Evaluating LLM-based Triple Extraction for Knowledge Graph Fact-Checking in Portuguese

Roney Lira de Sales Santos[1], **Lucas dos Santos**[1], **João Pedro Holanda Souza**[2]

[1]Federal University of Bahia (UFBA)
Camaçari – BA – Brazil

[2]Universidade Norte do Paraná (UNOPAR)
Londrina – PR – Brazil

roneysantos@ufba.br, lukas.santtos53@gmail.com, joaocmdt505@gmail.com

Abstract. *Automatic fake news detection in Portuguese is still often treated as a text classification task, without explicitly representing the factual veracity of the claims. In this work, we evaluate LLM-based triple extraction in a knowledge graph (KG)-based fact-checking system. We replace the Open Information Extraction component of a previous approach with triples generated by Sabiá 4, while keeping the remaining pipeline unchanged. The graph is built only from triples extracted from true news articles and is used as factual support to verify new instances. The experiments use true and fake news articles across four evaluation settings. In the closed setting with the complete KG, LLM-based extraction achieves an F1 score of 0.9992, outperforming the OIE-based configuration. However, in more restrictive evaluation settings, expressive triples require entity normalization, relation standardization, and semantic consolidation to provide factual support beyond direct evidence.*

1. Introduction

The spread of misinformation in digital environments has motivated the development of automatic methods for fake news detection. Although supervised approaches achieve strong results, many of them rely on linguistic and statistical cues from the text, without explicitly modeling the factual veracity of the content [Vosoughi et al. 2018].

Content-based methods can address this limitation by representing claims as semantic relations in knowledge graphs (KGs) [Ciampaglia et al. 2015]. In this context, the work of [Santos et al. 2026] proposed a system for Portuguese in which news articles are decomposed into triples in the form of subject, relation[1], and object. Triples extracted from true news articles are used to build a KG, which serves as a structured source of factual support for evaluating new instances.

In this approach, triple extraction is a critical step. In [Santos et al. 2026], this process is performed with Open Information Extraction (OIE) [Banko et al. 2007], a technique with known limitations in Portuguese, including difficulties with implicit subjects, complex syntactic structures, and relations distributed across the text [Stanovsky and Dagan 2018]. Errors at this stage affect which facts are represented in the KG and, consequently, the veracity scores assigned by the system.

[1]We use *relation* to refer to the textual predicate extracted from the news article. This choice emphasizes that relations are open textual expressions generated by the extractor, rather than predicates from a fixed schema.

Large language models (LLMs) can offer a promising alternative for this step, with the potential to extract relations with greater contextual coherence and better handle long-range dependencies [OpenAI et al. 2024]. However, it remains unclear how this substitution affects KG-based factual verification systems, especially for news in Portuguese.

In this work, we evaluate whether triple extraction with LLMs strengthens a KG-based fake news detection system for Portuguese. To this end, we take the approach of [Santos et al. 2026] as reference, replacing its OIE component with triples generated by an LLM while keeping the remaining pipeline components unchanged. This setting allows us to analyze, in a controlled way, how changes in factual representation affect the graph structure, the retrieval of factual support, and veracity inference.

The results show that LLM-based extraction outperforms the OIE-based approach in the closed setting with the complete KG, achieving an F1 score of 0.9992 compared with 0.9396. In more restrictive evaluations, such as *leave-one-news-out*, *leave-one-triple-out*, and train/test split, the system behaves conservatively when direct evidence is removed or when the evaluated facts fall outside the KG coverage. This result indicates that more expressive triples do not, by themselves, guarantee greater factual support beyond the evidence directly represented in the graph.

The main contributions of this work are: (i) the evaluation of an LLM as a triple extractor in a KG-based fact-checking pipeline for Portuguese; (ii) a controlled comparison with an OIE-based approach in the closed evaluation setting; and (iii) an analysis of the impact of triple extraction on graph structure, retrieval of factual support, and the veracity scores assigned to news articles.

The remainder of this paper is organized as follows. Section 2 reviews related work on fake news detection, fact-checking, and triple extraction. Section 3 describes the datasets, extraction procedure, KG construction, truth-value computation, and evaluation protocols. Section 4 presents and discusses the results, and Section 5 concludes the paper and outlines directions for future work.

