Towards Effective Fake News Detection

Name: Nilkant Kamble

Country: India

Abstract:

The emergence of misinformation, especially fake news, in the digital era has presented serious obstacles to social cohesiveness, public confidence, and the accuracy of information distribution[1]. While human oversight offers a more sophisticated knowledge, modern automated techniques frequently lack it.[2]. Traditional fact-checking methods are labour-intensive and slow.[3] In order to enhance the identification of false news, we describe in this study a hybrid strategy that combines machine learning with human verification.[4] We examine the approaches and constraints of current tools, like FactCheck.org, CaptainFact, and ClaimBuster, and we suggest a new framework that makes use of both human expertise and the scalability of automated systems.This hybrid model combines the contextual awareness and interpretive skills of human reviewers with a fast and efficient way to process huge datasets[5]. Our hybrid approach performs better in terms of accuracy and scalability than entirely automated or manual solutions, according to comparative analysis.[6] We conclude by discussing the limitations and future work needed to extend this system to mitigate fake news and thus disinformation.

Keywords: Disinformation, Fake news, Detection, Machine Learning, Fact-checking.

Introduction:

The quick spread of information in the era of digital connection has drastically changed how we get and distribute news. Disinformation, on the other hand, is a ubiquitous and sneaky issue that has been brought about by this unparalleled information flow [7]. The intentional dissemination of incorrect or misleading information with the goal of deceiving or controlling public opinion is known as disinformation, and it has grown to be a major problem for preserving the accuracy of news and information.[8] This phenomenon, which is sometimes called "fake news," has a significant impact on public opinion, political processes, and society's trust.[9]

Disinformation is not a novel idea; it has always been used to influence and cause disruption in a variety of ways. The word first appeared in Soviet vernacular, where it denoted deliberate lies used to destabilise enemies and manipulate public opinion[10]. However, the prevalence and effect of misinformation have significantly increased since the introduction of social media and digital media [1]. Fake news spreads now at a never-before-seen velocity and volume, taking advantage of the enormous reach of internet platforms to permeate and influence public conversation.[12]

Fake news operates on complex mechanisms that frequently take use of emotional responses to increase their impact. In order to trick viewers, false narratives are written to seem like real news articles. They achieve this by copying the structure and tone of authentic journalism. People find it difficult to distinguish between manufactured content and true information because of this purposeful resemblance. Fake news spreads mostly because of its emotional appeal, which can be achieved through sensationalism, inciting fear, or inciting indignation.[13] Stories that evoke strong emotions are more likely to be shared and reinforced on social media.

Widespread misinformation has far-reaching effects. Fake news has the potential to increase social polarization, weaken public faith in democratic processes, and diminish trust in media organizations. For example, the 2016 U.S. presidential election demonstrated how false news might be used as a weapon to stoke division and affect political outcomes[14].Similarly, during the COVID-19 pandemic, erroneous information about the virus and its treatment was quickly disseminated, hampering attempts to promote public health and putting lives in danger[15][16]

A diversified strategy is needed to address the false news problem. The dynamic nature of disinformation has proven to be too much for rule-based computers and traditional fact-checking techniques to keep up with. These methods frequently fail to convey the subtleties of sentiment and context that define fake news. Consequently, there is an increasing demand for more advanced solutions that can improve the dependability and accuracy of fake news identification.

Literature Review

One of the most urgent problems facing modern digital society is the proliferation of fake news, which has a big impact on how people consume and trust information. As social media platforms and user-generated material have grown in popularity, it has become more difficult to distinguish between factual news, skewed opinions, and blatant lies[9][11]. The topic of fake news has been the subject of much research, which has looked at its traits, causes, and the state of detection techniques.[17] This review of the literature examines the definitions and categories of fake news, investigates the technological and cognitive elements that contribute to its spread, and assesses the advantages and disadvantages of current detection models. It also emphasises the need for further advancements, particularly hybrid systems that combine machine learning and human judgement.[18]

Determining what fake news is is a major obstacle in combating it. Although "fake news" is frequently used as a general phrase to refer to inaccurate or deceptive information, its meaning has broadened to include a variety of manipulative content.[19] Five major categories are included in Tandoc et al. (2018)'s typology of fake news: news satire, news parody, fabrication, photo manipulation, and news-like advertising material. This classification demonstrates that fake news encompasses content that purposefully or inadvertently distorts reality in addition to entirely made-up stories[19]. Even though they are not intended to be misleading, satirical and parodic news can still perplex viewers, particularly if it is disseminated outside of its intended context.[20]

The distinction between disinformation and misinformation is another crucial one. Disinformation is the purposeful fabrication and transmission of lies in order to deceive, whereas misinformation is the accidental spread of inaccurate or misleading information [9]. Additionally, Wardle and Derakhshan (2017) present the idea of mal-information, which is when factual information is misrepresented or taken out of context in order to cause harm to people or organisations[21]. Because different forms of fake news may call for different approaches to detection and mitigation, these distinctions are crucial for developing detection systems.

