https://doi.org/10.33472/AFJBS.6.9.2024.5073-5086

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

# Assessing the Reliability of News Reports using Human-in-the-Loop to Fight False Information

**[1]Dr.N.Menaga, [2]Dr. R. Shanthi, [3]Dr. J. Jeya A Celin, [4]Sugitha Deivasigamani,[5]D.Gokila**

Assistant Professor, Department Of Computer Science, Hindusthan College Of Arts & Science, Coimbatore.
Assistant Professor, Department of computer Applications, Faculty of science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Chennai.
Professor, Department of Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil.
University College of Engineering, (A Constituent College of Anna University), Thirukkuvalai, Nagapattinam.
Assistant Professor, Department of Cybersecurity, Dr SNS Rajalakshmi College of Arts and Science, Coimbatore.
**Corresponding Author Email ID:**[1]drnmenaga@gmail.com

**Email ID**:[2]sha.raju2003@gmail.com, [3]jeyacelin@klu.ac.in, [4]sugi47@gmail.com.

Article History

Volume 6,Issue 9, 2024

Received:20 May 2024

Accepted : 15 Jun 2024

doi: 10.33472/AFJBS.6.9.2024.5073-5086

**Abstract**

In order to train Natural Language Processing models, annotated corpora are essential. Unfortunately, resources for training Machine Learning and Deep Learning algorithms are scarce due to the high cost, labor-intensive nature, and lengthy nature of complicated semantic annotation processes. Our work presents a framework for semi-automating difficult tasks based on the human-in-the-loop approach to address this challenge. used this process to produce a Spanish news reliability dataset in order to effectively counteract misinformation and false news. Using this approach produced a high-caliber resource that increases annotator productivity and speed while requiring fewer instances. The RUN dataset is the product of three incremental phases of the process. We evaluate the consistency of the annotations, the agreement between the annotations, the time reduction (about 64% less than fully manual annotation), and the model performance after training on the RUN semi-automatic dataset in order to determine the quality of the annotations. Our results demonstrate the suitability and effectiveness of our proposal, achieving an Accuracy of 97% and an F-Score of 96%.

**Keywords:** "Human-in-the Loop, Machine Learning and Deep Learning, Semi Supervised Learning, Fake News Detection";

## 1. Introduction

Disinformation, a crucial component of the more general notion of 'information disorder,' is the intentional production and dissemination of incorrect information with the aim of causing harm; it includes things like hoaxes and fake news. Although the phrase "fake news" is well-known, this study uses the more inclusive word "disinformation" to emphasize the problem's global scope. Disinformation is spreading much more quickly in the modern world thanks to the digital era, which poses serious risks to society, public health, and ideology. People are now more linked than ever thanks to technological improvements, but this connectivity has also made room for abuse. Massive volumes of information are spreading quickly and virally, making manual processing impractical. This calls for automatic detection. Current automatic identification methods frequently necessitate the involvement of human experts for annotation and feedback, resulting in a laborious and resource-intensive process.

From an engineering point of view, the research community investigates multiple approaches, such as automated fact-checking, polarization, credibility, and stance identification. "This work focuses on the automatic detection of document trustworthiness through the application of Artificial Intelligence (AI) and Natural Language Processing (NLP). The method is content-based; it doesn't require any outside information; it just depends on the content of the document. Lack of training data, which is essential for AI and NLP, is one of the main problems. Compiling datasets for AI system training and evaluation requires a significant quantity of human feedback. Effective dataset generation requires knowledge and effort due to its complexity and range of annotation difficulty. Creating customized, high-quality datasets that are chosen according to predetermined standards can improve accuracy while conserving. Creating customized, high-quality datasets that are chosen according to predetermined standards would improve accuracy and save time and effort. Updates to the dataset are required continuously to avoid obsolescence. The organization of the study comprises a review of pertinent scientific literature, information regarding the annotation guideline (RUN-AS), the semi-automatic annotation methodology, its application, an assessment framework, and, lastly, the research findings and future directions.

## 2. Literature Survey

Xie et al. [2020] covered a variety of governance-related subjects. Research, both nationally and internationally, has started to identify and emphasize the need of having different entities in positions of authority. A growing number of people now understand "governance" to include non-governmental forces including the public, social groups, and the market in addition to political authority. Some academics contend that in the context of new media, users ought to be seen as engaged members of civil society. They support allowing users to participate in the regulation of the media and integrating people from different backgrounds into formal or informal organizations. This will allow them to express their opinions in public, participate in politics, and help to realize a democratic society. In addition, certain domestic studies also suggest that modernizing governance and improving its efficiency are not possible with the notions of "management" and "supervision" applied by a single institution alone, due to the changing social circumstances. Rather, it calls for the participation of different stakeholders in different governance areas, such as corporations, governments, industry associations, and internet users. The state can effectively adapt and improve its capacities while increasing political participation by using this multidimensional strategy.