2. Related Work

Automatic fake news detection in Portuguese has been studied mainly through linguistic cues, stylometric features, and supervised models. The Fake.Br Corpus [Monteiro et al. 2018, Silva et al. 2020] advanced this line of research by providing true and fake news articles aligned by topic, enabling controlled experiments for Portuguese. Broader reviews and surveys of the area report strategies based on linguistic attributes, machine learning, textual structure, and content-based methods [Santos 2022]. Although relevant, these approaches tend to classify the news article as a textual unit, without explicitly representing the events described in it as verifiable facts.

In automatic fact-checking, the work of [Ciampaglia et al. 2015] formulates factual verification as a connectivity problem in knowledge graphs. In this approach, a claim is represented as a triple in the form of subject, relation, and object, and its truth value is estimated from the semantic proximity between entities in a graph derived from DBpedia [Auer et al. 2007]. In a complementary direction, [Thorne et al. 2018] consolidated tasks in which claims are classified based on evidence retrieved from external sources. Recent surveys show that automatic fact-checking combines claim extraction, evidence retrieval, knowledge representation, and inference [Guo et al. 2022].

For Portuguese, the work of [Santos and Pardo 2020] adapted knowledge graph-based verification to resources derived from Portuguese DBpedia and also explored strategies based on search engine results. More recently, [Santos et al. 2026] proposed a graph-based system built from true news articles in Portuguese. Unlike encyclopedic graphs, this method represents news articles as factual events extracted into triples and computes the veracity of new instances from the support found in the graph. Since this step depends on Open Information Extraction (OIE), a technique designed to extract relations without a predefined schema [Banko et al. 2007], its quality depends strongly on syntactic analysis and on how well linguistic resources fit the target language. For this reason, extraction in Portuguese is identified as one of the main bottlenecks of this approach.

More recently, LLMs have been explored in information extraction and knowledge graph construction. Some studies review their use in generative extraction of entities, relations, and events [Xu et al. 2024], while others investigate strategies for triple extraction and graph construction from free text [Van Cauter et al. 2024, Mihindukulasooriya et al. 2025]. These works suggest that LLMs can reduce some limitations of traditional extractors, but they also point to challenges related to standardization, redundancy, hallucination, and consistency in the extracted triples.

This work differs from previous studies by evaluating the impact of LLM-based triple extraction in a knowledge graph-based factual verification approach for Portuguese, considering final performance, graph structure, retrieval of factual support, and the truth values assigned to news articles.

3. Methodology

This section describes the methodology used to evaluate the impact of LLM-based triple extraction on a knowledge graph-based fake news detection system. In this paper, we use KG, from *Knowledge Graph*, to refer to the graph built from triples extracted from true news articles. The approach follows the principle of factual support verification in graphs adopted in [Santos et al. 2026], but replaces OIE-based extraction with a large language model, while keeping the remaining pipeline components unchanged.

3.1. Approach Overview

The pipeline consists of five steps: (i) consolidation of the news datasets; (ii) extraction of triples structured as `(subject, relation, object)`; (iii) construction of the KG with triples from true news articles; (iv) computation of the truth value of triples extracted from new articles; and (v) final classification based on the aggregation of the obtained values.

Each news article is represented as a set of triples in the form of subject, relation, and object. The central hypothesis is that the KG works as a structured memory of previously observed factual events and relations. Thus, the larger and more representative the KG is, the greater its expected capacity to retrieve factual support for new articles. Low values indicate absence or insufficiency of retrievable support in the available KG, rather than absolute proof of falsity.

3.2. Datasets

The experiments use three public datasets for fake news detection in Portuguese: Fake.Br, FakeRecogna [Garcia et al. 2022], and FakeTrueBR [Chavarro et al. 2023].

These datasets were consolidated into a single collection of 22,684 news articles, balanced between true and fake instances.

Fake.Br and FakeRecogna were used in already preprocessed versions, with texts normalized through steps such as tokenization, lemmatization, and/or stemming. We then standardized the fields and mapped the labels of the three datasets to a common binary representation, in which fake news articles receive label 0 and true news articles receive label 1. Table 1 summarizes the data.

Dataset	True	Fake	Total
Fake.Br	3,600	3,600	7,200
FakeRecogna	5,951	5,951	11,902
FakeTrueBR	1,791	1,791	3,582
Total	11,342	11,342	22,684

Table 1. Composition of the datasets used in the experiments.

