

# ELAINE: rELiAbility and evidence-aware News vErifier

Carlos Badenes-Olmedo<sup>1</sup>, Estela Saquete<sup>2,\*</sup>, Robiert Sepúlveda-Torres<sup>2</sup> and Alba Bonet-Jover<sup>2</sup>

<sup>1</sup>*Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

<sup>2</sup>*Department of Software and Computing Systems, University of Alicante, Spain*

## Abstract

Disinformation is one of the main problems of today's society, and specifically the viralization of fake news. This research presents ELAINE, a hybrid proposal to detect the veracity of news items that combines content reliability information with external evidence. The external evidence is extracted from a scientific knowledge base that contains medical information associated with coronavirus, organized in a knowledge graph created from a COVID-19 corpus. The information is accessed using Natural Language Question Answering and a set of evidences are extracted and their relevance measured. By combining both reliability and evidence information, the veracity of the news items can be predicted, improving both accuracy and F1 compared with using only reliability information. These results prove that the approach presented is very promising for the veracity detection task.

## Keywords

Disinformation, Reliability, Question Answering, Knowledge Graphs

## 1. Introduction

The proliferation of mis- and dis-information is one of the main problems associated with today's digital society. Disinformation needs misleading content, readers who are likely to believe it, and a powerful means to achieve viralization. [1] presents an analogy between an epidemiological concept and the phenomenon of disinformation: the disinformation triangle. This conceptual model is based on George McNew's model that reflects the interaction between the three factors that cause disease: pathogen, host, and environment. [1] transfers this concept to that of disinformation to show that, just as the presence of the three factors mentioned above is necessary for a disease to occur, the simultaneous presence of three factors is also necessary for information to go viral: the fake news item (pathogen), the reader (host), and the digital platform (environment).

Currently, there have been several attempts to disprove fake news with manual fact-checking

---

*NLP-MisInfo 2023: SEPLN 2023 Workshop on NLP applied to Misinformation, held as part of SEPLN 2023: 39th International Conference of the Spanish Society for Natural Language Processing, September 26th, 2023, Jaen, Spain*


✉ carlos.badenes@upm.es (C. Badenes-Olmedo); stela@dlsi.ua.es (E. Saquete); rsepulveda@dlsi.ua.es (R. Sepúlveda-Torres); alba.bonet@dlsi.ua.es (A. Bonet-Jover)

🌐 <https://portalcientifico.upm.es/en/ipublic/researcher/308607> (C. Badenes-Olmedo)

🆔 0000-0002-2753-9917 (C. Badenes-Olmedo); 0000-0002-6001-5461 (E. Saquete); 0000-0002-2784-2748 (R. Sepúlveda-Torres); 0000-0002-7172-0094 (A. Bonet-Jover)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

techniques<sup>1</sup> performed by journalists and communication professionals [2], but the speed of viralization, made possible by the digital ecosystem, makes it necessary to apply Artificial Intelligence (AI) algorithms, specifically Natural Language Processing (NLP) techniques and knowledge-based representations that provide evidence and indicators of reliability and veracity. The purpose is to alert the reader to a possible case of fake news. Assessing the veracity of information in a news item implies necessarily assessing both content and context signals [3]. Content signals can be determined by only considering the text or content of an article, whereas context signals require using external and world knowledge.

Content-based approaches currently focus on the study of linguistic aspects of falsehood by identifying different types of falsehoods [4] [5] [6]. Context-based approaches, on the other hand, tend to search for news on official fact-checking websites to determine the veracity of the news. This approach to the problem has the disadvantage that if the news has not been previously debunked for any reason, even if false, it cannot be detected by the system [7] [8]. Within context approaches, the relationships between publishers, news items and readers are also exploited to predict fake news [9]. Hybrid approaches combine content and context from different perspectives to detect disinformation [10].

This paper presents a novel hybrid methodology for the detection of disinformation and, more specifically, for the prediction of the veracity of a news item. The main contribution of this paper is to demonstrate that this evidence-aware methodology improves on a strategy that is based purely on the content reliability of a news item when detecting veracity since veracity highly depends on external world knowledge. The source code of the algorithms and datasets used in the experiments are publicly available<sup>2</sup>.

