KG-A-VLM: Knowledge Graph-Augmented Vision-Language Models for Patient Vital Assessment in Triage Robotics

Rutvik N. Joshi*
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
`rutvikj@andrew.cmu.edu`

John Galeotti
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
`jgaleott@andrew.cmu.edu`

Abstract

Triage robotics systems must rapidly and accurately assess patient vitals to determine urgency levels. However, vision-language models (VLMs) alone – even powerful ones like DINOv2 – often struggle with domain-specific knowledge and may hallucinate or misinterpret critical cues. We propose a framework that enhances VLMs with Knowledge Graphs (KGs) and Retrieval-Augmented Generation (RAG) to improve patient vital assessment. In our approach, a medical knowledge graph representing patients, symptoms, and observations is tightly integrated with a VLM via a RAG pipeline. Relevant structured facts about vital signs and symptoms are retrieved from the KG and encoded with a graph neural network (GNN), then used to augment the VLM's reasoning. This KGRAG (Knowledge Graph + RAG) strategy injects up-to-date medical context into the model's decision-making, reducing errors and improving interpretability. We outline the proposed K-RAGRec-style architecture – including adaptive KG retrieval, subgraph re-ranking, and knowledge-conditioned generation – tailored for triage scenarios. Experiments and results will be detailed in future work, but our methodology aims to demonstrate that combining visual analysis with structured knowledge can significantly advance autonomous triage assessment.

1 Introduction

In emergency care, triage is a critical process that involves gathering initial patient information, measuring vital signs, understanding symptoms, medical history, and assigning an urgency level to prioritize treatment. (1). Conventional triage relies on human expertise and judgment, which can be subjective and error-prone. Recent studies have shown growing interest in leveraging artificial intelligence (AI) to assist triage, using patient data (vital signs, symptoms, history) to accurately classify patients into risk categories. AI-driven triage systems have even demonstrated higher accuracy than traditional methods in some cases. Robotic triage assistants could further enhance this process by performing contact-free vital monitoring (e.g., via camera sensors) and reducing healthcare workers' exposure to hazards.

Vision-Language Models (VLMs) offer a promising foundation for triage robotics because they can interpret visual inputs and produce descriptive or analytical text. Models such as CLIP and DINOv2 learn rich visual representations from large-scale data, achieving all-purpose feature extraction across diverse images. For example, DINOv2 is a state-of-the-art vision transformer that produces robust

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

features without supervision (2), which could be applied to patient imagery (e.g. facial temperature cues, respiratory motions). However, domain-specific understanding remains a challenge: a general VLM-LLM may lack medical knowledge needed to interpret abnormal vital signs or subtle symptoms. Large language models can hallucinate or produce plausible-sounding but incorrect assessments when confronted with unfamiliar scenarios due to the inherent limitations of LLMs(3). Ensuring accurate and up-to-date medical reasoning is essential for patient safety, explainability and to build trust in VLM generation.

To overcome these limitations, we propose to augment VLM-based triage systems with external medical knowledge. In particular, we integrate a Knowledge Graph (KG) capturing relationships between patients, symptoms, and observations (vital signs) into the VLM's inference process. Knowledge graphs are structured networks of medical facts that can provide context (e.g., which combinations of symptoms indicate high risk, normal ranges for vitals given patient history). By coupling the VLM with a KG, we enable the system to cross-reference visual observations with established medical knowledge, reducing the reliance on the model's parametric memory alone.

Our approach adopts a Retrieval-Augmented Generation (RAG) paradigm to inject relevant knowledge into the model on-the-fly. RAG has recently gained traction as a solution for hallucination and knowledge update issues in LLMs (3). It works by retrieving pertinent external information (traditionally text documents) based on the input, and conditioning the model's output on this retrieved content. Here, we extend RAG to operate over a structured knowledge graph. By retrieving subgraphs of medical facts rather than free text, we preserve the structural relationships in the knowledge (which vanilla RAG might neglect). This structured retrieval is crucial in medicine, where the context (e.g., temporal trends in vitals, symptom co-occurrence) affects interpretation.