3.3. LLM-based Triple Extraction

Triple extraction was performed with Sabiá 4 [Laitz et al. 2026], a large language model developed by Maritaca AI[2] with a focus on Brazilian Portuguese. This choice is motivated by its alignment with the evaluated corpora, all of which are in Brazilian Portuguese. According to [Laitz et al. 2026], Sabiá 4 underwent continual pretraining on Portuguese and Brazilian legal corpora, supports a context window of 128 thousand tokens, and was tuned with instruction and preference data.

Each news article was submitted to the model with a fixed instruction to extract triples in the form of subject, relation, and object. The prompt was designed to reduce hallucinations, avoid unsupported inferences, and control the granularity of the relations. The model was instructed to extract only explicit information, produce a single semantic relation per triple, decompose objects containing embedded clauses or relations, preserve modality and speech attribution when relevant, and return only valid JSON in the format `[["subject", "relation", "object"], ...]`.

The calls were made through the Maritaca AI API[3], using a configuration compatible with the OpenAI Chat Completions API[4]. We used temperature 0 and a maximum limit of 1,000 tokens per response. Processing was executed in parallel, with request-per-minute control and retries in case of failure. The responses were cleaned and validated to remove Markdown markers and identify the list of triples. We consider a triple valid when it is extracted in a structured format with exactly three non-empty textual elements, corresponding to subject, relation, and object.

3.4. Knowledge Graph Construction

The KG was built from the JSONL file generated during extraction. Each record contains the news article identifier, its label, and a list of triples in the form of subject, relation,

[2]Available at: `https://www.maritaca.ai/`. Accessed: 12 May 2026.

[3]Available at: `https://docs.maritaca.ai/pt/visao-geral`. Accessed: 12 May 2026.

[4]Available at: `https://developers.openai.com/api/reference/chat-completions/overview`. Accessed: 12 May 2026.

and object. Before insertion, each triple was validated to ensure three non-empty textual elements.

Only triples extracted from true news articles were used to build the KG. Thus, the graph works as a structured memory of events and relations observed in news articles considered reliable, while fake news articles are used only for evaluation.

Formally, $KG = (V, E)$, where V represents subjects and objects, and E represents relations between them. Each triple (s, r, o) adds the nodes s and o to the graph, when not already present, and an edge labeled with relation r.

The implementation was carried out with NetworkX [Hagberg et al. 2008]. In the main experiments, we use an undirected multigraph, `nx.MultiGraph()`, allowing multiple relations between the same pair of entities. The choice of an undirected version follows the semantic proximity formulation based on paths in knowledge graphs [Ciampaglia et al. 2015].

3.5. Truth Value Computation

To verify a news article, the method computes a truth value for each triple based on the support found in the KG. Let $t = (s, r, o)$ be a triple. The value $\tau(t) \in [0, 1]$ is estimated from the connectivity between s and o, following the semantic proximity intuition of [Ciampaglia et al. 2015]. The relation r is preserved as the edge label, but the computation considers only the connectivity between subject and object.

If s and o correspond to the same node, or if there is a direct edge between them, we assign $\tau(t) = 1$. If either node is absent from the graph, or if there is no path between them, we assign $\tau(t) = 0$. In all other cases, the shortest weighted path between s and o is computed with Dijkstra's algorithm[5] [Dijkstra 1959]. The path cost is the sum of $\log(k(v))$ over the intermediate nodes, where $k(v)$ is the degree of node v considering relation multiplicity. The triple value is defined by Equation 1 [Ciampaglia et al. 2015]:

$$\tau(t) = \frac{1}{1 + \sum_{v \in P_{s,o}} \log(k(v))} \tag{1}$$

where $P_{s,o}$ contains the intermediate nodes in the path between s and o. Thus, direct connections or paths mediated by specific nodes receive higher values, while long or less informative paths receive lower support.

The final value of a news article N is computed as the mean truth value of its triples, as shown in Equation 2 [Santos 2022, Santos et al. 2026]:

$$Score(N) = \frac{1}{|T_N|} \sum_{t \in T_N} \tau(t) \tag{2}$$

where T_N is the set of triples extracted from the news article. News articles with no valid triples receive $Score(N) = 0$, since they provide no retrievable factual support in the KG.

Classification follows three decision ranges: $Score(N) \geq 0.60$ indicates a true news article; $Score(N) < 0.40$ indicates a fake news article; and values in $[0.40, 0.60)$ indicate an uncertain case, as in [Santos et al. 2026]. For binary comparisons, the uncertain class is conservatively mapped to fake.

