KGFakeNet: A Knowledge Graph-Enhanced Model for Fake News Detection

Anuj Kumar[1], **Pardeep Kumar**[2], **Abhishek Yadav**[1], **Satyadev Ahlawat**[2], **Yamuna Prasad**[1]
[1]Department of Computer Science & Engineering, Indian Institute of Technology Jammu, India
[2]Department of Electrical Engineering, Indian Institute of Technology Jammu, India
{anuj,pardeep.kumar,2021pcs1015,satyadev.ahlawat,yamuna.prasad}@iitjammu.ac.in

Abstract

The proliferation of fake news on social media has intensified the spread of misinformation, promoting societal biases, hate, and violence. While recent advancements in Generative AI (GenAI), particularly large language models (LLMs), have shown promise, these models often need more structured representation for accurate verification, as they rely on pre-trained data patterns without access to real-time or validated information. This study presents a framework that utilizes Open Information Extractor 6 (OpenIE6) to extract triplet relationships (subject-predicate-object) from statements and justifications to compute the cosine similarity between the Knowledge Graphs (KGs) of the statements and their supporting justification to precisely measure the relevance and alignment between them. This similarity feature is integrated with an attention mechanism over GenAI-generated embeddings to enhance the model's ability to capture semantic features accurately. In addition, a Multi-Layer Perceptron (MLP) classifier is employed to integrate all features, resulting in a 4% improvement in accuracy and a 5% increase in F1-score over state-of-the-art LLM-based approaches.

1 Introduction

"A lie gets halfway around the world before the truth has a chance to get its boots on."

Attributed to Winston Churchill

The rapid spread of misleading and factually inaccurate information, commonly called fake news, has become a critical issue in the digital age. Misinformation disrupts democratic processes, distorts public discourse, and misguides individual decision-making (Sharma et al., 2019). With digital platforms serving as the primary source for news consumption, the influence of fake news has grown exponentially, leading to significant societal consequences. These platforms allow for the swift dissemination of information, creating an environment where fake news can influence elections, deepen societal divides, and, in extreme cases, incite violence.

As the volume of information expands, traditional manual fact-checking methods cannot keep pace, highlighting the need for automated detection systems. Moreover, early models for fake news detection(Girgis et al., 2018; Trueman et al., 2021; Long et al., 2017) primarily focused on statement-level analysis, classifying statements as true or false. However, real-world misinformation often blends truth and falsehood, defying simple categorization. This complexity has led to the use of intermediate truth classifications—such as "half true," "barely true," and "mostly false"—which better capture the nuanced nature of many news items and emphasize the need for more sophisticated detection models.

To address this, the LIAR dataset by (Wang, 2017), sourced from POLITIFACT[1], introduced a more granular classification approach, categorizing statements across six levels of truthfulness, from "true" to "pants on fire". Models using the LIAR dataset have utilized linguistic features such as emotional tone, hedging, and speaker-related attributes (Thorne and Vlachos, 2018) to improve detection. However, while the LIAR dataset supports nuanced classification, it lacks external evidence. To address this gap, the LIAR-PLUS dataset (Alhindi et al., 2018) extends LIAR by providing additional contextual information, including justifications and detailed fact-checking verdicts for each labeled statement.

With the growing prominence of Generative AI (GenAI) models (Brown et al., 2020; AI, 2024) in Natural Language Processing (NLP) tasks such as machine translation, text classification, and data extraction, these models have also been explored

[1]https://www.politifact.com/

Proceedings of the Generative AI and Knowledge Graph Workshop (GenAIK), pages 109–122
January 19, 2025.

for applications in fake news detection (Hu et al., 2022). Although powerful, GenAI models, particularly large language models (LLMs), are limited by their lack of structured representation, as they largely rely on patterns from pre-trained datasets without access to real-time, validated information sources. This limitation raises concerns about the reliability of GenAI models in fact-checking tasks, as they lack the capability to cross-reference real-world facts dynamically. Recent research has addressed this limitation by integrating knowledge graphs (KGs) with GenAI models (Gu and Krenn, 2024), utilizing KG-structured, entity-based representations to enrich models with factual and contextually relevant information.

Knowledge Graph (KG) embeddings allow models to capture relationships and concepts in a structured format, storing entities (like "Paris" or "Einstein") and their relationships (e.g., "is the capital of," "was born in") in a graph form that enables efficient retrieval of factual information. By leveraging KG embeddings capture structured relationships that enhance verification, helping reduce hallucinations and improve accuracy. Moreover, Open Information Extraction (OpenIE) (Banko et al., 2007) has become essential for transforming unstructured text into structured knowledge. OpenIE6[2] (Kolluru et al., 2020), the latest research, enables the extraction of factual statements by identifying subject-predicate-object triplets, which form the backbone of knowledge graphs. OpenIE6 surpasses earlier versions with improved contextual accuracy and scalability, making it particularly effective for large-scale data sources. By allowing dynamic extraction of factual relationships directly from a wide array of sources, OpenIE6 equips models with up-to-date, contextually relevant information—a valuable trait for domains requiring dynamic knowledge updates.

Our proposed framework, KGNewsNet, leverages OpenIE6 to extract triplet relationships from statements and justifications, generating Knowledge Graphs (KGs) that capture structured semantic relationships. Afterwards, the cosine similarity between the statement and justification KGs is computed to generate a feature that quantifies alignment between statement and their justifications. The KG embeddings and OpenIE6 reduce hallucinations by grounding answers in structured data. Combined with GenAI-generated embeddings and enhanced by an attention mechanism, this feature enables our model to prioritize relevant aspects of the statement-justification pairs, improving overall detection accuracy. Additionally, a Multi-Layer Perceptron (MLP) classifier integrates these features, yielding substantial improvements in detection performance. Our model achieves a 4% increase in accuracy and a 5% boost in F1-score over existing LLM-based approaches, demonstrating the efficacy of KG-enhanced fact-checking. The key contributions of this work are as follows:

- **OpenIE6-Driven Knowledge Graph Integration:** We integrate KGs generated by OpenIE6 to provide structured, context-rich knowledge that strengthens the structured representation of the GenAI model, enhancing statement verification.
- **Enhanced statement-justification Alignment:** Our framework employs an attention mechanism that emphasizes critical aspects of statement-justification pairs, utilizing a specialized attention module within the GenAI model to improve semantic comprehension of misinformation. Additionally, a cosine similarity feature derived from the Knowledge Graph further refines the alignment, enhancing the model's ability to verify statements accurately.
- **Enhanced Detection Performance with Multi-Layer Perceptron (MLP) Classifier:** Integrating KG-based features, GenAI embeddings, and attention yields substantial performance improvements, achieving a 45.4% accuracy in six-class classification and outperforming established GenAI models.