## 2. Background

### 2.1. Automatic Disinformation Detection

An important part of disinformation detection research consists of building appropriate datasets to train Machine Learning (ML) and NLP models. In this sense, numerous corpora have been created so that models can learn from real examples of disinformation created by human experts. According to the literature reviewed, most corpora created to address this task use a binary classification, categorizing news as Fake or True, or Fake or Real [11][12]. Others, such as those focused on fact-checking tasks, use a fine-grained scale of labels that cover several degrees of veracity [13] [7]. However, in all cases, the annotation is global for the whole document, and the veracity or credibility of the different parts of the document are not considered. This single global classification of news, whether with binary or multiple values, depends on external knowledge, such as fact-checking platforms<sup>3</sup>. Furthermore, few datasets use a reliability classification and usually this classification is applied on the basis of the source's credibility [14] and not of purely textual or linguistic characteristics [15].

In our research proposal, Spanish news items related to the COVID-19 pandemic are used to address the lack of resources in languages other than English. To the author's knowledge,

---

<sup>1</sup>International Fact Checking Network: <https://www.poynter.org/ifcn/>

<sup>2</sup><https://github.com/cbadenes/elaine>

<sup>3</sup><https://toolbox.google.com/factcheck/apis>

approaches that address the disinformation task in Spanish [16] are scarce, since they are usually applied to English. Regarding COVID-19 disinformation detection, [17] built a detection system that uses ML to identify misleading information using reliable sources such as the World Health Organization, UNICEF, and the United Nations to collect ground-truth data and train ML models, concluding that collecting this information was effective for the task. [18] use MisRoBÆRTa, a novel transformer-based deep neural ensemble architecture for misinformation detection. They used a 100,000 real-world news article dataset labeled with 10 classes. [19] proposes a comprehensive approach that includes a two-phase detection and tracking. In the detection phase of the study carried out, several standalone and ensemble machine learning methods were applied on the Arcov-19 dataset. In the tracking phase, several similarity-based techniques were used to obtain the top 1% of similar tweets to a target tweet/post, which helped to find the source of the rumors.

Annotated datasets, as the ones presented in the literature, can help generate models that are capable of detecting indicators of veracity but will be limited by the constant creation of fake news that may not have appeared in the machine-learned examples and thus go undetected. In that sense, our proposal uses as a base a reliability annotated dataset, since this feature does not depend on external knowledge. However, it supports the veracity decision by combining the reliability of the content with external evidence extracted from scientific knowledge.

In summary, along with purely linguistic approaches [4] [5] [6], those based on the use of external knowledge have also been applied in the literature to determine the veracity of the elements within a news item. This knowledge can be structured in knowledge graphs, as subsequently indicated, or by retrieving information from websites or reliable documentary sources that provide the necessary data to verify or discredit information [20].

## 2.2. Evidence based on Knowledge Graphs

Knowledge Graphs (KG) have proven to be a valuable resource to structure data and provide knowledge extracted from document collections [21, 22]. Many Question Answering (QA) systems have been adapted to this graph-oriented context, leading to KGQA systems. Information is retrieved from graphs rather than documents or relational databases, and the methods commonly used in QA tasks are adapted to the particularities of graphs [23]. The entities or concepts mentioned in the question do not reside in a text or in a table row in a database but are resources in the KG with a unique identifier. Entity-linking techniques are then required to discover the graph resources, and the natural language question is usually transformed into a formal query to explore the knowledge graph (e.g. SPARQL, Cypher) [24].

Many KGQA systems emerged during the pandemic mainly due to the availability of clinical corpora, but also the need to obtain answers to questions oriented to coronaviruses. They aim to find relevant information within the massive literature [25]. However, their use to validate information has not been extended. The main challenges that must be overcome are to provide not only the answer, but also the evidence that supports it, and to infer the questions that allow a statement to be validated. Regarding the field of disinformation and fake news, different works in the literature have applied knowledge graphs to address the research problem [26, 27].

Our contributions are: firstly, a hybrid architecture that combines the content reliability annotation of news items with the evidence-aware context, where external knowledge is used to

support veracity classification of the news item by extracting evidences from scientific literature. This makes it unnecessary for the news item to be previously fact-checked. Secondly, the approach is applied to Spanish news, thereby addressing the lack of resources on languages other than English.

The next section presents the architecture, which has been developed in two layers: the reliability of the content and the evidence-aware context.