In summary, our contributions are as follows:

- Knowledge Graph-Integrated VLM Architecture: We propose a novel triage system architecture that tightly integrates a vision-language model with a medical knowledge graph via retrieval-augmented generation. To our knowledge, this is the first application of KG-augmented VLMs for autonomous patient vital assessment in robotics.
- Adaptive Knowledge Retrieval Mechanism: Building on the K-RAGRec framework, we develop an adaptive retrieval policy to fetch the most relevant portion of the knowledge graph for a given patient's data. This mechanism indexes and ranks subgraphs of the KG in real-time, providing the VLM with focused, contextual medical knowledge while filtering out noise.
- Graph Neural Network Encoding of Medical Subgraphs: We employ a GNN encoder to transform the retrieved subgraph (containing patient symptoms and observations) into a rich embedding that can interface with the VLM's language modality. This allows the model to incorporate structured medical facts (as a "graph prompt") into its reasoning process, improving interpretability and accuracy.
- Triage Knowledge Graph Design: We introduce a specialized knowledge graph schema for the triage domain, defining node types (Patient, Symptom, Observation) and their relationships (e.g., patient-observation, observation-symptom edges), including the temporal dimension of patient vitals. This schema enables capturing each patient's history and the general medical knowledge in a unified graph structure.

The rest of the paper is organized as follows: Section Related Work reviews prior efforts in VLMs, knowledge graphs, and retrieval augmentation in healthcare. Section Background provides necessary context on vision-language models, knowledge graphs, and RAG. In Section Proposed Method, we detail our KGRAG framework, including the system architecture and KG schema. Sections for Experiments, Results, and Discussion are placeholders for future work. Finally, Conclusion and Future Work outlines the implications of our approach and avenues for further development.

2 Literature Review and Background

Recent advances in AI systems have explored the integration of *structured knowledge representations*—such as Knowledge Graphs (KGs)—with perception modules and language models to enhance decision-making in real-world domains. This section summarizes key prior work relevant to our approach.

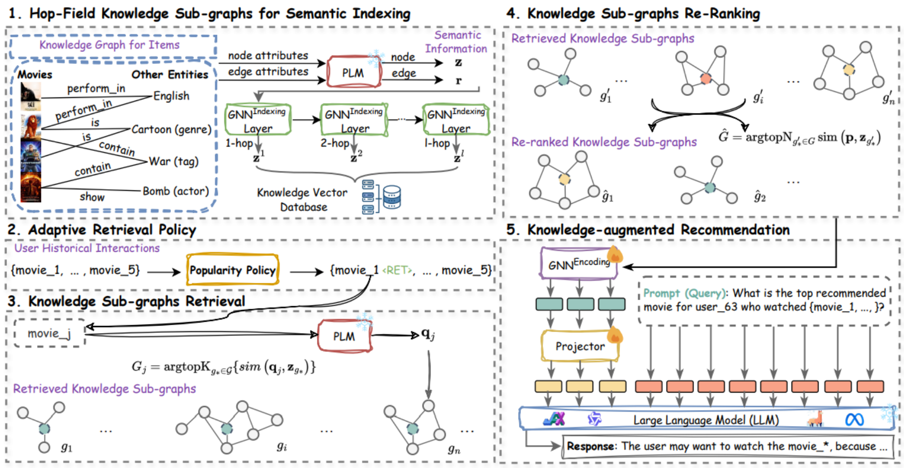


Figure 1: Proposed Architecture inspired from (3)

Earthquake Disaster Knowledge Graph Reasoning (4) Jiao and You proposed a rescue decision-making framework based on a domain-specific Earthquake Disaster Knowledge Graph (EDKG). Their architecture consists of three primary modules:

- A **Visual Perception module**, which leverages ResNet50 to classify structural damage and material types from images (achieving 83% accuracy on 2,500 training and 500 test images);
- A **Graph Mapping module**, which employs representation learning to embed knowledge graph entities for reasoning;
- A **Decision Reasoning module**, which outputs tailored rescue actions and equipment suggestions for disaster scenarios.

The EDKG itself was constructed using a top-down approach involving multi-source data collection, knowledge extraction, and fusion. This work highlights the feasibility of combining structured knowledge with visual perception to generate context-aware actions in high-stakes environments, and serves as an analog for our proposed triage-specific KG+VLM framework.

Automated KG Construction via VLMs and LLMs in E-commerce (5) Yang *et al.* introduced an end-to-end pipeline to automatically construct hierarchical knowledge graphs for e-commerce applications by jointly leveraging Vision-Language Models (VLMs) and Large Language Models (LLMs). Their method eliminates manual knowledge curation through:

- **Schema Initialization** to define target product properties;
- A cyclic pipeline of **Extraction, Formatting, Inference, Hierarchy Expansion**, and **Graph Pruning**; and
- The use of **Retrieval-Augmented Generation (RAG)** to provide accurate, up-to-date knowledge from external databases.