Because social media sites like Facebook, Twitter, and YouTube act as amplifiers for the quick spread of inaccurate or misleading information, they have been crucial in the emergence of fake news. On these platforms, fake news travels much more quickly and widely than true material, according to studies like those by Vosoughi, Roy, and Aral (2018). According to their research, false news's emotional appeal—particularly when it contains content that incites shock, terror, or outrage—contributes to its virality.[1] This result is consistent with Wardle and Derakhshan's (2017) finding that stories that evoke strong emotions in consumers tend to garner greater engagement.[21]

Furthermore, sensationalist information is frequently promoted by social media platforms, which are driven by algorithms that value interaction[11]. When people are constantly exposed to material that supports their prejudices and ideas, it produces the "echo chamber" effect. Fake news is harder to identify and combat because of the mix of cognitive biases and algorithmic content selection, which creates an atmosphere in which it can flourish.

A major factor in the allure and spread of fake news is cognitive bias. According to Lazer et al. (2018), people are more vulnerable to fake news because of cognitive processes including confirmation bias, which occurs when people favour information that supports their preconceived notions[1]. The phenomenon known as "motivated reasoning," which leads people to interpret information in a way that supports their worldview even in the face of contradicting facts, exacerbates this propensity.[18] Because people are psychologically prone to reject corrections that contradict their views, the cognitive foundations of fake news consumption make it more difficult to create detection systems that work.[22]

Another important cognitive hurdle to fighting fake news is the boomerang effect, which occurs when refuting a misleading claim might increase belief in the disinformation [23]. According to this field's research, fact-checking alone could not always be enough and occasionally even backfire. [23]

Researchers have looked to machine learning and artificial intelligence (AI) as crucial techniques for detection due to the scope and complexity of fake news propagation. To categorise news information according to linguistic and statistical characteristics, machine learning models—such as those based on Support Vector Machines (SVMs), Random Forests, and neural networks—have been used extensively[24]. An overview of the usage of these models to process huge datasets and find patterns that differentiate authentic information from fake news is given by Zhou et al [25]. These models are frequently trained using linguistic variables including word frequency, sentiment, and grammatical structure.

Despite their scalability, machine learning approaches face several challenges. One major issue is their inability to detect more nuanced forms of fake news, such as satire or biased opinion pieces that do not contain outright falsehoods but manipulate information to mislead.[26] These models often struggle with the subjective nature of biased reporting, where the facts may be presented truthfully but framed in a way that skews the reader's perception [22]. Additionally, machine learning systems are limited by the data they are trained on; they may fail to adapt to new types of disinformation or emerging forms of media manipulation, such as deepfakes.[27]

Natural Language Processing (NLP) has been investigated as an advanced method for detecting fake news in addition to conventional machine learning techniques. Naredla et al. suggest analysing sentiment, semantics, and narrative structures in news articles using natural language processing (NLP) approaches[28]. NLP systems can detect patterns that can point to prejudice or dishonesty by looking at how language is used to change facts or express emotions.[29]

But as Huffaker et al point out, NLP models have trouble comprehending context in real time[30]. Fake news frequently takes use of current sociopolitical or cultural settings that natural language processing (NLP) algorithms might not fully understand, particularly if those models are trained on static datasets [31]. Researchers propose combining human knowledge with machine-driven systems to overcome these drawbacks, resulting in hybrid models that capitalise on the advantages of both automation and human judgement.[32][33]

Due to the drawbacks of fully automated systems, hybrid detection techniques—which mix machine learning with human oversight—are becoming more and more popular[34]. Pennycook and Rand (2019) support an approach that combines machine-learning algorithms with user feedback and expert fact-checkers[18]. More sophisticated detection is possible with this method, especially when computers could have trouble picking up on subtle satire or manipulation[22]. AI systems frequently lack the deeper comprehension of context, cultural allusions, and sociopolitical subtleties that human specialists provide.[24]

Another idea for involving the public in the discovery of fake news is crowdsourcing. According to studies like those by Friggeri et al. (2014), when combined with expert verification, crowd-sourced fact-checking can be useful in identifying questionable content.[22] Similar approaches might be utilised for news verification, as crowdsourcing has been effectively used by websites such as Wikipedia]to maintain accurate information [35]. This strategy, however, raises questions regarding the veracity and bias of information gathered from crowdsourcing as well as the possibility of system manipulation by hostile parties.[36]

From this review, it is evident that while machine learning offers scalability, its limitations in understanding context and subtle cues remain a barrier to accurate fake news detection. This highlights the need for a hybrid system that combines the strengths of both human expertise and machine learning to address these challenges.[32][37]

Existing tools:

CaptainFact: is a community-driven, open-source platform that gathers fact-checking for web videos using crowdsourcing.[38][39] The platform combines multiple methods for real-time fact-checking, media annotation, and evidence gathering by utilizing user involvement.[39] With tools that allow users to collaboratively examine claims made in public discussions, interviews, and online media, CaptainFact.org is specifically designed to combat misinformation based on videos.

With a backend based on Node.js, CaptainFact.org's architecture is both modular and scalable, enabling effective asynchronous processing of user interactions and real-time data handling.[39] Data transfer between the user interface, the database, and external APIs is coordinated by this backend. Utilising React.js on the front end, CaptainFact.org offers a responsive and interactive user interface (UI) that allows users to fact-check video content in real-time. The platform supports browser extensions for Chrome and Firefox, enabling users to verify claims directly on video platforms like YouTube without leaving the content. PostgreSQL is used as the database for data management, where user information, claims, timestamps, and validated sources are stored. Particularly as the user base expands, its relational structure enables efficient evidence retrieval for every claim and optimized query formulation. Through the use of REST APIs for integration with external services, the platform may retrieve data for real-time source validation from reliable sources like academic libraries, news sources, and Wikipedia.

