

Injecting Knowledge Graph Embeddings into RAG Architectures: Scalable Fact-Checking for Combating Disinformation with LLMs

Orlando Abuanza Ubaque¹, Diego Rincon-Yanez^{2,3,*} and Declan O’Sullivan^{2,3}

¹Pontificia Universidad Javeriana, Bogota, colombia

²School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

³ADAPT Centre for Digital Content, Dublin, Ireland

Abstract

The growing threat of disinformation and misinformation across digital platforms has intensified the demand for systems capable of producing verifiable and trustworthy outputs. With the widespread adoption of Large Language Models (LLMs) for a variety of tasks, the requirement to provide accurate and fact-verifiable answers is increasing daily. GraphRAGs have become a powerful approach for solving complex tasks that require factual context to deliver accurate and explainable answers. However, Knowledge Bases (KB) used to provide factual and contextual knowledge are composed of thousands or millions of statements, which limits the size of inputs that an LLM can handle, typically by the number of input tokens supported by the model. This work addresses the problem of Fact-Checking by injecting Knowledge Graph Embedding (KGE) vector representations into LLMs using a Retrieval Augmented Generation (RAG) approach to obtain more accurate results. The results show a notable difference in the quality of the results with two different vector representations and two KB construction methods.

Keywords

Graph Retrieval Augmented Generation, Retrieval Augmented Generation, Knowledge Graphs, Knowledge Graphs Embeddings, Scalable AI

1. Introduction

The Disinformation phenomenon is a growing concern due to its ability to distort mass public perception of reality and undermine trust in valid information sources. The impact can be significant, as evidenced during the pandemic [1]. From the intent, disinformation[2] can be classified into two categories: Information or news (1) created to deceive or (2) produced without the intention of deceiving.

With the significant adoption of Large Language Models (LLMs) and their applications in many disciplines, particularly in the communication sciences, source validation and fact-checking are essential features. One of the biggest known constraints with Large Language Models is hallucinations [3]. These instances represent statistical drifting in the token generation by the language model, which sometimes corresponds to the generation of contexts that contradict or mislead real-world factual knowledge.

This work presents an approach to detecting disinformation by combining Knowledge Graphs and their vector representations (known as knowledge graph embeddings) with large language models. The approach utilises the WELFake dataset [4], a well-known fake news dataset. In this dataset, a Knowledge Graph (KG) is generated from a single news article using OpenIE techniques. Then, vector space representations are generated using low-dimensional embedding algorithms such as *RotatE* and *TransE*. To retrieve the context from the entire KB, the single article is decomposed into triples and compared via its vector representation and social network analysis techniques. Finally, the retrieved context is sent to a Large Language Model (LLM).

SGKi 2025: Scaling Knowledge Graphs for Industry Workshop, co-located with SEMANTiCS’25: International Conference on Semantic Systems, September 3–5, 2025, Vienna, Austria

*Corresponding author.

✉ abaunzao@javeriana.edu.co (O. A. Ubaque); diego.rincon@adaptcentre.ie (D. Rincon-Yanez); declan.osullivan@adaptcentre.ie (D. O’Sullivan)

✉ 0000-0002-8982-1678 (D. Rincon-Yanez); 0000-0002-8982-1678 (D. O’Sullivan)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This paper is structured as follows: Section 2 presents a brief overview of Knowledge Graphs and their embeddings, Retrieval Augmented Generation (RAGs) techniques and scenarios. Section 3 presents the overall procedure, with special focus on the retrieval part from the KG; Section 4 details the experimentation scenarios and the metrics utilised. Finally, Section 5 draws some conclusions.

2. Related work

Numerous studies have focused on detecting disinformation scenarios on social media using a diverse number of AI models. Normally, these studies extract linguistic features and train models[5], such as K-nearest neighbour (KNN), support vector machine (SVM), and stochastic gradient descent (SGD). In some cases, it is also considered to use contextual user information in addition to finding patterns that could help verify the validity of a particular story[6].

NeuroSymbolic AI (NeSyAI), as a combination of neural and symbolic methods, positions a promising candidate for many applications [7, 8]. One benefit of neuro-symbolic solutions is the integration of domain knowledge, such as in the form of Knowledge Graphs (KGs). Integrating KGs as a structured and symbolic knowledge representation into RAG-type applications offers a powerful approach to addressing the challenge of reducing hallucinations by combining the ability of language models to analyse text with the capability to retrieve relevant information from external sources, such as specialised knowledge bases [9].