2 Related Work

Identifying fake news has evolved through extensive research, progressing from traditional fact-checking approaches to advanced machine learning and deep learning techniques. Initial methods primarily relied on manual fact-checking and information retrieval, but as the volume of online misinformation increased, the demand for automated solutions became imperative. Research in this area has largely focused on combining natural language processing (NLP) with machine learning to identify linguistic and thematic patterns indicative of fake news. For example, Latent Dirichlet Allocation (LDA) (Casillo et al., 2021) has been employed

[2] https://github.com/dair-iitd/openie6

to reveal hidden topics within news content, highlighting patterns that suggest deceptive intent. This early content-based approach has provided a foundational technique for detecting inconsistencies in fabricated stories. Supervised learning algorithms, including support vector machines (SVMs) and random forests, have also demonstrated effectiveness in misinformation classification, leveraging labeled data to identify fake news.

The advent of deep learning methods, such as convolutional neural networks (CNNs) and bidirectional long short-term memory networks (BiLSTMs), further improved detection by capturing nuanced textual features. Introducing pre-trained language models (PLMs) like BERT (Devlin et al., 2019; Kotonya and Toni, 2020; Shu et al., 2019; Yang et al., 2022a; Atanasova et al., 2020) marked a significant advancement in the field, as these models harness vast corpora to recognize complex language structures, capturing subtler cues of misinformation. Recent studies have also explored the role of influential users in amplifying misinformation, contributing valuable perspectives for detection systems targeting social network effects.

Label	Train	Validation	Test
Barely True	1654	237	212
False	1995	263	249
Half True	2114	248	265
Mostly True	1962	251	241
Pants on Fire	839	116	92
True	1676	169	208
Total	**10240**	**1284**	**1267**

Table 1: Dataset Statistics showing the distribution of labels across training, validation, and test splits.

The automated fact-checking systems have emerged as essential tools for combating misinformation, scaling up the verification process by cross-referencing claims with authoritative sources. Complementary approaches, including source verification, metadata analysis, and digital forensics, enhance these systems by assessing the credibility of information sources. A prominent advancement in this area involves integrating external knowledge bases (KBs) with PLMs to improve claim verification. Models like ERINE (Zhang et al., 2019) and TagMe (Ferragina and Scaiella, 2010) leverage structured factual data from repositories such as WikiData (Vrandečić and Krötzsch, 2014), allowing for more robust fact-checking by anchoring statements in verified, external data. However, despite improved accuracy, challenges remain in ensuring that relevant knowledge is effectively applied to specific statements, with issues often arising from the overgeneralization or irrelevance of incorporated knowledge. Addressing these limitations requires a balance between leveraging external data and maintaining relevance to the context of the claims being verified.

The rise of Generative AI (GenAI) models has introduced new possibilities for scalable misinformation detection by utilizing advanced language understanding and generation capabilities (Hu et al., 2022). Instruction-following models like InstructGPT (Ouyang et al., 2022) and Self-Instruct (Wang et al., 2023) have demonstrated efficacy in validating content by following structured prompts, combining data analysis with instruction-based guidelines to enhance claim verification. ChatGPT (OpenAI, 2024) adds a conversational aspect to fact-checking, enabling real-time, interactive validation through human-like dialogue, though its proprietary constraints limit customization for broader research applications.

Open-source alternatives, such as Stanford's Alpaca (Taori et al., 2023) built on the LLaMA framework (Touvron et al., 2023), offer more flexible options, allowing researchers to integrate external knowledge sources and customize models for specific applications. Recent research continues to explore instruction-following GenAI for misinformation detection, as seen in (Cheung and Lam, 2023), where external evidence retrieval is combined with instruction-based models, and in (Wang et al., 2024), which employs prompt-based modules to generate claim justifications. However, these models still face significant challenges with hallucinations, particularly in cases lacking structured, factual grounding. Integrating GenAI models with knowledge-rich databases can help mitigate this issue by supporting accuracy and consistency in generated responses, providing a clearer factual foundation for misinformation detection.

Our work, KGNewsNet, builds upon these advancements by addressing key limitations in GenAI-based misinformation detection models, particularly the insufficient integration of external evidence in models like (Cheung and Lam, 2023). KGNewsNet enhances fake news detection by combining knowledge graphs (KGs) with the LIAR-PLUS dataset, leveraging KG embeddings and attention mechanisms to incorporate structured,

Index	Column	Liar-Plus
1	ID	11972.json
2	Label	TRUE
3	Statement	Building a wall on the U.S.-Mexico border will take literally years.
4	Subject	Immigration
5	Speaker	Rick Perry
6	Job Title	Governor
7	State Info	Texas
8	Party Affiliation	Republican
9	True Counts	30
10	Mostly True Counts	30
11	Half True Counts	42
12	False Counts	23
13	Pants on Fire Counts	18
14	Context	Radio interview
15	Justification	Meantime, engineering experts agree the wall would most likely take years to complete. Keep in mind, too, that it took more than six years to build roughly 700 miles of fence and barriers along the roughly 2,000-mile U.S.-Mexico border.

Table 2: Example entry from the LIAR-PLUS dataset, showcasing metadata such as speaker details, historical truthfulness counts, and a justification for the claim.

external data into the model more effectively. This approach provides a grounded, contextually relevant framework that addresses gaps in existing PLM-based models. Our experimental results indicate a substantial improvement, with KGNewsNet achieving an accuracy of 0.454, outperforming comparable models and demonstrating its effectiveness in misinformation detection.

3 Preliminaries

3.1 Problem Definition

This work aims to develop a model that can accurately classify statements into multiple truthfulness categories by leveraging external knowledge and justifications. Let $S = \{s_1, s_2, \ldots, s_n\}$ and $J = \{j_1, j_2, \ldots, j_n\}$ represent the set of statements and justification to be classified, where each statement s_i and j_i is associated with a truthfulness label, the goal is to predict its truthfulness label $y_i \in \{c_1, c_2, \ldots, c_k\}$, with $k = 6$ corresponding to the six truthfulness categories (e.g., true, mostly true, half-true, barely true, false, and pants on fire) by considering; The textual content of the statement s_i. The corresponding justification J_i provides factual support or context for the statement. External knowledge K is derived from a knowledge graph (KG) that includes relevant factual information. Metadata M, such as the speaker information and context.