Claim identification is the first step in the fact-checking process, where users can utilize timestamps to mark particular video segments and extract written or spoken assertions.[2] Natural Language Processing (NLP) techniques are applied to these assertions in order to automatically classify the content (e.g., health, politics, economics)[40]. After that, there is source validation, in which users or moderators affix proof to confirm or deny the assertion. The software makes use of URL verification algorithms and web scraping to make sure the sources are reliable domains with pertinent data.

To further ensure that high-quality evidence is promoted, a citation scoring system gives preference to peer-reviewed academic articles, government reports, and other credible sources. After the claim and evidence are submitted, a peer verification process is initiated, in which users vote on the accuracy of the claim. CaptainFact.org uses a reputation-based voting system, in which users are assigned reputation scores based on their contribution history and accuracy in previous fact-checking. These scores are calculated using trust algorithms for the verification process. Lastly, a consensus algorithm makes sure that a claim is only marked as verified when a sufficient number of high-reputation users concur on the validity of the evidence, guaranteeing transparency and lowering the possibility of biased or inaccurate conclusion.

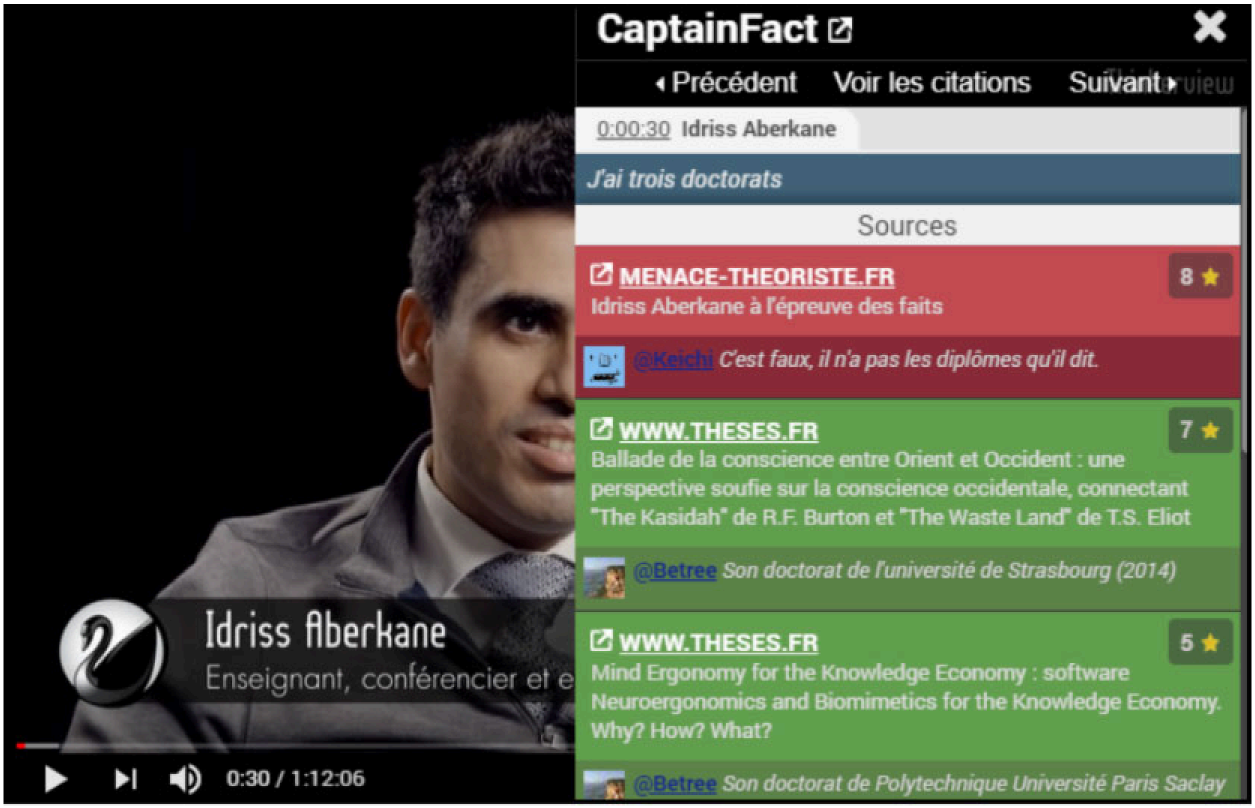


Figure a. Captain Fact user interface

source:https://captainfact.io/

FactCheck.org: As a centralized, nonpartisan fact-checking organisation, FactCheck.org verifies claims mainly in the political and public policy domains by utilising a sophisticated web infrastructure, digital technologies, and editorial processes[41] [43]. Traditional content management systems (CMS), like WordPress, are used in the construction of the platform's backend. These systems are optimised for scalability and searchability, enabling the platform to handle heavy traffic and store articles, fact checks, and references in an effective manner[42]. FactCheck.org employs a PHP-based backend architecture to dynamically generate web pages, while the frontend is designed using HTML5, CSS, and JavaScript, creating a user-friendly interface with comprehensive search functionality[42][44]. The site uses a relational MySQL database to store fact-checks, claim metadata, references, and sources, ensuring efficient retrieval of fact-checked information for users. To verify statements about legislative and political issues, FactCheck.org interacts with APIs and data sources from the Congressional Budget Office, government archives, and public documents.

