

# Uncovering the Semantics of Political Fake News: A Latent Semantic Analysis Approach with NLP and SVD

Supriya Mishra<sup>1</sup>, Roop Ranjan<sup>2</sup>, Bajrang Lal<sup>3</sup>

<sup>1,3</sup>Department of Computer Science and Engineering, Singhania University, Jhunjhunu, Rajasthan, India

<sup>2</sup>Department of Computer Science and Engineering, Buddha Institute of Technology, Gorakhpur, U. P., India

ORCID- 0009-0001-3064-0477<sup>1</sup>

ORCID- 0000-0001-8616-0563<sup>2</sup>,0009-0002-9462-6831<sup>3</sup>

## KEYWORD

## ABSTRACT

Latent semantic analysis, Social Media, Internet News, Fake News, natural language-processing

According to recent studies, fake news is often created with the intention of misleading readers, spreads quickly, and causes significant harm to human civilization through social media. Using machine learning or deep learning approaches, the majority of studies in the current literature concentrate on fake news sentiment analysis, as well as detection and identification. Nonetheless, comparatively few scholars have focused on the analysis of false news. For fake political news, this is particularly true. In contrast to previously published works in which fraud media detection models were developed from the perspective of computer scientists, an efficient method is aimed to be developed in this study that combines latent based semantic analysis (LSA) and natural language processing (NLP) using singular value decomposition (SVD) techniques to assist social scientists in the analysis of fake news in order to pinpoint its precise components. Furthermore, the researchers examine the traits of fake media and real media. A natural-world example, such as the 2016 US presidential election campaign, is used to illustrate the power of our strategies. The experimental findings may help future researchers identify bogus news. According to this study, the five ideas adopted from LSA are indicative for fake political news during an election.

## 1. Introduction:

Over the past ten years, fake news has spread across various democratic nations, including Indonesia, Kenya, Lebanon, and many others worldwide. This widespread dissemination of false information has created challenges for people globally, leading to confusion and difficulty in identifying what constitutes real news (Obreja, D. M. ,2022). The nature of the fake news can be understood, which will simplify the identification. Researchers consider that the following strategy is the best to facilitate empathy by engaging in online activity, i.e., informed involvement (Vafeiadis, M., and Xiao, A., 2021). However, fake news continues to spread rapidly via personal messaging applications and social media.

Although social media has some advantages, it is also very dangerous. Some have argued that social media is not the right medium through which social change can be created. It has been observed by various researchers to have a dark side and particularly the ease with which it can be manipulated. Regrettably, Southeast Asia is the country where the cases of fake news are the most (Lee ,2019). Simultaneously, the United States being a big, democratic, and technologically developed country has not been an exception since the potent impact of fake news, especially when it comes to presidential elections. Technological advancement in part influences the regulation of the mass media, as well as other forms of communication. In case traditional mass media is still in control, other digital platforms, including the Internet, are likely to enjoy the present media environment. Mass media integration has given rise to new forms of network-based news which has gained popularity in the media studies. A significant number of people find the news on the Internet (Hallin, D. C et al., 2021).

With the advent of online news, accessing information without relying on newspapers has become increasingly easy. Online news platforms often provide free content, as their production costs are relatively lower than those of print newspapers. Despite its convenience, internet news also has the ability to transmit disinformation. While individuals like the convenience of digital news, they are concerned about the ramifications of

communication in the digital era. Fake news can alter the public's views, incite fear or anger, and even provoke unrest and violence in certain communities. Misleading information, when spread deliberately, can be a powerful tool for sowing division. Most published studies focus on identifying fake news (Patel et al. , 2022). Various detection algorithms have been developed depending on the target and context of detection (Shah, M. N., & Ganatra, A.,2022). In recent years, social scientists, particularly in the field of communication, have shown a preference for artificial methods over understanding technical algorithms. As a result, IT professionals and computer scientists are actively working to find the most effective strategies for detecting fake information.

Historically, social scientists have relied on traditional methods such as surveys, observations, and interviews to study fake news (Delaney et al., 2022), (Pérez-Escoda, A. 2022). However, these techniques are often limited to traditional sources, such as newspapers and broadsheets. In the current digital environment, identifying fake news on online platforms is relatively easier than through conventional means. Given the limitations of traditional approaches, there is a growing need for more effective tools that can help social scientists achieve their objectives. Modern research methods are increasingly regarded as the most effective way to detect fraudulent online news. Natural language processing (NLP) and text mining techniques are widely employed in studies of fake news. Deep learning models such as CNN, RNN, BERT, and LSTM are frequently employed to derive insights based on specific research goals (Tashtoush, Y.,et al , 2022). While a majority of studies focus on fake news detection and classification ((Capuano et al. ,2023),( Song et al., 2022),( Zhang et al. , 2023)),((Iwendi et al., 2022),( Lin et al., 2022 ), (Pratama et al., 2021)) only a limited number of researchers have concentrated on comprehensive fake news analysis ((Kapusta et al., 2020),( Kozik et al., 2022),( Lim, S. 2020)) especially in the context of political misinformation.

Additionally, many researchers approach the problem from a computer science perspective, relying on artificial intelligence techniques. In contrast to prior studies rooted in the technical views of computer scientists, the present research provides the objective to develop an effective approach by integrating NLP and LSA using SVD(Singular Value Decomposition) techniques. This integration is designed to support social scientists in identifying the core components of fake news through analytical methods. The researchers also investigate the differences between authentic and fake news.

The study uses the 2016 US election as a real-world example to show the efficacy of this methodology. The results of the experiment may also signify valuable insights and recommendations for upcoming research on distinguishing false information from credible news. The rest of the paper is structured as follows: Section 2 discusses the related work by various authors. Section 3 discusses the methodology. Section 4 discusses the experimental results. Section 5 presents a discussion of various parameters and findings from the proposed work. Section 6 outlines the limitations of the proposed work, and Section 7 presents the conclusion and future work.

## 2. Related Works

Theories of journalism are typically characterized as the activity of transforming information dissemination in public society. The news report process is influenced by four ideologies in journalism, which are as follows: social doctrine (Dupuis, I. (2021), liberalism (Pickard, V. (2013), authoritarianism (Krzyzanowski, M., & Ekström, M. (2022), and Soviet communism (Tomaselli, K. G. (2009). The impact of journalistic ideologies is growing, yet great ideologies persist, according to studies conducted by scholars over the years. Four fundamental ideologies usually implement new concepts like these (Arqoub et al., 2022).

In reality, the definition of online news and traditional print news is rather similar. The news is delivered and disseminated on various platforms. Nevertheless, we must consider that its conceptual growth cannot be disregarded (Khan et al., 2022) Despite having identical definitions, web news can nevertheless be described uniquely. Fact-based information is disseminated online via social media (the most popular medium) and technological platforms like websites and mobile applications.