Thus, the classification function can be defined as:

$$\hat{y}_i = f(s_i, J_i, K, M)$$

Where f is the model that learns to map the combination of the statement, justification, external knowledge, and metadata to the correct truthfulness category.

The model aims to minimize the classification error across all statements in the dataset:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(f(s_i, J_i, K, M), y_i)$$

Where $\mathcal{L}$ is the loss function (e.g., categorical cross-entropy loss) and n is the number of statements.

3.2 Dataset

The development of our advanced fact-checking model began with the use of the LIAR-PLUS dataset(Alhindi et al., 2018), an enhanced version of the original LIAR dataset(Wang, 2017). Compiled by Alhindi et al.(Alhindi et al., 2018), LIAR-PLUS consists of approximately 12.8K annotated short statements. This dataset extends beyond the original LIAR dataset by incorporating justifications, which provide essential context and explanations for each statement's truthfulness classification.

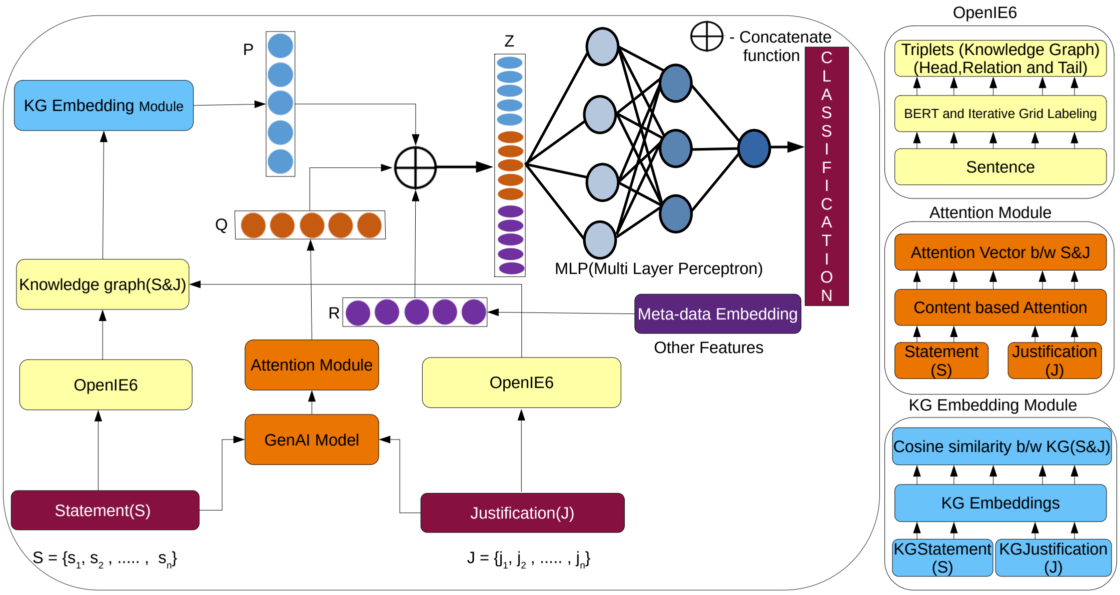

Figure 1: This framework represents the architecture of KGNewsNet computes value P (from TransE embeddings) and Q (from GPT (Black et al., 2022) embeddings) using attention mechanisms. These, along with metadata vector R, are concatenated into Z and passed through an MLP for final classification.

The LIAR-PLUS dataset is divided into three distinct subsets as shown in table **??**, each of which plays a crucial role in model training, validation, and testing. A key feature of the LIAR-PLUS dataset is the inclusion of a "Justification" column, which offers textual explanations or evidence supporting each statement's verdict. This addition enhances the fact-checking process by providing contextual information that the model can leverage when determining the truthfulness of a statement. By incorporating these justifications, the model can make more informed and accurate decisions based on the statement and the rationale behind the verdict.

Table 2 outlines the feature structure in the LIAR dataset, where rows 1 to 15 represent various data points. Feature 1 provides the label, while Feature 2 includes the main statement or news text, which forms the primary content analyzed for truthfulness. Contextual details are provided by features $\{4, 5, 6, 7, 8, 15\}$: feature 4 specifies the subject, offering insight into the topic; feature 5 identifies the speaker, establishing the source; feature 6 includes the speaker's job title, adding professional background; feature 7 provides state information, offering geographical context; feature 8 denotes the speaker's party affiliation; and feature 15 offers additional context to enrich the background of the statement. Additionally, features $\{9, \ldots, 13\}$ detail the speaker's historical truthfulness record by counting previous statements categorized by veracity, which is crucial for assessing credibility and adds an essential dimension to the analysis.

The wealth of information in the LIAR-PLUS dataset offered us a unique opportunity to delve deeper into fact-checking. It allowed our model to harness the statement's words and underlying justifications, producing a more nuanced and accurate understanding of truth in an age where misinformation often spreads unchecked.

4 Methodology

This section describes the methodology used to develop our proposed model, KGNewsNet, shown in Figure 1. The model integrates multiple layers of textual analysis, knowledge graph embeddings, and metadata, including sentiment analysis, to improve the accuracy of fake news detection.

4.1 Attention Module

Let S_i and J_i represent the i-th statement and its corresponding justification, respectively, where each has up to n_i tokens. Each token in S_i and J_i is using GPT-NeoX(Black et al., 2022) to get token embeddings. Let $\mathbf{s}_{i,t} \in \mathbb{R}^D$ and $\mathbf{j}_{i,t} \in \mathbb{R}^D$ denote the embedding vectors for the t-th tokens in the i-th statement S_i and justification J_i, respectively.

Thus, we have

$$S_i = \{\mathbf{s}_{i,1}, \mathbf{s}_{i,2}, \ldots, \mathbf{s}_{i,n_i}\}$$

$$J_i = \{\mathbf{j}_{i,1}, \mathbf{j}_{i,2}, \ldots, \mathbf{j}_{i,n_i}\}$$

where $n_i = 512$ is the number of tokens in S_i and J_i and padding is applied if S_i or J_i contains fewer than n_i tokens.