At FactCheck.org, the process of fact-checking starts with a thorough claim identification step, during which editorial staff members search speeches, interviews, news articles, and social media for noteworthy assertions. The selection of claims is predicated upon their potential impact and relevance[45]. Instead of employing automation for this procedure, FactCheck.org uses a journalistic review paradigm in which editors and researchers manually compile claims and start the verification process[46]. Validating sources requires in-depth investigation, cross-referencing statements with official statistics, government papers, expert interviews, and reliable databases, among other primary sources.[47] Manual content analysis is used by FactCheck.org to guarantee that its sources are reliable, current, and impartial. The editorial staff of the platform then creates thorough reports that fully disclose the assertion, the supporting data, and the conclusion while properly attributing all references. Senior editors evaluate fact-checks before to publication as part of a multi-tiered peer review process that guarantees objectivity and accuracy for every story. In contrast to crowdsourcing models, FactCheck.org uses professional inspection to ensure factual accuracy rather than depending on user contributions or voting mechanisms. It does, however, include user feedback tools so that readers can propose revisions or more claims for examination, guaranteeing community involvement while preserving editorial control.

Figure b. FactCheck.org user interface.

Source: https://www.factcheck.org/

ClaimBusters: is a sophisticated, automated fact-checking tool that combines machine learning and natural language processing (NLP) methods to recognise and validate assertions in speeches, news stories, and other textual material. By offering automated tools for real-time claim identification and validation, the platform seeks to solve the problem of disinformation.[4]

Python and a number of machine learning packages, such as TensorFlow and Scikit-learn, which make it easier to create and implement NLP models, power the ClaimBusters backend.[48][49] The architecture of the platform makes use of a microservices paradigm, in which various services independently manage functions including claim detection, evidence collection, and verification.[50.] This method improves scalability and enables frequent modifications to individual parts. The main database system, PostgreSQL, is used to manage claim records, metadata, and user data. Large text volumes and verification results require effective data administration, which is made possible by the relational structure of the database, which also facilitates complicated searches[51]. In order to assess claims and offer context, ClaimBusters also connects with outside sources using REST APIs. Information from academic libraries, news databases, and fact-checking organizations are accessed.

The claim detection step of the fact-checking workflow involves the platform's NLP models analyzing incoming text to locate and extract particular claims. Large datasets of labeled claims and fact-checks are used to train these models, which then use methods like topic modeling and named entity recognition (NER) to comprehend the claims' context and content. Following the detection of claims, the system retrieves essential information from reliable sources through automated web scraping and API integration. This stage entails evaluating the validity of the sources and making sure the information received either correctly confirms or contradicts the claims.

During the verification stage, ClaimBusters assesses the consistency and reliability of the evidence using machine learning techniques. The software processes and analyses the data using neural networks and support vector machines (SVMs), then utilises classification models to group the assertions according to their accuracy. A confidence score that indicates the probability of the claim being true or false is then displayed alongside the results. the machine learning models.

Here's a focused comparative matrix that highlights the strengths and limitations of **CaptainFact.org**, **FactCheck.org**, and **ClaimBusters**:

Tool	**Strengths**	**Limitations**

CaptainFact.org	- **Real-time Fact-Checking**: Provides immediate verification of claims within video content. - **Crowdsourced Verification**: Leverages community involvement for a broad range of perspectives. - **Modular Architecture**: Scalable and adaptable to growing user base. - **Browser Extensions**: Integrates directly with video platforms like YouTube for seamless fact-checking.	- **Potential Bias**: Crowdsourced nature may introduce bias depending on the user base. - **Quality Control**: Variability in the accuracy of contributions may affect the reliability of fact-checking. - **Scalability Challenges**: As user volume grows, maintaining the quality and speed of verification can be challenging.
FactCheck.org	- **Professional Oversight**: Expert-driven fact-checking ensures high accuracy and reliability. - **Established Reputation**: Well-regarded for its thorough, unbiased reviews. - **Comprehensive Coverage**: Provides in-depth analysis and extensive evidence for claims. - **Detailed Reports**: Offers clear, well-documented fact-checking results.	- **Slower Updates**: Manual processes may result in slower response times to emerging misinformation. - **Limited to Textual Content**: Primarily focuses on articles and speeches, missing video content and other media. - **Centralized Model**: Relies on a smaller team of experts, which may limit the breadth of covered claims.

ClaimBusters	- **Automated Claim Detection**: Uses advanced NLP and machine learning to quickly identify and assess claims. - **Scalable Architecture**: Microservices model supports large-scale data processing and real-time analysis. - **Data-Driven**: Leverages large datasets for training and validation, enhancing accuracy. - **Efficiency**: Streamlines the fact-checking process with minimal manual intervention.	- **Contextual Limitations**: Automated systems may struggle with nuanced or context-dependent claims. - **Machine Learning Accuracy**: Reliance on algorithms may result in occasional errors or missed claims. - **Lack of Human Oversight**: Absence of manual review may reduce the ability to handle complex cases or ensure nuanced understanding.