Furthermore, as stated by Caplan et al. [2020], platforms also aim to regulate, shape, and exert influence over the cultural contributions and content production made by users. Some platforms, like YouTube's "Partner Program" (YPP), collaborate with content providers through shared advertising revenue plans. In order to encourage users to actively engage in content creation, the platform strikes deals with prominent users and makes use of its material and traffic resources. This methodology arranges and transforms the contributions of dispersed, capricious, novice, or expert users into expendable "labor." Because the platform controls the processes of exchange and circulation, users frequently don't fully comprehend the value that their information creates. Although every user is visible to the platform, users themselves usually have very basic and imprecise impressions of the platform, particularly when it comes to its algorithmic workings. Professional content creators face a great deal of uncertainty in this scenario because their access to audiences depends on the whims of the algorithms, unless they have the ability to bargain with platform administrators. Although the platform has the ability to monitor every user, consumers themselves usually have very basic and imprecise views of the platform, particularly with regard to its algorithmic workings. For content creators who are not able to bargain with platform managers, this presents a great deal of uncertainty because it will determine whether or not they can reach audiences based on the erratic algorithms.

According to Xiao et al. [2020], the introduction of the Internet has accelerated the independent growth of social systems, resulting in a change in the public domain. This movement has shifted from a single national governmental structure to professional organizations, especially big businesses in the economy. As private power becomes more prevalent on social media, the government's hegemonic role in policing information in the public domain is under threat, leading to a situation known as "parallel" governance. The government uses two main tactics to get involved in the private management of these platforms in order to remedy this. Issuing administrative directives is one way to exercise direct authority. The other approach is more subtle; it entails putting pressure on platforms from without to gradually introduce efficient self-regulation.

As noted by Xie et al. [2020], mobilizing significant civil society organizations is crucial to combating the issue of commercial platforms stealing public influence. Through the independent acts of trade associations and citizen groups, these groups can systematically define the corporate duties of commercial platforms, creating clear and visible norms for content governance. In order to fulfill their obligations, platforms must first fulfill two essential requirements. Firstly, from the standpoint of personal growth, maintaining a well-managed platform environment and promoting a favorable social image are fundamental requirements to draw in top-tier users and guarantee long-term expansion. Second, platforms ought to fulfill public duties in proportion to their private authority when evaluating the alignment of rights and responsibilities. It is only fair that platform owners who profit from illegal activity on the platform, whether directly or indirectly, take the appropriate action. In 2019, the European Union approved the "Copyright Directive on the Single Digital Market," which emphasized the roles that internet platforms played in providing information resources and copyright

protection. It emphasizes that platforms need to make money, but they also need to manage material ethically and in a way that makes sense given their size and income.
Zhou et al. [2020]; The interaction between the government, platform, and users is characterized by a triangle in the digital world of platforms. The government works with the platform, giving it some authority, and the platform, in turn, uses both explicit and implicit regulations to manage its private user supervision. Given that it takes into account the coexistence of private, public, and individual rights, this novel model for content control goes beyond the conventional binary approach. Under this approach, both public and private power entities may pose a threat to users' private rights. Ensuring that these power entities have a relatively stable operational limit becomes imperative when taking into account the rights and interests of all players within the platform ecosystem. This serves as the foundation for maintaining the stability of the triangular structure and ensuring the legitimacy of governance.

Song and others [2020] To differentiate itself from health knowledge, health rumors are unsubstantiated health-related tidbits that get around the public. These false pieces of information frequently cause people to make poor decisions that have a negative impact on their physical and mental health. As such, dispelling health rumors has emerged as a crucial component of improving the caliber of online health services. There are two main strategies to stop health rumors from spreading further: disseminating the facts and protecting against rumors. Because of the internet's inherent transparency, directly suppressing rumors can be difficult and even harmful. Encouraging the truth produces more long-lasting results than trying to stop rumors. It is possible to reduce public mistrust and reliance on misleading information to some extent by utilizing corrective material to clarify and refute health rumors and disseminating true information to oppose their spread.

As per Tong et al's [2020] assertion, the phrase "rumor refuting" first surfaced in the Internet domain and then became a recognized scientific notion after a series of intense discussions within the academic community. The topics of conversation included social issues like the spread of rumors and their source, as well as the difficulties in dispelling them. Refuting rumors (disseminating the truth) and blocking rumors are two widely accepted tactics for avoiding the spread and amplification of falsehoods. Considering the Internet's inherent transparency, research has shown that disseminating the truth is more effective over the long term than trying to dispel rumors. Three main areas of interest exist in the field of rumor debunking research: the qualities of information that contradicts rumors, the behavior of users engaged in debunking rumors, and the platforms designed for rumor refutation.

Li and associates [2021]; Examining the efficacy of rumor debunking on the platform is crucial, thirdly. Certain researchers evaluate the efficacy of rumor rebuttal by analyzing the information processing capabilities of the platform, the reliability of its sources, and user participation. Li et al. developed a regression model to look into the relationship between REI, content, and contextual factors. They also presented the REI (Rumor refuting Effectiveness Index) as a measure to assess the effectiveness of rumor refuting. Fourthly, it's important to investigate the platform's technical improvements. Global academics have demonstrated exceptional proficiency in creating intelligent systems for dispelling rumors on various platforms. For example, they have combined crowdsourcing with machine learning algorithms to develop platforms that can recognize and dispel claims about global health automatically.