### 3. ELAINE's Architecture

Reliability content information is combined with context information to uncover technical evidence from scientific publications on the health domain. Specifically, Spanish news items related to the COVID-19 pandemic were selected to provide much needed resources in languages other than English. Our approach is based on a QA system that combines knowledge graphs and documentary corpora to provide answers in natural language from natural language questions [28]. We have taken advantage of a knowledge graph created from the CORON-19 corpus that semantically represents associations between drugs and diseases or symptoms based on scientific publications on coronaviruses in the last 50 years [29]. The QA system provides not only answers, but also evidence in the form of textual paragraphs that supports the answer.

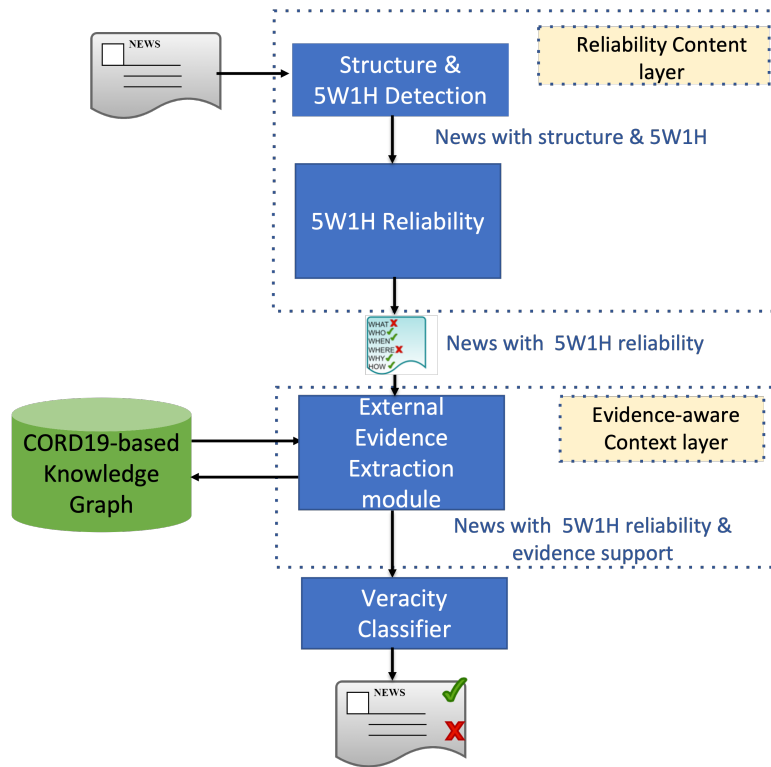
A two-layer architecture is proposed in a pipeline. Figure 1 depicts the two main components on which ELAINE's architecture is based and whose outputs will be used to determine the veracity of the news. First, the content reliability layer determines the news item's essential elements of information, assigning a reliability value to them according to the reliability criteria of the annotation schema presented in [30]. Based on the reliability of the elements of the news item, a global reliability of the news item is also obtained. Second, the evidence-aware context layer uses the elements of information previously extracted to infer questions that assess the veracity of the news item. It then retrieves evidence from the answers and measures its relevance to verify the facts. Finally, using both content and context features, the veracity of each news item is obtained.

#### 3.1. Content Reliability Layer

Considering the definition of veracity as “the quality of being true” and the definition of reliability as “the quality of being likely to be correct or true”<sup>4</sup>, and using as a starting point previous research in which only news content is used [30], this layer determines the reliability of the news item. Veracity and reliability are closely related concepts, as fake news often includes a mix of reliable and unreliable information items. Thus, determining the content reliability/unreliability signals is a feasible approach to be used in a veracity detection pipeline since there is a direct relationship between the veracity of a news item and the reliability of the information contained in it [31]. In view of this connection, this first layer focuses on detecting reliable and unreliable information in Spanish news texts, without using external knowledge, and based exclusively on the news content and a set of reliability criteria.

---

<sup>4</sup><https://www.oxfordlearnersdictionaries.com/>



**Figure 1:** ELAINE’s architecture.

The reliability content layer contains two components based on previous research [30] concluding that the structural parts and essential content of a news piece have specific reliability values, which influence the overall reliability and veracity value of the news story. This scheme is based on the idea that, although there are different ways of writing a news story, there are two key principles on which all well-built news should be based: neutrality and the inverted pyramid structure [32]. Thus, the objectivity of a news piece may depend on these two factors, so the detection of unusual deviations from these accepted journalistic norms could provide a clue to detect fake news.