Regarding social media, numerous styles and methods of disseminating news are present, but what counts is which social media platform is most widely used by customers (Baptista, J. P., & Gradim, A. (2022). Basic features of (ICT) are the process of digitizing and the ability to transform whole text (symbol-wise meaning in whole recorded and coded forms) to binary code, with the same generation, allocation, and storage procedures being shared. In actuality, the news age was at its height when web journalism first emerged. Although radio and television were the intended uses for the infrastructure, the Internet adopted regulations to create online news.

Detik.com was Indonesia's first internet news source. Although this company provides information about the nation, it is common knowledge in the Indonesian mass media that the most significant type of information is available in the country's capital, Jakarta (McQuail, D., 2010). However, because online news transcends national boundaries, people in other nations can get information instantly.

There has been an attempt to define "fake news" (Tandoc et al, 2018). The underlying philosophy and diversity of journalism are used in this study to apply this concept. It offers a notion and typology of news in social media. Satire, parody, and photo manipulation are examples of typological media ideas, including journalism and publicity in journalism and advertising. The definition and application of the phrase "fake news" in earlier studies have been analyzed by the research. An examination of 34 scholarly works published between the early 2000s and late 2010s that employed the term "fake-news" provided a classification of categories of fake news being produced, including propaganda, satire, parody, modification, manipulation, and advertisement. These criteria are based on the degree of deception and the degree of fact. Based on earlier researchers' explanations of fake news, this paper presents a conceptual framework for the numerous varieties of false news reported in the literature. A research developed a classification of fake news definitions based on authenticity and purpose factors. Fake news includes parody, satire, and fabrication (Hangloo, S. & Arora, B. 2022).

According to recent guidelines, misinformation, clickbait, rumours, propaganda, and disinformation can all be classified as fake news; however, researchers continue to debate this classification. We must, therefore, carefully determine the type of news by the new paradigm. Fake news does have some influence and advances civilization, but society is also facing challenges. It is difficult to distinguish between false news published elsewhere and fake news generated by online media. "Consequences of fake news being spread by mainstream media" are noted by Tsfati et al. (Tsfati et al., 2020). The fake news website claims that their findings have had a minor impact on the general audience. Consequently, the data implies that a sizable percentage of people are aware of and trust major false news items. Official news outlets or mainstream media disseminate news that could contribute to the spread of fake news. The motivations behind the reporting of fake news and its effects on viewers are contributed to by the Internet's mainstream media. Thus, the spread of false information is exacerbated by major news outlets. Research shows that fake news is concentrated and relatively limited on social media, and that some stories are remembered, endorsed, and even believed by a significant portion of the audience. This is in spite of the lack of empirical estimates on the number of people exposed to fake news stories through mainstream news outlets, which suggests that more people learn about the most covered stories.—at least when it comes to them—from major news sources. Since the evidence is circumstantial, it is highly persuasive, even if it is based on logic rather than concrete proof. Evidence showing that the mainstream media are very credible in their journalism should be presented. In five countries—Saudi Arabia, Lebanon, Qatar, the United Arab Emirates, and Tunisia—the exposure to online fake news varies depending on the type of exposure (Martin, J. D., & Hassan, F. 2020).

This occurs at a time when political momentum is rising due to election-related events, politicians' efforts to socialize public policy to the general public, and other factors. Fake political news exposure (FNE) is a word that is understandably used in the study of those countries. Sometimes, the mainstream media inadvertently disseminates false information. However, the mainstream media also frequently disseminates inaccurate information. Given that the public must get news in an impartial and confirmed manner, this is the worst journalistic practice. A term that has historically applied to journalism is journalistic credibility. Today's public also takes into account the mainstream media's reputation and credibility.

## 2.1 Identifying false news

The majority of published studies have focused on detecting fake news in related works, employing a computational method. A content-based fake news detection technique that can work effectively across various datasets and themes was discovered by (Capuano, N., et al , 2023). The drawbacks of current propagation-based false news detection systems should be addressed by the congeal. To categorize fake news, (Song, C., et al., 2022) introduced a dynamic propagation graph-based technique for detecting fake news. Research has focused on the sentiment categorization of false news, as well as the detection of fake news. For example, Iwendi (Iwendi et al., 2022) identified false material about COVID-19 on social media using deep learning models. SVM and NB are two ML techniques that Pratama and Tjahyanto (Pratama, Ret al, 2021). utilized to investigate how bogus accounts affect sentiment analysis. False or damaging information is detected using Linetal.

## 2.2 Semantic Latent Analysis

Meaningful concepts that have semantic meaning are more effectively attracted by LSA Kwon, H., et al , 2017). To describe the concept–word document relationship, LSA applies singular value decomposition (SVD) to the text data (TDM) to create a semantic vector space ((Yu et al., 2008), (Chen et al., 2021) , (R. M., & Korkontzelos, I. (2021)).

Therefore, LSA has been effectively used to extract valuable information from a large number of documents. In livestreaming, for instance, K-means and LSA were used by Chenatel (Chen et al., 2021) to arrange the chosen keywords into concepts that were easy to comprehend. The authors of Hsiao and Hsiao's work (Hsiao, Y. H., & Hsiao, Y. T. (2021) similarly employed LSA to identify aspects of hotel quality from internet reviews. Additionally, this study utilized LSA to comprehend the concepts of Fake News.

### 3. Methodology:

Prior studies often employed deep learning or machine learning methods to detect bogus news. Nonetheless, the primary goal of this study is to examine the elements of fake news:

Latent Semantic Analysis (LSA) was utilized to extract pertinent sentences from the input material in order to provide a summary by revealing latent or hidden semantic structures (D'Silva et al., 2023). To find important ideas in the gathered documents, the LSA approach models the term-document matrix (TDM) and uses Singular Value Decomposition (SVD). The study's methodology is illustrated in Figure 1, and the steps of the process are outlined below. The following is a description of the technical parameters of implementation:

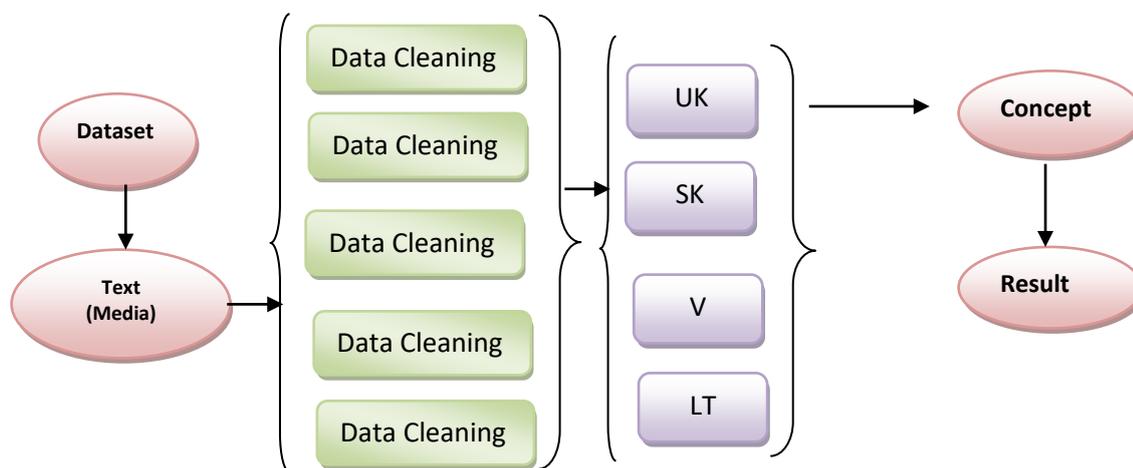


Figure 1 The implementation flow of the study

The system used is manufactured by Hewlett-Packard and is equipped with an Intel® Core i7-4790 processor running at 3.60 GHz with 8 logical cores, and it has 32 GB of RAM installed. The operating system used is Windows 10 Pro, 64-bit (Build 19044). The TF-IDF matrix is generated using Anaconda Notebook 6.5.4, and the LSA-SVD process is performed using MATLAB R2022a.

#### 3.1 Dataset Acquisition

First, the web news data were gathered based on the requirements of our experiment. This study uses a selection of fake news as its search object. A unique collection of online news data, specifically curated to examine false news during political campaigns, served as the basis for the investigation. Since this data is not a standard one, this study can be applied to analyze the fake news pertaining to political elections.

#### 3.2 Pre-processing of Data and NLP

Python was used to pre-process the data so that it could be analyzed for this study. In order to prepare the data for the creation of the TermDocument Matrix (TDM) for LSA, tokenization and lemmatization were crucial pre-processing steps. In text mining applications, natural language processing (NLP) is essential because it makes it possible to extract useful information from text. NLP allows complex statements, like those in analysis of tweets can help with objective communication (Martinelli, D. D., 2022). The implementation process is described in detail in the sections that follow.

### 3.2.1 Tokenization

Text tokenization was done using the Python Natural Language Toolkit (NLTK). In this study, texts were divided using "Unigram."

### 3.2.2 Cleaning the Data

In order to obtain clean data for future research, stop words like "the" and "or," non-English keywords, and unnecessary icons were removed at this stage.

### 3.2.3 Lemmatization

In this phase, complex word forms were simplified to their base or root form. For example, words like "ate" and "eaten" were reduced to "eat."

### 3.2.4 Word Frequency Counting

In this step, the frequency of each word was calculated, and those that appeared fewer than five times were removed.

### 3.2.5 Building the TDM (Term-Document Matrix)

In this step, the Term-Document Matrix (TDM) was built using the TF-IDF (Term Frequency-Inverse Document Frequency) weights, as described in Equation (1).

$$TF\text{-}IDF = TF(t_i, d_i) \times \log\left(\frac{N}{Nt_i}\right) \quad (1)$$

where,  $Nt_i$  indicates the number of total documents that include  $t_i$ , number of features,  $N$  is the count of all documents, and  $t_i$  is the  $i^{\text{th}}$  word; the  $j^{\text{th}}$  document is signified.

## 3.3 LSA (Latent Semantic Analysis)

In this stage, Singular Value Decomposition (SVD) was used to examine the link between words and their underlying ideas.

### 3.3.1 SVD Process

Figure 2 illustrates the Term-Document Matrix (TDM). The SVD function generates three matrices:

Term Matrix (U): An  $t \times r$  matrix

Singular Value Matrix (S): An  $r \times r$  orthogonal matrix

Document Matrix ( $V^t$ ): A  $r \times n$  matrix

In these matrices,  $t$  represents the number of terms,  $n$  refers to the count of documents, and  $r$  is the number of latent concepts. The SVD procedure is visually depicted in equation 2

$$\begin{matrix} A = & U & \times & S & \times & V^t \\ (t \times n) & (t \times r) & & (r \times r) & & (r \times n) \end{matrix} \quad (2)$$

### 3.3.2 Dimension Reduction

After performing SVD, a significant amount of irrelevant data may still be present; therefore, the dimensional space must be reduced. To preserve the original qualities, the feature value  $k$  must be carefully chosen. The  $k$  value is determined in this study using the screen test. The creep plot always shows the eigenvalues on a downward slope, with the eigenvalues arranged in order from smallest to largest. Concepts to the left of the graph's "elbow," where these eigenvalues appear to level out, should be preserved as important under the screen test. In equation 3, we only need to take the  $S_k$  matrix because we have chosen  $k$  ideas, which reduces the dimensionality.

$$\begin{matrix} A_k = & U_k & \times & S_k & \times & V^t \\ (t \times n) & (t \times r) & & (r \times r) & & (r \times n) \end{matrix} \quad (3)$$

### 3.3.3 Rotation of the axes orthogonally

To determine the concept load (LT), Equation (2) is used to calculate the concept  $S_k$ , which is then multiplied by the dimensionally restricted idea, UK. The feature words are then rated based on the load, and a name is given to the word concept. The dimensionally delimited concept (Uk) is multiplied by the concept  $S_k$ , as determined by Equation (4), to obtain the concept load (LT). Following the ranking of each feature word based on the load (LT).

$$LT=Uk \times S_k \quad (4)$$

### 3.4 Giving the extracted concepts names

The focus was given to assigning names to concepts identified in the subsequent steps. These concept names were derived from the keywords extracted in the previous phase. In addition to being removed, the significant concepts were carefully examined, which contributed to their definition and understanding.

### 3.5 Results Explanation

After presenting findings of the LSA, the authors found the concepts of Step 4. Moreover, our results were used to obtain an accurate examination of fake news.

### 3.6 Developing arguments and conclusions

According to the findings, the authors conducted a broad discussion and came up with a conclusive statement. The discussion section is outlined into four major sections, as follows: The first part explores the approaches to distinguishing between the genuine and the fake news, with the heavy focus on the analytical methods that can be used regarding the variants. The second element examines the structural properties of the accurate and incorrect news, and thus, the accurate demarcation of misinformation is improved. The third component deals with the popularity of fake news that spreads incomplete or outdated news. The last element considers the consequences of fake news on the political sphere and its relevance in the environment of a rapidly changing global space. The discussion is concluded by the proposal of resolution.