[5]Dijkstra's algorithm was chosen because it computes shortest paths in weighted graphs with non-negative weights, which is the case for the cost function based on logarithmic node degrees.

3.6. Evaluation Protocols

The evaluation was conducted in four settings:

- **Complete KG**: the KG is built with all triples extracted from true news articles, and verification is applied to the full dataset. This is a closed evaluation setting because true news articles evaluated by the system may also have contributed triples to the KG. Therefore, this setting measures the internal compatibility between articles and graph, indicating the maximum factual retrieval capacity when facts are represented in the KG.
- ***Leave-one-news-out***: when evaluating a true news article, all its triples are temporarily removed from the KG. This setting reduces the effect of direct memorization and tests whether other true news articles provide factual support for the evaluated events.
- ***Leave-one-triple-out***: the evaluated triple is removed from the KG before computing its truth value, when it appears as a direct connection. This protocol follows the intuition of removing direct evidence in knowledge graphs [Ciampaglia et al. 2015] and checks whether there is indirect support for the claim.
- **Train/test split**: the KG is built only with triples from true news articles in the training set, and evaluation is performed on the test set. We used an 80/20 split, empirically chosen to preserve most true factual evidence for KG construction while still reserving a representative portion of the data for evaluation. This is the setting closest to a real application, since it evaluates news articles not used to build the graph.

In addition to accuracy, precision, recall, F1, and AUC, we report *coverage* [Geifman and El-Yaniv 2017], defined as the proportion of news articles for which the system issues a true or fake decision. This measure is relevant because the method includes an uncertain class, used when retrievable factual support in the KG is insufficient. Thus, low support in the KG should be interpreted as absence of retrievable evidence in the available base, and not necessarily as absolute falsity.

4. Results and Discussion

This section presents the results of LLM-based triple extraction in the KG-based pipeline. The analysis covers the characterization of triples and the graph, the evaluation protocols, the comparison with OIE, and the main effects observed in factual retrieval.

4.1. Characterization of Triples and the Graph

Extraction with Sabiá 4 generated 296,362 valid triples from 22,684 news articles, with an average of 13.06 triples per article and a median of 11. In total, 1,170 articles had no valid triples, corresponding to 5.16% of the corpus. Table 2 summarizes the distribution of triples by class.

True news articles generated more triples on average, suggesting higher factual density in the texts used to build the KG. However, the presence of true articles with no valid triples indicates that extraction is still sensitive to textual structure and to how events are expressed.

Class	Articles	Triples	Mean	No triples
Fake	11,342	116,601	10.28	179
True	11,342	179,761	15.85	991
Total	22,684	296,362	13.06	1,170

Table 2. Statistics of the triples extracted by Sabiá 4.

The KG built from true news articles contains 253,464 nodes and 217,485 edges when multiplicity is considered, of which 209,188 are unique edges. The structure has 52,480 connected components, with the largest component containing 112,588 nodes, equivalent to 44.42% of the total. Table 3 presents the main structural statistics.

Metric	Value
Nodes	253,464
Edges with multiplicity	217,485
Unique edges	209,188
Connected components	52,480
Nodes in largest component	112,588
Largest component proportion	44.42%
Average degree	1.72
Nodes with degree 1	218,622
Distinct relations	76,681

Table 3. Structural statistics of the KG built from triples extracted from true news articles.

These results indicate that LLM-based extraction produces a broad and semantically varied graph, but also a sparse one. The large number of nodes with degree 1, corresponding to 86.25% of the nodes, shows that many entities appear in low-recurrence relations. Figure 1 reinforces this pattern, with a distribution concentrated in low degrees and a long tail formed by a small number of highly connected nodes. For factual verification, this means that the KG stores many events, but part of them remains in poorly connected regions.

4.2. Truth Value Evaluations

The four evaluations were designed to verify whether the KG retrieves facts present in the graph itself, whether the system depends on the direct presence of the evaluated article, whether indirect support exists when the corresponding edge is removed, and whether the KG evaluates unseen news articles from a previously built factual base. Table 4 presents the results of the four protocols. For binary comparison, instances classified as uncertain were mapped to the fake class, following a conservative policy.