**Structure and 5W1H Component:** organizes the text according to two levels of information representation. First, the news story is divided according to the journalistic structure (i.e. TITLE, SUBTITLE, LEAD, BODY and CONCLUSION), and then the 5W1H elements of each part of the structure are determined. After this procedure, news items are annotated with structure and 5W1H labels.

**5W1H Reliability Component:** determines the reliability of each of those parts previously detected and predicts the reliability of the news piece using the reliability of the different 5W1H items. The reliability criteria presented in this proposal focus on the principles of accuracy and neutrality. The more in depth criteria are presented in [30].

### 3.2. Evidence-aware Context Layer

Verification by context requires a knowledge base from which to find evidence that helps to validate a claim. The information available in the knowledge base is assumed to be reliable, and can be organized in a structured way in knowledge graphs or unstructured in documentary corpora. Specifically, our knowledge base [29] contains information on medications and symptoms associated with the coronavirus organized in a knowledge graph created from the CORD-19<sup>5</sup> corpus, which contains scientific publications on SARS-CoV-2, MERS, and COVID-19. The access interface is based on a natural language QA system [28] that retrieves answers combining the information of the knowledge graph with the content of scientific articles. However, the QA system cannot directly verify the validity of a claim (e.g. *“Is it true that the vaccine is effective?”*), since it does not support logical reasoning. We have extended the QA functionality in three steps to provide evidence that helps verify a claim: (1) dynamic question creation, (2) evidence validation, and (3) relevance measurement.

**Dynamic Question Creation:** The first step consists of the dynamic generation of questions from a statement. We use structure and reliability annotations and a generative language model<sup>6</sup> trained with the SQUAD 2.0 dataset [33] to create questions in a given context. From an assertion (e.g. *“Janssen asks the EU to approve a single-dose vaccine that will speed up immunization”*) and the 5W1H annotations (e.g. *“Janssen”* (Who), *“the EU”* (Who), and *“single-dose vaccine approval”* (What)), the context is constructed with the statement itself and removing its annotations (i.e. *“What”, “Who”, ..*) to create incomplete sentences (e.g. *“asks the EU to approve a single-dose vaccine that will speed up immunization”* or *“Janssen asks to approve a single-dose vaccine...”*). The questions may be specific enough to extract evidence (e.g. *“What is the name of the vaccine that will speed up immunization?”*), but others may be too general (e.g. *“What is Janssen asking the EU to approve?”*). We measure the specificity by calculating the cosine similarity of the vector representations, based on sentence transformers, of the questions with respect to the original statement. We define a similarity threshold to filter suitable questions that after the experiments is set to 0.6.

**Evidence Validation:** Once the questions are available, we use the coronavirus QA system<sup>7</sup> to retrieve evidence from the knowledge base. An evidence is a paragraph where the answer to the question is mentioned. To validate whether an evidence is related to the original statement, not to the question, we check that it covers the annotations (i.e. 5W1H). Specifically, we review that the entities, identified by Named Entity Recognition (NER) techniques, within the 5W1H annotations are also mentioned in the evidence (e.g. *“Janssen”* and *“EU”*). Evidence that does not contain these entities is discarded.

**Relevance Measurement:** Finally, we must measure the relevance of the evidence with respect to the statement. The content of the evidence is reduced using summarization techniques based on transformers to remove any noise present. Next, we calculate its relevance by applying a Textual Entailment (TE) procedure to determine the directional relation between text fragments. The relation holds whenever the truth of one text fragment follows from another text. In the TE framework, the entailing and entailed texts are termed *premise* and *hypothesis*, respectively.

