots and agents (Etesam et al. 2024). GPT-4V functions as a traffic assistant by recognizing and analyzing complex traffic events from video footage (Zhou and Knoll, 2024). Instead of solely using computer vision technologies or just vision models, GPT-Vision can be used as a validation process for the results inferred by computer vision. One limitation of using CNNs alone is that they may heavily rely on image quality. Using a large language model to validate low-confidence classifications may enhance the accuracy of the results. This paper also tests the plausibility of this method.

The Retrieval-Augmented Generation (RAG) is a crucial technique that connects chat models with information sources in databases or various third-party APIs. This technique allows users to directly assess the accuracy of large language models in answering specific questions. RAG typically involves an embedding algorithm that embeds specific information (e.g., vectors) within the data. By identifying relationships between these data points (e.g., the distance between two vectors), the chat model can infer results similar to previously stored data. This process heavily relies on the database. The knowledge graph database tool, Neo4j (Neo4j, https://www.neo4j.com) aids researchers in designing domain knowledge databases that are easily accessible to large language models and provides extensive libraries for building the RAG system. The ability to visualize data also helps make the database clear and intuitive.

The impact of GAI on supply chain management has been significant, with research highlighting a focus on conceptual studies and the need for more analytical models and empirical research, particularly in the delivery process and food and manufacturing supply chains (Ben-Daya et al. 2017). Addressing these issues, this paper proposes integrating the YOLO detection model with physical sensors in mobile devices to provide data for a real-time updating generative intelligent cognitive assistant (GenICA) platform. During data input, GPT-Vision models are used to validate the confidence of the YOLO image output. The data is stored in a knowledge graph database and embedded with vectors. A chat model is created to generate safety information based on data retrieved from the database. A RAG system specifically designed to answer domain knowledge of road safety information is built. The subsequent sections of this paper are structured as follows: Section 2 provides a review of related work, Section 3 outlines the framework of the GenICA platform and functional analysis of the system, and Sections 4 and 5 detail the construction of the case study simulation and the development of the backup algorithms and the knowledge graph database. Section 6 discusses the urban rideshare system interface, while Section 7 presents the conclusion and directions for future work.

2. Related Work

2.1 Safety Recommendation Systems

The safety problem of running the rideshare businesses addresses concern from both the companies and the government. Researchers have also put huge efforts into policy analysis and data analysis. The main safety issues include riding on the sidewalk, road defects like cracks and uneven edges (Mura et al. 2022). Largely investigated in computer vision research, the road information now can be detected through convolutional neural networks. YOLO detection is widely used in computer vision, providing abundant possibilities for various applications. Previous research has shown that the YOLO model is well-suited for road information detection. For instance, the RD-YOLO model, an improved version of YOLOv5, incorporates a generalized feature pyramid network (GFPN), coordinate attention (CA) mechanism, and Focal-EIOU Loss to enhance detection accuracy and efficiency in roadside perception systems (Huang and Huang 2022). Additionally, an improved YOLO network, named YOLO-R, is designed for pedestrian detection by adding three Passthrough layers and adjusting the network structure to enhance detection accuracy, reduce false detection rates, and achieve real-time performance at 25 frames per second (Lan et al. 2018). Beyond road information, the YOLO model can also detect vehicles on the road. The YOLO v5-Ghost model, an improved version of YOLO v5s, enhances detection speed and efficiency using Ghost modules to replace conventional convolution layers, making it suitable for embedded devices while maintaining high detection accuracy in a virtual environment (Wu et al. 2021). Therefore, with the support of the YOLO detection model, the proposed GAI-based platform is robust enough to ensure the system's effective operation.

URS recommendation services is a concept that connects physical devices and vehicles with cloud data, and gives feedback based on the data for end users. Research has also extensively explored its application in supply chain

management and express delivery. For instance, a GAI-based dynamic food supply chain management system incorporating the Bee Colony algorithm ensures food quality and fast delivery while providing real-time tracking and tracing of contaminated food products within the supply chain (Nagarajan et al. 2022). However, rarely does research focus on connecting the last mile of the supply chain, specifically the express delivery process, with the rideshare scooter system. To address the issue of insufficient information on small roads, building a GAI-based platform is a necessary and natural solution for building and deployment. Therefore, the proposed GenICA came to be a plausible idea.

2.2 GPT-Vision

Large language models are remarkable products of natural language processing technology, developed with the aid of big data. Trained on millions of datasets, these models are now capable of performing a wide range of tasks requested by users. While many of these models are used for text generation, processing image information has also been extensively investigated. The GPT-Vision model is one such model, trained to generate images based on text input. Tom et al. (2024) evaluated the effectiveness of GPT-4V in predicting human-assessed risk levels in traffic images by correlating its outputs with human ratings. Their study demonstrated that repeated and varied prompts improve predictive validity, highlighting the potential of GPT-4V in automated driving contexts. Although vision models are commonly used to generate new images, few studies have investigated using these tools as a validation process for the output of vision algorithms. Therefore, this paper explores the idea of using the GPT-Vision model as an enabling tool to validate the classifications made by YOLO models.

2.3 Knowledge Graph and RAG System

Knowledge graphs employ methods such as rule-based reasoning, which utilizes logical rules and ontologies to infer new knowledge, and representation-based reasoning, which uses distributed representations and neural networks to capture complex patterns and relationships within the graph (Chen et al. 2020). This underscores the knowledge graph's robust capability to store and analyze domain-specific information, traditionally reliant on human cognition.
With the advent of large language models, this traditional reliance on human cognition can now be supplemented and enhanced through retrieval augmentation generation (RAG). RAG systems leverage external data sources and APIs to enrich and expand their own databases. Examples such as BiomedRAG (Li et al. 2024) and RAGTAG (Arora et al. 2024) demonstrate the versatile applications of RAG systems in domains such as biomedicine and industrial sectors, highlighting their ability to enhance knowledge retrieval and generation across diverse fields.

