Bilkent University CS491
Project Specification Document
T2426

Group Members:

Ayberk Berat Eroğlu - 22103675

Alp Batu Aksan - 22103246

Pelin Öner - 22102409

İpek Sönmez - 22103939

Efe Tuna Can - 22102127

Table of Contents

1. Introduction

1.1. Description

In a world overwhelmed by rapidly spreading misinformation, ensuring that what we read online is accurate and reliable has become a pressing challenge. Our project CheckMate is a browser extension that provides fact-checking and political bias detection services. Currently, most fact-checking is done by humans researching the topic. For example, Meta, the company behind Facebook and Instagram, uses human resources for fact-checking [1]. This approach is often too slow to keep up with the overwhelming flood of information. This is because fact-checking is a long process and requires serious human resources for each article to be fact-checked [2]. Our project CheckMate aims to automate fact-checking with its Artificial Intelligence (AI) driven system and ultimately provide a quick and easy way to combat misinformation online. Slowly, the internet is becoming the primary source of news people see; in some countries, such as the UK, it is the biggest news source [3]. Thus, CheckMate will work with all major news and social media websites to provide its services.

For its services, CheckMate will evaluate each article and assign a reliability score for them. To do this, natural language processing (NLP) will be used to determine whether or not the language used is biased. It will analyze the article for any political bias and will use this information for its reliability score. Moreover, it will check other sources on the internet for confirmatory or contradictory information regarding the information in the selected news article through search Application Programming Interfaces (APIs). CheckMate will also use the metadata of the news article, if possible, such as the website on which it was published or its author. Using the metadata of the news article CheckMate will generate a credibility score for the news' source as well regarding whether the source released fake news before, has political inclination, uses biased language, etc. This credibility score will contribute to the overall reliability score of the news article. Additionally, CheckMate will perform reverse searches on visuals, such as photos, thumbnails, and videos, used in news articles to check for

inconsistencies or mismatched contexts through search APIs. This reverse search will also provide us with whether or not those visuals were used before in different contexts. After CheckMate completes its analysis, the system will show the reliability score of the article alongside the reasons for its score. These reasons will range from having a high political bias to providing contradictory information to other trustworthy sources.

1.2. High Level System Architecture & Components of Proposed Solution

In our system, there are 5 main components: Content Analyzer component, Political Analyzer component, News Validator component, Metadata Collector component, and Multimedia Analyzer component. When users use the system, the components make calls to servers accordingly. There is also the database component, which uses PostgreSQL to save each

information related to the system. The users use their devices to connect to the system and to be able to use its features. Their browsers make requests to the servers of the system's main components to be able to use their features. To see these relationships clearly, a high-level system architecture diagram is given below:

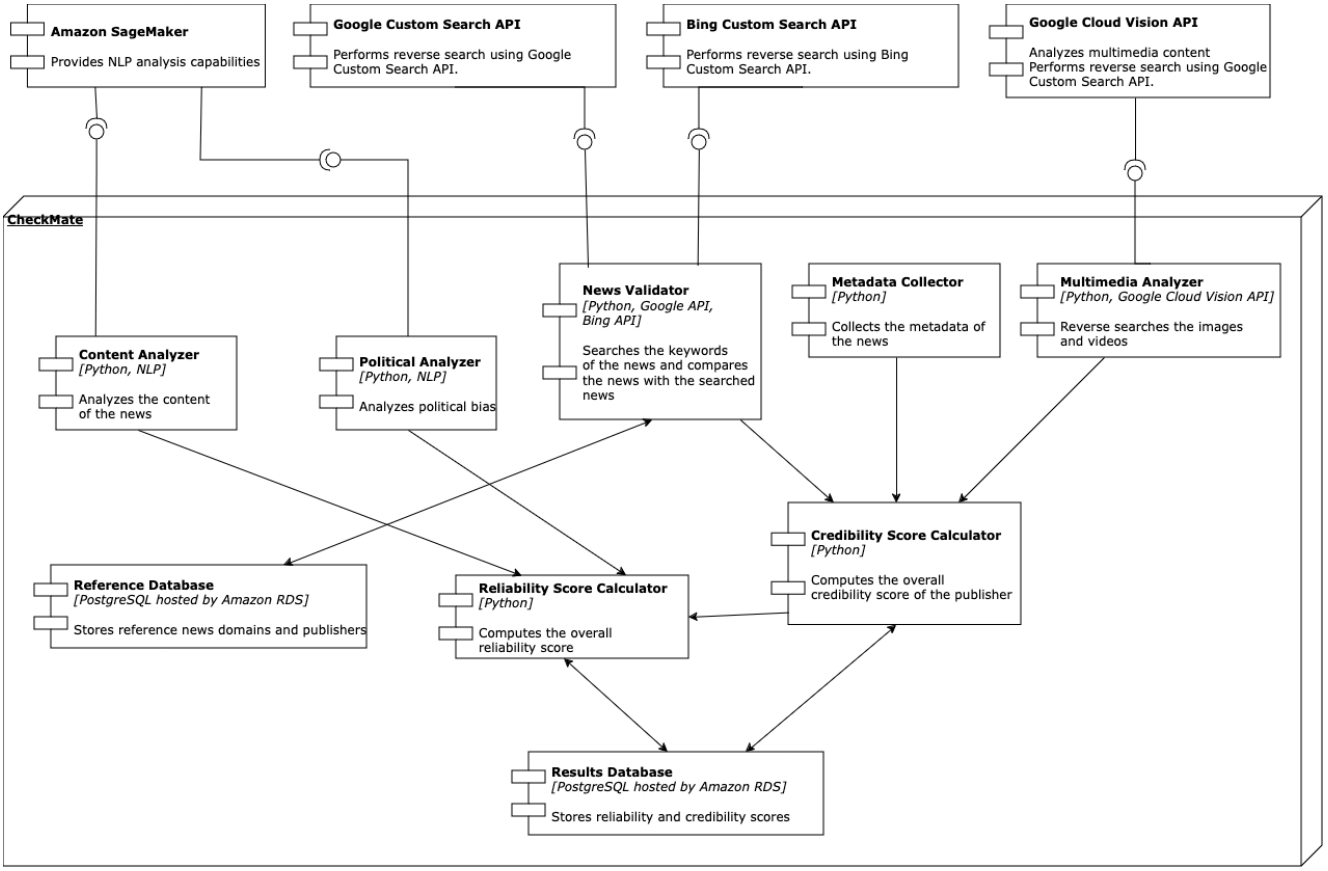

Figure 1: High Level System Architecture Diagram

1.3. Constraints

CheckMate is a browser extension that provides fact-checking and political bias detection services.

1.3.1. Implementation Constraints

- The browser extension will be compatible with Chromium browsers. Some chromium browser examples can be found at [4].
- GitHub and Jira will be used to track the deadlines, issues, and code.
- React.JS framework will be used for frontend development.
- Python will be used for machine learning (ML) development and backend development.
- PostgreSQL hosted by Amazon RDS server will be used for the database system.

1.3.2. Economic Constraints

- Our project requires several external libraries, frameworks, and models. Therefore, our group has opted to use as many open-source frameworks as possible.
- Google Vision API [5], the backend deployment on Amazon services, and Amazon Sagemaker for ML development will use a paid plan.
- If necessary, Bing's news search API [6] and News API [7] will be used for finding news and will use a paid plan

1.3.3. Ethical Constraints

- All interactions and data collected from the users will be handled within data protection law General Data Protection Regulation (GDPR).
- To enforce the law, the users must be well informed about the application's limitations and scope before registration.
- The program will be utilizing personal data and storing said data in a database hosted on Amazon RDS.
- No unnecessary user data will be collected.
- The system will clearly communicate to the user about its shortcomings.
- The system will clearly communicate reasons for reliability scores of the news articles and the reasons for credibility scores for news sources.
- The system will suggest trustworthy sources to the user about the news article that they are searching for.
- The system is designed to ensure objectivity and fairness, actively compensating for any biased outcomes to provide accurate labeling of news articles and news sources.

1.3.4. Language Constraints

- The system will work on news articles and new sources in English because of the lack of labeled Turkish news article datasets.

1.4. Professional and Ethical Issues

CheckMate is designed to address the critical challenge of misinformation while respecting professional and ethical principles. The system will provide users with reliable fact-checking and political bias detection for online content, but great care will be taken to ensure the process remains fair, transparent, and respectful of privacy.

Firstly, all user interactions with the application, including any articles submitted for analysis, will remain strictly confidential. Only necessary user information will be collected and will not be used in any way without the users' consent. Additionally, the system will be designed to provide feedback in a professional and unbiased manner, even when evaluating articles with highly controversial content. Inappropriate queries directed at the system will be met with neutral, instructive responses, guiding users toward productive interactions.