This KG construction paradigm significantly outperformed baseline approaches across multiple evaluation metrics and demonstrates the power of combining perception and retrieval for structured knowledge generation. Moreover, their use of RAG to mitigate hallucination and fill knowledge gaps in LLMs closely aligns with our motivation in the triage robotics domain.

Knowledge Graphs in Clinical AI Knowledge graphs encode medical knowledge in a structured form, linking entities such as diseases, symptoms, vital signs, medications, and patients. Notable examples include ontologies like SNOMED CT and UMLS, which enumerate clinical concepts and relationships. Prior works have used KGs for clinical decision support, such as diagnosing diseases based on symptoms or recommending treatments. In triage or diagnosis tasks, a KG can provide a priori connections (e.g., hypotension is related to shock, fever plus cough suggests infection).

Graph-based reasoning can improve transparency of AI decisions by tracing which facts led to a conclusion. In this work, we construct a lightweight knowledge graph tailored to triage, rather than relying on a vast general medical ontology. This domain-specific KG focuses on vitals and emergency symptoms to ensure relevant, high-precision knowledge retrieval.

Bayesian Networks and Knowledge Graphs for Triage Reasoning (Internal Discussion) In ongoing project meetings, the use of **Bayesian networks and KGs** has been identified as pivotal for expLiteraturelainable triage decision-making. Three key considerations emerged:

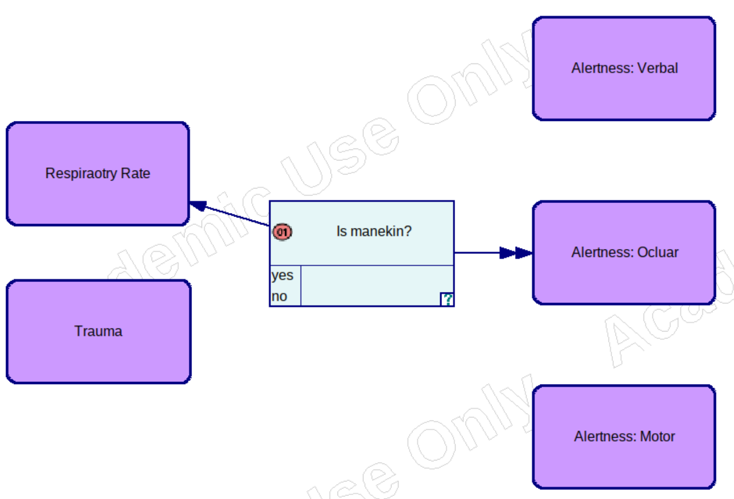
- **Dummy Detection**: The system must differentiate real humans from mannequins, as this impacts the decision logic (e.g., skipping vitals if no human is present).
- **Hierarchical Conditions for Absence of Trauma**: Determining the lack of trauma in regions like Head, Torso, or Limbs requires aggregating multiple fine-grained cues, aligning well with KG-based multi-hop reasoning.
- **Confidence Scores for Explainability**: Both the KG and Bayesian network modules must generate interpretable confidence scores to support verifiable clinical decisions.

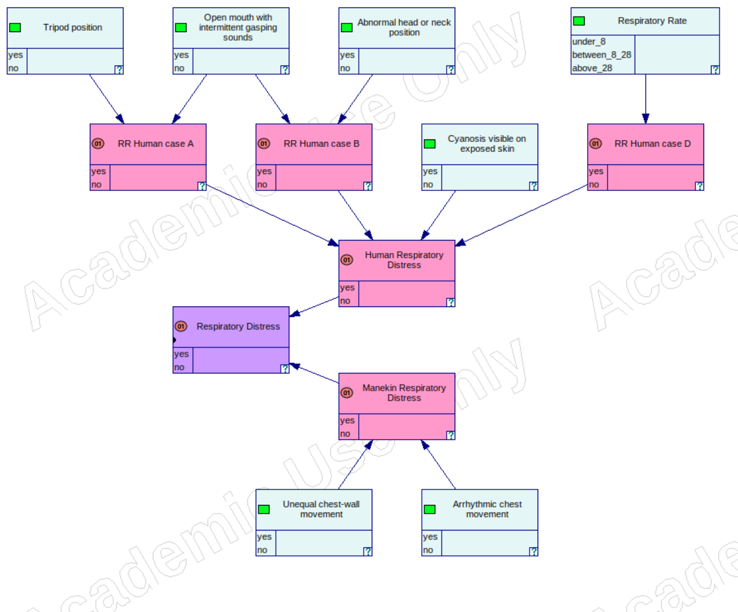
The next phase involves verifying the feasibility of these components within the current architecture, followed by strategy development to ensure an efficient and actively used system. Preliminary Bayesian network diagrams (color-coded by inference role) provide a starting point for transitioning from conceptual models to real-time triage implementations.