---

<sup>5</sup><https://allenai.org/data/cord-19>

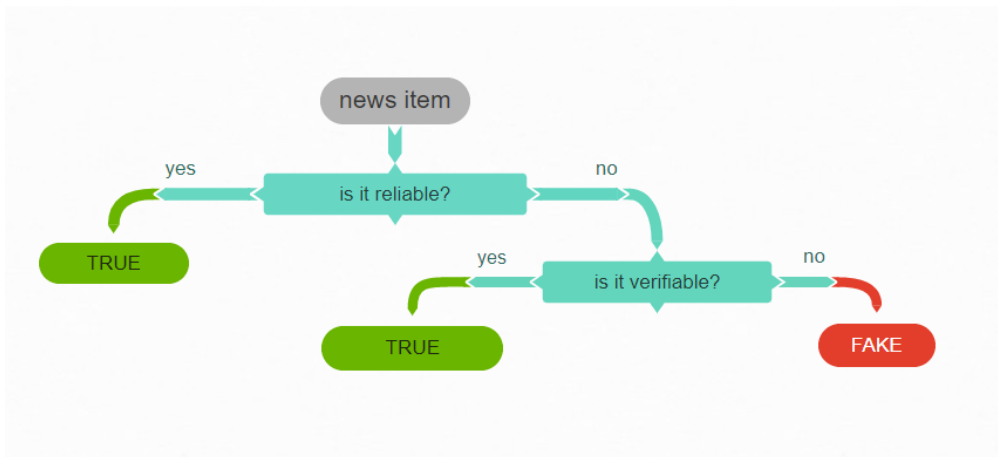
<sup>6</sup><https://huggingface.co/allenai/t5-small-squad2-question-generation>

<sup>7</sup><https://drugs4covid.oeg.fi.upm.es>

The evidence is our *premise*, and the statement is *hypothesis*. Thus, the evidence can be in contradiction to or in support of what is stated, or evidence can be neutral in the case of inconclusive evidence either way.

### 3.3. Veracity Classifier

The veracity of a news item depends on the reliability of its content and the evidence of its claims. The relationship between both aspects, based on the content and context of the news, is reflected in the following decision tree.



**Figure 2:** Veracity of a news item based on content reliability and evidence-aware context.

A news item is considered TRUE when it is reliable. However, when the content is not reliable, the evidence of the statement is verified. A news item is verifiable when it is possible to verify most of its statements. And a statement is verifiable when there is at least one *entailment* evidence or more than one *neutral* evidence. Otherwise, the statement is not verifiable. If a news item has more verifiable statements than unverifiable ones, it is considered TRUE. Otherwise, it is considered to be FAKE.

### 3.4. Evaluation Environment

The main contribution of this paper is to show that combining content reliability and evidence-aware context is more effective for detecting veracity than a strategy based purely on the content reliability of a news item. Because reliability detection is beyond the scope of this paper, we used a previously annotated dataset with: i) the structure and 5W1H items; ii) the reliability information of each item; and, iii) the news item's global reliability value. A detailed description of the input data set is provided, as well as evaluation measures and a detailed description of the experiments performed. The source code of the algorithms, the data sets used in the experiments and the evaluation logic is available in the following repository<sup>8</sup>.

<sup>8</sup><https://github.com/cbadenes/elaine>

### 3.5. Datasets

To evaluate the benefit of incorporating scientific evidence into the veracity detection architecture, the news items used in the experiments were annotated not only with the veracity value (true/false), which will be used to evaluate the predictions, but also with the reliability of the news item as well as that of each essential content item within it. The dataset used in these experiments comprises three subsets of news built with different strategies:

- **Coronavirus Pandemic News:** contains 100 non-technical news items related to coronavirus pandemic that were manually retrieved from several Spanish digital media (52/48 True/Fake).
- **COVID-19 News:** contains 116 news items related to COVID-19 automatically pre-annotated with the 5W1H elements (112/4 True/Fake).
- **Health News:** contains 55 news articles (and social posts) retrieved from the Spanish Fake News Corpus Version 2.0 [34, 35, 16] (27/28 True/Fake). This dataset covers several topics and includes variants of Spanish.

Table 1 describes the different subsets used in the experiments.

**Table 1**  
Corpora description

Corpus	Language	Format	Fake	True	Total
Pandemic News	Spanish (colloquial)	News	48	52	100
COVID-19 News	Spanish (technical)	News	4	112	116
Health News	Spanish (variants)	Social Posts	28	27	55
<i>Full Dataset</i>	—	—	<i>80</i>	<i>191</i>	<i>271</i>

### 3.6. Experiments Description

In order to evaluate the improvement of the hybrid proposal, the commonly used NLP measures (accuracy, precision, recall and F-measure) are used. The corpora used in the evaluation contain news items previously verified as true or false, which are usually paragraphs of text (e.g. “*With the outbreak of the coronavirus that has affected many countries around the world, not only in health...*”). A **Reliability performance** experiment has been performed to predict the veracity of the news items by using only the reliability information of the 5W1H elements and the global reliability of the news item. A baseline approach is applied considering as True news when it is a reliable news and Fake news when is unreliable. The reliability values were manually annotated for this work in order to measure the validity of evidences without possible errors derived from automatic reliability prediction. Moreover, a **Reliability + Evidence performance** experiment has been performed to predict the veracity by combining reliability and evidence as described in the decision tree in Figure 2. This strategy takes advantage of the 5W1H annotations and the reliability value of the previous strategy to analyze evidence that helps to evaluate the veracity of a news item. It proves to be a feasible approach for detecting veracity.