2.4 Intelligent Cognitive Assistant

The concept of an Intelligent Cognitive Assistant (Gong et al. 2021) highlights the shift to crowdsourced manufacturing through cyber platforms and intelligent cognitive assistants, enabling manufacturing as a service with a focus on mass customization, open manufacturing, and predictive adaptation of processes. While primarily applied in industrial settings, this concept can be adapted for urban rideshare systems. In this context, an intelligent cognitive assistant can provide safety recommendations, supported by generative algorithms and Neo4j knowledge graph databases. By leveraging advanced data processing and AI methodologies, this system can analyze real-time data from sensors and mobile devices to offer tailored safety advice and enhance the overall safety of urban rideshare web services.

2.5 Research Gaps and Motivations

Overall, four key research gaps are identified to motivate the development of an AGI-based platform for road safety information analysis: (a) Real-World Road Information is Strongly Time-Related: The complexity and diversity of real-world road situations are difficult to predict with a set of assumed expectations; (b) Powerful Computer Vision Tools, GPT-Vision Models, and Easily Reachable Mobile Device Sensors: The robust YOLO detection algorithms and GPT-Vision models can serve as the foundation for the GAI-based platform to gather and validate road data. With sensors on mobile devices, data can be efficiently uploaded to the cloud server and processed immediately to provide timely information to users; (3) Clear Knowledge Graph Database and Simple RAG Construction: Neo4j offers a convenient environment to manage and store data. It also provides access to various large language models and retrieval tools for effective data retrieval; and (d) An Innovative ICA Concept: The proposed GenICA incorporates large language models and knowledge graphs, utilizing generative algorithms to provide safety information without involving any decision-making processes.

By highlighting these considerations, this paper proposes a GenICA platform to facilitate the gathering of road information using rideshare scooters. By deploying the YOLO model powered by GPT-Vision models and utilizing sensors in mobile devices, the knowledge graph and RAG system can generate safety information for users.

3. System Analysis and Design for GenICA

In this section, we delve into the framework and functional analysis of the proposed GenICA. We will outline the architecture and elaborate on the workflow of the platform, highlighting the integration of computer vision, large language models, knowledge graph database that domain road safety analysis knowledge and the road information RAG system for providing the real time updating safety information.

3.1 The Generative ICA Framework

The GenICA proposed in this paper is designed to be adaptable to a wide range of industrial applications beyond the rideshare scooter system. To demonstrate the comprehensive benefits of this platform, a structured framework is essential, encompassing all three layers, as illustrated in Figure 1.

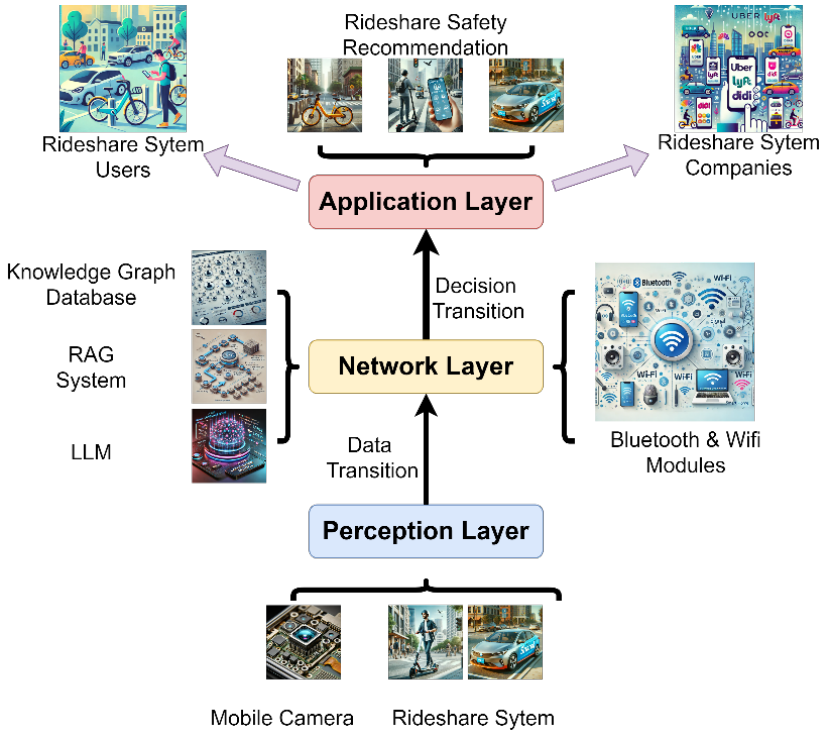

Figure 1. Real-time rideshare riding safety recommendations based on GAI

In the perception layer, which forms the foundational layer of the GenICA framework, all physical devices and sensors gather information. In this paper, these include sensors embedded in mobile devices and electric rideshare scooters. This layer is crucial as it collects raw data from various sources such as GPS data, accelerometer readings, and environmental sensors, which are essential for understanding real-time conditions. The network layer serves as the intermediary, responsible for transmitting all information gathered from the perception layer to cloud servers. This layer includes components such as the knowledge graph database storing road information and relationships, large language models (LLMs) supporting query generation and YOLO model results consolidation, and the retrieval-augmented generation (RAG) system for data retrieval and final suggestions. The network layer not only handles data transmission but also stores and analyzes data for further application. In this paper, gathered information includes road conditions, traffic flows, and other contextual data. Analyzed outcomes, such as safety suggestions and optimized traffic management strategies, are critical outputs of this layer, ensuring data integrity and real-time processing. The hardware of this layer includes Bluetooth and Wi-Fi modules.