In the complete KG setting, the system reaches an F1 score of 0.9992 and an AUC of 0.9993, indicating strong factual retrieval when the events from true news articles are represented in the graph. This result, however, should be interpreted with caution: since the KG is built with all available true news articles, part of the evaluated true articles also contributed to the construction of the graph itself. Therefore, this protocol mainly measures the internal consistency of the KG and its ability to retrieve facts already present in the base, rather than the system's performance on completely new articles.

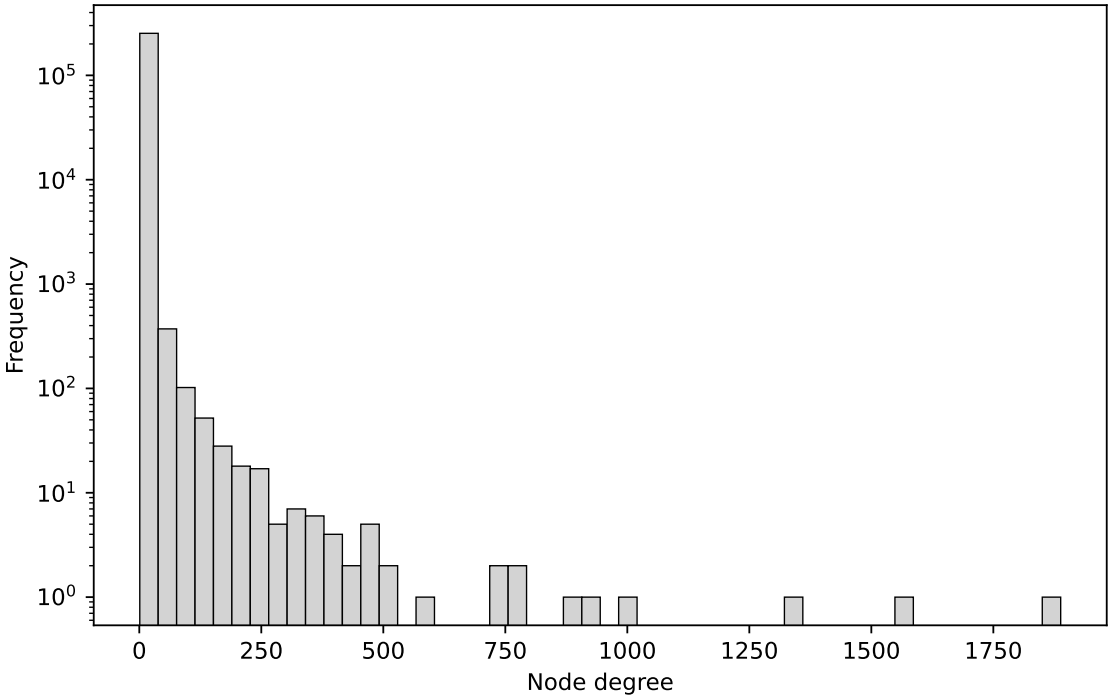

Figure 1. Degree distribution of the nodes in the KG built from triples extracted by Sabiá 4.

Protocol	Acc.	Prec.	Rec.	F1	AUC	Coverage
Complete KG	0.9992	0.9993	0.9991	0.9992	0.9993	0.9989
Leave-one-news-out	0.5201	0.9831	0.0410	0.0787	0.5475	0.9927
Leave-one-triple-out	0.5202	0.9832	0.0412	0.0790	0.5832	0.9919
Train/test split	0.5147	1.0000	0.0291	0.0566	0.5375	0.9938

Table 4. Results of the evaluation protocols with triples extracted by Sabiá 4.

In the restrictive[6] settings, the behavior changes substantially. The F1 score drops to 0.0787 in *leave-one-news-out*, 0.0790 in *leave-one-triple-out*, and 0.0566 in the train/test split. In all cases, precision remains high, but recall is very low. This indicates that the system avoids classifying an article as true without strong retrievable evidence, but fails to recognize many true articles when their events are not sufficiently connected to the KG.

Figure 2 confirms this behavior. In the complete KG setting, the means clearly separate fake and true news articles. In the restrictive protocols, the values of true articles shift to low ranges, approaching the values of fake articles. Thus, the performance drop is associated with KG coverage and connectivity, and not necessarily with the absence of factual differences between classes.

[6]We call these protocols restrictive because, unlike the complete KG setting, they reduce the evidence available for truth value computation by removing the article itself, the evaluated triple, or by splitting the data into training and test sets.