CheckMate's reliability scoring system will be based on rigorous and impartial criteria. Articles will be analyzed through multiple methods, including NLP for tone and professionalism, cross-referencing with other trusted sources for consistency, and evaluating metadata for additional context. The ML algorithm will also detect political bias without endorsing or discrediting any particular viewpoint. Visual content, such as images or thumbnails, will undergo reverse searches to detect any possible misrepresentation or use of visuals out of context.

While CheckMate will provide users with an estimated reliability score and reasoning behind its evaluation, it will not position itself as the ultimate authority on truth. Instead, it will encourage users to critically engage with the information presented and explore the detailed explanations for its findings. The goal is to empower individuals with the tools to make informed decisions, not to dictate what they should believe.

1.5. Standards

- The system will abide by the European Fact-Checking Standards Network (EFCSN) [8] as a benchmark in fact-checking news articles.
- The system will abide by Google Cloud Vision API Terms of Service, Bing Search API License Agreement, and other third-party API usage policies.
- The system will abide to General Data Protection Regulation (GDPR) [9]

2. Design Requirements

2.1. Functional Requirements

- The system will maintain a reference database of known news websites and social media platforms commonly used for journalism, as demonstrated in [10], which will be regularly updated to incorporate new and emerging platforms.
- The system will compare the Uniform Resource Locators (URLs) of each website that the user enters with that reference database.
- The system will collect metadata (e.g., publication date, location, and author) of the news article on the new source.
- The system will validate the time and place of publication against credible sources to verify accuracy.
- The system will update the reliability score of news websites periodically based on newly available information.
- The system will analyze the publisher of each news article to check whether they are deemed reliable.
- The system will allow users to access news domains that have published false information to update the publisher's reliability.
- The system will analyze the content of each news article to assess its factual accuracy by performing reverse searches using APIs such as Google Custom Search API [11], Bing News Search API [12], and Bing Custom Search API [13].

- The system will use an NLP model to determine whether the article is biased or not.
- The system will flag statements with no credible references or with conflicting information.
- The system will analyze the language and tone of political news articles to determine political polarity using an NLP model.
- Based on sentiment analysis, the system will classify political slants along a compass (e.g., left-leaning, right-leaning, centrist).
- The system will display the political slant of each article on a visual compass, enabling users to see the degree of bias.
- The system will check if the news article has appeared on other reliable news websites.
- The system will compare the information across multiple sources to detect discrepancies and increase the accuracy score if verified by multiple independent sources.
- The system will lower the accuracy score if conflicting information is found in other credible sources.
- The system will detect if images or videos are present in the news article.
- The system will analyze multimedia content to identify instances of use in unrelated contexts by utilizing tools such as Google Cloud Vision API's label detection feature [14].
- The system will verify the relevance and accuracy of multimedia content by comparing it against known sources through reverse search methods using Google Cloud Vision's web detection feature [15] and Bing Image Search API [16].
- The system will compute an overall accuracy score for each news article based on the metadata, source reliability, publisher credibility, content accuracy, political slant, cross-verification, and multimedia verification.
- The system will display the accuracy score to users, showing a breakdown of how each criterion contributes to the final score.

- The system will allow users to submit feedback if they believe a news article is inaccurate or biased.
- The system will provide a visual interface to show each news article's accuracy score and political compass.
- The system will allow users to view detailed information for each accuracy criterion, including explanations and sources used in the analysis.
- The system will allow administrators to update the databases of publishers and websites with new information as it becomes available.

2.2. Non-Functional Requirements

2.2.1. Usability

- The system will provide an intuitive and user-friendly interface, allowing users to easily access each article's reliability score and political compass.
- The system will display clear visual indicators (e.g., color-coded scores, compass graphics) to effectively convey the reliability and political slant of articles.
- The system will allow users to provide feedback on accuracy and bias with minimal steps.
- The system will be accessible on major Chromium browsers (e.g., Chrome, Opera, Edge) and optimized for responsive display across different screen sizes.

2.2.2. Reliability

- The system will ensure accuracy and consistency in reliability scores and political spectrum classification by regularly updating the reference databases and verifying data sources.
- The system will have mechanisms to flag and handle inconsistencies or errors in analysis.