## 4. Results and Discussion

Table 2 shows all the datasets trained. The first subset (Pandemic News) contains news about COVID-19 but its main characteristic is that the verifiable element is not necessarily the coronavirus. That is, the critical information in the news may be a non-scientific-technical agent (e.g. *“Some members of royalty have suffered the effects of the coronavirus...”*), or contain indeterminate temporary information (e.g. *“...which already affects more than 1500 people in Spain”*) or is focused on non-clinical effects of the pandemic (e.g. *“...a trucker tells us that he carries masks, respirators and all kinds of sanitary material and how...”*). The main difference between the two strategies lies in the volume of correctly identified news. While the two approaches show similar behavior when considering false news as true (i.e., similar precision), the evidence-based approach discards less true news than the reliability-based approach (i.e., higher recall). In this way, the volume of true news increases without reducing the quality of identifying true news.

The second subset (COVID-19 News) also contains news about coronavirus in Spanish. More technical language is used in this subset, and the core of the information to be verified is mainly clinical (e.g. *“...it was announced in China that Egyptian scientists have discovered a vaccine that allows the healing...”*). The technical and specific feature of the dataset supports the high performance for all the strategies, specifically the use of evidence, since our knowledge graph is built from scientific publications. The key point lies in coverage, that is, handling false negatives, or when news that is true is considered Fake. Using evidence better covers that scenario and adds to reliability with better performance, as our hybrid approach demonstrates. For news that was already considered true through reliability, it is not necessary to resort to evidence.

In the third subset (Health News), with 55 news items, after an exhaustive analysis of the news, some peculiarities were detected. It contains five news items from social networks (12.73% of the total of this third dataset), more specifically Facebook in this case, and two news items from a satirical website (ELMundoToday<sup>9</sup>). Anomalies may arise when processing news items automatically from these sources as the search is based on scientific evidence. Moreover, as explained in subsection 3.5, this dataset contains generic health news, some of them related to COVID-19, but not from a clinical perspective. This means that a search for evidence in our specific COVID-19 knowledge base will not result in these types of news items being fully corroborated. This same situation is relevant to the first subset, with general-domain news regarding COVID-19. These types of news items were related to situations that are less scientific and more collateral to the pandemic itself.

The joint analysis of the three subsets (Full Dataset) demonstrates, as it is also appreciated separately, that the combined use of reliability and evidence to verify the news is the most robust strategy. This hybrid process almost completely reduces false negatives (true news that is considered fake), as demonstrated by its recall value, and also reduces false positives (fake news that is considered true), as demonstrated by its precision value. The capability to correctly identify true news and fake news is also shown with a higher accuracy value. In short, it seems that the weakness of the reliability strategy when identifying true news as fake news is mitigated by incorporating evidence analysis based on knowledge graphs. Furthermore, given

---

<sup>9</sup><https://www.elmundotoday.com/>

the difference in the evaluations using the different types of news sources, future lines of research could benefit from including source credibility or source features to improve performance in veracity detection. This would enable evidence to be discovered in more suitable knowledge bases.

**Table 2**

Performance in pandemic news

<i>Pandemic News</i>				
<b>Strategy</b>	<b>Precision</b>	<b>Recall</b>	<b>Acc</b>	<b>F_1</b>
Reliability	<b>0.52</b>	0.90	0.51	0.66
Reliability + Evidence	<b>0.52</b>	<b>1.0</b>	<b>0.52</b>	<b>0.68</b>
<i>COVID-19 News</i>				
Reliability	<b>0.98</b>	0.54	0.55	0.70
Reliability + Evidence	0.97	<b>0.99</b>	<b>0.96</b>	<b>0.98</b>
<i>Health News</i>				
Reliability	<b>0.77</b>	0.89	<b>0.82</b>	<b>0.83</b>
Reliability + Evidence	0.49	<b>1.0</b>	0.49	0.66
<i>Full Dataset</i>				
Reliability	<b>0.72</b>	0.69	0.59	0.70
Reliability + Evidence	0.70	<b>0.99</b>	<b>0.70</b>	<b>0.82</b>