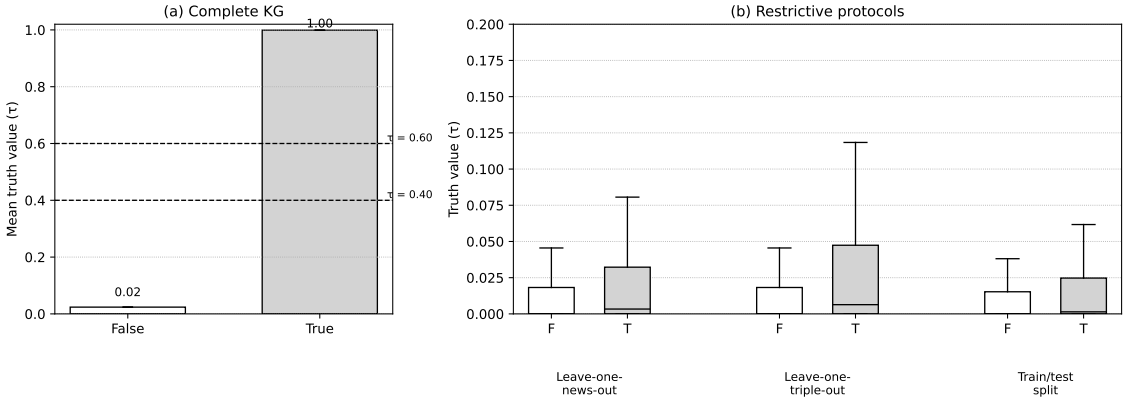

Figure 2. Truth values by class and protocol. Panel (a) shows the mean value in the complete KG setting, while panel (b) presents the distribution of values in the restrictive protocols.

4.3. Comparison with OIE

Table 5 compares extraction with Sabiá 4 to the OIE-based system reported in [Santos et al. 2026]. The comparison is made only in the complete KG setting, since this is the closest to the configuration presented in the reference work; the other protocols have no direct equivalent.

Extraction	Acc.	Prec.	Rec.	F1
OIE [Santos et al. 2026]	0.9359	0.8887	0.9966	0.9396
LLM (Sabiá 4)	0.9992	0.9993	0.9991	0.9992

Table 5. Comparison between OIE and LLM (Sabiá 4) in the closed evaluation setting.

Because the OIE results are available only for the complete KG setting, this comparison supports conclusions only about factual retrieval in the closed setting. It does not establish that Sabiá 4 is more robust than OIE when direct evidence is removed or when previously unseen articles are evaluated. Reproducing the restrictive protocols with OIE-based extraction is therefore necessary for a broader comparison.

4.4. Discussion and Qualitative Analysis

The results should be interpreted in light of the role of the KG in the method. The graph works as a structured memory of events and relations extracted from true news articles. Thus, the score expresses the degree of factual support available in the KG, rather than absolute proof of truth or falsity. When factual evidence is removed or was never inserted into the graph, many true articles receive low truth values due to lack of retrievable support.

In a practical deployment, this factual memory would require continuous maintenance. Newly verified news articles could be periodically ingested with source provenance and temporal metadata attached to each triple. Such metadata would help distinguish current evidence from outdated or conflicting information and support the revision or removal of facts that are no longer valid.

The qualitative analysis of triples helps explain this behavior. Sabiá 4 preserves information relevant to journalistic texts, such as speech attribution, modality, and contextual relations, distinguishing statements such as *"o ministro afirmou que a medida será revista"* (the minister stated that the measure will be reviewed), *"o ministro negou a medida"* (the minister denied the measure), and *"a medida foi revista"* (the measure was reviewed). At the same time, we observed long subjects, objects, and relations, as well as textual variation for semantically close entities. These phenomena enrich the representation, but also fragment the graph.

For example, forms such as *"presidente Lula"* (President Lula), *"Lula"*, and *"Luiz Inácio Lula da Silva"* may refer to the same entity, while relations such as *"informou"* (reported), *"afirmou"* (stated), *"declarou"* (declared), and *"disse"* (said) may play similar discourse roles. Since they appear as distinct nodes and relations in the KG, they reduce recurrence and make it harder to form paths between related facts.

Thus, the comparison with OIE indicates that LLMs improve factual retrieval in the closed setting. However, the restrictive protocols reveal a broader limitation of KG-based approaches: their dependence on graph coverage and connectivity. This limitation may affect triples extracted by both LLMs and OIE. In the case of LLMs, the greater textual richness of the triples may increase fragmentation when there is no entity and relation normalization.