The application layer is the topmost layer, encompassing user interfaces and communications. It enables real-time updates about road safety, traffic conditions, and personalized service recommendations for end users. For businesses and enterprises, it involves monitoring data trends, operational efficiency, and customer satisfaction metrics. Unlike traditional approaches, GenICA integrates both types of end users—service users and providers—within each component of the application phase. By integrating these layers effectively, the proposed GenICA ensures seamless data flow, processing, and actionable insights tailored to various stakeholders. This multi-layered approach facilitates a holistic view of system operations, driving improvements in safety, efficiency, and user satisfaction across different industrial applications.

3.2 Functional Analysis of the GenICA

The functional analysis of the GenICA is depicted in a diagram (IDEF0) as shown in Figure 2. Initially, image processing (A1) is a process that is established by the visual data (I1) gathered from mobile camera sensors during the riding process of e-scooter users. The data is classified by the pretrained YOLO model (C1) and validated by the GPT-vision model (C2). The processed and consolidated visual data involves all the road information including the defects and whether it is on the road or sidewalk.

The data abstraction from A1, along with the pretrained image data that used to train the YOLO (I2) and the entity relationships (C3), will be categorized in the construction of knowledge graph database (A2). Out from the A2, the query retrieved information is then delivered to the request from information retrieving and feedback process (A3). During this phase, generative algorithms (M2) and vector embedding algorithms (M1) will play an important part in building the connection between safety problem entities and the real time images. With the RAG (C4), hallucinations will be avoided. Finally, a chat agent (C5) is built, used to give feedback in human language in the road safety suggestion interface (A4) to give users the final safety suggestions.

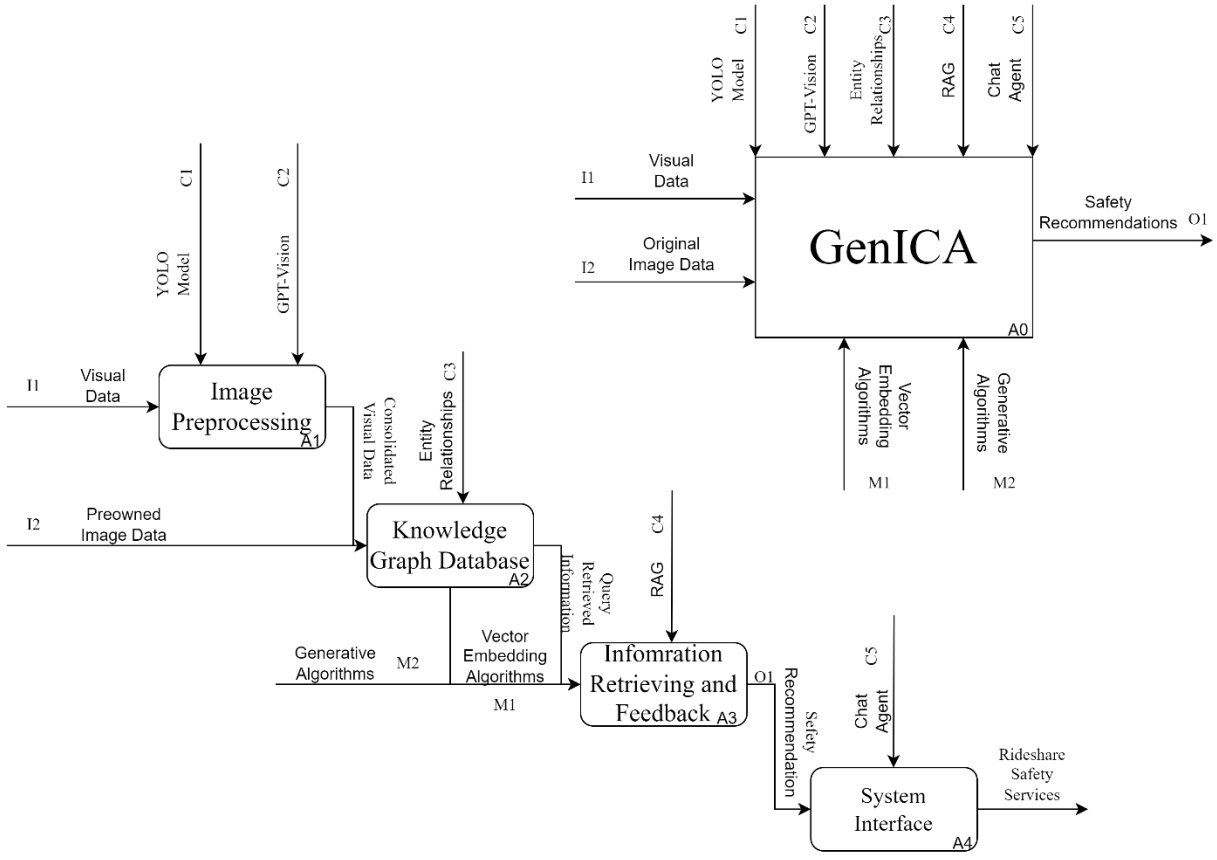

Figure 2. GenICA functional model depicted in IDEF0 diagrams

3.3 RAG System Architecture and Knowledge Graph Database

A knowledge graph is used to represent relationships between different knowledge bases. In a knowledge graph, nodes represent the knowledge bases, and edges represent the relationships between them. These relationships can be one-to-many, many-to-one, or one-to-one, effectively visualizing the connections and logic between them. Traditional methods of constructing knowledge graphs often face challenges with data management. Neo4j stands out by addressing these challenges, providing robust database management and representing relationships between nodes in a graphical format. Figure 3 illustrates the architecture of a Neo4j knowledge graph database. The construction of the knowledge graph begins at level one, where all relationships and domain knowledge require initial human input to establish a foundational structure. At level two, the database connects with large language models to enrich each node with specific information and support model understanding and training. Finally, at level three, the database gains the capability to generate new information based on the trained domain knowledge database, facilitated by generative algorithms.