## 4. Experimental Results

LSA's findings on fake news during the US election campaign. To extract the Term Document Matrix (TDM) from the dataset for MATLAB analysis, Python was used for this study. Fraudulent news articles from the 2016 US presidential election campaign are included in it. The dataset contains a total of 17,903 news article records dring March of 2016 to December of 2017, with each entry providing a title, text, subject, and date. For further analysis, 5,009 instances of fake news were selected based on the "text" variable. This dataset was chosen due to the significant political context surrounding the US presidential election at that time.

A sample of the bogus news used is shown in Table 1. Several unnecessary icons from the original data appear in this table. Step 2 outlines the implementation of the NLP process. Lastly, we use TF-IDF weights to build a  $5,009 \times 10,019$  TDM for LSA.

Table1.A sample of fraud news

Title	Text	Subject	Data
SheriffClarke David makes a funny moment in internet threatening to stab people.	It was revealed on day after Thursday that former SDC, who had been under consideration for the role of household personal secretary in the Trump administration, was embroiled in an electronic mail controversy. The controversy stemmed from an incident that occurred in January, when Clarke had a brief altercation with fellow passenger Dan Black on an aircraft. Black was later arrested by police for reasons that remain unclear, though some speculated it was due to hurt feelings. Clarke reportedly instructed law enforcement to detain Black after he disembarked from the plane. In response, the FBI executed a search warrant to examine the discussions surrounding the incident. Despite copies of the warrant being available online, Clarke dismissed the reports as "fake news." Until they get it, I'll keep poking these scumbags and sticking a	Media	30 December 2017

	<p>sharp stick in their eye. The sentence mentioning that they were assaulted by people who are superior to them, is another form of a statement of personal grievance, which Clarke closed his post with the slogan that he used repeatedly: MAGA, meaning Make America Great Again. This was followed by another tweet which he referred to as NEWS FLASH in which he claimed that the LIB (latin; left-leaning or liberal) media habitually manufactures fake news. Putting his comments into a larger context of what Clarke believed was a policy long running effort by the media to delegitimize him, Clarke stated that these attempts to silence his voice could only succeed in strengthening his will. This was quickly met with praise and criticism amidst online communities with his comments which contained a stern promise to continue addressing those who he says are spreading fake stories.</p>		
--	---	--	--

Then, the best k-value was identified. The screen plot that helps in this identification is depicted in Figure 3.. Based on the analysis presented in Figure 2, the k-value was selected as outlined in Section 5.

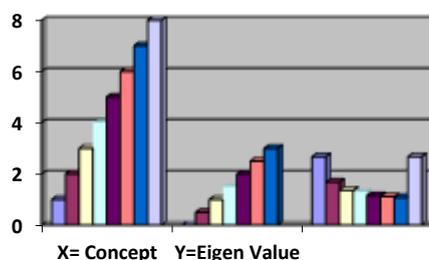


Figure 2. Screen plot

During the Singular Value Decomposition (SVD) process, the concepts with the highest loadings were identified after extracting the data. Four key terms were then selected from each loading, which the authors grouped into precise thematic ideas. This process resulted in the identification of five distinct ideas. The priority keywords and corresponding loadings for each topic are presented in Table 2. These extracted concepts were then named based on the representative terms associated with each idea. The Table 3. Shows the concepts and their representative lexicon.

Table 2. The loadings and keywords in every concept were extracted.

Concept 1	Concept 2	Concept 3	Concept 4	Concept 5
<b>Keywords loading</b>				
Administration 0.392	Republican 0.923	Great 0.287	Clinton 0.235	Campaign 0.162
Political 0.377	Senator 0.194	Fact 0.362	Tweeted 0.258	Report 0.221
America 0.256	Political 0.149	Justice 0.175	Talking 0.113	Medium 0.137
Much 0.273	Election 0.144	Life 0.127	Say 0.098	World 0.112
Republican 0.184	Campaign 0.121	Idea 0.085	Claim 0.092	America 0.066
Election 0.182	Clinton 0.078	Medium 0.081	Decided 0.084	United 0.061
Breaking 0.159	Senate 0.076	Public 0.088	Great 0.071	Senate 0.058
Accused 0.177	Donald 0.072	Time 0.079	Medium 0.069	FBI 0.055
Senate 0.188	Message 0.064	Message 0.084	Justice 0.059	Pie 0.151

The LSA data were used to build the concepts. However, it was inevitable that the problem of constructing coherent stories from the concepts by mixing keywords and their related meanings would arise. All terms considered to have similar meanings were based on contextual similarities, which were determined by the

frequency of their co-occurrence within documents. This approach is consistent with the observations of Wang et al. (Wang et al., 2012) and does not indicate any syntactic relationships. In other words, substantial similarity between two keywords was deemed to exist if they appeared in various contexts that described the same area. For instance, terms like "Administration," "Public," "Muore and "America" were grouped under a single category. Following this process, the extracted ideas related to fake news were then explained in detail.

Table 3. Concepts retrieved

No.	Concept Name	Attorney
1	Consortium	Administration, America, Political, Much
2	Diplomacy	Election ,Political, Campaign ,Senator
3	Ensuing	Great, Fact, Life, Justice
4	Assertion	Say, Clinton, Tweeted, Talking,
5	Circulation	Campaign, Report, Medium, Pic

### Idea#1: Cooperation

"Coalition" is the name of the notion of the representative glossary that includes the words "Political," "Administration," "Much," and "America." Coalition building is a political technique that politicians and/or parties require. Furthermore, in politics, coalitions are typically employed to gain power. However, because the parties' relationship is not long-term, the alliance is only short (Hochman, E. R. 2014). Depending on the objectives of the parties throughout the election, the candidate who wins the election will also be in charge of the government.

### Idea#2: Politic

It is a commonplace among scholars to draw up politic along bifurcated conceptual lines: connotative and denotative. When used in connotative context, the word politic can denote a negative consequence, which is the strategic exploitation of identifiable groups of people, an ethnic group or a tribal group, in an attempt to achieve particular partisan goals. These groups can be taken to represent interest groups, thus creating a more strategically minded or exploitative conceptualization of that term. The denotative framework, on the other hand, offers a stronger affirmative interpretation, which is the custodial role of judicious resource management of the wider public interest including the education of the citizens, the protection of natural gifts, and the drafting of policies aimed at the promotion of the general good.

### Idea #3: Future

"Future" is the notion derived from the representative lexicon that includes "Fact," "Great," "Justice," and "Life." The term "future" refers to people who live in optimism about the future. Hope may be what motivates people to make investments in a better future. Especially with coalitions formed to build positive ties, there can be much hope in the political landscape. The things that are anticipated to occur shortly are also referred to as the "future."