5. Conclusion

This paper evaluated the use of LLMs for triple extraction in a knowledge graph-based fake news detection system for Portuguese. Starting from a previously proposed approach, we replaced the OIE-based extraction step with triples generated by Sabiá 4, while keeping the remaining pipeline components unchanged. We then analyzed how the factual representation extracted from news articles affects KG construction, factual support retrieval, and veracity inference.

The results indicate that the approach is promising for factual verification in Portuguese when the relevant events are represented in the KG. In the complete KG setting, LLM-based extraction provides strong factual retrieval and outperforms the OIE-based configuration in the closed evaluation setting. In the restrictive protocols, however, the greater expressiveness of the triples does not guarantee greater factual support when direct evidence is removed or when the evaluated facts fall outside the KG coverage. This behavior may also occur with OIE and should be investigated in future studies that replicate the same protocols with different extractors.

These findings reinforce that extraction is a decisive step in graph-based fact-checking systems, as it determines which facts are stored, how they are connected, and which evidence can be retrieved. Future work should compare LLMs from different families and sizes, including standard instruction-following and reasoning-oriented configurations, under the same extraction and evaluation protocols. We also plan to investigate entity and relation normalization, coreference resolution, entity disambiguation, and the use of ontologies or lightweight semantic schemas to reduce KG fragmentation. In addition, the graph should be continuously expanded with newly verified news articles, while preserving source provenance and temporal metadata to distinguish current evidence from outdated or conflicting information. Finally, we intend to explicitly incorporate the rela-

tion into the truth value computation, allowing the method to distinguish triples with the same pair of entities but different relations, such as *"o ministro anunciou a medida"* (the minister announced the measure) and *"o ministro negou a medida"* (the minister denied the measure).

Limitations

This work has some limitations. First, the truth value computed by the method should be interpreted as the degree of factual support retrievable from the available KG, rather than as absolute proof of truth or falsity. Thus, true news articles whose facts are not yet represented in the graph may receive low support. Second, although extracted relations are stored as edge labels, the current computation mainly considers connectivity between subject and object. This may assign similar support to triples with the same pair of entities but different relations. Third, LLM-based extraction may generate textual variation, redundancy, or long relations, contributing to KG fragmentation. Fourth, the experiments use a single LLM, Sabiá 4. Therefore, the results do not isolate the effects of model family, size, training data, or reasoning-oriented inference and should not be generalized to LLM-based extractors as a whole. Finally, the restrictive protocols were applied to the LLM-based configuration, but still need to be reproduced with OIE-based configurations for a broader comparison.

Ethics Statement

This work uses only public datasets already made available for research on fake news detection in Portuguese. We did not collect personal data directly from users, nor did we conduct interventions with human participants. The news articles are handled in aggregate form and used exclusively for experimental evaluation. We acknowledge, however, that automatic misinformation detection systems may produce errors, especially when factual coverage in the KG is limited. Therefore, the method's outputs should be interpreted as support for human analysis, not as definitive decisions about the veracity of a news article.

Generative AI Use

Generative AI tools were used to support textual revision of the manuscript, including grammatical and spelling corrections, and limited assistance with LaTeX formatting, such as emphasis, tables, and section structure. Generative AI was not used to produce experimental results, modify data, generate metrics, or replace the authors' scientific analysis. All methodological decisions, interpretations of results, and final versions of the text were reviewed and validated by the authors.

Acknowledgments

This research was fully supported by the Federal University of Bahia (UFBA) through the JovemPesq Program, under Call for Proposals PRPPG No. 023/2025. The authors also gratefully acknowledge the financial support provided by the Brazilian National Council for Scientific and Technological Development (CNPq).

References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). DBpedia: A nucleus for a web of open data. In *The Semantic Web*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735, Berlin, Heidelberg. Springer.

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., and Etzioni, O. (2007). Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 2670–2676. Morgan Kaufmann Publishers Inc.

Chavarro, J. P., Carvalho, D. A., and Rezende, S. O. (2023). FakeTrueBR: Um corpus brasileiro de notícias falsas. In *Anais da XVIII Escola Regional de Banco de Dados*, pages 53–62, Porto Alegre, RS, Brasil. SBC.

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. *PLOS ONE*, 10(6):e0128193.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271.