Knowledge Graphs used to reduce disinformation have been notable for their ability to capture and represent complex semantic relationships between entities. However, manual information extraction and annotation can be employed to generate these KGs; these are still fundamentally useful for smaller datasets or when precision in relationship capture is required [10]. In the same pathway, hybrid approaches to disinformation are performed [11] by integrating a KG through heterogeneous representation ensembles, and the use of neural networks to combine representations from Language Models, allowing deeper context understanding and relationship understanding between the mentioned entities.

Integrating KG into Retrieval-Augmented Generation (RAG) ensures the ability of Language Models to analyse text and retrieve relevant information from external sources, thereby enhancing accuracy and reliability, while mitigating hallucinations [12]. Techniques such as *RotatE*, *TransE*, or *DisMult* can enrich knowledge representation[13], combined with the efficient construction of KG, can provide a solid foundation in knowledge representation, improving pre-trained language models, thereby contributing to the accuracy, effectiveness and fact-checking capabilities[14].

3. Injecting KGE into the GraphRAG Workflow

To move closer to a scenario where LLMs can be confidently used for fake news detection. The LLMs must be equipped with tools for this purpose. Specifically, the use of RAGs (Retrieval-Augmented Generation) with knowledge graph embeddings is explored. The proposed method constructs the knowledge graphs (KG), supported by OpenIE methods, where entities and relationships are extracted from unstructured text data (the news text). For this task, the well-known Stanford CoreNLP, through the Stanza library, provides pre-trained models and utilises a pipeline-based approach to structure data into subject-predicate-object (S, P, O) triplets. KGs were also generated using the REBEL model[15]. REBEL reformulates the task as a sequence-to-sequence (seq2seq) problem within a pre-trained language model, BART (Bidirectional and Auto-Regressive Transformer) [16]. Once the news articles are generated into KGs, additional triples are added, inserting the *TrueNews* or *FakeNews* label to the news article representation, following the representation shown in 1.

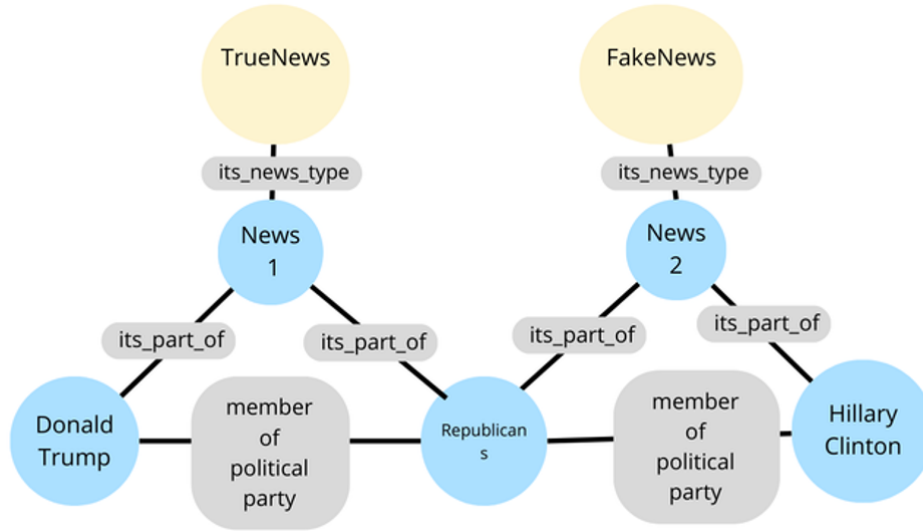


Figure 1: Knowledge Graph structure from two articles: Entities are connected to their respective news articles and parent nodes.

3.1. Knowledge Graph Embedding Techniques

Knowledge graph embeddings (KGEs) learn dimensional representations of labelled, directed multigraph nodes and edges to predict missing parts of a triple (entities or relations). These have been utilised in various tasks, including fact-checking, question-answering, link prediction, and entity linking[17].