To capture the alignment between each statement-justification pair, we compute a matrix of attention weights α_i between S_i & J_i. The attention weight $\alpha_{i,t,u}$ between the t-th token in S_i and the u-th token in J_i is given by

$$\alpha_{i,t,u} = \frac{\exp\left(\frac{\mathbf{s}_{i,t} \cdot \mathbf{j}_{i,u}}{\sqrt{D}}\right)}{\sum_{t=1,v=1}^{n_i} \exp\left(\frac{\mathbf{s}_{i,t} \cdot \mathbf{j}_{i,v}}{\sqrt{D}}\right)} \quad (1)$$

where $\mathbf{s}_{i,t} \cdot \mathbf{j}_{i,u}$ denotes the dot product between the t-th token embedding in S_i and the u-th token embedding in J_i, computed as

$$\mathbf{s}_{i,t} \cdot \mathbf{j}_{i,u} = \sum_{p=1}^{D} s_{i,t,p}\, j_{i,u,p} \quad (2)$$

and $D = 150$ is the dimensionality of the embeddings. The softmax normalization ensures that $\alpha_{i,t,u}$ forms a probability distribution over the tokens in J_i for each token in S_i, capturing the relative alignment of each justification token with respect to each statement token.

Next, we construct a context-aware representation $\mathbf{c}_{i,t}$ for each token $\mathbf{s}_{i,t}$ in the statement S_i by computing a weighted sum of the token embeddings in J_i based on the attention weights:

$$\mathbf{c}_{i,t} = \sum_{u=1}^{n_i} \alpha_{i,t,u}\, \mathbf{j}_{i,u} \quad (3)$$

where $\mathbf{c}_{i,t} \in \mathbb{R}^D$ is the attended representation of the t-th token in S_i, taking into account its alignment with each token in J_i.

To obtain a single content-based attention vector Q_i for each statement S_i that incorporates the context from the justification J_i, we aggregate the $\mathbf{c}_{i,t}$ vectors across all tokens in S_i. We use average pooling over the $\mathbf{c}_{i,t}$ vectors to produce Q_i:

$$Q_i = \frac{1}{n_i} \sum_{t=1}^{n_i} \mathbf{c}_{i,t} \quad (4)$$

Where $Q_i \in \mathbb{R}^D$ is the content-based attention vector that summarizes the alignment between each statement S_i and its corresponding justification J_i across all tokens. This process effectively captures token-level alignment for multiple statement-justification pairs, yielding a context-aware representation for each statement based on its justification.

4.2 Knowledge Graph Extraction and Embedding Module

Our dataset contains n statement-justification pairs, where each statement S_i has a corresponding justification J_i. For each i-th statement-justification pair, we use OpenIE6(Kolluru et al., 2020) to extract structured knowledge graphs in the form of triplets (h, r, t), where h is the head entity, r is the relation, and t is the tail entity. For more information about OpenIE6, please refer to Appendix A.3.

Since a single statement or justification can yield multiple triplets, we limit the number of extracted triplets to a maximum of m_1 for statements and m_2 for justifications, where m_1 = 3 and m_2 = 6 as expressed in (Kolluru et al., 2020) it reflects the length of statements and justifications. Let the sets of triplets extracted from the i-th statement S_i and justification J_i be represented as:

$$\mathrm{KG}_{S_i} = \{(h_{i,k}^S, r_{i,k}^S, t_{i,k}^S)\}_{k=1}^{\min(n_{S_i}, m_1)}$$

$$\mathrm{KG}_{J_i} = \{(h_{i,l}^J, r_{i,l}^J, t_{i,l}^J)\}_{l=1}^{\min(n_{J_i}, m_2)}$$

where n_{S_i} and n_{J_i} are the total number of possible triplets extracted by OpenIE6 from S_i and J_i, respectively.

Each triplet (h, r, t) is then represented by embeddings $\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$ using TransE. The TransE model ensures that the relationship holds by approximating the translation:

$$\mathbf{h} + \mathbf{r} \approx \mathbf{t} \quad (5)$$

where $\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$ are the vector representations for the head, relation, and tail entities, respectively, typically in $\mathbb{R}^D$ where D is the embedding dimensionality.

For each possible pair of triplets from KG_{S_i} and KG_{J_i}, we compute the cosine similarity (CS) between their embeddings. This results in a total of $m_1 \times m_2 = M$ cosine similarity values for each statement-justification pair. Each pairwise cosine similarity is computed as follows:

$$\mathrm{CS}((\mathrm{h}_{i,k}^S, r_{i,k}^S, t_{i,k}^S), (h_{i,l}^J, r_{i,l}^J, t_{i,l}^J)) = \frac{\mathbf{h}_{i,k} \cdot \mathbf{t}_{i,l}}{\sqrt{\|\mathbf{h}_{i,k}\|^2} \cdot \sqrt{\|\mathbf{t}_{i,l}\|^2}} \quad (6)$$

Models	Accuracy	F1-Score
LSTM(Girgis et al., 2018)	0.224	0.217
Hybrid CNN(Girgis et al., 2018)	0.247	0.274
SNN (LM + KG + KG-ENTITY)(Koloski et al., 2022)	0.267	0.267
KnowBert-W+W(Whitehouse et al., 2022)	0.294	0.289
CofCED(Yang et al., 2022b)	0.294	0.295
AC-BiLSTM(Trueman et al., 2021)	0.338	0.351
P_Bi_LSTM(Alhindi et al., 2018)	0.374	0.361
CapsulNet(Goldani et al., 2021)	0.395	-
Hybrid LSTM(Long et al., 2017)	0.407	0.415
DSNDM + Att.(Chernyavskiy and Ilvovsky, 2020)	0.412	0.402
Generative AI Model Performances		
ChatGPT(OpenAI, 2024)	0.263	0.251
FactLLaMA(Cheung and Lam, 2023)	0.304	0.299
FactLLaMA$_{know}$(Cheung and Lam, 2023)	0.313	0.304
L-Defense$_{LLaMA2}$(Wang et al., 2024)	0.328	0.314
L-Defense$_{ChatGPT}$(Wang et al., 2024)	0.311	0.305
Proposed KGNewsNet without KG	0.441	0.436
Proposed KGNewsNet	**0.454**	**0.450**

Table 3: Comparison of the proposed KGNewsNet with previous state-of-the-art models. Results are evaluated based on Accuracy and F1-Score. The proposed KGNewsNet achieves the best performance.

where $\mathbf{h}_{i,k}$ and $\mathbf{t}_{i,l}$ are the embeddings of the head and tail entities in each pair of triplets from KG_{S_i} and KG_{J_i}, respectively, and $\|\mathbf{h}_{i,k}\|$ denotes the Euclidean norm of vector $\mathbf{h}_{i,k}$.