### Idea #4: Declaration

The representative lexicon for this idea includes the terms "Tweeted," "Clinton," "Talking," and "Say." The concept is known as a "statement." A "Statement" is a representation of key political communications and a form of formal correspondence between the public and politicians or other high-ranking officials. Speech and written communication are the two forms of "statement." An official statement, also referred to as a statement, typically refers to extremely important messages communicated to the broader public. Mass media is typically used to deliver official pronouncements; however, these days, social media is more often used. VIPs often utilize social media accounts to engage with the public.

### Idea #5: Problems

Problems are considered rumors that lack factual support. This term is used to describe the noise in a political setting. Issues, however, are typically exploited in campaign activities for some reason. Issues must, therefore, not be trusted until they can be verified, and the real story is established. Since the media occasionally uses the news to report on concerns, they have also turned into "big fans" of rumors.

## 5. Discussion

How to Tell the Difference between False and Real News. We identified and named five key concepts related to fake news based on our findings. They are "issues," "coalition," "politics," "future," and "statement." These news categories are typically used to sway readers and influence election outcomes, as is well known. The content that can sway readers having "precision" and is fabricated by a fake newsmaker can be categorized based on our findings.

These ideas, however, are only for the 2016 US election campaign. We tried to use these findings to identify elements of fake news. Based on the LSA's findings and observations, we contrasted fake and authentic news. Following that, the distinctions between fake and authentic news were identified. A summary of the findings, along with some of the variations, is presented in Table 4. Initially, false news is not grounded in reality; rather, it is based on social media and fabricated facts. Second, while true news typically comes from a variety of sources that cover both sides of the story, fake news often originates from a single source. Thirdly, fake news is often considered one-sided, whereas real news presents both positive and negative perspectives. Fourth, websites or URLs that are used to publish or transmit news or information containing misleading or inaccurate details are known as identified and unidentified domains. Additionally, trustworthy news sources, such as official websites or reputable media outlets, can help identify regions associated with erroneous or misleading news. Official declarations regarding false information that circulates on social media or other websites are sometimes provided by these news providers.

Table 4. Differences between fake news and accurate news

No.	Precise news	Fraud news
1	based on reality or facts	Considering social media
2	According to the facts	based on fictitious facts
3	has a lot of sources	has a single source.
4	has the opposite data.	includes biased information.
5	has a domain that has been recognized.	includes an unknown domain.

Furthermore, some of the standards of accurate news are used by fake news. The requirements for newsworthiness, as seen through the lens of journalism, are given in Table 5. The following table lists the essential components of accurate news. However, investigation reveals that bogus news often contains the same components as legitimate news. The example of fake and real news is provided in Table 6. The structure of real and fraudulent news will also be examined.

Table 5.The criteria for distinguishing the systematic of accurate news and fake news.

No.	Precise and fraud news
1	5W+1H(what, when, why, where, who+How)
2	Erection
3	Fame
4	Presence
5	Dissension
6	Encounter
7	Consequence

Table 6 provides a framework for analyzing the structure of fake and true news. The headline, lead, and data section are included in the overall structure of both kinds of news. In addition to showcasing compelling language, headlines occasionally aim to capitalize on the news issue. Furthermore, clickbait titles are frequently used to entice readers to read online news, whether it is factual or not. Additionally, both accurate and fake news can follow similar structures for presenting content. Two key types of leads commonly used in news articles are the "What Lead," which explains what is happening, and the "Who Lead," which details who is involved in the event. Both authentic and fraudulent news sources use these leads. Both forms of news also include the article's body, which offers more information about the events. Both trustworthy and fraudulent news usually include some facts in their conclusion to persuade the reader of the news' veracity. This tactic has the potential to deceive readers because, even when fake news contains factual information, the overall story may still be inaccurate or skewed. As a result, even when fake news is presented as legitimate and convincing, readers may still be duped.

structure	Precise news	Fraud news
subject	Shepard Smith, who dispels the network's favourite Hillary Clinton "scandal," enrages fans on FoxNews.	Watch as Delusional Trump Supporters Go Wild on FoxHost to Report Uranium One Deal Facts
primacy	When Fox News anchor Shepard Smith publicly denied what many on his own network had called the "Hillary Clinton crisis," viewers were incensed. Viewers reacted negatively to his remarks because they ran counter to the narratives that other Fox personalities were pushing. Given their more liberal editorial stances, some even conjectured that Smith ought to think about working for CNN or MSNBC instead, considering his apparent departure from the network's usual tone. Fox's more conservative audience criticized Smith's position on the matter, underscoring the conflict between the network's various factions and its audience.	It is evident that many Trump supporters expect Fox News to share their opinions, and tensions increased when Donald Trump's allegations regarding Hillary Clinton and the Uranium One deal with Russia turned out to be true. Shepard Smith, the anchor for Fox News, checked. The Republican allegations that, while serving as Secretary of State in 2010, a "pay-to-play" scheme related to the sale of uranium to Russian interests was approved by Clinton were thoroughly debunked by Smith. His fact-based refutation of the claims angered many in Trump's base, who felt that Fox News should be supporting the narrative pushed by the President and his allies. Smith's rebuttal highlighted the growing divide between the network's more straightforward reporting and the expectations of certain segments of its audience.
build	Shepard Smith's criticism, which called President Trump's charges against Hillary Clinton "inaccurate," came after Republicans on Capitol Hill once again demanded that a special counsel look into Clinton. The Uranium One deal was the subject of a renewed push for an investigation, with GOP lawmakers accusing wrongdoing. Smith responded by going on the air to directly address these allegations, stressing that there was no justification for the appointment of a special counsel and pointing out that the evidence did not support the charges. His fact-based response exposed the continued partisan split on the matter and contrasted sharply with the narrative being advanced by some Republican leaders. During an appearance on Fox News last month, former White House adviser Sebastian Gorka intensified the rhetoric even more. He compared the case to the espionage case of Julius and Ethel Rosenberg in the 1950s, when they were found guilty of providing the Soviet Union with access to American nuclear secrets. In reference to the Rosenbergs' execution for their alleged treasonous acts, Gorka noted, "those people got the chair," underscoring the seriousness of the comparison.	Nine agency chiefs had to first approve the agreement, and they all agreed. According to a State Department assistant, Clinton did not intervene.  Third, Trump's claim that Hillary transferred 20% of the uranium to Russia was refuted by the treaty, which stipulates that the uranium must be sold to civilian reactor operators in the United States.  Fourth, nine Uranium-related individuals One gave the Clinton Foundation all but \$4 million of the \$140 million