Proposed Solution.

The growing sophistication and volume of disinformation make it essential to design a system that combines both the speed of machine learning algorithms and the contextual understanding of human expertise.[52][53] This section outlines a **hybrid fake news detection system** that integrates machine learning, natural language processing (NLP), and human intervention to accurately detect fake news.[53][54] The proposed solution operates in several layers to ensure rapid processing and accurate detection.[55][56]

Methodology

The proposed fake news detection system is built on a five-layer architecture designed to combine the strengths of machine learning with human expertise to accurately identify disinformation[57]. Each layer plays a unique role in the detection and validation process, ensuring efficiency and reliability.

1. Input Layer

The Input Layer is responsible for gathering data from various online sources such as social media platforms, news websites, and online forums. The data collected is pre-processed to remove noise, irrelevant information, and special characters. This layer ensures that the system only analyzes clean, structured data, which is crucial for the subsequent machine learning steps.[58]

2. Machine Learning Layer

Once the data is prepared, the Machine Learning Layer processes it using advanced Natural Language Processing (NLP) algorithms and machine learning models like **Support Vector Machines (SVM)** and **Random Forests**.[59] This layer extracts key features from the input data, such as sentiment, linguistic patterns, and emotional tone. The models are trained to identify disinformation patterns by comparing these features against pre-existing datasets of verified and fake news articles.[60] This automated analysis helps to quickly classify content and flag suspicious items.

3. Output Layer

The results generated by the machine learning models are then delivered through the Output Layer. The system provides a preliminary decision on whether the news is “Fake” or “True” while offering additional contextual data.[61] This includes similar articles, emotion analysis, and other supporting features that help users gain a comprehensive understanding of the content.

4. Human Review Layer

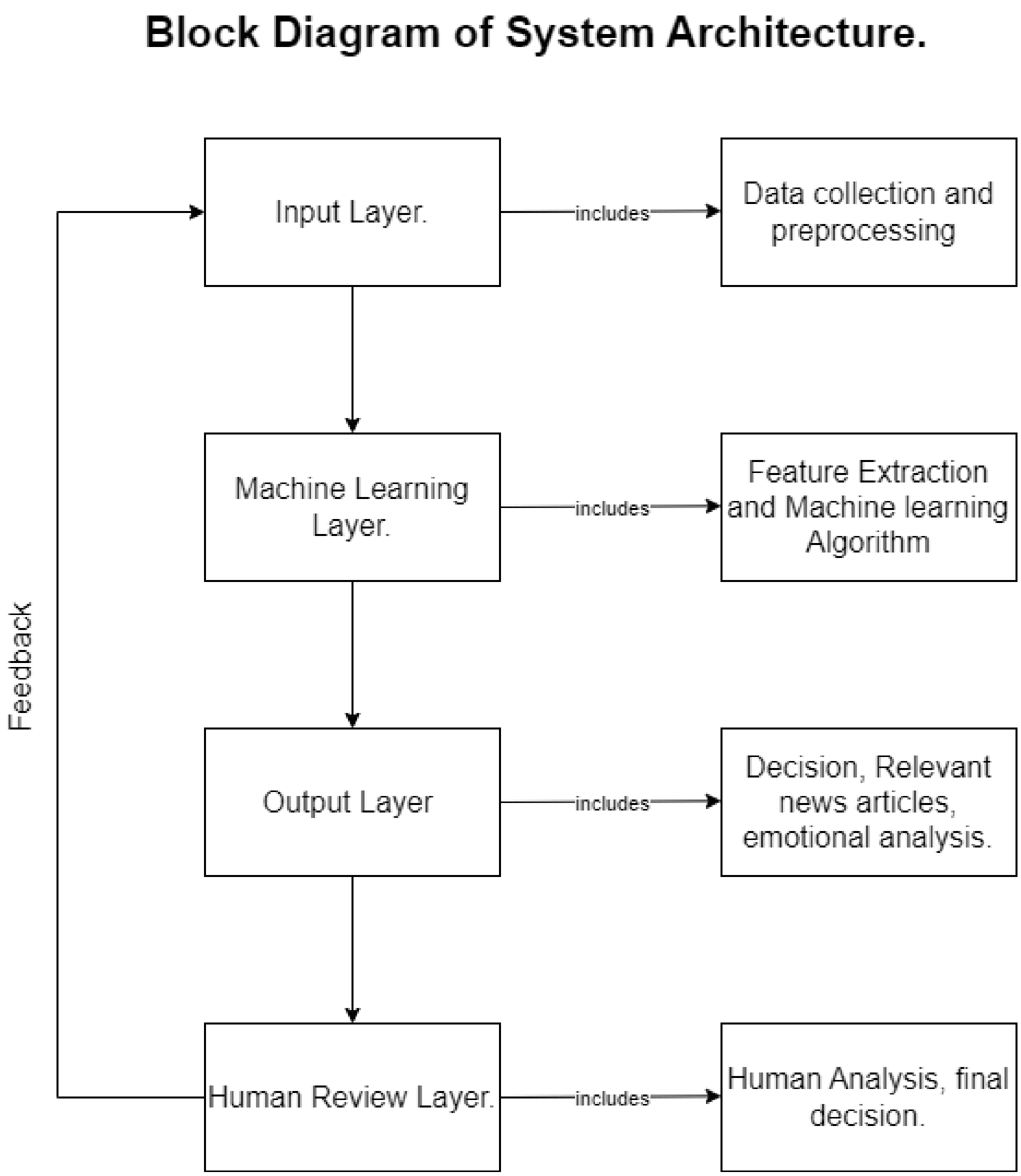
Flagged content is escalated to the Human Review Layer, where expert reviewers analyze it further. Human intervention is essential for understanding subtleties that algorithms may overlook, such as cultural context, intent, and nuance. Reviewers provide a more in-depth analysis, ensuring that content classified as disinformation receives thorough vetting before any final decision is made[62]