The final output of this module for each i-th statement-justification pair is a vector $\mathbf{P}_i \in \mathbb{R}^M$, containing cosine similarity scores:

$$\mathbf{P}_i = \left[\text{CS}((h_{i,k}^S, r_{i,k}^S, t_{i,k}^S), (h_{i,l}^J, r_{i,l}^J, t_{i,l}^J))\right]_{k=1,l=1}^{m_1,m_2} \tag{7}$$

This vector $\mathbf{P}_i$ captures the alignment between each combination of triplet pairs across the statement and justification. By iterating through each of the n statement-justification pairs, we maintain consistency and manageability in the knowledge graph representations while capturing detailed relational alignment within the text.

4.3 Feature Vector Construction and Classification

To build the final feature vector, we concatenate the attention module vector $\mathbf{Q}$, KG module vector $\mathbf{P}$, and metadata features $\mathbf{R}$. The metadata features $\mathbf{R}$ include the information shown in Table 2, such as speaker information, party affiliation, and count information that details the speaker's historical truthfulness record. This information, derived by counting previous statements categorized by veracity, is crucial for assessing credibility and adds an essential dimension to the analysis. The final feature vector $\mathbf{Z}$, which is a concatenation of all the above vectors, is defined as:

$$\mathbf{Z} = [\mathbf{Q}\|\mathbf{P}\|\mathbf{R}] \tag{8}$$

The feature vector $\mathbf{Z}$ is then passed into a Multi-Layer Perceptron (MLP) for classification, where the veracity prediction output $\hat{y}$ is calculated as:

$$\hat{y} = \frac{\exp(\mathbf{w}^T\mathbf{Z} + b)}{\sum_k \exp(\mathbf{w}_k^T\mathbf{Z} + b_k)} \tag{9}$$

where $\mathbf{w}$ and b represent the weight vector and bias for each class in the MLP, and $\mathbf{w}_k$ and b_k are the weight vector and bias for each potential class k. To optimize the model, we use categorical cross-entropy loss:

$$L = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log(\hat{y}_{i,k}) \tag{10}$$

where y_i is the true label for sample i and $\hat{y}_i$ is the predicted probability for the true class.

The KGNewsNet algorithm, as outlined in Algorithm 1, provides a detailed implementation of the methodology described in this work. The algorithm highlights how features from attention and KG alignment are fused with metadata and processed through a Multi-Layer Perceptron (MLP)

Algorithm 1 KGNewsNet: Fake News Detection

Data: $S = \{S_i\}$: Statements, $J = \{J_i\}$: Justifications, $\mathbf{R}$: Metadata, Knowledge Graph Triplets (Head, Relation, Tail)

Result: Veracity predictions $\hat{y} = \{\hat{y}_i\}$

Initialize Embed, attention mechanism, KG embedding lookup, and MLP classifier Set loss function L

for *epoch* $= 1$ *to* N **do**

foreach *batch* $(S_i, J_i, \mathbf{R}_i)$ **do**

Step 1: Compute token embeddings for statements and justifications:

$\mathbf{S}_{i,t} = \text{Embed}(s_{i,t}), \quad \mathbf{J}_{i,t} = \text{Embed}(j_{i,t})$

Step 2: Compute attention weights between tokens in S_i and J_i:

$\alpha_{i,t,u} = \frac{\exp(\frac{\mathbf{s}_{i,t} \cdot \mathbf{j}_{i,u}}{\sqrt{D}})}{\sum_{t=1,v=1}^{n_i} \exp(\frac{\mathbf{s}_{i,t} \cdot \mathbf{j}_{i,v}}{\sqrt{D}})}$

Step 3: Compute context-aware representation of S_i:

$Q_i = \frac{1}{n_i} \sum_{t=1}^{n_i} \sum_{u=1}^{n_i} \alpha_{i,t,u} \mathbf{j}_{i,u}$

Step 4: Retrieve KG embeddings for S_i and J_i from the lookup dictionary & Compute cosine similarity between all triplet pairs:

$\text{CS}((h^S_{i,k}, r^S_{i,k}, t^S_{i,k}), (h^J_{i,l}, r^J_{i,l}, t^J_{i,l})) = \frac{\mathbf{h}_{i,k} \cdot \mathbf{t}_{i,l}}{\sqrt{\|\mathbf{h}_{i,k}\|^2} \cdot \sqrt{\|\mathbf{t}_{i,l}\|^2}}$ Construct the cosine similarity vector:

$\mathbf{P}_i = \left[\text{CS}((h^S_{i,k}, r^S_{i,k}, t^S_{i,k}), (h^J_{i,l}, r^J_{i,l}, t^J_{i,l}))\right]_{k=1,l=1}^{m_1,m_2}$

Step 5: Concatenate features from attention, KG embeddings, and metadata:

$Z_i = [Q_i \| \mathbf{P}_i \| \mathbf{R}_i]$

Step 6: Perform classification using MLP:

$\hat{y}_i = \text{Softmax}(\mathbf{W} Z_i + \mathbf{b})$

Step 7: Compute loss and update model parameters:

$L = -\sum_{i=1}^{N} \sum_{k=1}^{K} y_{i,k} \log(\hat{y}_{i,k})$

end

Step 8: Evaluate metrics on validation data

end

Result: Compute $\hat{y}_i$ for unseen S_i, J_i using trained parameters.

for final classification. By following the step-by-step process.

Time Complexity: The overall time complexity of KGNewsNet algorithm 1 is $O(N \cdot B \cdot (T \cdot E + T^2 \cdot D + T^2 \cdot D + m_1 \cdot m_2 \cdot D + D + L \cdot D^2))$, where N is the number of epochs, B is the batch size, T is the token sequence length, E is the embedding computation cost, D is the embedding dimensions, m_1 and m_2 are the numbers of triplets in statements and justifications, and L is the number of MLP layers.