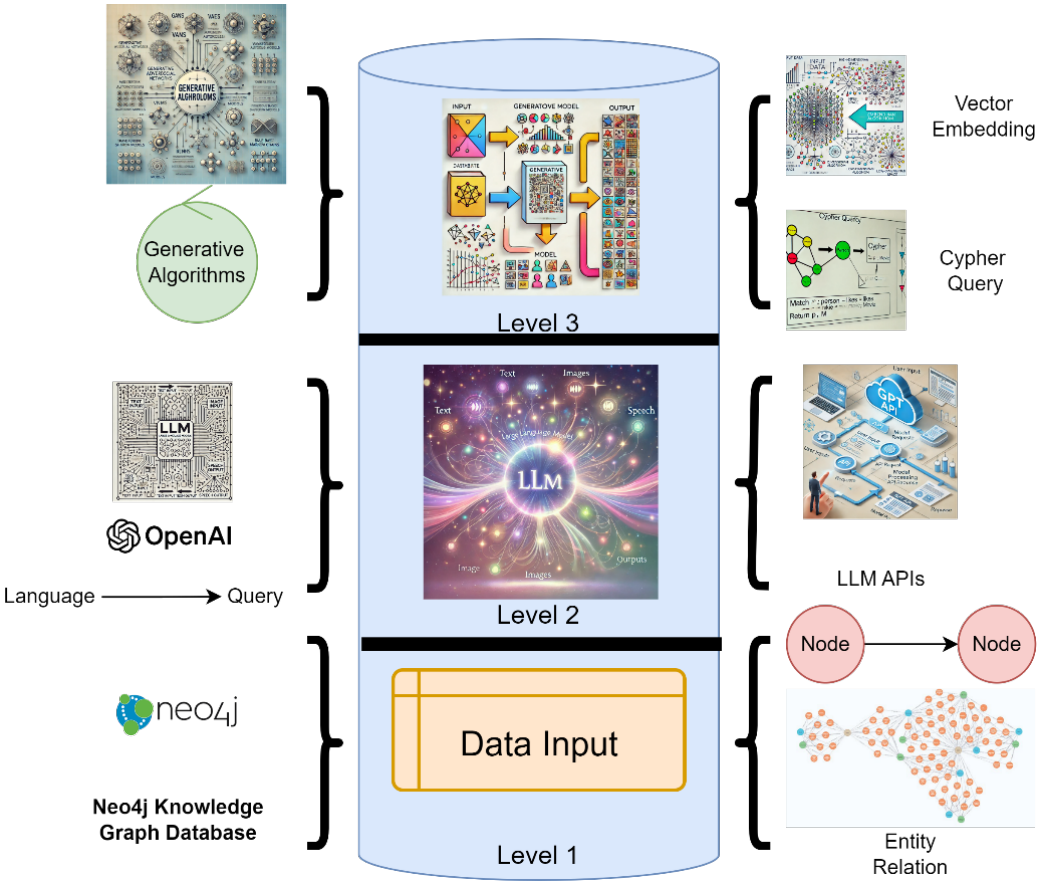

Figure 3. RAG system implementation deploying a knowledge graph database

The RAG system plays a crucial role in generating appropriate responses directly to users. It consists of three main components: understanding queries, retrieving information, and providing human understandable feedback. The function of understanding queries typically involves a text-prompting language model, such as the GPT-3.5-instruct model, to comprehend human language queries. Retrieving information is complex as it involves generating queries based on human language to retrieve data from the database. The retrieval process utilizes generative algorithms and vector embedding techniques to access knowledge stored in nodes. The Langchain library ensures accuracy in retrieval and query generation. The function of giving human understandable feedback part connects to a chat agent and utilizes the retrieved information to generate human language responses. The workflow of the RAG system is illustrated in Figure 4, depicting how these components interact to provide effective communication and information retrieval for users (Figure 5).