The TransE model is one of the most well-known in this category. It uses a translation-based approach to model relationships. For a valid triplet (h, r, t) , the embedding of the tail entity should be close to the embedding of the head plus a vector representing the relation r in the embedding space: $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$. This model naturally captures hierarchical and structural relationships. However, its simplicity presents limitations when modelling more complex relationships, such as non-transitive or many-to-many relations [18].

RotatE, also part of the **Geometric Models**, models relationships as rotations in the complex plane. For the triplet (h, r, t) , RotatE uses a rotation operation to transform the embedding of entity h into that of t : $\mathbf{t} \approx \mathbf{h} \circ \mathbf{r}$, where \circ denotes the Hadamard product. This strategy is particularly useful for modelling complex patterns such as symmetry, asymmetry, transitivity, and inversion [19].

Table 1

Experiment Results: Ablation study vs the OpenIE + KGE techniques evaluation. The best results are highlighted.

Metrics	Base	Stanford		REBEL	
		RotatE	TransE	RotatE	TransE
Precision	3.7	3.7	4.1	3.2	4.1
Accuracy	3.7	3.2	4.0	3.0	3.8
Contextual Understanding	3.5	3.3	4.4	3.3	3.9
Compelling Misinformation	3.7	3.1	4.0	3.1	3.6
Transparency and Traceability	3.2	3.0	3.9	2.7	3.7
Source Retrieval Accuracy	2.7	2.9	3.3	2.8	4.0
Total	3.4	3.2	4.0	3.8	3.0

3.2. Enhanced Retrieval Augmented Scenario

Once the KG is constructed, various manipulations can be performed to extract information from it. A particularly effective approach is the use of a neighbourhood function, which searches for the

nearest neighbours in proximity using algorithms such as Dijkstra’s algorithm. By incorporating such information, we can bridge the gap between surface-level entity recognition and a deeper, more accurate understanding of news. The graphs generated using Stanford NLP and REBEL were processed to obtain embeddings in the vector space. The two Knowledge Graph Embedding (KGE) models were then applied to capture the semantic relationships and structure of the graphs

For the search of nearest entities, the unique entities extracted from the subject and object nodes of the news have been mapped into the vector space of the global generated KGE, which is the knowledge base. Cosine similarity was computed between the vector of the target entity and the vectors of all other entities in the vector space. This step allowed for the identification of entities most similar to the target entity in the news.

Subsequently, for each combination of the subject entity from the news and its nearest neighbours, the pretrained KGE model was used to predict the most probable relation between them. This step enabled the identification of potential connections between the news and the existing knowledge base. Finally, the extracted entities, relationships, and triplets were used as augmented knowledge. This context was fed into GPT-4o (LLM) during the final stage to generate predictions or grounded responses based on the provided knowledge.

4. Experiment and Evaluation

The evaluation was designed to assess both the RAG model’s performance and the quality of the generated text, specifically in terms of factual accuracy, contextual understanding, and transparency. Six metrics were applied to evaluate the performance of the tested models: precision, accuracy, contextual understanding, compelling misinformation, transparency and traceability, and source retrieval accuracy. The average of the evaluations for each scenario in each metric was used to rate the proposed approach, as shown in the Figure 2. The results are presented in Table 1, where it is evident that the best-performing model is TransE with REBEL del for Knowledge Graph Construction.