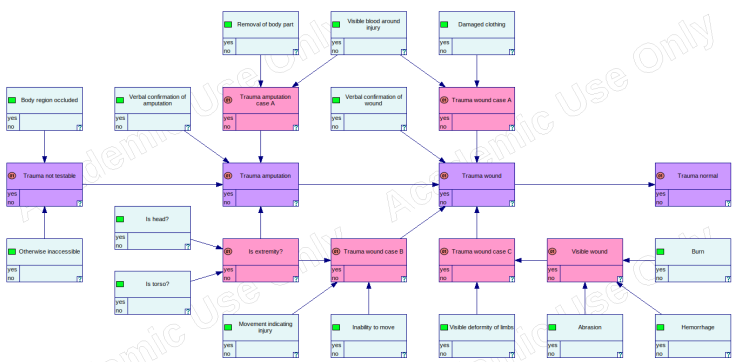
3 Methodology

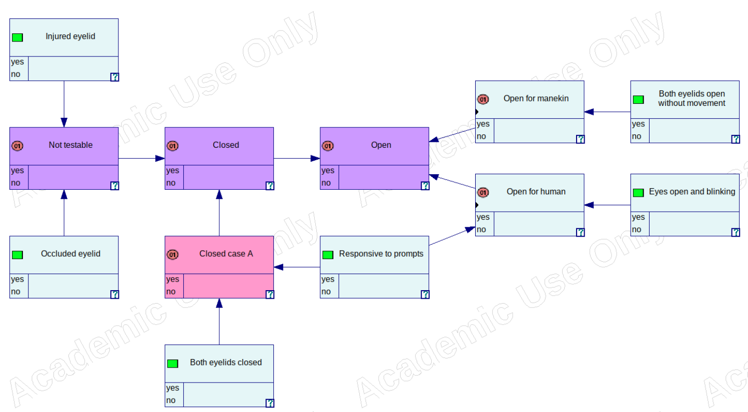
3.1 Current Bayesian network models

As discussed with Airlab member Szymon Rusiecki, below are the current Bayesian network architectures which serve as a starting point for further project work.
Color Legend:
Purple -> Our Targets
Blue -> Data that robot should be able to predict
Pink -> Grouped inputs for readability (those do not tell us if there are AND or OR statements in documentation!)

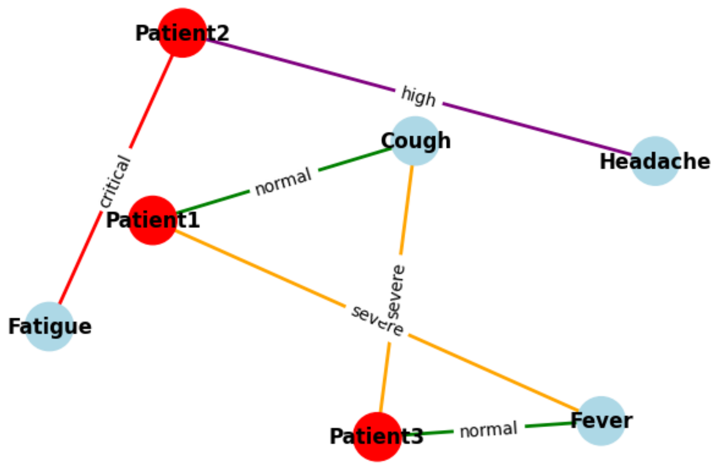


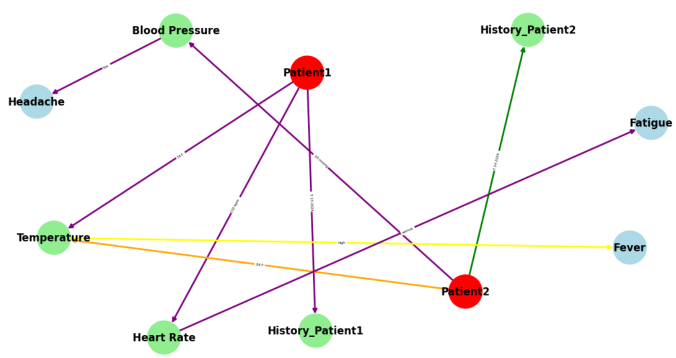

Tripod position
Open mouth with intermittent gasping sounds
Abnormal head or neck position
Respiratory Rate
under_8
between_8_28
above_28
RR Human case A
RR Human case B
Cyanosis visible on exposed skin
RR Human case D
Human Respiratory Distress
Respiratory Distress
Manekin Respiratory Distress
Unequal chest-wall movement
Arrhythmic chest movement