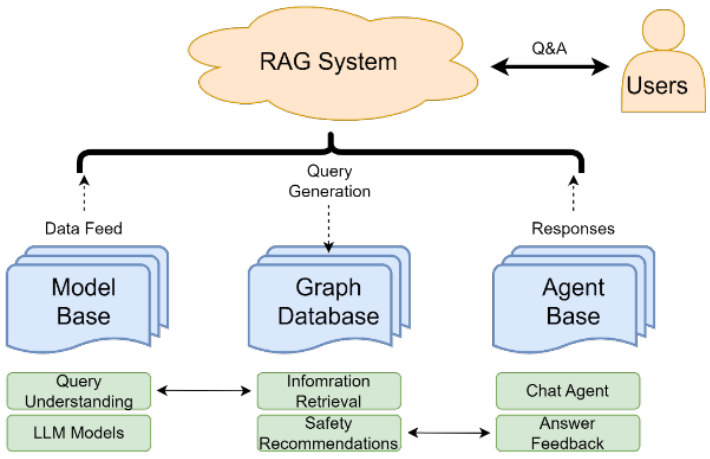

Figure 4. RAG system architecture

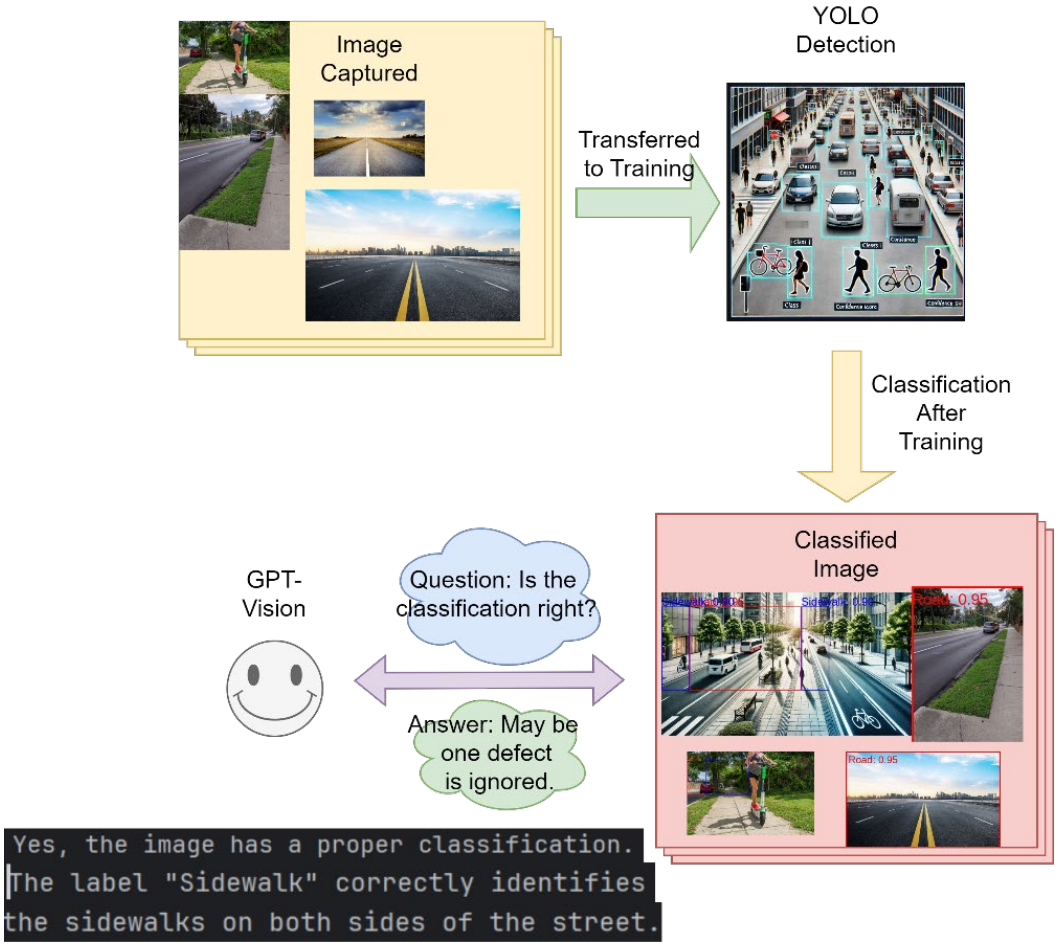

Figure 5. GPT-Vision powered YOLO detection procedure

4. GPT-Vision Powered YOLO Model

The YOLO (You Only Look Once) model has been extensively applied across various research areas. This technique is proficient in generating road information, such as classifying whether an object is a road or sidewalk, or detecting defects on the road. However, the YOLO model may encounter situations where it lacks confidence in its classifications, particularly in specific frames where it provides low confidence scores.

To mitigate these issues, leveraging a GPT-vision model to validate the results of the YOLO model can be beneficial. The GPT-vision model can help in verifying and enhancing the confidence of classifications made by the YOLO model, thereby reducing instances of low-confidence classifications. Additionally, integrating such validation processes can help prevent database pollution by ensuring that only accurate and reliable information enters the database, enhancing overall data quality and reliability. Figure 5 depicts the procedure of the GPT-vision powered YOLO model.

The YOLO model is trained to detect two types of information: the presence of roads and sidewalks, and the detection of defects such as potholes and cracks. This information forms the basis of the safety knowledge to be stored in the subsequent database construction. To ensure the accuracy of classifications, the GPT-vision model is employed to answer inbuilt questions that verify whether the images are correctly classified. If the images are properly classified, they are directly transferred to the database for further analysis. If not, the images are reprocessed by the GPT-vision model to add correct labels before being transferred to the database. This process ensures that only accurately classified images are stored, enhancing the quality and reliability of the database.

5. Rideshare Safety Recommendation Knowledge Graph Database

Road safety information in this paper considers two aspects: defects on the road and the proper identification of roads or sidewalks. Consequently, the safety suggestions outputted in the final interfaces will be closely related to these two pieces of information. This section provides details on building the relationship between the nodes in the database and the conversion from image data to node data. It also introduces the embedding algorithm used in this process, explaining its utility and underlying principles. Figure 6 gives a detailed description of the knowledge construction using Neo4j (Figure 6).

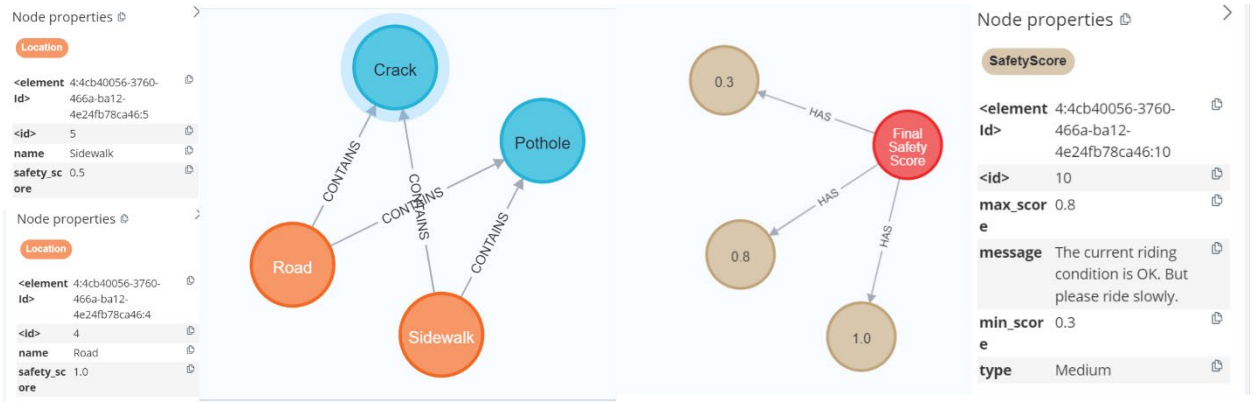

Figure 6. Safety recommendation knowledge graph database using Neo4j

5.1 Modeling Knowledge Node and Links

The nodes in our GenICA system represent safety conditions, with two main types: those for the chosen roads and those for road defects. Nodes representing roads and sidewalks include the "road" node, which has a safety score of 1, and the "sidewalk" node, which has a safety score of 0.5. Defects, such as potholes and cracks, are represented as sub-nodes linked to the "roads chosen" nodes with a "contains" relationship. Each defect node has a safety score of -0.1.