5. Feedback Layer

The final decision made by the human reviewer is passed back to the system as feedback. This feedback is used to refine the machine learning models, improving their ability to detect fake news over time.[63]. The system continuously learns from the outcomes of human reviews, enhancing its accuracy in identifying disinformation patterns.[64]

System Architecture

1. **Input Layer:** Collects data from social media, news websites, and online forums.
2. **Machine Learning Layer:** Utilizes NLP algorithms and machine learning models (e.g., Random Forests, SVM) to detect disinformation patterns in the input data.
3. **Output Layer:** The system outputs the decision as well as relevant news articles, emotion analysis and other features for human review.
4. **Human Review Layer:** Flagged content is analyzed by human reviewers who evaluate the context, intent, and subtlety of the article. Make a comparative and educated final decision based on the information provides
5. **Feedback:** The final decision made by user is then provided as feedback to the system for updates.



Future Work:

The future work for this research will concentrate on implementing the proposed hybrid system, creating and performing tests to assess its usefulness, and investigating the dynamics of human-computer interaction[65]. In order to enable human engagement, the system will first be constructed by integrating machine learning models—such as NLP algorithms, Random Forests, and SVMs—with an intuitive user interface.[66] Social media, news websites, and online discussion boards will all provide data that will be used to train and test the algorithm in scenarios including the genuine identification of deception.

Experiments will be created after the system is constructed to assess its accuracy and effectiveness.[67] Through the consideration of various material types and situations, these tests will assess the system's capacity to identify misinformation on a variety of platforms. The system will be compared against current systems using key metrics like precision, recall, and response speed, with an emphasis on maximising the blend of machine learning insights and human judgements. [68]

Future research will also examine human-computer interaction in great detail. It's critical to examine how users engage with the system because it depends on human reviewers to reach final, informed conclusions.[69] To improve the overall design and functionality, the efficiency of the human review layer, the cognitive burden on users, and their confidence in the system will be assessed. It will be essential to comprehend how consumers use machine-generated insights to make well-informed judgements in order to improve the system for real-world use.

Conclusion:

In this paper, a hybrid approach that combines human experience and machine learning techniques for the detection of fake news is introduced. Our method aims to address the drawbacks of current approaches, including the labour-intensive, sluggish process of manual fact-checking observed in FactCheck.org and the contextual shortcomings of automated systems such as ClaimBusters. The approach combines speed and accuracy by combining the capacity of machine learning to process large datasets with the interpretive skills of human reviewers. When possibly fraudulent content is swiftly identified by the machine learning layer, human reviewers offer in-depth examination to guarantee that subtle content is accurately read.

The system's automated data processing allows for scalability, and human monitoring improves accuracy. Over time, the system's performance can be improved by continuously improving the machine learning models through the use of feedback loops. Compared to machine-only or manual methods, this hybrid approach to fake news detection offers a quicker and more accurate answer.

The system is not without its drawbacks, though. Because some content is subjective, using human reviewers exclusively could lead to bottlenecks during periods of widespread disinformation. Additionally, bias could be introduced into the human review process. Furthermore, in order to update machine learning models, the feedback system needs regular human input, which could impede real-time decision-making.

In subsequent research, we'll concentrate on streamlining the human review procedure, possibly leveraging crowdsourcing to expand the system, and further automating comments to lighten the workload on reviewers. Even though the suggested approach shows promise in combating the growing problem of misinformation, continuous improvements are required to preserve accuracy, speed, and scalability in a changing media environment.

References:

1.Vosoughi, S., Roy, D. and Aral, S., 2018. The spread of true and false news online. *science*, *359*(6380), pp.1146-1151.

2. Graves, D., 2018. Understanding the promise and limits of automated fact-checking. *Reuters Institute for the Study of Journalism*.

3.Bulganová, D., 2020. Fact-Checking as a Key Competence in Infodemia. *Marketing Identity*, *8*(1), pp.54-61.

4.Hassan, N., Zhang, G., Arslan, F., Caraballo, J., Jimenez, D., Gawsane, S., Hasan, S., Joseph, M., Kulkarni, A., Nayak, A.K. and Sable, V., 2017. Claimbuster: The first-ever end-to-end fact-checking system. *Proceedings of the VLDB Endowment*, *10*(12), pp.1945-1948.

5. Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T.H., Ding, K., Karami, M. and Liu, H., 2020. Combating disinformation in a social media age. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(6), p.e1385.