Garcia, G. L., Afonso, L. C. S., and Papa, J. P. (2022). FakeRecogna: A new brazilian corpus for fake news detection. In Pinheiro, V., Gamallo, P., Amaro, R., Scarton, C., Batista, F., Silva, D., Magro, C., and Pinto, H., editors, *Computational Processing of the Portuguese Language*, volume 13208 of *Lecture Notes in Computer Science*, pages 57–67, Cham. Springer.

Geifman, Y. and El-Yaniv, R. (2017). Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30.

Guo, Z., Schlichtkrull, M., and Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10:178–206.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, pages 11–15.

Laitz, T., Almeida, T. S., Abonizio, H., Malaquias Junior, R., Bonás, G. K., Piau, M., Larcher, C., Pires, R., and Nogueira, R. (2026). Sabiá-4 technical report.

Mihindukulasooriya, N., Tiwari, A., Galkin, M., Costabello, L., and Simperl, E. (2025). Automatic prompt optimization for knowledge graph construction with large language models. In *Proceedings of the Workshops of the VLDB 2025 Conference*.

Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., Almeida, T. A., Ruiz, E. E. S., and Vale, O. A. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *Computational Processing of the Portuguese Language*, volume 11122 of *Lecture Notes in Computer Science*, pages 324–334, Cham. Springer.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R.,

Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H. W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S. P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S. S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan, T., Łukasz Kaiser, Kamali, A., Kanitscheider, I., Keskar, N. S., Khan, T., Kilpatrick, L., Kim, J. W., Kim, C., Kim, Y., Kirchner, J. H., Kiros, J., Knight, M., Kokotajlo, D., Łukasz Kondraciuk, Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C. M., Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew, B., McKinney, S. M., McLeavey, C., McMillan, P., McNeil, J., Medina, D., Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing, D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak, R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino, A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M., Peng, A., Perelman, A., de Avila Belbute Peres, F., Petrov, M., de Oliveira Pinto, H. P., Michael, Pokorny, Pokrass, M., Pong, V. H., Powell, T., Power, A., Power, B., Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C., Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama, K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F. P., Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M. B., Tillet, P., Tootoonchian, A., Tseng, E., Tuggle, P., Turley, N., Tworek, J., Uribe, J. F. C., Vallone, A., Vijayvergiya, A., Voss, C., Wainwright, C., Wang, J. J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T., Zhuang, J., Zhuk, W., and Zoph, B. (2024). Gpt-4 technical report.

Santos, L. d., Santos, M. R. E., Souza, Y. S., Sousa, J. P. H., and Santos, R. L. d. S. (2026). Exploring knowledge graphs for automatic fake news detection in Portuguese. In *Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026) - Vol. 1*, pages 446–455, Salvador, Brazil. Association for Computational Linguistics.

Santos, R. L. d. S. (2022). *Detecção automática de notícias falsas em português*. Tese (doutorado em ciências – ciências de computação e matemática computacional), Universidade de São Paulo, São Carlos.

Santos, R. L. d. S. and Pardo, T. A. S. (2020). Fact-checking for portuguese: Knowledge graph and google search-based methods. In Quaresma, P., Vieira, R., Aluísio, S.,

Moniz, H., Batista, F., and Gonçalves, T., editors, *Computational Processing of the Portuguese Language*, volume 12037 of *Lecture Notes in Computer Science*, pages 195–205, Cham. Springer.

Silva, R. M., Santos, R. L., Almeida, T. A., and Pardo, T. A. (2020). Towards automatically filtering fake news in portuguese. *Expert Systems with Applications*, 146:113199.

Stanovsky, G. and Dagan, I. (2018). Supervised open information extraction. In *Proceedings of NAACL-HLT*, pages 885–895.

Thorne, J., Vlachos, A., Christodoulopoulos, C., and Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana. Association for Computational Linguistics.

Van Cauter, Z., Demeester, T., Vercruyssen, V., and Ongenae, F. (2024). Ontology-guided knowledge graph construction from maintenance short texts. In *Proceedings of the 1st Workshop on Knowledge Augmented Large Language Models*, pages 79–88. Association for Computational Linguistics.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Xu, D., Chen, W., Peng, W., Zhang, C., Xu, T., Zhao, X., Wu, X., Zheng, Y., and Wang, E. (2024). Large language models for generative information extraction: A survey. *Frontiers of Computer Science*, 18(6):186357.