Li and associates [2021]; Examining the efficacy of rumor debunking on the platform is crucial, thirdly. Certain researchers evaluate the efficacy of rumor rebuttal by analyzing the information processing capabilities of the platform, the reliability of its sources, and user participation. Li

et al. developed a regression model to look into the relationship between REI, content, and contextual factors. They also presented the REI (Rumor refuting Effectiveness Index) as a measure to assess the effectiveness of rumor refuting. Fourthly, it's important to investigate the platform's technical improvements. Global academics have demonstrated exceptional proficiency in creating intelligent systems for dispelling rumors on various platforms. For example, they have combined crowdsourcing with machine learning algorithms to develop platforms that can recognize and dispel claims about global health automatically. In an effort to improve the automation and intelligence of information quality assessment, a few international researchers have begun experimenting with deep learning algorithms to create classifiers for automated health information assessment.

LO and associates [2020] Live shopping is a new kind of e-social commerce-based purchasing experience. It makes use of live broadcasts' dynamic elements to provide an engaging, information-rich, and interactive online purchasing environment. Live shopping effectively closes the perception gap, resolves information opacity difficulties, and radically changes the conventional e-commerce model by enabling real-time interaction between hosts and viewers. Notably, it makes it possible for hosts and their audience to interact in real time, both ways. The total amount of merchandise sold through live e-commerce during the "Double Eleven" shopping festival in 2021 was an astounding 131.86 billion yuan, an 80.9% rise over the 72.9 billion yuan recorded during the same period in 2020. A considerable amount of online suppliers and sellers have entered the live streaming shopping space due to its huge benefits. Because of this, live streaming for shopping has become a popular marketing avenue for international businesses and has received a lot of attention from academics.

Xu and associates [2020] At the moment, most live shopping research focuses on how audience incentives, psychological processes, and anchors' qualities as information sources affect customers' intents to buy and engagement patterns. The most common type of research methodology is empirical research using questionnaire surveys. But examining the information that real live content conveys—including things like subtitles—has received less attention, and there aren't any clear benchmarks for judging how good live content is. Essentially, live streaming rooms that provide top-notch live material with items are the ones that routinely land spots in the top ten sales rankings. Their content qualities can practically act as industry standards. Given this, the author has conducted research on the characteristics of high-quality live shopping broadcasts' content and their dissemination methods.

Xue and associates [2020] At the moment, most live broadcast content research focuses on the interactive element. Researchers such as Xue have investigated how real-time interactions affect participation in social commerce, taking into account aspects like control, interaction, responsiveness, customisation, and entertainment. In order to evaluate the depth and intensity of interaction and analyze the dynamic consequences of live shopping interactions, Kang et al. employed interactive responsiveness and personalization. In addition, Fan Xiaojun et al. have defined and investigated interactivity in the context of mobile video live broadcasts by integrating aspects such as synchronization, responsiveness, mindfulness, and interaction frequency.

Zhang and associates [2021] The author has assessed the quality of live broadcast content using three primary criteria: vividness, utility, and credibility. One argument for live shopping is that it gives customers an intuitive way to explore things, with the anchor offering knowledgeable product explanations. It is still unknown which factors are fundamental to excellent live shopping broadcasts, despite the fact that numerous research have suggested numerous elements linked to the features of live broadcast material. Moreover, it's possible that some of these variable definitions don't fit the particular context of live shopping because they were taken from other contexts. In order to close this disparity, the author has used text mining methods on subtitle texts to identify the distinct content attributes of top-notch live shopping broadcasts. The purpose of this study is to define these distinctive variables with accuracy and relevance to their respective contexts. The general trends in the distribution of these attributes of high-quality live broadcast content are then determined by statistical analysis.

According to a study by Wong et al. [2020], responsiveness, customization, visualization, professionalism, dependability, and enjoyment are some of the major elements that affect how effective live broadcasts are. The high degree of reactivity, in accordance with the social contagion hypothesis, generates a sense of urgency that starts a herd effect and pushes the audience to make impulsive purchases. Personalized recommendations that are in line with audience tastes are the main emphasis of live broadcast commerce, given the abundance of information available. This lowers the cost of information processing for the audience and improves the effectiveness and caliber of decision-making. Viewers experience social recognition and a reduction in psychological distance when they believe that the anchor's recommendations are in line with their wants. In live broadcasts, visualization heightens social presence, shortens psychological distances, and offers an immersive experience.

Notably, there is little external input, such as comments and word-of-mouth information about the products, and product links are only active during the live session. Customers may perceive more dangers as a result of this. But the anchor's expert justifications and dedication to high-quality products help to increase customer confidence and lower perceived dangers. According to user participation theory, participants in live broadcast activities who actively participate get completely absorbed in the process, which induces a sense of flow. As a result, anchor-organized online activities contribute to the audience's experience of closeness and connection, as well as their sense of warmth and belonging, all of which are ultimately enjoyable.