6. Thorne, J. and Vlachos, A., 2018. Automated fact checking: Task formulations, methods and future directions. *arXiv preprint arXiv:1806.07687*

7. Wardle, C. and Derakhshan, H., 2017. *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1-107). Strasbourg: Council of Europe

8. Jack, C., 2017. Lexicon of lies: Terms for problematic information. *Data & Society*, *3*(22), pp.1094-1096.

9. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D. and Schudson, M., 2018. The science of fake news. *Science*, *359*(6380), pp.1094-1096.

10. Rid, T., 2020. *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.

11. Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C.M., Brugnoli, E., Schmidt, A.L., Zola, P., Zollo, F. and Scala, A., 2020. The COVID-19 social media infodemic. *Scientific reports*, *10*(1), pp.1-10.

12. Allcott, H. and Gentzkow, M., 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), pp.211-236.

13.Martel, C., Pennycook, G. and Rand, D.G., 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications*, *5*, pp.1-20.

14. Howard, P.N. and Kollanyi, B., 2016. Bots,# strongerin, and# brexit: Computational propaganda during the uk-eu referendum. *arXiv preprint arXiv:1606.06356*.][Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. and Lazer, D., 2019. Fake news on Twitter during the 2016 US presidential election. *Science*, *363*(6425), pp.374-378.

15..Greene, C.M. and Murphy, G., 2021. Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. *Journal of experimental psychology: Applied*, *27*(4), p.773

16. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G. and Rand, D.G., 2020. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, *31*(7), pp.770-780.

17. Di Domenico, G., Sit, J., Ishizaka, A. and Nunan, D., 2021. Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, pp.329-341.

18. West, J.D. and Bergstrom, C.T., 2021. Misinformation in and about science. *Proceedings of the National Academy of Sciences*, *118*(15), p.e1912444117.

19 Tandoc Jr, E.C., Lim, Z.W. and Ling, R., 2018. Defining “fake news” A typology of scholarly definitions. *Digital journalism*, *6*(2), pp.137-153.

20.Naeem, S.B., Bhatti, R. and Khan, A., 2021. An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*, *38*(2), pp.143-149.

21. Wardle, C. and Derakhshan, H., 2017. *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27, pp. 1-107). Strasbourg: Council of Europe.

22. Friggeri, A., Adamic, L., Eckles, D. and Cheng, J., 2014, May. Rumor cascades. In *proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 101-110).

23. Nyhan, B. and Reifler, J., 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), pp.303-330.

24. Kumar, S. and Shah, N., 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*.

25.Zhou, X. and Zafarani, R., 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), pp.1-40.

26. Gros, D., Li, Y. and Yu, Z., 2021. The RUA-robot dataset: Helping avoid chatbot deception by detecting user questions about human or non-human identity. *arXiv preprint arXiv:2106.02692*.

27. Wang, W.Y., 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

28. Naredla, N.R. and Adedoyin, F.F., 2022. Detection of hyperpartisan news articles using natural language processing technique. *International Journal of Information Management Data Insights*, *2*(1), p.100064.

29. Oshikawa, R., Qian, J. and Wang, W.Y., 2018. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

30. Huffaker, J.S., Kummerfeld, J.K., Lasecki, W.S. and Ackerman, M.S., 2020, April. Crowdsourced detection of emotionally manipulative language. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-14).

31. Meesad, P., 2021. Thai fake news detection based on information retrieval, natural language processing and machine learning. *SN Computer Science*, *2*(6), p.425.

32. Shabani, S. and Sokhn, M., 2018, October. Hybrid machine-crowd approach for fake news detection. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)* (pp. 299-306). IEEE.

33. Adib, Q.A.R., Mehedi, M.H.K., Sakib, M.S., Patwary, K.K., Hossain, M.S. and Rasel, A.A., 2021, October. A deep hybrid learning approach to detect bangla fake news. In *2021 5th international symposium on multidisciplinary studies and innovative technologies (ISMSIT)* (pp. 442-447). IEEE.

34. Albahar, M., 2021. A hybrid model for fake news detection: Leveraging news content and user comments in fake news. *IET Information Security*, *15*(2), pp.169-177.

35. Ghezzi, A., Gabelloni, D., Martini, A. and Natalicchio, A., 2018. Crowdsourcing: a review and suggestions for future research. *International Journal of management reviews*, *20*(2), pp.343-363.

36. Pinto, M.R., de Lima, Y.O., Barbosa, C.E. and de Souza, J.M., 2019, May. Towards fact-checking through crowdsourcing. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 494-499). IEEE.

37.Kaliyar, R.K., Goswami, A. and Narang, P., 2021, January. A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset. In *ICAART (2)* (pp. 1066-1072).

38. Mykytiuk, A., Vysotska, V., Markiv, O., Chyrun, L. and Pelekh, Y., 2023. Technology of Fake News Recognition Based on Machine Learning Methods. In *COLINS (1)* (pp. 311-330

39. https://captainfact.io/

40. Thorne, J., Vlachos, A., Christodoulopoulos, C. and Mittal, A., 2018. FEVER: a large-scale dataset for fact extraction and VERification. *arXiv preprint arXiv:1803.05355*.