2.2.3. Performance

- The system will perform efficient data processing, using caching and optimized algorithms to minimize latency in news verification and NLP model usage.
- The system will handle high traffic loads by distributing requests and limiting non-essential checks when necessary to maintain a seamless user experience.

2.2.4. Supportability

- The system will be easy to update to accommodate changes in the reference databases, external APIs, or NLP models.
- The system's codebase will be modular and well-documented, enabling maintainers to make modifications or add new functionalities with minimal risk of disruption.
- The system will log significant events and errors, allowing developers to diagnose issues and track the system's health and performance.

2.2.5. Scalability

- The system will be able to scale to accommodate an increasing number of users and a higher frequency of requests without impacting performance.
- The system's architecture will allow for integration with additional NLP models and new APIs for extended capabilities (e.g., enhanced multimedia analysis or additional political slant dimensions).
- The system will support future expansion of reference databases, maintaining efficient operations as the number of sources and articles grows.

3. Feasibility Discussions

3.1. Market & Competitive Analysis

The market for a tool like CheckMate is strong due to growing concerns about misinformation, especially on social media, where rapid news dissemination often lacks verification. Several key market factors bolster CheckMate's relevance:

Information Overload: Automated solutions like CheckMate come into play during the rise in highly proliferating news. Contemporary solutions must focus on automation rather than manual validations due to the sheer number of news published daily on news outlets and social media.

Bias and Reliability: Increasing political divides have heightened the demand for tools that can assess ideological slants in news stories, making CheckMate's bias detection feature particularly appealing. Identifying the political agenda of each news outlet and informing users of these political alignments are crucial.

Potential target audiences include:

- **Individual consumers** for personal use and awareness.
- **Educational institutions** for teaching media literacy.
- **Media organizations** to enhance reporting integrity.
- **Businesses** for brand monitoring and competitive analysis.
- **Government agencies** to manage public communication.
- **Social media platforms** to control misinformation on their networks.

Competitive Analysis

CheckMate faces limited direct competition, but relevant players do exist in both manual and automated fact-checking:

- **Manual Competitors:**

There are platforms that are dedicated to fact-checking through crowd-sourced work, such as Teyit.org and some work done by Reddit users [17,18]. These platforms are based on human knowledge and teamwork to verify the news, informing users of the specifics and sophisticated analysis. However, they have a problem with scalability, especially during the misinformation high activity periods such as during elections or globally sensitive issues. Due to the fact that they use human verification to respond to posts, they are likely to take long to provide quick responses hence slowing down the process of containing the spread of fake news. Thus, CheckMate has no such shortcomings, as it uses AI to generate an immediate and fully automated analysis of news content. Furthermore, manual platforms do not consider ideological bias, which is one of CheckMate's features, and provide users with information about the ideological stance of the content they consume.

- **Automated Competitors:**

Currently, the automated fact-checking field is filled with more specific platforms such as InVID, WeVerify, SurfSafe, and Google Fact Check Explorer [19,20,21,22]. All these solutions are very useful in different ways, but none of them has the all-in-one and multiple-function features that CheckMate has. InVID and WeVerify are designed to tackle video and image content for users who are concerned about multimedia misinformation. SurfSafe focuses on identifying the source of images for the purpose of determining their legitimacy, but it does not consider stories or the text messages that come with these pictures. Although these tools are useful in their specific fields, they do not offer the textual and multimedia analysis that is integrated into CheckMate. While Google Fact Check Explorer recompiles and indexes fact-checks reports from other reputed sources, it also enables users to search through previously fact-checked claims. Nonetheless, its reactive character does not allow using it for identifying fake news in real-time or in relation to new topics. On the other hand, the real-time analysis provided by CheckMate comes with the added bonus of detection,

so users have credibility assessments at their disposal before news breaks. Likewise, tools such as ClaimBuster and Hoaxy are designed to find out which claims may be fact-checked or to track the flow of misinformation, aimed at researchers and organizations. CheckMate fills this gap through the integration of similar features in a simple-to-use platform intended for popular use.

3.2. Academic Analysis

The study of political bias in news and the challenge of falsifiability has been significantly advanced by academic research. The insights derived from these studies are directly relevant to the development of projects like CheckMate, which aim to analyze news credibility and detect ideological slants. By integrating theories from psychology, linguistics, and computational methods, these studies provide a comprehensive framework to address the complexities of misinformation and bias.