	<p>Although the facts and context are very different, conservative figures have compared this dramatic framing to past instances of betrayal and espionage, reflecting the heightened political tensions surrounding the issue.</p>	<p>that was donated. Long before Hillary Clinton even started thinking about running for secretary of state, in 2007, the man in question had already sold his stake in the company. This occurred long before Clinton was named president by Barack Obama after he was elected. Furthermore, a committee comprising several government agencies made the decision, and Clinton's role was primarily advisory. As a result, she lacked the power to approve or veto the agreement herself. The Justice Department's attempt to name a special prosecutor based on the accusations is essentially undermined by the fact that conservatives have no solid case against Clinton. Charges that just don't stand up to scrutiny are the foundation of the special prosecutor movement. Ultimately, the facts back Clinton, and any court would likely dismiss these fabricated claims if brought before them. The charges against her, driven largely by partisan narratives, have little grounding in reality, and efforts to pursue them further appear politically motivated rather than fact-based.</p>
<p>features</p>	<p>Several fact-checkers, including <i>The Washington Post</i>, have already dismantled the core of the charges surrounding the Uranium One deal. In light of this, Fox News' debunking, particularly that of Shepard Smith, was surprising. Smith echoed many of the fact-checkers' points in his broadcast. "Now, here's the charge," he said in response to the accusations. The Clinton Foundation was accused of receiving approximately \$140 million in donations from nine people who were involved in the uranium deal. It was alleged that as part of a quid pro quo agreement, Secretary of State Hillary Clinton authorized the sale to Russian interests. However, as fact-checkers, including <i>The Washington Post</i>, have pointed out, there is no evidence supporting the idea that Clinton's approval of the deal was influenced by these donations. Smith's careful dissection of the story, emphasizing the lack of direct connections between the donations and the</p>	<p>This is the video on YouTube. Trump supporters responded by throwing a fit and demanding that Fox fire Smith for revealing the truth. Get that hack, @ShepNewsTeamoutoffox. Just now, he offered excuses for #CrookedHillary and is obviously a double agent. Hi, Shep. Regardless of when the 145M was sent, Hillary took it. Lab Lover</p>

	<p>approval process, further undermined the narrative being pushed by Trump supporters and conservative commentators. Peter Schweizer, senior editor-at-large of <i>Breitbart</i> and author of the book that brought the Uranium One allegations to the forefront, has been one of the primary figures behind these claims. However, as Smith and other fact-checkers have demonstrated, the evidence simply does not back up the accusations of wrongdoing.</p> <p>initially made the accusation in his 2015 book "Clinton Cash." Donald Trump used the charge as evidence of Clinton's corruption the following year.</p>	<p>(@Dutchistheballs) 14 November 2017 Pay for play, you idiot</p>
--	--	--

Table6.The structure of precise and fraud news.

Fake news is produced to manipulate audiences, according to journalistic sensibilities (Wahutu, J. S. 2019). The term "manipulation" is sometimes used to describe a Western perspective on media dynamics. For example, the spread of "fake news" is closely tied to traditional (Western) media norms, which can be problematic. As a result, the problem is not essentially "fake news"; rather, it emphasizes the necessity for the global media landscape to embrace new normative frameworks. Asia and Africa may have quite different media norms than the US, Canada, Europe, and Australia.

Maintaining this Western-centric viewpoint is becoming more and more untenable in the post-Facebook, Cambridge Analytica era, as it only serves to exacerbate the "fake news" problem. Social media platforms are now a major means of disseminating fake news, facilitating its quick spread and increasing its impact. The concepts presented in this article are not thought to be mutually exclusive; in fact, interactions that are frequently missed in the ongoing "fake news" debate are often seen. These connections highlight how complicated the problem is and imply that combating "fake news" calls for a more thorough and sophisticated strategy. In order to shield the traditional press from harsh criticism of its flaws, the narrative frequently places the entire blame on social media companies and their users. It is more fruitful to concentrate on the larger news culture influencing the African media landscape rather than becoming enmeshed in the frequently simplistic discussion of "fake news," as these three factors are intricately linked. The well-defined idea of "information morality panic" has made its way into African conversations about "fake news," despite the fact that regional differences in news consumption patterns may be structural. Instead of blaming social media alone, this change emphasizes the necessity of critically analyzing the media landscape as a whole.

The concept of delayed facts is the one that is least commonly used to describe online news. While some journalists might disagree, others might be in favor of delayed facts. Online news is distinguished by its rapid dissemination. It is also the main problem with lobbying messages. Journalists often leave out facts when reporting the news. When news is first released, it is always not fully complete. The idea of "delayed facts" has emerged as a crucial component in traditional newspaper news discovery. Newspapers usually concentrate on providing factual information, in contrast to online news. However, because of its infinite column length and space, online news has the advantage of offering "continuous" updates, which enables a more dynamic flow of information. This flexibility, though, may also cause online journalists to become preoccupied with delayed facts, which could divert them from making sure their reporting is accurate. Confusion and misunderstanding may result from readers processing inaccurate or partial information.

The effect of false information on the current state of the information era although the impact of fake news has been acknowledged, researchers have systematically examined the impact of fake news in a disruptive environment. Generally speaking, disruptive times are characterized by the emergence of numerous intangible innovations that established institutions, companies, agencies, or organizations are not aware of, upending and potentially upending the current system order. Even if they lack the skills of a professional journalist, anyone can become a citizen journalist today by using social media platforms to collect, edit, and disseminate information. Information subversion is a "gift" of uncertain times. It is simple for anyone to create information that the media does not cover, which is the benefit of subversive information.

However, people can also transmit information that the mass media has not. According to the journalistic criteria discussed in the previous section of the article, this is a catastrophe for information dissemination. Unreliable information is a significant issue for the general public, who struggle to distinguish between

potentially false information and accurate information. This false information may become fake news and is now referred to as a hoax. The analysis's most obvious conclusion is that if people with a solid understanding of information can identify and disseminate just the most relevant information, information disruption can help foster a healthy democracy. The mass media now must integrate functionality. It is essential to link the verification modality to the fourth traditional function of mass media. In an era where information is abundant, the mass media may be primarily focused on validation. Figure 3 depicts an information disruption.