The system also includes final safety score nodes, which aggregate the safety scores from the roads and defect nodes to provide an overall safety assessment. These nodes are categorized into three types: low safety score nodes, with scores between 0 to 0.3, median safety score nodes, with scores between 0.3 to 0.8, and high safety score nodes, with scores between 0.8 to 1.0.

For example, the low safety score node suggests: "Please be careful about the current conditions. Ride slowly or choose another road." The median safety score node advises: "The current riding condition is OK. But please ride slowly." The high safety score node reassures: "Your current riding condition is very safe. Enjoy your ride."

This structured approach ensures that safety conditions are clearly represented and categorized, providing users with actionable safety suggestions based on the aggregated safety scores.

5.2 Image information Conversion

While the input to the database is image data, the data stored must be in text form, necessitating the conversion of image data into text. This conversion is performed by the GPT-4.0 model, which first categorizes the image as either a road or sidewalk. It then identifies and describes any defects, such as cracks and potholes, present on the road or sidewalk. The model generates a CSV file that includes the riding path name (road or sidewalk), the number of cracks, and the number of potholes. These details are used to calculate the final safety score. Based on this final score, users can retrieve safety suggestions by referencing the corresponding score ranges in the database. Figure 7 gives an example of the image input and out as a csv file.

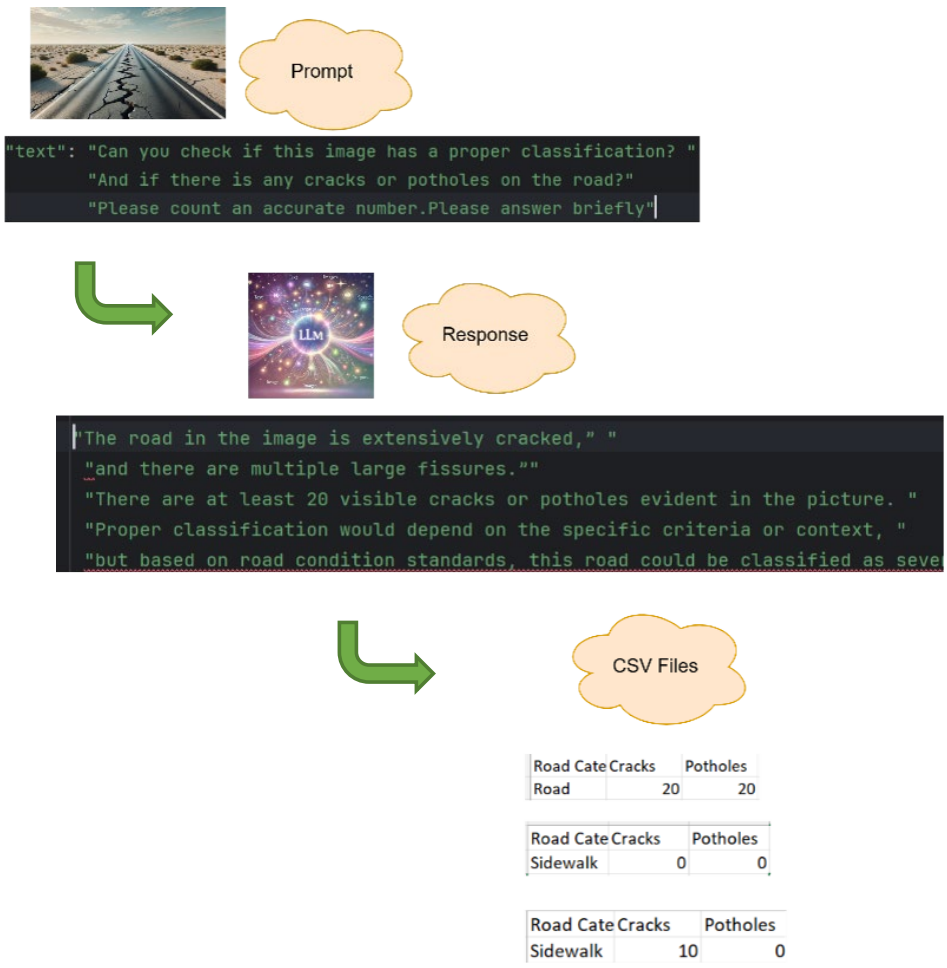

Figure 7. Procedure of image conversion to a Csv file

5.2 Image information Conversion

While the input to the database is image data, the data stored must be in text form, necessitating the conversion of image data into text. This conversion is performed by the GPT-4.0 model, which first categorizes the image as either a road or sidewalk. It then identifies and describes any defects, such as cracks and potholes, present on the road or

sidewalk. The model generates a CSV file that includes the riding path name (road or sidewalk), the number of cracks, and the number of potholes. These details are used to calculate the final safety score. Based on this final score, users can retrieve safety suggestions by referencing the corresponding score ranges in the database. Figure 7 gives an example of the image input and out as a csv file.