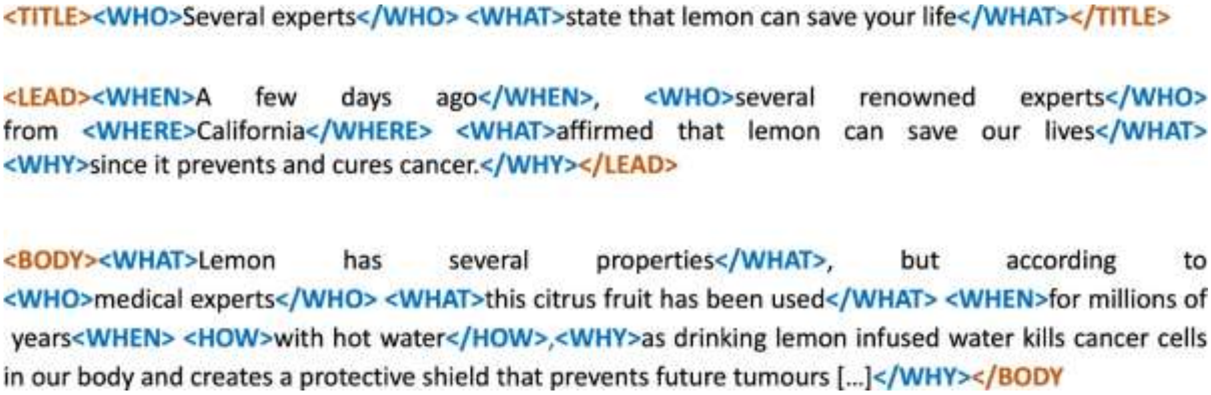
According to Huang et al. (2021), people usually buy things during live shopping when they actually need them. The very fact that customers frequently contact with anchors deters them from making impulsive purchases, in contrast to traditional impulse buying. When evaluating products, consumers frequently turn to the purchasing histories and product reviews of previous users on established e-commerce sites. As a result, anchors do not always need to give a thorough explanation of the features of the product. Skilled anchors would be better served by devoting their time to thoroughly exhibiting products, emphasizing features that will draw in viewers with particular needs. This strategy meets the needs of prospective clients who are actively looking for products and gives them the knowledge they need to make wise choices.

## 3. Proposed Research Methodology

The present study centers on the application of Artificial Intelligence (AI) in the identification of disinformation, primarily highlighting the crucial function of the dataset. In this particular scenario, two major issues emerge. For the purpose of training statistical models for the automated identification of disinformation, the first obstacle relates to the scarcity of high-quality datasets containing labeled instances in the Spanish language. The second problem is the significant time and effort needed to obtain these instances with labels. Improved annotation techniques that can maximize the quantity and quality of annotated data have not received much attention, as a research by Bonet-Jover (2023) points out. This paper's next sections include a thorough analysis of the state-of-the-art in disinformation datasets and a thorough overview of the literature on the "Human-in-the-Loop" idea. Lastly, a comprehensive analysis of the many approaches that are frequently used in the corpus building process in the Natural Language Processing (NLP) domain is conducted.

### 3.1 New Annotation Technique

The major focus of this work is to annotate Spanish news articles from digital newspapers in different sectors by using two well-known journalistic techniques: the FWOH (Five W and One H - FWOH) and the Inverted Pyramid. The Inverted Pyramid highlights objectivity and organization as characteristics of hard news by classifying news stories into TITLE, SUBTITLE, LEAD, BODY, and CONCLUSION categories. By addressing the fundamental issues of WHO?, WHAT?, WHEN?, WHERE?, WHY?, and HOW? in the first phrases, the FWOH technique guarantees consistency and clarity. a sample that is depicted in Figure 1. We present the RUN-AS (Reliable and Unreliable News Annotation Scheme), which relies only on textual, linguistic, and semantic analysis without external dependencies for fine-grained annotation and reliability evaluation. It takes into account factors including vagueness, subjectivity, lack of evidence, and emotionally charged content. It includes the Inverted Pyramid structure, FWOH content analysis, and Elements of Interest, allowing for a nuanced reliability assignment based on textual and linguistic analysis. Bonet-Jover et al. (2023) provide a full description of the annotation guidelines, which include signs of emotional charge.



<TITLE><WHO>Several experts</WHO> <WHAT>state that lemon can save your life</WHAT></TITLE>

<LEAD><WHEN>A few days ago</WHEN>, <WHO>several renowned experts</WHO> from <WHERE>California</WHERE> <WHAT>affirmed that lemon can save our lives</WHAT> <WHY>since it prevents and cures cancer.</WHY></LEAD>

<BODY><WHAT>Lemon has several properties</WHAT>, but according to <WHO>medical experts</WHO> <WHAT>this citrus fruit has been used</WHAT> <WHEN>for millions of years<WHEN> <HOW>with hot water</HOW>,<WHY>as drinking lemon infused water kills cancer cells in our body and creates a protective shield that prevents future tumours [...]</WHY></BODY

**Figure. 1. Sample Pyramid Structure of FWOH Annotation System**

Our proposal does not rely on conventional world knowledge for news story validity prediction. Finding the salient characteristics that define whether a piece of news is considered reliable or not is the aim, not determining its accuracy. This method offers journalists and users assistance by presenting important details in a first-level, text-only annotation. To guarantee coordination, uniformity, and adherence to the given annotation guidelines, the annotation technique entailed enlisting the assistance of a linguistic annotator. An efficient and user-friendly annotation tool, Brat allowed for the annotation of news stories quickly and effectively.