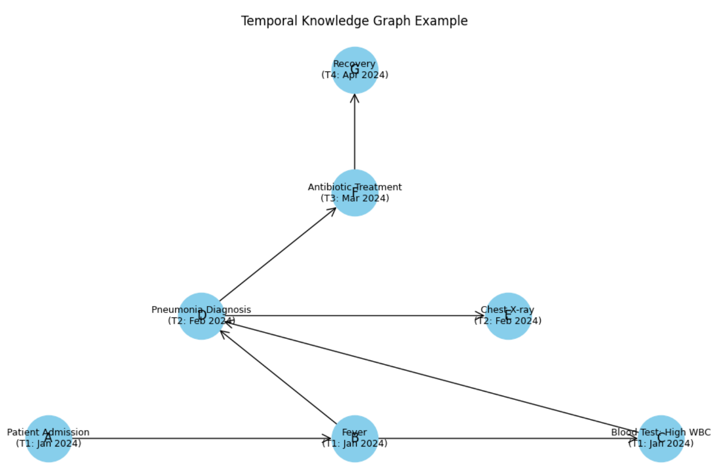

Removal of body part
Visible blood around injury
Damaged clothing
Body region occluded
Verbal confirmation of amputation
Trauma amputation case A
Verbal confirmation of wound
Trauma wound case A
Trauma not testable
Trauma amputation
Trauma wound
Trauma normal
Is head?
Otherwise inaccessible
Is extremity?
Trauma wound case B
Trauma wound case C
Visible wound
Burn
Is torso?
Movement indicating injury
Inability to move
Visible deformity of limbs
Abrasion
Hemorrhage


Injured eyelid
Open for manekin
Both eyelids open without movement
Not testable
Closed
Open
Open for human
Eyes open and blinking
Occluded eyelid
Closed case A
Responsive to prompts
Both eyelids closed