5 Experiments and Results

In this section, we present the experimental setup and results of our proposed model, KGNewsNet, as well as its comparison with other state-of-the-art models for fake news detection. The experiments are conducted on the LIAR-PLUS dataset, and the results demonstrate how the integration of statement, justification, and external knowledge representations leads to significant performance improvements. We use accuracy and F1-score as the evaluation metrics to benchmark KGNewsNet's performance.

5.1 Experimental Setup

The experiments were conducted in a cloud environment with 40 vCPUs, a Tesla V100-PCIE GPU with 32GB of memory, and 256GB of RAM, providing ample resources for efficient model training. We used the LIAR-PLUS dataset (Alhindi et al., 2018) for veracity prediction, leveraging tokenization, padding or truncating to a fixed length, and embedding generation as outlined in the Methodology section. KGNewsNet was trained using the LIAR-PLUS dataset (Alhindi et al., 2018), with preprocessing and embedding techniques outlined in the Methodology section. Additional details regarding the parameter details for result replication are provided in Appendix A.1.

5.2 Results

Table 3 presents the performance of KGNewsNet compared with previous state-of-the-art models, including traditional models, advanced architectures, and recent Generative AI approaches. Traditional models such as LSTM (Girgis et al., 2018), Hybrid CNN (Girgis et al., 2018), and KnowBert-W+W (Whitehouse et al., 2022) achieve moderate accuracy scores of 0.224, 0.247, and 0.294, respectively. Their limited performance can be attributed to the absence of structured knowledge integration, which restricts their ability to capture contextual and relational nuances in statements and justifications.

Advanced architectures, such as CapsuleNet (Goldani et al., 2021) and Hybrid LSTM (Long et al., 2017), introduce richer representational techniques, achieving accuracy scores of 0.395 and 0.407, respectively. Generative AI models like FactLLaMA$_{know}$ (Cheung and Lam, 2023) and L-Defense$_{LLaMA2}$ (Wang et al., 2024) show incremental gains, with accuracies of 0.313 and 0.328. However, these models struggle to match KGNewsNet's performance due to their lack of explicit knowledge integration. Generative models rely on pre-trained contextual embeddings but lack mechanisms to align statements with external knowledge, making it difficult to validate claims effectively. Furthermore, their probabilistic nature and sensitivity to prompt design often result in inconsistent performance, particularly for claims requiring nuanced reasoning or factual grounding.

KGNewsNet demonstrates the effectiveness of integrating Knowledge Graph (KG) embeddings to address these limitations. By leveraging KG triplets, the model achieves an accuracy of 0.454 and an F1-score of 0.450, outperforming all other methods. This improvement underscores the importance of knowledge grounding in aligning statements and justifications. The KG module enhances token-level alignment and enriches the content-based attention vector, enabling the model to capture complex relationships effectively.

As outlined in Algorithm 1, KGNewsNet's computational complexity. Unlike traditional models with linear operations or generative models relying on token embeddings, KGNewsNet introduces additional computational cost through explicit pairwise alignment between statements and justifications. This higher complexity enables superior performance in tasks requiring structured support and nuanced veracity detection.Additionally, Appendix A.2 illustrates triplet alignment and prediction results (Tables 4 and 5), showing strong alignment in "true" cases and partial alignment in "barely-true" or "half-true" cases. These examples highlight KGNewsNet's ability to capture contextual relationships while revealing challenges in distinguishing closely related truthfulness categories, pointing to potential refinements for interpreting nuanced distinctions.

6 Conclusion

This paper presents KGNewsNet, a model for fake news detection that harnesses statements, justifications, metadata, and external knowledge graph embeddings to enhance classification performance. The results indicate that incorporating external knowledge sources and meticulously extracting features from both statements and justifications are pivotal in advancing fake news detection accuracy. While the model achieves strong overall performance, there remain opportunities for improvement, particularly in addressing complex financial statements and nuanced claims requiring intricate reasoning.

7 Limitations

KGNewsNet demonstrates significant potential in leveraging Knowledge Graph (KG) triplet alignment for veracity assessment but faces several limitations. The reliance on OpenIE6 for triplet extraction often generates lengthy or overly detailed triplets, which can dilute focus on critical information and complicate alignment. The evaluation, conducted exclusively on the LIAR-PLUS dataset, aligns well with the model's capabilities but limits its generalizability to datasets with less structured justifications or evidence-based fact-checking (e.g., FEVER). Extending evaluations to diverse datasets and optimizing the computational overhead of KG embedding and triplet alignment processes remain key areas for future work. Additionally, further improvements in explainability, such as visualizing triplet alignment or providing user-friendly insights into the model's decisions, would enhance its applicability in real-world fact-checking scenarios.

Acknowledgments

IHUB NTIHAC FOUNDATION partially funds this research under project numbers IHUB-NTIHAC/2021/01/14 & IHUB-NTIHAC/2021/01/15.

References

Google AI. 2024. Palm: scaling language modeling with pathways. *J. Mach. Learn. Res.*, 24(1).

Tariq Alhindi, Savvas Petridis, and Smaranda Muresan. 2018. Where is your evidence: Improving fact-checking by justification modeling. In *Proceedings of the First Workshop on Fact Extraction and Verification (FEVER)*, pages 85–90.

Pepa Atanasova, Jakob Grue Simonsen, Christina Lioma, and Isabelle Augenstein. 2020. Generating fact checking explanations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7352–7364. Association for Computational Linguistics.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, page 2670–2676.

Sid Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, Michael Pieler, USVSN Sai Prashanth, Shivanshu Purohit, Laria Reynolds, Jonathan Tow, Ben Wang, and Samuel Weinbach. 2022. GPT-NeoX-20B: An open-source autoregressive language model. In *Proceedings of the ACL Workshop on Challenges & Perspectives in Creating Large Language Models*.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

Mario Casillo, Francesco Colace, Brij B. Gupta, Domenico Santaniello, and Carmine Valentino. 2021. Fake news detection using lda topic modeling and k-nearest neighbor classifier. In *Computational Data and Social Networks: 10th International Conference, CSoNet 2021*, pages 330–339. Springer International Publishing.

Alexander Chernyavskiy and Dmitry Ilvovsky. 2020. Dsndm: Deep siamese neural discourse model with attention for text pairs categorization and ranking. In *Proceedings of the First Workshop on Computational Approaches to Discourse*, pages 76–85.