### 3.2 Semi-automatic Annotation using Human-in-the-loop Technique

A Human-in-the-Loop (HITL) methodology designed to semi-automate the dataset construction process is described in this section. By gradually introducing automation into the dataset development process, the HITL technique helped to reduce the workload of human annotators and enable the construction of larger and more economically viable datasets. The three stages of the methodology involved gradually incorporating automation into the annotation process while preserving a fixed amount of news items every batch. Gathering a wide variety of news pieces from different sources was the first step.

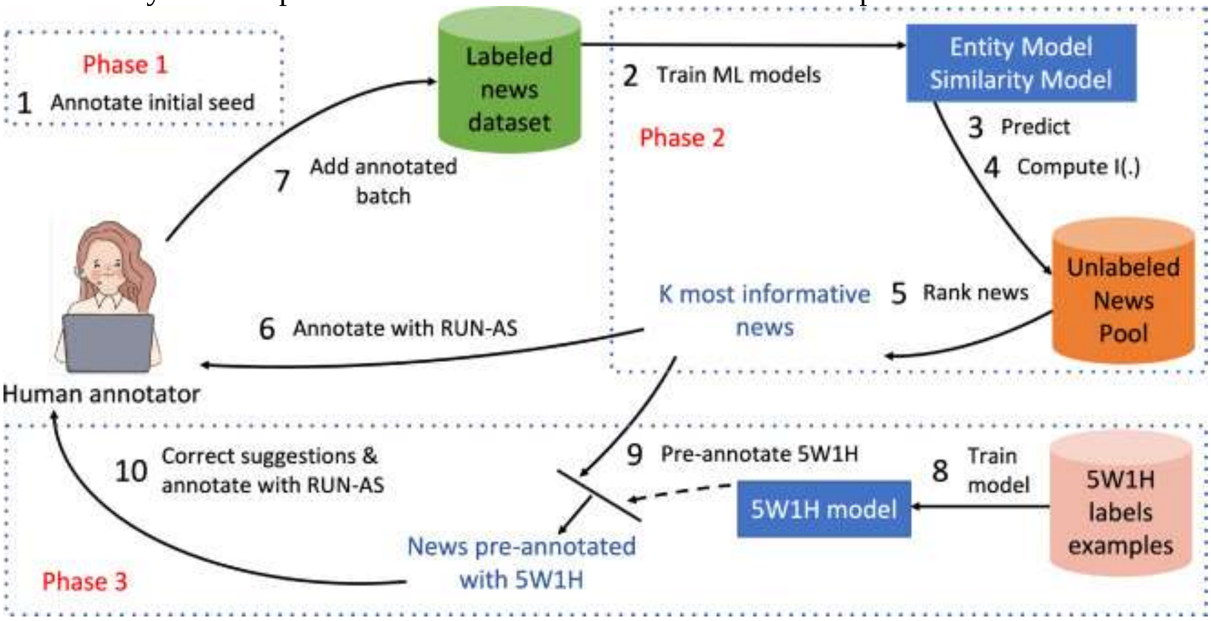


**Figure. 2. Phases of Human-in the-Loop Model**

A high-level summary of the three HITL-based phases is provided in Figure 2. Phase 1 comprised manual corpus compilation and annotation, Phase 2 automated compilation combined with human annotation, and Phase 3 automated compilation combined with semi-automatic annotation.

### 3.2 Semi – Automatic Annotation Algorithm using HITL

Phase 1: The first stage of the method entailed manual collecting and annotation, which involved obtaining forty news stories in Spanish from nine distinct sources, following the steps shown in Figure 2's first step. First, each item had to be individually retrieved and annotated; this proved to be a labor- and time-intensive process that produced the first dataset iteration, represented by the green cylinder in Figure 2.

Phase 2: In order to improve annotation efficiency, the second phase implemented active learning as a Human-in-the-Loop (HITL) technique. After training on previously annotated news from Phase 1, a supervised model ranked unlabeled content according to its informativeness. A skilled annotator whittled down the list by eliminating inappropriate material. Forty additional articles were added to the corpus as a result of repeating this method four times with ten news pieces each.

Phase 3: The third stage was the use of human-machine interaction to speed up the annotation process. Pre-annotation was developed by a system with assistance from machine learning, and was later reviewed and improved by human annotators. With the FWOH model pre-annotating a few news items, this procedure concentrated on FWOH labels. The addition of the RUN-AS annotations improved both detection reliability and machine learning retraining after human assessment and completion of pre-annotations. In line with the effectiveness of the HITL technique, this phase entailed annotating 90 news items spread across nine batches.

### 3.3 Human-in-the-Loop to Combat Fake News

Briefly put, there are three main stages to our suggested process. Annotating and manually compiling a collection of news items is Phase 1. Phase 2 introduces active learning, which enables human annotators to work with a machine learning model that automatically identifies the most instructive news articles to annotate. Annotation is still a human operation even though this phase simplifies the process and lowers compilation costs. Last but not least, Phase 3 includes human-machine interaction. A machine learning-assisted labeling system provides automatic pre-annotation, which people review and improve to create a machine-human-machine loop. The three phases of this methodology, which was first applied to dependability annotation in the context of disinformation, can be applied to a variety of complicated annotation situations. To confirm its adaptability, research on other annotation techniques is still in progress. Active learning uses entropy and informativeness formulas that are applicable to various annotation settings since they are not dependent on annotated items or relationships. It is also possible to modify the pre-annotation module to fit the particular issue. Additionally, since the machine learning models employed are language-agnostic, generalization to other languages and annotation schemes should be possible, albeit with varying baseline F1 scores based on the complexity of the learning problem. In this case, the language of the corpus, Spanish, has no bearing on the outcomes.