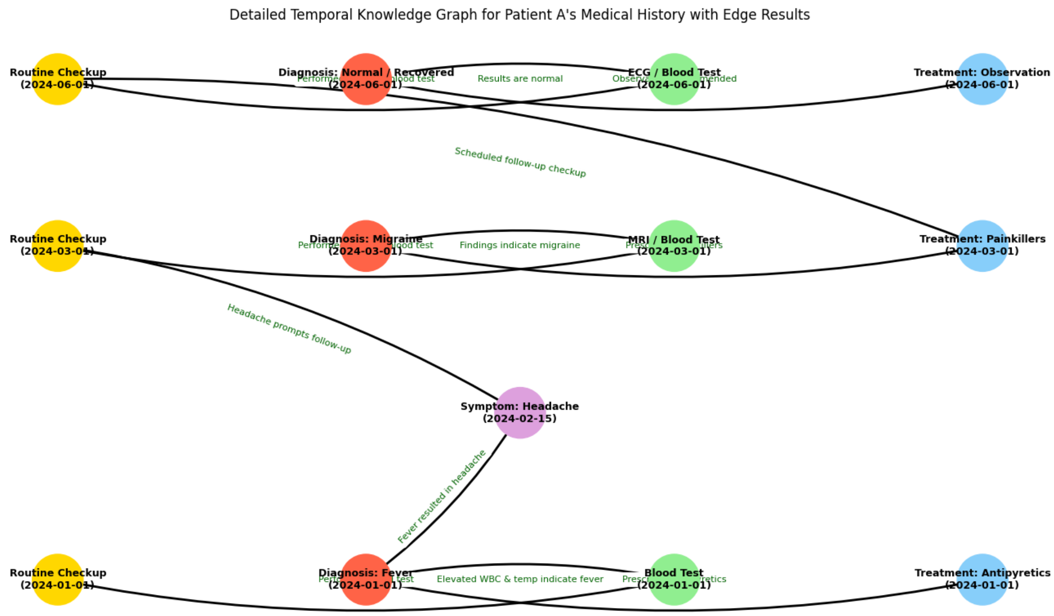
3.2 Initial KG proposed





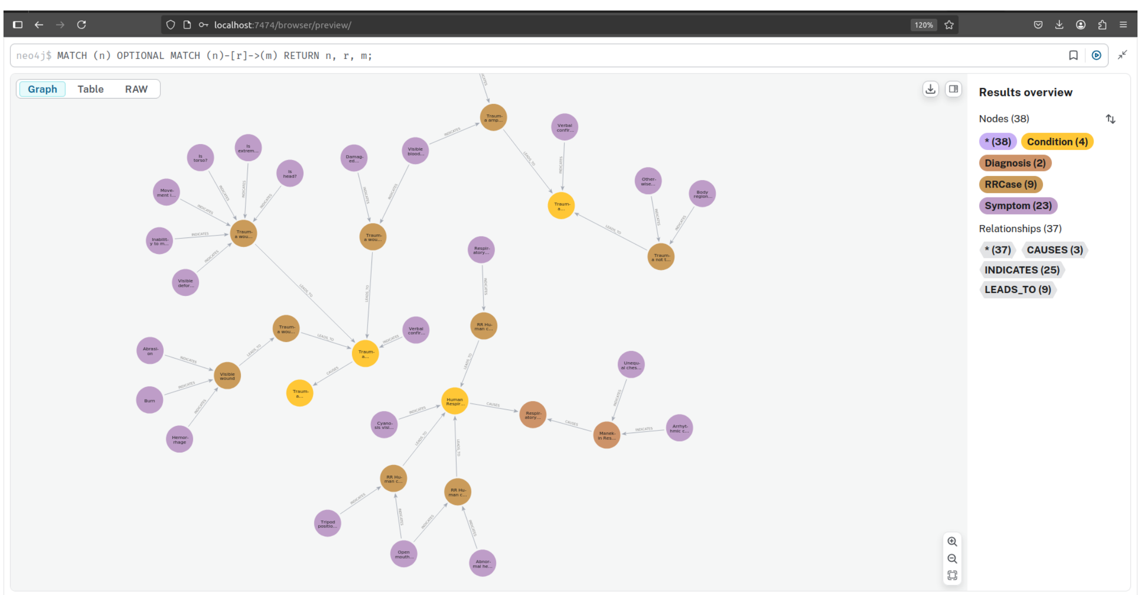


A combination of the above with a temporal dimension connecting different patients and symptoms which will look like



3.3 Final KG proposed

Below is a visualisation of the KG combining the architectural style of the team sugguestions and initial KG. This is a sample output since the actual graph is very complex and difficult to visualise



4 Baselines and Extensions

To ground our experiments and ensure fair comparisons, we selected two VLM models as baselines:

- **Dino V2:** A general-purpose Visual language model that serves as our non-coding baseline.
- **Qwen:** Another general-purpose Visual language model

.

As mentioned in previous section, we used original LLM-RAG framework from the (3) paper as the baseline for our experimentation. This provides a consistent foundation for evaluating the impact of alternative inference-time prompting strategies relative to the base VLM model, which is the primary extension we are proposing in this work.

5 Results and Analysis

5.1 Results Overview

TODO

6 Future Directions

TODO

7 Conclusion

TODO

8 Acknowledgments

TODO

9 GitHub Repo

Code for KGs

References

[1] A. Defilippo, P. Veltri, P. Lió, and P. H. Guzzi, "Leveraging graph neural networks for supporting automatic triage of patients," *Scientific Reports*, vol. 14, no. 1, p. 12548, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-63376-2

[2] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "Dinov2: Learning robust visual features without supervision," 2024. [Online]. Available: https://arxiv.org/abs/2304.07193

[3] S. Wang, W. Fan, Y. Feng, X. Ma, S. Wang, and D. Yin, "Knowledge graph retrieval-augmented generation for llm-based recommendation," 2025. [Online]. Available: https://arxiv.org/abs/2501.02226

[4] Y. Jiao and S. You, "Rescue decision via earthquake disaster knowledge graph reasoning," *Multimedia Systems*, vol. 29, no. 2, pp. 605–614, Apr 2023. [Online]. Available: https://doi.org/10.1007/s00530-022-01002-9

[5] S. Fang, K. Ma, T. Zheng, X. Du, N. Lu, G. Zhang, and Q. Tang, "Karpa: A training-free method of adapting knowledge graph as references for large language model's reasoning path aggregation," 2024. [Online]. Available: https://arxiv.org/abs/2412.20995