41. Graves, L. and Glaisyer, T., 2012. The fact-checking universe in Spring 2012. *New America*.]

42. Johnson, A.M., Willenborg, A., Heckman, C., Whitacre, J., Reynolds, L., Sterner, E.A., Harmon, L., Lunsford, S. and Drerup, S., 2018. Library instruction and information literacy 2017. *Reference Services Review*, *46*(4), pp.628-734.

43. Amazeen, M.A., 2015. Revisiting the epistemology of fact-checking. *Critical review*, *27*(1), pp.1-22.

44. Ünver, A., 2023. Emerging technologies and automated fact-checking: Tools, techniques and algorithms. *Techniques and Algorithms (August 29, 2023)*

45. Graves, L. (2016). *Deciding What's True: The Rise of Political Fact-Checking in American Journalism*. Columbia University Press

46. https://www.factcheck.org/

47. Mena, P., 2019. Principles and boundaries of fact-checking: Journalists' perceptions. *Journalism practice*, *13*(6), pp.657-672

48. Majithia, S., Arslan, F., Lubal, S., Jimenez, D., Arora, P., Caraballo, J. and Li, C., 2019, July. ClaimPortal: Integrated monitoring, searching, checking, and analytics of factual claims on twitter. In *Proceedings of the 57th annual meeting of the association for computational linguistics: system demonstrations* (pp. 153-158).

49. Ganegedara, T., 2018. *Natural Language Processing with TensorFlow: Teach language to machines using Python's deep learning library*. Packt Publishing Ltd.

50. Hassan, N., Arslan, F., Li, C. and Tremayne, M., 2017, August. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1803-1812).

51. Cazalens, S., Lamarre, P., Leblay, J., Manolescu, I. and Tannier, X., 2018, April. A content management perspective on fact-checking. In *Companion Proceedings of the The Web Conference 2018* (pp. 565-574).

52. Okoro, E.M., Abara, B.A., Umagba, A.O., Ajonye, A.A. and Isa, Z.S., 2018. A hybrid approach to fake news detection on social media. *Nigerian Journal of Technology*, *37*(2), pp.454-462

53. Das, A., Liu, H., Kovatchev, V. and Lease, M., 2023. The state of human-centered NLP technology for fact-checking. *Information processing & management*, *60*(2), p.103219.

54. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T. and He, L., 2022. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, pp.364-381

55. Zhu, J.Y., Tang, B. and Li, V.O., 2019. A five-layer architecture for big data processing and analytics. *international Journal of big Data intelligence*, *6*(1), pp.38-49

56. Howard, J. and Gugger, S., 2020. Fastai: a layered API for deep learning. *Information*, *11*(2), p.108

57. La Barbera, D., Roitero, K. and Mizzaro, S., 2023. Combining human intelligence and machine learning for fact-checking: Towards a hybrid human-in-the-loop framework. *Intelligenza Artificiale*, (Preprint), pp.1-10.

58. Kalra, V. and Aggarwal, R., 2017. Importance of Text Data Preprocessing & Implementation in RapidMiner. *ICITKM*, *14*, pp.71-75.

59. Su, Q., Wan, M., Liu, X. and Huang, C.R., 2020. Motivations, methods and metrics of misinformation detection: an NLP perspective. *Natural Language Processing Research*, *1*(1-2), pp.1-13

60.Hamed, S.K., Ab Aziz, M.J. and Yaakub, M.R., 2023. Fake news detection model on social media by leveraging sentiment analysis of news content and emotion analysis of users' comments. *Sensors*, *23*(4), p.1748

61. Lai, C.M., Chen, M.H., Kristiani, E., Verma, V.K. and Yang, C.T., 2022. Fake news classification based on content level features. *Applied Sciences*, *12*(3), p.1116.

62. .Plotkina, D., Munzel, A. and Pallud, J., 2020. Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research*, *109*, pp.511-523.

63. Bauer, K., Heigl, R., Hinz, O. and Kosfeld, M., 2024. Feedback loops in machine learning: A study on the interplay of continuous updating and human discrimination. *Journal of the Association for Information Systems*, *25*(4), pp.804-866

64. Zhang, Z., Rudra, K. and Anand, A., 2021, October. FaxPlainAC: A fact-checking tool based on explainable models with human correction in the loop. In *Proceedings of the 30th ACM international conference on information & knowledge management* (pp. 4823-4827).

65. Qudus, U., Röder, M., Saleem, M. and Ngonga Ngomo, A.C., 2022, October. Hybridfc: A hybrid fact-checking approach for knowledge graphs. In *International Semantic Web Conference* (pp. 462-480). Cham: Springer International Publishing.

66. Hamsheen, E.S. and Flah, L.R., 2023. Improvement performance by using Machine learning algorithms for fake news detection. *Zanco Journal of Pure and Applied Sciences*, *35*(2), pp.48-57.

67. Pavleska, T., Školkay, A., Zankova, B., Ribeiro, N. and Bechmann, A., 2018. Performance analysis of fact-checking organizations and initiatives in Europe: a critical overview of online platforms fighting fake news. Social media and convergence, 29, pp.1-28

68. Powers, D.M., 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

69. Amershi, S., Cakmak, M., Knox, W.B. and Kulesza, T., 2014. Power to the people: The role of humans in interactive machine learning. *AI magazine*, *35*(4), pp.105-120.