## 4. Experimental Results and Discussions

This section includes a number of experiments to evaluate the quality and efficacy of the selected methodology. It begins by outlining the features of the dataset. Second, it provides two dataset quality metrics, labeling consistency and inter-annotation agreement, respectively. In terms of methodology, it measures efficacy as well as efficiency. The section ends by discussing the shortcomings of the suggested strategy.

### 4.1 Dataset

Employed the Spanish Reliable and Unreliable News (RUN) Dataset using the HITL-based technique. This dataset includes 170 news stories from Latin American and Spanish digital media, rated as either Unreliable or Reliable according to the neutrality and accuracy of its semantic components. It was developed step-by-step, with 40 news items in Phase 1, 40 in Phase 2, and 90 in Phase 3. Phase 3, which was more effective, made it possible to annotate more things faster. Even with its original small size, the purpose of creating the dataset was to verify the semi-automatic approach. Compared to random manual selection, the adoption of human-in-the-loop techniques guarantees the representativeness and relevance of the dataset (Monarch, 2021).

Table 1. Reliable and unreliable FWOH in dataset

| FWOH | Unreliable items | Reliable items | Total items |
|---|---|---|---|
| **WHAT** | 687 | 1600 | 2296 |
| **WHEN** | 117 | 573 | 690 |
| **WHERE** | 685 | 58 | 747 |

| FWOH | Unreliable items | Reliable items | Total items |
|---|---|---|---|
| **WHO** | 326 | 1525 | 1856 |
| **WHY** | 142 | 241 | 384 |
| **HOW** | 165 | 358 | 529 |
| **TOTAL dataset** | 2122 | 4355 | 6502 |

This effort aims to speed up the process of creating datasets without sacrificing accuracy. In light of the length and complexity of annotation, as well as the potential for subjectivity due to the semantic nature of the guideline, the approach simplifies the process, as the following sections illustrate. Rather than depending only on dataset size, the approach's validation is based on a high-quality dataset with carefully selected examples that improve accuracy.

### 4.2 Measuring the Quality of Annotation

We looked at the distribution of FWOH annotations in Phase 1 (without pre-annotation) and Phase 3 (with pre-annotation) in order to validate the resource produced using the semi-automatic methods and evaluate the quality difference between batches with and without pre-annotations. Furthermore, inside the final RUN dataset, we assessed the inter-annotator agreement. Using the formula IAA = number of matches / (number of matches + number of non-matches), two experienced linguistic annotators evaluated the quality of the dataset in additional detail.

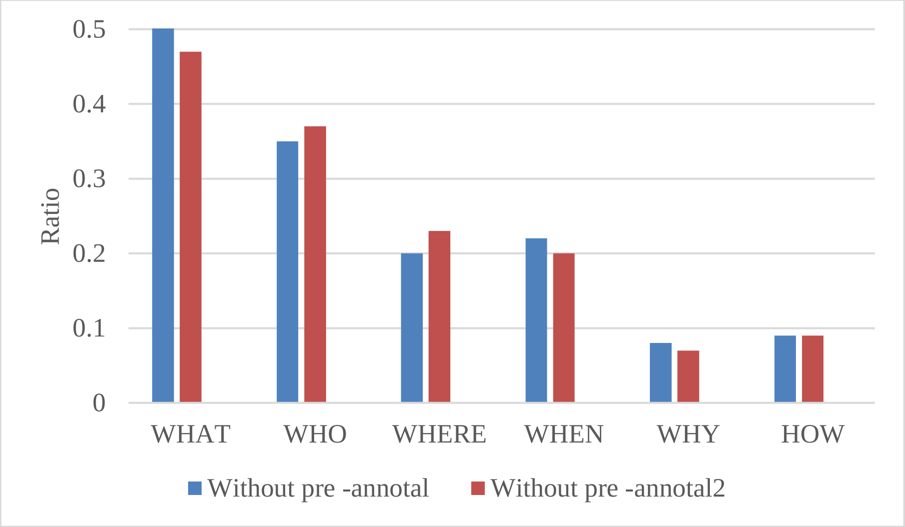


Figure. 3. Comparison of the annotations batches with and without pre-annotations in FWOH

### 4.3 Time Reduction

The amount of time necessary to locate, read, and save each news item was the basis for the compilation task's time calculations. In a similar vein, the annotation job was chosen based on how much time was allotted to each news item's annotation. There is a presentation of the overall average time per news item and the average time per news item in minutes for each phase.

The first phase of the compilation process required fifteen minutes on average for each news item. This required manual labor for things like finding sources, evaluating content for appropriateness, storing, and integrating into the user interface. During this phase, each news item required manual annotation, which took an average of 16.7 minutes.