## 5. Conclusions and Future Work

This paper presents ELAINE, a novel hybrid proposal that combats disinformation in Spanish news by predicting the veracity of news items. The proposal combines the content reliability information of news items with external context information to uncover technical evidence from scientific publications related to the coronavirus. Our approach demonstrates how the extraction of external knowledge can enable veracity detection more robust than that based solely on news content. Thus, the proposed architecture is constructed with two layers. The first one includes extracting the essential information of the news (Journalistic news structure and the 5W1H) and the reliability of each of these essential information items, as well as the overall reliability of the news item. Using the output of this first layer, the second layer extracts external evidence of the essential parts of the news item and measures their relevance. Finally, using both the reliability and the evidence, the veracity of the news item is obtained using a decision tree. The proposal was evaluated with three different corpora on health news, significantly improving accuracy and F1 results by comparing the approach that uses only news content reliability with the hybrid approach that combines this information with evidence extraction (from 0.59 to 0.70 Acc, 0.70 to 0.82 F1). These results demonstrate the feasibility of our proposal. Future work could use other external sources to detect false information in more general domains and not only in the technical coronavirus domain.

## References

- [1] V. L. Rubin, Disinformation and misinformation triangle: A conceptual model for “fake news” epidemic, causal factors and interventions, *Journal of documentation* 75 (2019) 1013–1034.
- [2] W. Lowrey, The emergence and development of news fact-checking sites, *Journalism Studies* 18 (2017) 376–394. arXiv:<https://doi.org/10.1080/1461670X.2015.1052537>.
- [3] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, M. Palomar, Fighting post-truth using natural language processing: A review and open challenges, *Expert systems with applications* 141 (2020) 112943.
- [4] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, in: *Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, 2018*, pp. 3391–3401.
- [5] R. Mihalcea, C. Strapparava, The lie detector: Explorations in the automatic recognition of deceptive language, in: *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Association for Computational Linguistics, 2009*, pp. 309–312.
- [6] S. Feng, R. Banerjee, Y. Choi, Syntactic stylometry for deception detection., in: *ACL (2), The Association for Computer Linguistics, 2012*, pp. 171–175.
- [7] A. Vlachos, S. Riedel, Fact checking: Task definition and dataset construction, in: *Proceedings of the ACL 2014 workshop on language technologies and computational social science, 2014*, pp. 18–22.
- [8] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, 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), Association for Computational Linguistics, 2018*, pp. 809–819.
- [9] K. Shu, S. Wang, H. Liu, Beyond news contents: The role of social context for fake news detection, in: *WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining, WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Association for Computing Machinery, Inc, 2019*, pp. 312–320. 12th ACM International Conference on Web Search and Data Mining, WSDM 2019 ; Conference date: 11-02-2019 Through 15-02-2019.
- [10] S. Volkova, K. Shaffer, J. Y. Jang, N. Hodas, Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter, in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, 2017*, pp. 647–653.
- [11] F. K. A. Salem, R. Al Feel, S. Elbassuoni, M. Jaber, M. Farah, Fa-kes: A fake news dataset around the syrian war, in: *Proceedings of the international AAAI conference on web and social media, volume 13, 2019*, pp. 573–582.
- [12] R. M. Silva, R. L. Santos, T. A. Almeida, T. A. Pardo, Towards automatically filtering fake news in portuguese, *Expert Systems with Applications* 146 (2020) 113199.
- [13] W. Y. Wang, “liar, liar pants on fire”: A new benchmark dataset for fake news detection, in: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*