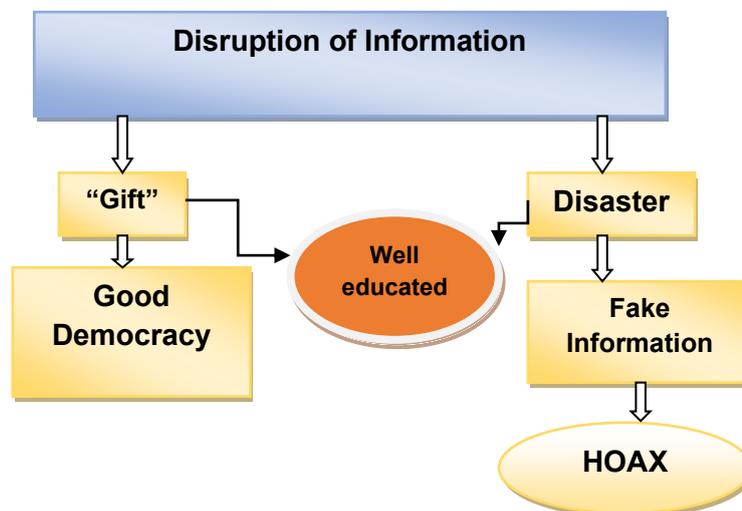


Figure 3. Fake news in the disruption of information era.

## 6. Limitation

The fact that political fake news is the exclusive focus of this study, specifically in the context of a high-profile presidential election that attracted significant attention from fake news producers, means that the findings may not be directly applicable to other domains, such as health, social issues, culture, finance, and so on. The unique nature of political events, with their heightened visibility and polarization, may create a different dynamic for the spread and impact of fake news compared to other topics. Therefore, caution should be exercised when attempting to generalize the study's conclusions to other areas. As a result, the generalization that this approach can be helpful for other news types is not the aim of this study. Additionally, the computational efficiency of SVD, which was used for the LSA investigation, is limited.

## 7. Conclusion and future work

Using LSA, this study sought to analyze fake news. We identified five key terms from bogus political news: "coalition," "politic," "future," "statement," and "issues." To sway readers and influence the election's outcome, these sets of ideas were used to fabricate fake news. One could use these ideas to spot phony election reporting.

Additionally, examining the gathered fake news beforehand from a journalistic standpoint was crucial. The writers contrasted real news and fake news. Some notable distinctions were discovered. First, false information is spread via social media. Real news sources typically cover all sides of a story, whereas fake news often originates from a single source. Third, phony news presents one side of the story, whereas real news presents both advantages and disadvantages.

Last but not least, bogus new software originates from URLs with unknown domains. Additionally, our research examined the article structures of true and fake news and identified these proven discriminant components of accurate news content. It was discovered that the elements of fake and real news are the same in terms of article architecture, implying that fake news authors will attempt to pass off phony news as factual. As a result, it is impossible to discern the veracity of news from the format and composition of news articles. Sentences that are difficult for readers to recognize.

The study's conclusions help us understand false news in several ways and provide a foundation for distinguishing between genuine and misleading news. It is worth noting that LSA is a well-established methodology, and the dataset used is also available.

In contrast to the more scientific, data-based approaches, the goal of this study is to provide social scientists—who typically describe what they observe using observational methods—with simple tools. Our proposed method can assist social scientists in gleaned useful information from the deluge of fake news. Furthermore, despite being straightforward and unambiguous, the proposed method has never been used in the analysis of fake news. This is the study's main contribution in practice as well.

This study's scope of the study focussed on fake election news. Future research could cover a wider range of subjects, which would help us better understand the dynamics of fake news and enhance readers' capacity to discern it from reliable sources. In light of the ongoing information war, such research may also be essential to reducing the impact of fake news on democracies. Also, not only fake news but real news detection will also be implemented on several legitimate datasets.

Furthermore, this study's dataset only contained fake news. The development of techniques for identifying fake news and comparing the outcomes with those for identifying true news would be an intriguing avenue for future research. This could provide important information about how well readers can distinguish between false and accurate information, thus strengthening our defenses against misinformation.

## 8. References:

1. Obreja, D. M. (2022). Mapping the political landscape on social media using bibliometrics: A longitudinal co-word analysis on Twitter and Facebook publications published between 2012 and 2021. *Social Science Computer Review, in press*. <https://doi.org/10.1177/08944393221077925>
2. Vafeiadis, M., & Xiao, A. (2021). Fake news: How emotions, involvement, need for cognition and rebuttal evidence (story vs. informational) influence consumer reactions toward a targeted organization. *Public Relations Review, 47*, 102088. <https://doi.org/10.1016/j.pubrev.2021.102088>
3. Lee, T. (2019). The global rise of “fake news” and the threat to democratic elections in the USA. *Public Administration and Policy, 22*, 15–24. <https://doi.org/10.1108/PAP-06-2019-0011>
4. Hallin, D. C., Mellado, C., & Mancini, P. (2021). The concept of hybridity in journalism studies. *International Journal of Press/Politics, 28*, 219–237. <https://doi.org/10.1177/19401612211017237>
5. Patel, M., Padiya, J., & Singh, M. (2022). Fake news detection using machine learning and natural language processing. *Studies in Computational Intelligence, 1001*, 127–148. [https://doi.org/10.1007/978-3-030-83204-5\\_9](https://doi.org/10.1007/978-3-030-83204-5_9)
6. Shah, M. N., & Ganatra, A. (2022). A systematic literature review and existing challenges toward fake news detection models. *Social Network Analysis and Mining, 12*, 168. <https://doi.org/10.1007/s13278-022-00932-6>
7. Delaney, C. J., Bacon, H. R., & Matson, S. F. (2022). From fake news to racism: A study of change in a reading intervention class. *Journal of Adolescent & Adult Literacy, 65*, 419–429. <https://doi.org/10.1002/jaal.1214>
8. Pérez-Escoda, A. (2022). Infodemic and fake news turning shift for media: Distrust among university students. *Information, 13*, 523. <https://doi.org/10.3390/info13110523>
9. Tashtoush, Y., Alrababah, B., Darwish, O., Maabreh, M., & Alsaedi, N. (2022). A deep learning framework for detection of COVID-19 fake news on social media platforms. *Data, 7*, 65. <https://doi.org/10.3390/data7070065>
10. Capuano, N., Fenza, G., Loia, V., & David, F. (2023). Content-based fake news detection with machine and deep learning: A systematic review. *Neurocomputing, 530*, 91–103. <https://doi.org/10.1016/j.neucom.2022.11.112>
11. Song, C., Teng, Y., Zhu, Y., Wei, S., & Wu, B. (2022). Dynamic graph neural network for fake news detection. *Neurocomputing, 505*, 362–374. <https://doi.org/10.1016/j.neucom.2022.07.087>
12. Zhang, Y., Hu, Y., Jiang, N., & Yetisen, A. K. (2023). Wearable artificial intelligence biosensor networks. *Biosensors and Bioelectronics, 219*, 114825. <https://doi.org/10.1016/j.bios.2022.114825>
13. Iwendi, C., Mohan, S., Khan, S., Ibeke, E., Ahmadian, A., & Ciano, T. (2022). COVID-19 fake news sentiment analysis. *Computers & Electrical Engineering, 101*, 107967. <https://doi.org/10.1016/j.compeleceng.2022.107967>
14. Lin, S. Y., Kung, Y. C., & Leu, F. Y. (2022). Predictive intelligence in harmful news identification by BERT-based ensemble learning model with text sentiment analysis. *Information Processing & Management, 59*, 102872. <https://doi.org/10.1016/j.ipm.2021.102872>
15. Pratama, R. P., & Tjahyanto, A. (2021). The influence of fake accounts on sentiment analysis related to COVID-19 in Indonesia. *Procedia Computer Science, 197*, 143–150. <https://doi.org/10.1016/j.procs.2021.12.122>
16. Kapusta, J., Hájek, P., Munk, M., & Benko, L. (2020). Comparison of fake and real news based on morphological analysis. *Procedia Computer Science, 171*, 2285–2293. <https://doi.org/10.1016/j.procs.2020.04.247>
17. Kozik, R., Kula, S., Choraś, M., & Woźniak, M. (2022). Technical solution to counter potential crime: Text analysis to detect fake news and disinformation. *Journal of Computational Science, 60*, 101576. <https://doi.org/10.1016/j.jocs.2022.101576>
18. Lim, S. (2020). Academic library guides for tackling fake news: A content analysis. *Journal of Academic Librarianship, 46*, 102195. <https://doi.org/10.1016/j.acalib.2020.102195>
19. Krzyżanowski, M., & Ekström, M. (2022). The normalization of far-right populism and nativist authoritarianism: Discursive practices in media, journalism and the wider public sphere/s. *Discourse & Society, 33*, 719–729. <https://doi.org/10.1177/09579265221079820>
20. Pickard, V. (2013). Social democracy or corporate libertarianism? Conflicting media policy narratives in the wake of market failure. *Communication Theory, 23*, 336–355. <https://doi.org/10.1111/comt.12017>
21. Tomaselli, K. G. (2009). (Afri)Ethics, communitarianism and libertarianism. *International Communication Gazette, 71*, 577–594. <https://doi.org/10.1177/1748048509339794>