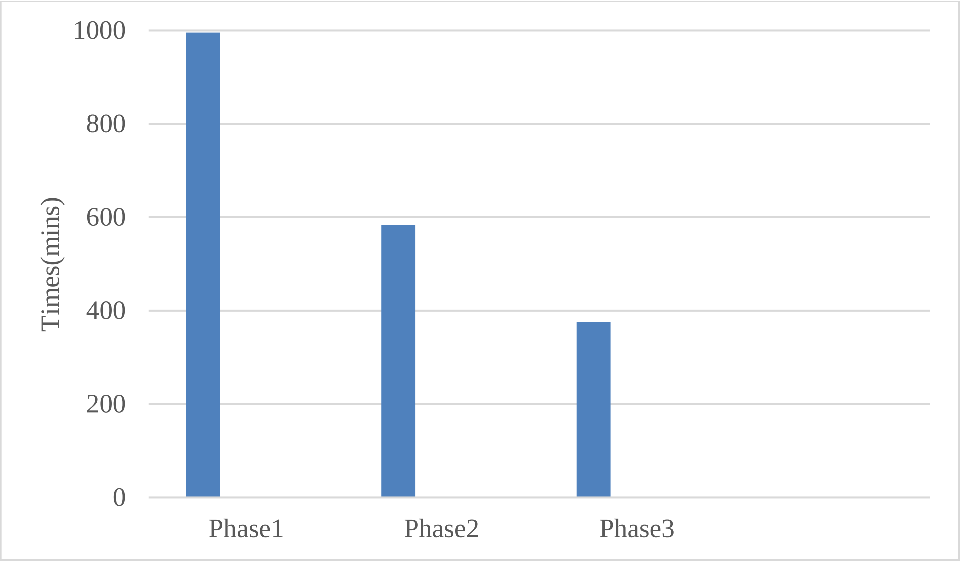
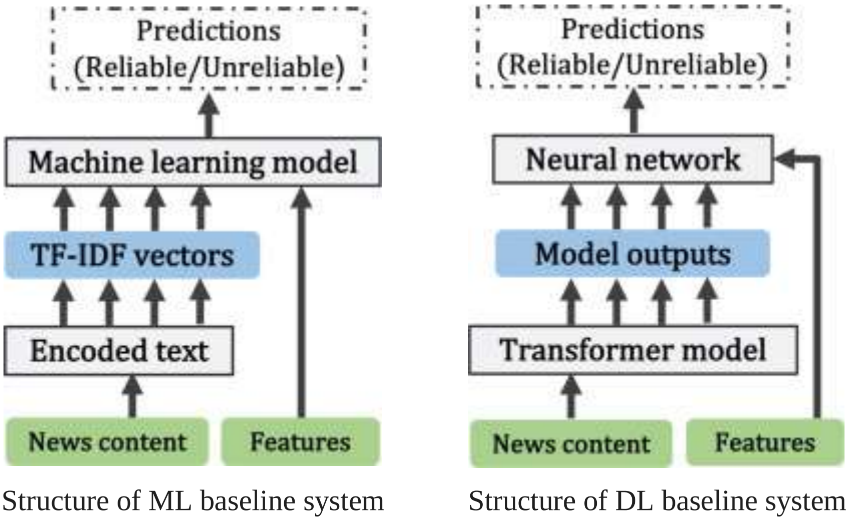


**Figure. 4. Time Reduction Performance**

### 4.4 Prediction Performance

By employing two baseline systems to forecast a random test set, we assessed the semi-automatically created dataset in order to verify the efficacy of the suggested methodology for dataset development. Figure 5 shows the internal organization of the baseline systems for ML and DL, respectively. To forecast the test set, these baseline systems underwent training using the RUN dataset. The input for the baselines was always the news material (TITLE and BODY text), with the addition in certain cases of features from fine-grained annotation. Each baseline produced a categorization of Dependable or Unreliable.



Structure of ML baseline system Structure of DL baseline system

**Figure 5. Baseline Structure**

Figure 5 shows the baseline based on DL. Here, five prediction models were used to encode the news material before it was evaluated. A Multilayer Perceptron (MLP) was the neural network's unique training and prediction model, and its inputs were the encoded vector (model output) and the annotated features. The baseline embraced the Sepúlveda-Torres et al. (2021) classification architecture, which blends the transformer model with external features. Because the BETO pre-trained model is a Spanish language model with a reputation for performing well on a variety of NLP tasks, we chose it as the transformer model for the baseline system in this case (Canete et al., 2020). The BETO model is in the state of fine-tuning. Table 2 displays the suggested models' prediction performance.

Table 2. Performance of ML and DL models for predicting test set with features and without them.

| **Model** | **Without features** | | **With features** | |
|---|---|---|---|---|
| | **F-Score** | **Accuracy** | **F-Score** | **Accuracy** |
| **Proposed Method** | **0.88** | **0.89** | **0.96** | **0.97** |
| **TRANSFORMERS [10]** | **0.82** | **0.84** | **0.92** | **0.93** |
| **BERT [11]** | **0.82** | **0.83** | **0.89** | **0.90** |
| **GPT [12]** | **0.79** | **0.81** | **0.83** | **0.85** |
| **REGRESSION** | **0.74** | **0.74** | **0.949** | **0.95** |

Table 7 shows a comparison of the accuracy and F-Score of several models with and without the use of extra features. Interestingly, adding more characteristics to the "Proposed Method" shows a notable improvement. It shows a significant increase in Accuracy (from 0.89 to 0.97), as well as F-Score (from 0.88 to 0.96). This improvement implies that the suggested approach is quite successful in making the most of these extra features to boost performance.