Tsun-Hin Cheung and Kin-Man Lam. 2023. Factllama: Optimizing instruction-following language models with external knowledge for automated fact-checking. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 846–853.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186.

Paolo Ferragina and Ugo Scaiella. 2010. Tagme: On-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1625–1628.

Sherry Girgis, Eslam Amer, and Mahmoud Gadallah. 2018. Deep learning algorithms for detecting fake news in online text. In *Proceedings of the 13th International Conference on Computer Engineering and Systems (ICCES)*, pages 93–97.

Mohammad Hadi Goldani, Saeedeh Momtazi, and Reza Safabakhsh. 2021. Detecting fake news with capsule neural networks. *Applied Soft Computing*, 101:106991.

Xuemei Gu and Mario Krenn. 2024. Interesting scientific idea generation using knowledge graphs and llms: Evaluations with 100 research group leaders. *Preprint*, arXiv:2405.17044.

Linmei Hu, Siqi Wei, Ziwang Zhao, and Bin Wu. 2022. Deep learning for fake news detection: A comprehensive survey. *AI Open*, 3:133–155.

Keshav Kolluru, Vaibhav Adlakha, Samarth Aggarwal, Mausam, and Soumen Chakrabarti. 2020. Openie6: Iterative grid labeling and coordination analysis for open information extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, Seattle, U.S.A.

Boshko Koloski, Timen Stepišnik Perdih, Marko Robnik-Šikonja, Senja Pollak, and Blaž Škrlj. 2022. Knowledge graph informed fake news classification via heterogeneous representation ensembles. *Neurocomputing*, 496:208–226.

Neema Kotonya and Francesca Toni. 2020. Explainable automated fact-checking for public health claims. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online. Association for Computational Linguistics.

Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. 2017. Fake news detection through multi-perspective speaker profiles. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 252–256.

OpenAI. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc.

Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3):1–42.

Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. 2019. defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 395–405. Association for Computing Machinery.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford alpaca: An instruction-following llama model.

James Thorne and Andreas Vlachos. 2018. Automated fact checking: Task formulations, methods and future directions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3346–3359, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Tina Esther Trueman, Ashok Kumar, Narayanasamy P., and Vidya J. 2021. Attention-based c-bilstm for fake news detection. *Applied Soft Computing*, 110:107600.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Bo Wang, Jing Ma, Hongzhan Lin, Zhiwei Yang, Ruichao Yang, Yuan Tian, and Yi Chang. 2024. Explainable fake news detection with large language model via defense among competing wisdom. In *Proceedings of the ACM Web Conference 2024*, WWW '24, page 2452–2463. Association for Computing Machinery.

William Yang Wang. 2017. Liar, liar pants on fire: A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13484–13508. Association for Computational Linguistics.

Chenxi Whitehouse, Tillman Weyde, Pranava Madhyastha, and Nikos Komninos. 2022. Evaluation of fake news detection with knowledge-enhanced language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1425–1429.

Zhiwei Yang, Jing Ma, Hechang Chen, Hongzhan Lin, Ziyang Luo, and Yi Chang. 2022a. A coarse-to-fine cascaded evidence-distillation neural network for explainable fake news detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2608–2621. International Committee on Computational Linguistics.

Zhiwei Yang, Jing Ma, Hechang Chen, Hongzhan Lin, Ziyang Luo, and Yi Chang. 2022b. A coarse-to-fine cascaded evidence-distillation neural network for explainable fake news detection. *arXiv preprint arXiv:2209.14642*.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. Ernie: Enhanced language representation with informative entities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1441–1451, Florence, Italy. Association for Computational Linguistics.

A Appendix

A.1 Parameter Details

The experiment is conducted on LIAR-PLUS dataset (Alhindi et al., 2018) for veracity prediction, leveraging tokenization, padding or truncating to a fixed length, and embedding generation as outlined in the Methodology section. For classification, an MLPClassifier with the Adam solver was configured to ensure effective optimization. The learning rate initialization was set to 0.001, and the learning rate scheduling was adaptive, reducing the rate if no improvement was observed in validation performance, aiding convergence. The network architecture consisted of a single hidden layer with 50 neurons, balanced for computational efficiency and model complexity. ReLU activation was used to expedite training, and the batch size was set to 'auto', adjusting based on available memory. Additional controls included a tolerance level (tol) of 0.0001 to set a minimum threshold for performance improvement, and an early stopping criterion n_iter_no_change=10, halting training if no

Statement triplets: (Americans, working, now), (Americans, working, 70s), (Americans, working, less than in the 70s) **Justification triplets**: (Hartzler, talking about, decade of the 70s), (first eight years of the 70s, employment-population ratio, lower than 2015), (first eight years of the 70s, labor force participation rate, lower than 2015), (employment-population ratio, comparison, 2015 vs. 70s), (labor force participation rate, comparison, 2015 vs. 70s), (decade of the 70s, employment-population and labor force participation, lower than 2015) **Explanation**: The label is "barely-true." Partial alignment occurs as the justification triplets provide historical employment data in the 70s, but there's no direct comparison with "now," supporting only a partial truth.
Statement triplets: (Republicans, attacks, programs in stimulus plan), (programs in stimulus plan, not stimulative, less than 1 percent), (programs, account for, less than 1 percent of package) **Justification triplets**: (Obama, point, perspective in order), (legislators, quibbling over, small portion of spending), (publicized projects, represent, small portion of spending), (Republicans, said, large percentages of stimulus plan not stimulative), (stimulus plan, criticized by, Republicans), (spending, ineffective use of, taxpayer money) **Explanation**: The label is "half-true." Some alignment occurs as the justification acknowledges the criticism but lacks specifics about "less than 1 percent," resulting in partial support consistent with the "half-true" label.
Statement triplets: (Canada, created, more jobs), (time period, January), (Canada, created, more jobs than U.S.) **Justification triplets**: (November 2010, U.S. economy created, 93,000 jobs), (December 2010, U.S. created, 121,000 jobs), (recent months, U.S. job creation, exceeded Canada only in October), (January, U.S. job creation, especially low), (January, Canadian job creation, especially high), (comparison, job creation, Canada vs. U.S.) **Explanation**: The label is "true." Strong alignment as the justification confirms high Canadian job creation relative to the US in January, fully supporting the statement and matching the "true" label.