22. Dupuis, I. (2021). The mass media's systemic contribution to political transformation: Coverage of the 1956 uprising in Hungarian print media (June 1988–June 1989). *Central European Journal of Communication*, 14, 305–320. [https://doi.org/10.51480/1899-5101.14.2\(29\).10](https://doi.org/10.51480/1899-5101.14.2(29).10)
23. Arqoub, O. A., Elegu, A. A., Özad, B. E., Dwikat, H., & Oloyede, F. A. (2022). Mapping the scholarship of fake news research: A systematic review. *Journal of Practice*, 16, 56–86. <https://doi.org/10.1080/17512786.2022.2045689>
24. Khan, S. A., Shahzad, K., Shabbir, O., & Iqbal, A. (2022). Developing a framework for fake news diffusion control (FNDC) on digital media (DM): A systematic review 2010–2022. *Sustainability*, 14, 15287. <https://doi.org/10.3390/su142215287>
25. Baptista, J. P., & Gradim, A. (2022). Who believes in fake news? Identification of political (a)symmetries. *Social Sciences*, 11, 254. <https://doi.org/10.3390/socsci11060254>
26. McQuail, D. (2010). *McQuail's mass communication theory* (6th ed.). Sage Publications. ISBN: 9781849202916
27. Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news”: A typology of scholarly definitions. *Digital Journalism*, 6, 137–153. <https://doi.org/10.1080/21670811.2017.1360143>
28. Hangloo, S., & Arora, B. (2022). Combating multimodal fake news on social media: Methods, datasets, and future perspective. In *International Conference on Innovative Computing and Communications* (Vol. 28, pp. 425–437). Springer. ISBN: 0123456789
29. Tsfti, Y., Boomgaarden, H. G., Strömbäck, J., Vliegthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: Literature review and synthesis. *Annals of the International Communication Association*, 44, 157–173. <https://doi.org/10.1080/23808985.2020.1759443>
30. Martin, J. D., & Hassan, F. (2020). News media credibility ratings and perceptions of online fake news exposure in five countries. *Journalism Studies*, 21, 2215–2233. <https://doi.org/10.1080/1461670X.2020.1790025>
31. Kwon, H., Kim, J., & Park, Y. (2017). Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. *Technovation*, 60–61, 15–28. <https://doi.org/10.1016/j.technovation.2017.01.003>
32. Yu, B., Xu, Z. B., & Li, C. H. (2008). Latent semantic analysis for text categorization using neural network. *Knowledge-Based Systems*, 21, 900–904. <https://doi.org/10.1016/j.knsys.2008.01.006>
33. Chen, W. K., Chen, L. S., & Pan, Y. T. (2021). A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. *Applied Soft Computing*, 111, 107704. <https://doi.org/10.1016/j.asoc.2021.107704>
34. Suleman, R. M., & Korkontzelos, I. (2021). Extending latent semantic analysis to manage its syntactic blindness. *Expert Systems with Applications*, 165, 114130. <https://doi.org/10.1016/j.eswa.2020.114130>
35. Hsiao, Y. H., & Hsiao, Y. T. (2021). Online review analytics for hotel quality at macro and micro levels. *Industrial Management & Data Systems*, 121, 268–289. <https://doi.org/10.1108/IMDS-05-2020-0273>
36. D'Silva, J., Sharma, U., & More, C. (2023). Automatic text summarization of Konkani texts using latent semantic analysis. In D. Gupta et al. (Eds.), *International Conference on Innovative Computing and Communications* (pp. 425–437). Springer Nature.
37. Martinelli, D. D. (2022). Evolution of Alzheimer's disease research from a health-tech perspective: Insights from text mining. *International Journal of Information Management Data Insights*, 2, 100089. <https://doi.org/10.1016/j.ijime.2022.100089>
38. Wang, C., Guo, X., & Han, H. (2012, August 20–24). Crime detection using latent semantic analysis and hierarchical structure. *2012 IEEE International Conference on Computer Science and Automation Engineering* (pp. 337–340). <https://doi.org/10.1109/CSAE.2012.6272843>
39. Hochman, E. R. (2014). Ein Volk, ein Reich, eine Republik: Großdeutsch nationalism and democratic politics in the Weimar and first Austrian republics. *German History*, 32, 29–52. <https://doi.org/10.1093/gerhis/ght103>
40. Wahutu, J. S. (2019). Fake news and journalistic “rules of the game”. *African Journal of Studies*, 40, 13–25. <https://doi.org/10.1080/02589001.2019.1588213>