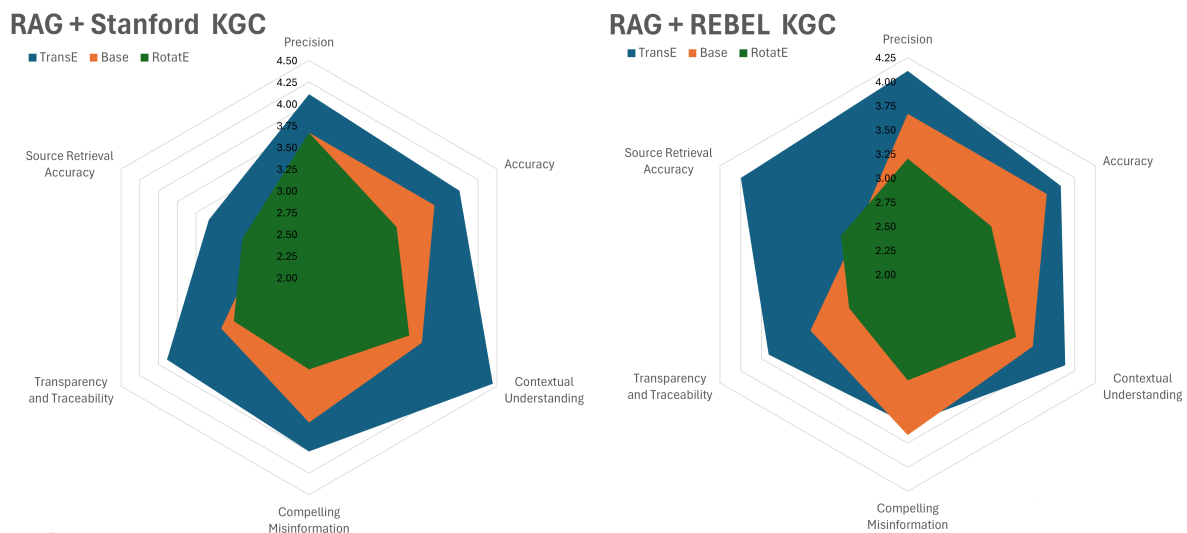


Figure 2: Evaluation Results using two embedding models *RotatE* y *TransE*: On the left, the Stanford KGC and the LSTM REBEL on the right.

However, precision, specifically, accurately predicting whether a news item is false, ties in with TransE and Stanford, which also achieved the best performance in retrieving useful information for decision-making. On the other hand, the worst-performing model overall was RotatE with REBEL, although it predicted truthfulness just as well as the baseline model.

The most computationally and time-intensive model was RotatE with Stanford, requiring more than 8 hours of training on a Google Colab T4 GPU. In all cases, the models tended to predict true news

correctly but often misclassified false news.

5. Conclusions and Future Work

While the LLMs can model the veracity of news, they rely on the assumption that if the supposed source is a real organisation, such as a magazine, government, or a recognised person, everything they say must be true. This leads to hallucinations when concluding these false facts. A notable example of this was a case involving WikiLeaks, a news item supposedly generated by the Washington Examiner. The model erroneously claims the news is true because it mentions Clinton, Podesta, and the source, the Washington Examiner (WE). In contrast, the complete RAGs, embeddings, and KG neighbours prioritised the entities mentioned in the case, such as Clinton, Podesta, and Doug Band, since they were found in both retrievals.

News with high semantic content but few entities (people, places, or things) tends to perform worse, sometimes causing hallucinations due to lack of context, particularly in REBEL. This issue is significantly reduced with Stanford due to a higher granularity in the extracted triplets. However, this comes at the cost of generating embeddings, as it quadruples the number of entities obtained compared to REBEL.

Future work will validate the proposed approach using evaluation metrics such as F1-score, accuracy, and area under the ROC curve (AUC-ROC). These metrics will comprehensively assess the model's performance and ability to accurately differentiate between classes. Additionally, hyperparameter tuning and the inclusion of additional data will be explored to further optimise the model's effectiveness.

Acknowledgments

The ADAPT Centre for Digital Content Technology also partially supports the project under the Research Ireland Research Centres Programme (Grant 13/RC/2106_P2).

Declaration on Generative AI

The author have not employed any Generative AI tools in creating the paper.

References

- [1] G. M. Nieves-Cuervo, E. F. Manrique-Hernández, A. F. Robledo-Colonia, E. K. A. Grillo, Infodemia: noticias falsas y tendencias de mortalidad por COVID-19 en seis países de América Latina, *Revista Panamericana de Salud Pública* 45 (2021) 1. doi:10.26633/RPSP.2021.44.
- [2] H. Allcott, M. Gentzkow, Social Media and Fake News in the 2016 Election, *Journal of Economic Perspectives* 31 (2017) 211–236. doi:10.1257/jep.31.2.211.
- [3] N. Dziri, S. Milton, M. Yu, O. Zaiane, S. Reddy, On the origin of hallucinations in conversational models: Is it the datasets or the models?, in: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Seattle, United States, 2022, pp. 5271–5285. doi:10.18653/v1/2022.naacl-main.387.
- [4] P. K. Verma, P. Agrawal, I. Amorim, R. Prodan, WELFake: Word Embedding Over Linguistic Features for Fake News Detection, *IEEE Transactions on Computational Social Systems* 8 (2021) 881–893. doi:10.1109/TCSS.2021.3068519.
- [5] Y. Wang, L. Wang, Y. Yang, Y. Zhang, Detecting fake news by enhanced text representation with multi-EDU-structure awareness, *Expert Systems with Applications* 206 (2022) 117781. doi:10.1016/j.eswa.2022.117781. arXiv:2205.15139.