Table 4: Case Study of Triplet Alignment Between Statements and Justifications for Veracity Labels by highlighting aligned justification triplets. For "true" labels, strong alignment with multiple highlighted triplets provides clear support, while partial alignment in "half-true" or "barely-true" cases indicates incomplete support. The use of a Knowledge Graph (KG) structures these comparisons, capturing subtle distinctions and improving the accuracy of veracity assessment.

improvement was observed for 10 iterations. Early stopping was applied to prevent overfitting, and validation performance was monitored throughout training. This setup, with adaptive learning rates, controlled complexity, and early stopping, was optimized to achieve stable convergence and reliable generalizability on the LIAR-PLUS dataset.

A.2 Case study

Table 4 presents case studies that illustrate how our methodology, KGNewsNet, uses structured triplet alignment between statements and justifications to assess veracity accurately. In each example, KGNewsNet extracts key entities and relationships from both statements and justifications, creating triplets that are compared to determine factual alignment. By leveraging Knowledge Graph (KG) embeddings, our model captures not only the semantic content of each entity but also its contextual relationship within the statement, enabling nuanced verification.

For "true" labels, strong alignment is observed, with multiple justification triplets highlighted in green, providing direct and clear support for the statement. For instance, in the "Canada created more jobs than the U.S." example, both statement and justification triplets align on key factors like "job creation," "January," and "comparative performance," resulting in a high degree of factual support. This alignment showcases KGNewsNet's

Statement	Justifications	Label	Prediction
Pregnant women are at an increased risk of pre-term pregnancy by 80 percent.	"The statement attributes the statistic to the Women's Fund of Rhode Island".	barely-true	barely-true
Elizabeth Warren lied about wanting to abolish the Federal Minimum Wage.	"Trump said, "Elizabeth Warren lied about abolishing the Federal Minimum Wage." Yet, when Trump was asked if he would have a federal floor with states going higher, he replied, "No." She simply used Trump's own words".	barely-true	false
Every dollar the state spent on audits last year delivered $64 in cost savings.	"Brown said that for every dollar the Secretary of State spent on audits last year, it found $64 in cost savings. However, the total potential savings might be underestimated".	true	barely-true
Public display of a long rifle is perfectly legal in Texas.	"Texas law explicitly restricts handguns and some other weapons from being openly carried around. However, the law is silent on long rifles, meaning that their public display is legal".	true	true
We were the last flag flying in Benghazi.	"The meaning of the phrase "last flag flying" shifted from its original meaning as politicians used it as a rhetorical talking point. In his testimony, the phrase was used more rhetorically than literally".	false	false

Table 5: Prediction Results of KGNewsNet

ability to interpret context-sensitive data accurately, supported by the structured comparison of triplets that validates the statement comprehensively.

In contrast, examples with "half-true" or "barely-true" labels show only partial alignment, with fewer highlighted triplets in the justification. For the statement "Republicans attack the stimulus plan for programs that account for less than 1 percent of spending," some alignment is achieved as the justification acknowledges similar criticisms. However, it lacks explicit confirmation of the "less than 1 percent" detail, reflecting partial support for the statement. This partial alignment, captured through KG-guided triplet comparison, helps KGNewsNet differentiate between full and partial truths.

By structuring comparisons with KG triplets, KGNewsNet effectively reduces ambiguity in cases with close but distinct veracity labels. Table 5 further illustrates KGNewsNet's performance, where it accurately captures veracity by aligning entities, relationships, and contexts in diverse examples, including statements about economic data, policy claims, and public figures. This structured approach allows KGNewsNet to distinguish between factual alignment levels, refining its predictions with a greater degree of accuracy than conventional models. Through KG triplet alignment, our model benefits from enhanced tructured representation, yielding consistent and reliable veracity assessments across challenging, context-dependent statements.

A.3 OpenIE6

Our methodology leverages OpenIE6 for extracting structured triplets from statements and justifications, which enhances the accuracy and efficiency of Open Information Extraction (OpenIE) through its novel Iterative Grid Labeling (IGL) approach. OpenIE6 frames extraction as a 2-D grid labeling task, where rows represent potential extractions, and columns correspond to words in the sentence. This design enables rapid extraction processing without compromising on quality, as it reduces the need for repeated encoding steps common in earlier OpenIE systems.

To improve extraction comprehensiveness, OpenIE6 imposes constraints during training to ensure high recall, incorporating penalties for omitted information. Furthermore, it addresses complex coordination structures, such as conjunctive phrases,

Parameter	Description
`-mode splitpredict`	Enables prediction mode, allowing the model to split conjunctive structures for better extraction.
`-inp sentences.txt`	Specifies the input file containing sentences for which triplet relations are extracted.
`-out predictions.txt`	Sets the output file where extracted triplets will be saved.
`-rescoring`	Applies a rescoring mechanism to enhance prediction accuracy.
`-task oie`	Defines the task as Open Information Extraction (OIE).
`-gpus 1`	Configures the process to run on one GPU for computational efficiency.
`-oie_model models/oie_model/epoch=14_eval_acc=0.551_v0.ckpt`	Path to the pre-trained OpenIE model used for relation extraction.
`-conj_model models/conj_model/epoch=28_eval_acc=0.854.ckpt`	Path to the conjunction handling model that processes compound structures.
`-rescore_model models/rescore_model`	Path to the rescoring model to refine extraction accuracy.
`-num_extractions` $m_1 = 3$ `&` $m_2 = 6$	Limits the number of extractions per sentence to a maximum of $m_1 = 3$ for statements and $m_2 = 6$ for justifications.

Table 6: Parameters used to configure OpenIE6 for triplet relation extraction tasks.

through a specialized coordination analyzer built on the same grid-based framework. This unique combination of constraints and coordination handling allows OpenIE6 to set new standards in OpenIE performance, achieving notable improvements in recall and extraction quality at speeds up to 10 times faster than prior models.

Table 6 outlines the key parameters used to configure OpenIE6 for our triplet extraction tasks. These settings include options for mode, input/output file handling, rescoring, GPU usage, and model paths for specific tasks, ensuring optimized processing for our experimental setup. We limited extractions to a maximum of $m_1 = 3$ triplets for statements and $m_2 = 6$ triplets for justifications to maintain extraction relevance and computational efficiency.