The "TRANSFORMERS [10]," "BERT [11], and "GPT [12]" models also demonstrate enhanced functionality upon integration of more features. This pattern emphasizes how well these elements work together to improve model performance overall. In contrast, the F-Score and Accuracy of the "REGRESSION" model initially lag behind. That performs much better with extra features, though, especially when it comes to F-Score.

Overall, the findings demonstrate how adding more features can enhance the efficacy and accuracy of the model. With attributes taken into account, the suggested technique in particular shines out as it earns the greatest F-Score and Accuracy, demonstrating its effectiveness in the specific work context. These results imply that feature augmentation can be an effective tactic for improving models' efficacy in tasks linked to misinformation detection.

## 5. Conclusion and Further work

The creation and implementation of a human-in-the-loop methodology for the semi-automatic annotation of complex datasets is the approach's primary novelty. There are two key components to this process. To determine which news articles are most suited for annotation, it first uses the tried-and-true method of Active Learning (AL). It uses a diversity sampling approach, namely. Second, it presents a process of human-machine interaction in which human annotators improve and assess the pre-annotations automatically produced by a labeling system with assistance from machine learning. An intelligent selection of news items for annotation and the provision of highly confident pre-annotated suggestions are the two key components of this methodology that lead to an improvement in the annotation process. Future research should look into expanding cross-lingually, diversifying domains, real-time detection, sophisticated semantics, hybrid models, ethics, international cooperation, transparency, and custom solutions.

## References

[1] Xie Xinzhou, Zhu Yaoying. Research on the development trend and response strategies of online content governance [J]. News and Writing, 2020(4):76-82.

[2] Caplan R, Gillespie T. Tiered governance and demonetization: The shifting terms of labor and compensation in the Platform economy[J].Social Media Society,2020(4-6):1-13.

[3] Xiao Mengli. Research on power generation and regulatory selection of platform enterprises [J]. Hebei Law, 2020, 38(10):75-89.

[4] Xie Xinzhou, Zhu Yaoying. Research on EU digital copyright protection from the perspective of information resources management [J]. Journal of Information Resources Management, 2020, 10(6): 60-70.

[5] Zhou Hui. Ideal types of Internet platform governance and good governance—from the perspective of the relationship between the government and platform enterprises [J]. Journal of Law, 2020, 41(9):24-36.

[6] Song Xiaokang, Zhao Yuxiang, Song Shijie, et al. Research on factors affecting willingness to share health rumors based on MOA theory [J]. Journal of Information Science, 2020, 39(5):511-520

[7] Tong Wensheng, Yi Baihui. Network rumor refutation: Domestic research progress and theoretical analysis framework [J]. Journal of Intelligence, 2020, 39(6):128-134, 202.

[8] Li Z， Zhang Q， Du X， et al. Social media rumor refutation effectiveness: evaluation， modelling and enhancement [J]. Information Processing & Management， 2021， 58(1):102420.

[9] Anderson P F， Allee N， Grove S， et al. Web site evaluation checklist， University of Michigan [EB/OL]［2021-06-25］.http:// www personal.umich.edu/~pfa/pro/courses/WebEvalNew.pdf

[10] Lo P S, Dwivedi Y K, Tan G W H, et al. Why do consumers buy impulsively during live streaming? A deep learning based dual-stage SEM-ANN analysis[J]. Journal of Business Research, 2022, 147: 325-37.

[11] Xu Y, Ye Y X. Who watches live streaming in China? Examining viewers' behaviors, personality traits, and motivations[J]. Frontiers in Psychology, 2020, 11: 1607.

[12] Xue J L, Liang X J, Xie T, et al. See now, act now: How to interact with customers to enhance social comer ceengagement? [J]. Information & Management, 2020, 57(6): 103324.

[13] Zhang M, Sun L, Qin F, et al. E-service quality on live streaming platforms: Swift guanxi perspective [J]. Journal of Services Marketing, 2021, 35(3): 312-24.

[14] Wong kitrungrueng A, Assarut N. The role of live streaming in building consumer trust and engagement with social commerce sellers [J]. Journal of Business Research, 2020, 117: 543-556.

[15] Huang Minxue, Ye Yuqian, Wang Wei. The impact of live broadcast anchor types on consumer purchase intentions and behaviors under different types of products [DB/ OL]. (2021-09-15) [2022-12-05]. http:// / kns. cnki.net/ kcms/ detail/12. 1288. F. 20210915. 0954. 002. html.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in NIPS, 2017.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for languageunderstanding,” in NAACL, 2019, pp. 4171–4186.

[18] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving language understanding by generative pretraining,” 2018.

[19] Bonet-Jover, A., Sepúlveda-Torres, R., Saquete, E., Martínez-Barco, P., Piad-Morffis, A., & Estevez-Velarde, S. (2023). Applying Human-in-the-Loop to construct a dataset for determining content reliability to combat fake news. Engineering Applications of Artificial Intelligence, 126, 107152.

[20] Laskar, M. T. R., Chen, C., Fu, X. Y., & Tn, S. B. (2022). Improving named entity recognition in telephone conversations via effective active learning with human in the loop. arXiv preprint arXiv:2211.01354.