5.3 Cypher Query

After the image information is saved as a CSV file, the Cypher query tool is used to generate queries that retrieve data from the database. For example, given a specific condition such as "road with 5 cracks," the database provides feedback on these specific nodes and their final scores. Using the calculated final score, the database identifies the corresponding final score node to provide safety suggestions. To ensure accurate data retrieval, specific prompt information is crucial for Cypher queries. In this paper, the Cypher prompt template is as follows:

> You are a skilled Neo4j Developer. Your task is to create Cypher queries that answer user questions about safety information.
> Make sure to focus on extracting key details such as safety_score, max_score, min_score, name.
> Consider the relationships and properties provided in the schema below.
> Fine Tuning:
> The relationships query should be like: (f:FinalSafetyScore)-[:HAS]->(s:SafetyScore).
> The relationships query should be like: (r:Road)-[:CONTAINS]->(d:Defect).
> The relationships query should be like: (s:Sidewalk)-[:CONTAINS]->(d:Defect).
> This template outlines the specific database structure and is highly effective for generating Cypher queries that retrieve safety suggestions based on image inputs.

Figure 8 shows an example of the query generation process using the above cypher query.

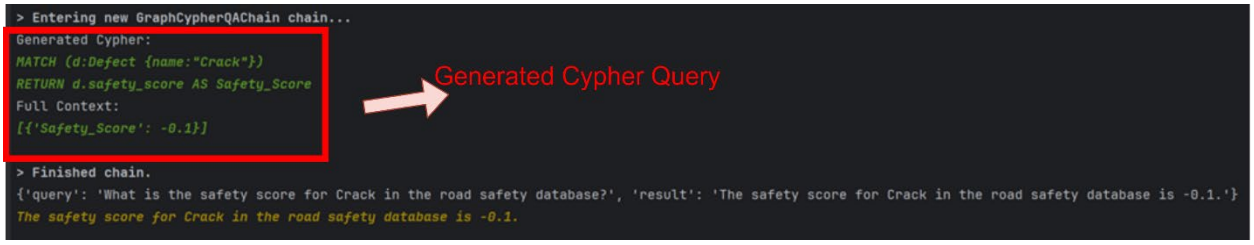

Figure 8. Cypher query process

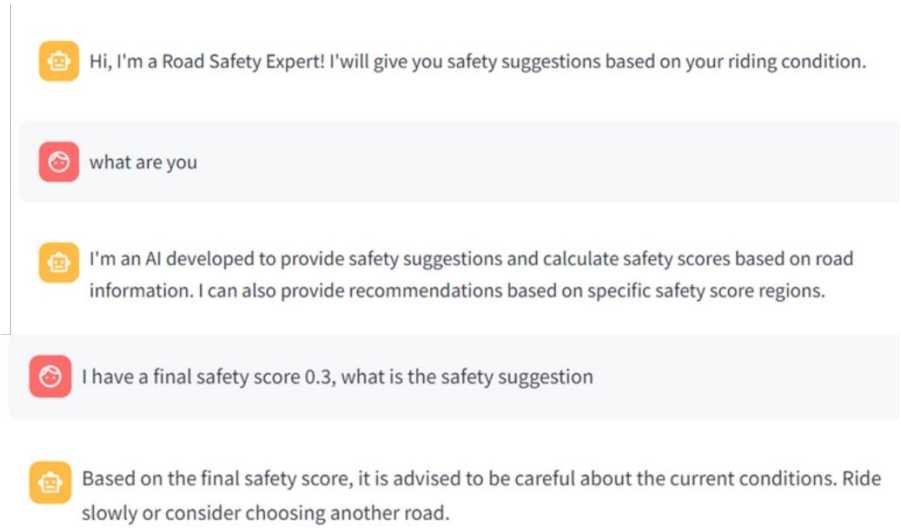

Figure 9. User interface of the Chatbot generating a real-time rideshare safety recommendation web service

6. Rideshare Safety Recommendation Chatbot Interface

After setting up the database and RAG system, a final chatbot interface is essential for providing user feedback. The chatbot retrieves safety information from the database and offers safety suggestions based on the calculated safety score. Once a real-time image is provided and the safety score is calculated, the chatbot automatically generates and displays the safety suggestions to the user. This interface ensures users receive timely and relevant safety information in an easily accessible manner. Figure 9 shows an example of the user interface.

7. Conclusions

This paper proposes the GenICA as a new concept designed to generate road safety information and test its feasibility. The GenICA, embedded with a knowledge graph database and generative algorithms for information retrieval, has been examined and compared to traditional online recommendation systems. The GenICA provides users with enhanced situational awareness and serves as a reference for self-generation. It offers deeper insights into data features for designers monitoring the system, highlighting the generative capabilities of the model. The GenICA allows for the generative application of industrial knowledge, showcasing its potential to improve road safety and provide valuable information to both users and designers.

However, there are limitations to this approach. One significant issue is hallucination, where large language models generate incorrect or misleading answers. This can result in significantly wrong advice being provided during riding, even when performing the same function or building the same model. To mitigate this, more specific knowledge should be generalized and standardized to minimize discrepancies. Additionally, real-life environmental conditions are crucial to consider, as real traffic is more complex and can dramatically affect the analysis process, potentially causing instability. Therefore, future work should focus on training the model with more specific knowledge and refining the learning process to enhance its reliability and accuracy.

References

Arora, C., Herda, T., and Homm, V., Generating Test Scenarios from NL Requirements using Retrieval-Augmented LLMs: An Industrial Study. *arXiv preprint arXiv:* 2404.12772, 2024.

Ashqar, H. I., Almannaa, M. H., Elhenawy, M., and Rakha, H. A., & House, L., Smartphone transportation mode recognition using a hierarchical machine learning classifier and pooled features from time and frequency domains. *IEEE Transactions on Intelligent Transportation Systems*, vol.20(1), pp.244–252. ,2019.

Ben-Daya, M., Hassini, E., and Bahroun, Z., Internet of things and supply chain management: a literature review. *International Journal of Production Research*, vol.57(15–16), pp. 4719–4742, 2017.