- (Volume 2: Short Papers), Association for Computational Linguistics, 2017, pp. 422–426.
- [14] S. Dhoju, M. Main Uddin Rony, M. Ashad Kabir, N. Hassan, Differences in health news from reliable and unreliable media, in: Companion Proceedings of The 2019 World Wide Web Conference, 2019, pp. 981–987.
- [15] R. Assaf, M. Saheb, Dataset for arabic fake news, in: 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT), 2021, pp. 1–4.
- [16] J.-P. Posadas-Durán, H. Gómez-Adorno, G. Sidorov, J. J. M. Escobar, Detection of fake news in a new corpus for the spanish language, *Journal of Intelligent & Fuzzy Systems* 36 (2019) 4869–4876.
- [17] M. K. Elhadad, K. F. Li, F. Gebali, Detecting misleading information on covid-19, *Ieee Access* 8 (2020) 165201–165215.
- [18] C.-O. Truică, E.-S. Apostol, Misrobæarta: Transformers versus misinformation, *Mathematics* 10 (2022) 569.
- [19] S. N. Qasem, M. Al-Sarem, F. Saeed, An ensemble learning based approach for detecting and tracking covid19 rumors, *Computers, Materials and Continua* (2021) 1721–1747.
- [20] G. Karadzhov, P. Nakov, L. Màrquez, A. Barrón-Cedeño, I. Koychev, Fully automated fact checking using external sources, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, INCOMA Ltd., 2017, pp. 344–353.
- [21] D. N. Nicholson, C. S. Greene, Constructing knowledge graphs and their biomedical applications, *Computational and Structural Biotechnology Journal* 18 (2020) 1414–1428.
- [22] K. Fecho, C. Bizon, F. Miller, S. H. Schurman, C. P. Schmitt, W. Xue, K. Morton, P. Wang, A. Tropsha, A biomedical knowledge graph system to propose mechanistic hypotheses for real-world environmental health observations: Cohort study and informatics application, *JMIR Medical Informatics* 9 (2021).
- [23] M. Yani, A. A. Krisnadhi, Challenges, techniques, and trends of simple knowledge graph question answering: A survey, *Information* 12 (2021).
- [24] A. Pereira, A. Trifan, R. P. Lopes, J. L. Oliveira, Systematic review of question answering over knowledge bases, *IET Software* 16 (2022) 1–13.
- [25] R. Shaina, S. B, R. LC, Coquad: a covid-19 question answering dataset system, *BMC Bioinformatics* (2022).
- [26] P. Lin, Q. Song, Y. Wu, J. Pi, Discovering patterns for fact checking in knowledge graphs, *J. Data and Information Quality* 11 (2019).
- [27] G. Ciampaglia, P. Shiralkar, L. Rocha, J. Bollen, F. Menczer, A. Flammini, Computational fact checking from knowledge networks, *PloS one* 10 (2015).
- [28] C. Badenes-Olmedo, Ó. Corcho, Muheqa: Zero-shot question answering over multiple and heterogeneous knowledge bases, *Semantic Web* (2023).
- [29] C. Badenes-Olmedo, O. Corcho, Lessons learned to enable question answering on knowledge graphs extracted from scientific publications: A case study on the coronavirus literature, *Journal of Biomedical Informatics* 142 (2023) 104382. URL: <https://www.sciencedirect.com/science/article/pii/S153204642300103X>. doi:<https://doi.org/10.1016/j.jbi.2023.104382>.
- [30] A. Bonet-Jover, R. Sepúlveda-Torres, E. Saquete, P. M. Barco, Annotating reliability to enhance disinformation detection: annotation scheme, resource and evaluation, *Proce-*

samiento del Lenguaje Natural 70 (2023) 15–26.

- [31] A. X. Zhang, A. Ranganathan, S. E. Metz, S. Appling, C. M. Sehat, N. Gilmore, N. B. Adams, E. Vincent, J. Lee, M. Robbins, et al., A structured response to misinformation: Defining and annotating credibility indicators in news articles, in: Companion Proceedings of the The Web Conference 2018, 2018, pp. 603–612.
- [32] E. A. Thomson, P. R. White, P. Kitley, “objectivity” and “hard news” reporting across cultures: Comparing the news report in english, french, japanese and indonesian journalism, *Journalism studies* 9 (2008) 212–228.
- [33] P. Rajpurkar, J. Zhang, K. Lopyrev, P. Liang, SQuAD: 100,000+ questions for machine comprehension of text, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2016, pp. 2383–2392.
- [34] H. Gómez-Adorno, J. P. Posadas-Durán, G. B. Enguix, C. P. Capetillo, Overview of fakedes at iberlef 2021: Fake news detection in spanish shared task, *Procesamiento del Lenguaje Natural* 67 (2021) 223–231.
- [35] M. E. Aragón, H. J. Jarquín-Vásquez, M. Montes-y Gómez, H. J. Escalante, L. V. Pineda, H. Gómez-Adorno, J. P. Posadas-Durán, G. Bel-Enguix, Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in mexican spanish., in: IberLEF@ SEPLN, 2020, pp. 222–235.