[23] also directly argues that the problem with fake news is not political bias alone, but also fundamental cognitive heuristics like familiarity and source. The conclusions of the authors correspond to the purpose of CheckMate, the tool designed to help people become more thoughtful and critical when interacting with news. This is well supported by [24], which presents the LIAR dataset, a large-scale benchmark with short human-labeled statements categorized by their veracity. The dataset shows that incorporating external metadata like the affiliation of the speaker or the historical truthfulness of the statement is beneficial for creating better machine learning models for fake news detection. By these benchmarks, CheckMate can build algorithms that give credible scores that involve metadata within context with language analysis.

The linguistic aspect of fake news detection is also supported by [25], who continues the topic of how fake or biased content tends to use shocking words, relative terms, and overlapping claims. In another step, [26]. carry this approach forward in the case of hyperpartisan news, using stylometric methods that indicate features that are different from

regular news reporting. These studies show that it is possible to use metadata in parallel with the qualitative analysis of the texts to distinguish between ideological poles without focusing on mere factual distortions. For CheckMate, this insight can be incorporated in the design of a hybrid system that makes use of both stylistic and substantive indicators to alert users to potentially biased content.

Another important layer in fake news detection, as it is shown by [27], is the social engagement data. The sharing patterns and the interaction of the audience proposing a post often describe the diffusion of fake news. This additional information can then be fed back into CheckMate's structure in order to improve the accuracy of its credibility scoring. Likewise, the graded approach to deception detection is presented in [25]. is a similar manner because it suggests using PolitiFact's factuality scale to represent truthfulness on a continuum. This is in line with the operation of the CheckMate framework that does not aim to classify posts and profiles as black and white true or false, but rather in shades of grey taking into account the half-truths and/or manipulation, both of which are recognized features of the new forms of misinformation.

Hybrid CNN architecture used in [24] gives specific details of how features from text and metadata can be adopted for fake news detection. His approach increases accuracy by including not only shallow syntactic features of the language but also the syntactic and semantic features such as speaker credibility, history, and statement context. This methodology is directly relevant to CheckMate as a combined model can make use of real-time social interaction data, historical credibility history, and linguistic signals to offer multiple perspectives to the users for evaluating news.

All these academic findings together provide the theoretical basis for CheckMate and make use of validated approaches. The project will be able to use the LIAR dataset to develop effective training models, linguistic findings in [25] for the detection of bias and the social engagement data for the credibility check in real-time. The integration of all these interdisciplinary approaches can help CheckMate to create an effective tool in the analysis of

news credibility and political bias for the user to be knowledgeable and wise in the emergent complicated media environment.

4. Glossary

API: An API, or Application Programming Interface, is a collection of rules and protocols that allow software applications to interact and share data, features, or functionality. API communication can be understood as an exchange of requests and responses between a client and a server. The client is the application sending the request, while the server processes the request and sends back a response. The API acts as the intermediary, facilitating this connection between the two. [28]

NLP: Natural Language Processing (NLP) is a branch of computer science and artificial intelligence focused on enabling computers to comprehend human language. It combines computational linguistics, which examines the mechanics of language, with statistical methods, machine learning, and deep learning models. These techniques equip computers to analyze and interpret text or speech data, understanding their overall meaning, as well as the speaker's or writer's intentions and emotions. [29]

ML: Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy. [30]

URL: The location of a webpage or file on the Internet. [31]

5. References

[1] Meta, "How fact-checking works," Meta Transparency Center. [Online]. Available: https://transparency.meta.com/en-us/features/how-fact-checking-works/. [Accessed: 16-Nov-2024].

[2] Knight Science Journalism, "Fact-checking science journalism: How to make sure your stories are true," KSJ Handbook. [Online]. Available: https://ksjhandbook.org/fact-checking-science-journalism-how-to-make-sure-your-stories-are-true/the-fact-checking-process/. [Accessed: 16-Nov-2024].

[3] J. Waterson, "Internet overtakes TV as UK's most popular news source for first time," The Guardian, 10-Sep-2024. [Online]. Available: https://www.theguardian.com/media/article/2024/sep/10/internet-tv-uk-most-popular-news-source-first-time. [Accessed: 16-Nov-2024].

[4] Mihir J, "Browser Comparison Finale Chromium-based Browsers", Medium. [Online]. Available:https://medium.com/@mihirgrand/browser-comparison-finale-chromium-based-browsers-2b6063e74165 [Accessed: 16-Nov-2024].