Chen, X., Jia, S. and Yang, X., A review: Knowledge reasoning over knowledge graph, *Expert Systems with Applications*, vol.141, pp.112948, 2020.

Driessen, T., Dodou, D., Bazilinskyy, P., and de Winter, J., Putting ChatGPT vision (GPT-4V) to the test: Risk perception in traffic images. *Royal Society Open Science*, vol.11. pp.231676, 2024.

della Mura, M., Failla, S., Gori, N., Micucci, A., and Paganelli, F. E-Scooter Presence in Urban Areas: Are Consistent Rules, Paying Attention and Smooth Infrastructure Enough for Safety? *Sustainability*, vol.14, pp.14303, 2022.

Etesam, Y., Yalçın, Ö. N., Zhang, C., and Lim, A., Contextual Emotion Recognition using Large Vision Language Models, *arXiv preprint, arXiv*: 2405.08992, 2024.

Gong, X., Jiao, R., and Jariwala, A., and Morkos, B., Crowdsourced manufacturing cyber platform and intelligent cognitive assistants for delivery of manufacturing as a service: Fundamental issues and outlook. *International Journal of Advanced Manufacturing Technology,* vol. 117, pp. 1997–2007, 2021.

Huang, L., and Huang, W., RD-YOLO: An Effective and Efficient Object Detector for Roadside Perception System. *Sensors (Basel),* vol.22, no.21, 8097, 2022.

Lan, W., Dang, J., Wang, Y., and Wang, S., "Pedestrian Detection Based on YOLO Network Model," *IEEE International Conference on Mechatronics and Automation*, Changchun, China, pp. 1547-1551, 2018.

Li, M., Kilicoglu, H., Xu, H., and Zhang, R., BiomedRAG: A Retrieval Augmented Large Language Model for Biomedicine. *arXiv preprint, arXiv*:2405.00465, 2024.

Ma, Q., Yang, H., Mayhue, A., Sun, Y., Huang, Z., and Ma, Y., E-Scooter safety: The riding risk analysis based on mobile sensing data. *Accident Analysis & Prevention*, vol.151, pp. 105954, 2021.

Nagarajan, S. M., Deverajan, G. G., Chatterjee, P., Alnumay, W., and Muthukumaran, Integration of IoT based routing process for food supply chain management in sustainable smart cities. *Sustainable Cities and Society*, vol. 76, pp. 103448, 2022.

Riggs, W., Kawashima, M., and Batstone, D., Exploring best practice for municipal e-scooter policy in the United States. *Transportation Research Part A: Policy and Practice*, vol. 151, pp. 18-27, 2021.

Wu, T., Wang, T., and Liu, Y., "Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network," *The 3rd World Symposium on Artificial Intelligence*, Guangzhou, China, 2021, pp. 24-28, 2021.

Zakhem, M., Smith-Colin,An, J., E-scooter route assignment framework to improve user safety, comfort and compliance with city rules and regulations. *Transportation Research Part A: Policy and Practice*, vol 179, pp. 103930, 2024.

Zhang, L., Yang, F., Daniel Zhang, Y., and Zhu, Y. J., Road crack detection using deep convolutional neural network., *IEEE International Conference on Image Processing*, Phoenix, AZ, USA, pp. 3708-3712, 2016.

Zhou, X. and Knoll, A.C., GPT-4V as Traffic Assistant: An In-depth Look at Vision Language Model on Complex Traffic Events. *arXiv preprint, arXiv*:2402.02205, 2024.

Biographies

Chenxi Tao is currently a graduate student in School of Mechanical Engineering at Georgia Institute of Technology, USA. His research interests include large language models for industrial applications, artificial intelligence, robot learning, design optimization, operation planning, decision making and data analysis. He has recently graduated from Columbia University in New York City, USA with M.S. in Mechanical Engineering. He holds B.Eng. in Mechanical Engineering from Shanghai Maritime University, China.

Roosan Liyons has recently graduated from Georgia Institute of Technology, USA with M.S. in Mechanical Engineering. His research interests include decision making, industrial AI solutions, machine learning, data analytics, and advanced manufacturing systems. He holds B.Eng. in Mechanical Engineering from Georgia Institute of Technology, USA. He currently works in industry as an R&D engineer. More info about him: https://www.linkedin.com/in/rliyons.

Jianxin Jiao is the editor-in-chief of Journal of Engineering Design and an associate professor of mechanical and industrial engineering at Georgia Tech, USA. Prior to joining the School of Mechanical Engineering at Georgia Tech in December 2008, he was an Assistant Professor and then Associate Professor in the School of Mechanical and Aerospace Engineering at Nanyang Technological University, Singapore. Before his career in Singapore, he was a Visiting Scholar in the Department of Industrial Engineering and Engineering Management at Hong Kong University of Science and Technology from 1998 to 1999. From 1993 to 1994, he was a Lecturer of Industrial Engineering in the School of Management at Tianjin University, China, and from 1988 to 1990, he worked as an Associate Lecturer in the Department of Industrial Design at Tianjin University of Science and Technology, China. More info about his research: https://scholar.google.com/citations?user=9yikEHAAAAAJ&hl=en&oi=ao.

Seung-Kyum Choi directly began at Georgia Tech in Fall 2006 as an Assistant Professor. Prior to joining Georgia Tech, he was a research assistant at Wright State University, conducting research on uncertainty quantification techniques for the analytical certification of complex engineered systems. Dr. Choi's research interests include structural reliability, probabilistic mechanics, statistical approaches to the design of structural systems, multidisciplinary design optimization, and the decision support process for complex engineered systems. At Georgia Tech, he pursues challenging research issues to quantify uncertainty for the analytical certification of practical engineering systems. The results of this effort include the development of state-of-the-art numerical techniques in statistical methods, and structural analysis methods that incorporate innovative uncertainty quantification techniques.