- [6] K. Shu, L. Cui, S. Wang, D. Lee, H. Liu, DEFEND, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '19*, ACM, New York, NY, USA, 2019, pp. 395–405. doi:10.1145/3292500.3330935.
- [7] K. Karel-Batista, D. Rincon-Yanez, S. Senatore, Human-oriented fuzzy-based assessments of knowledge graph embeddings for fake news detection, in: *20th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer Nature Switzerland, 2024, p. 12. doi:10.1007/978-3-031-73997-2_10.
- [8] D. Rincon-Yanez, M. H. Gad-Elrab, D. Stepanova, K. T. Tran, C. Chu Xuan, B. Zhou, E. Karlamov, Addressing the Scalability Bottleneck of Semantic Technologies at Bosch, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 13998 LNCS, 2023, pp. 177–181. doi:10.1007/978-3-031-43458-7_33.
- [9] W. J. Schmidt, D. Rincon-Yanez, E. Kharlamov, A. Paschke, Scaling Scientific Knowledge Discovery with Neuro-Symbolic AI and Large Language Models, in: *CEUR Workshop Proceedings*, volume 3759, 2024, pp. 1–7. URL: <https://ceur-ws.org/Vol-3759/workshop4.pdf>.
- [10] A. Jaradeh, M.-B. Kurdy, ArEmotive Bridging the Gap: Automatic Ontology Augmentation Using Zero-Shot Classification for Fine-Grained Sentiment Analysis of Arabic Text, *IEEE Access* 11 (2023) 81318–81330. doi:10.1109/ACCESS.2023.3300737.
- [11] B. Koloski, T. S. Perdih, M. Robnik-Šikonja, S. Pollak, B. Škrlj, Knowledge graph informed fake news classification via heterogeneous representation ensembles, *Neurocomputing* 496 (2022) 208–226. doi:10.1016/j.neucom.2022.01.096.
- [12] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela, Retrieval-augmented generation for knowledge-intensive nlp tasks, in: *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20*, Curran Associates Inc., Red Hook, NY, USA, 2020, p. 16.
- [13] L. Fu, H. Peng, S. Liu, KG-MFEND: an efficient knowledge graph-based model for multi-domain fake news detection, *The Journal of Supercomputing* 79 (2023) 18417–18444. doi:10.1007/s11227-023-05381-2.
- [14] B. Yaman, L. McKenna, A. Randles, L. Kilgallon, P. Crooks, D. O'Sullivan, Digital prosopographical information in the virtual record treasury of ireland's knowledge graph for irish history, in: *SemDH@ESWC*, 2024. URL: <https://ceur-ws.org/Vol-3724/paper2.pdf>.
- [15] D. D. S. a. Roos Bakker a b, M. de Boer, Knowledge graph extraction in a practical context, in: *Neurosymbolic Artificial Intelligence*, 2022, pp. 1–14.
- [16] L. Berglund, M. Tong, M. Kaufmann, M. Balesni, A. C. Stickland, T. Korbak, O. Evans, the Reversal Curse: Llms Trained on “a Is B” Fail To Learn “B Is a”, in: *12th International Conference on Learning Representations, ICLR 2024*, 2024, pp. 1–21. arXiv:2309.12288.
- [17] D. Rincon-Yanez, A. Mouakher, S. Senatore, Enhancing downstream tasks in Knowledge Graphs Embeddings: A Complement Graph-based Approach Applied to Bilateral Trade, *Procedia Computer Science* 225 (2023) 3692–3700. doi:10.1016/j.procs.2023.10.364.
- [18] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, in: *Advances in Neural Information Processing Systems, NIPS'13*, Curran Associates Inc., Red Hook, NY, USA, 2013, pp. 2787–2795.
- [19] Z. Sun, H. Deng, X. Nie, J. Tang, Y. Yang, RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space, in: *(EMNLP-IJCNLP)*, 2019, pp. 119–128. doi:10.18653/v1/d19-1013.