[5] Google, "Vision AI," Google Cloud. [Online]. Available: https://cloud.google.com/vision/?hl=en. [Accessed: 16-Nov-2024].

[6] Microsoft, "Bing News Search API," Microsoft. [Online]. Available: https://www.microsoft.com/en-us/bing/apis/bing-news-search-api. [Accessed: 16-Nov-2024].

[7] NewsAPI, NewsAPI.org. [Online]. Available: https://newsapi.org/. [Accessed: 16-Nov-2024].

[8] EFCSN, "Code of standards," *EFCSN*. [Online]. Available: https://efcsn.com/code-of-standards/. [Accessed: 16-Nov-2024].

[9] Intersoft Consulting, "General data protection regulation (GDPR)," General Data Protection Regulation (GDPR) [Online], Available: https://gdpr-info.eu/ [Accessed: 16-Nov-2024].

[10] GitHub, "U.S. News Domains," GitHub. [Online]. Available: https://github.com/ercexpo/us-news-domains. [Accessed: 16-Nov-2024].

[11] Google Developers, "Custom Search API," Google Developers. [Online]. Available: https://developers.google.com/custom-search/v1/overview. [Accessed: 16-Nov-2024].

[12] Microsoft, "Bing News Search API," Microsoft. [Online]. Available: https://www.microsoft.com/en-us/bing/apis/bing-news-search-api. [Accessed: 16-Nov-2024].

[13] Microsoft, "Bing Custom Search API," Microsoft. [Online]. Available: https://www.microsoft.com/en-us/bing/apis/bing-custom-search-api. [Accessed: 16-Nov-2024].

[14] Google Cloud Vision, "Label detection," Google Cloud. [Online]. Available: https://cloud.google.com/vision/docs/labels. [Accessed: 16-Nov-2024].

[15] Google Cloud Vision, "Web detection," Google Cloud. [Online]. Available: https://cloud.google.com/vision/docs/detecting-web#vision_web_detection-python. [Accessed: 16-Nov-2024].

[16] Microsoft, "Bing Image Search API," Microsoft. [Online]. Available: http://www.microsoft.com/en-us/bing/apis/bing-image-search-api. [Accessed: 16-Nov-2024].

[17]“Teyit,” *teyit.org*. https://teyit.org/

[18] Reddit, “Reddit,” *Reddit*, 2005. https://www.reddit.com/

[19]“InVID Project - Video Verification,” *InVID project*. https://www.invid-project.eu/

[20]“Home,” *WeVerify*. https://weverify.eu/

[21]”Surfsafe,” *Chrome Web Store.* *https://chromewebstore.google.com/detail/surfsafe-join-the-fight-a/hbpagabeiphkfhbboacggckhkkipgdmh*

[22]”Fact Check Tools,” *toolbox.google.com*. https://toolbox.google.com/factcheck/explorer/search/list:recent

[23] G. Pennycook, and D. G. Rand, “The Psychology of Fake News,” *Trends in Cognitive Sciences*, vol. 25, no. 5, pp. 388–402, May 1 2021, doi: https://doi.org/10.1016/j.tics.2021.02.007. [Accessed: 11 Nov 2024].

[24] W. Wang, “‘Liar, Liar Pants on Fire’: A New Benchmark Dataset for Fake News Detection.” Accessed: Jul. 20, 2024. [Online]. Available: https://arxiv.org/pdf/1705.00648v1

[25] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, “Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking,” *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, doi: https://doi.org/10.18653/v1/d17-1317.

[26] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, “A Stylometric Inquiry into Hyperpartisan and Fake News,” *arXiv:1702.05638 [cs]*, Feb. 2017, Available: https://arxiv.org/abs/1702.05638

[27] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, Sep. 2017, doi: https://doi.org/10.1145/3137597.3137600.

[28] IBM, "What is an API?," IBM Topics. [Online]. Available: https://www.ibm.com/topics/api. [Accessed: 16-Nov-2024].

[29] GeeksforGeeks, "Natural language processing overview," GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/natural-language-processing-overview/. [Accessed: 16-Nov-2024].

[30] IBM, "What is machine learning?," IBM Topics. [Online]. Available: https://www.ibm.com/topics/machine-learning. [Accessed: 16-Nov-2024].

[31] Google, "URL," Google Support. [Online]. Available: https://support.google.com/google-ads/answer/14095?hl=en. [Accessed: 16-Nov-2024].