International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 10, Volume 08 (October 2021) https://www.irjcs.com/archives

# NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS

**Aryaman Sharma**
Undergraduate student, Computer Engineering
Delhi Technological University
aryamansharma0408@gmail.com

**Publication History**
Manuscript Reference No: **IRJCS/RS/Vol.08/Issue10/OCCS10080**
http://www.irjcs.com/volumes/Vol8/iss-10/01.OCCS10080.pdf
Received: 03, October 2021
Accepted: 21, October 2021
Published: 01, November 2021
**Citation:** Aryaman (2021). NATURAL LANGUAGE PROCESSING AND SENTIMENT ANALYSIS. International Research Journal of Computer Science, X- October Issue, 2021, Pages 237-242
**doi:** https://doi.org/10.26562/irjcs.2021.v0810.001
Editor: Dr.A.Arul Lawrence Selvakumar, Chief Editor, IRJCS, AM Publications, India


**Abstract:** Natural Language Processing is one of the branches of Artificial Intelligence that has only recently entered the spotlight. Apple Siri, Amazon Alexa, and, more recently, Google Duplex are just a few of the most well-known instances of NLP in work, with the technology delivering outstanding human-machine interactions. By 2023, there are estimated to be eight billion digital voice assistants in use, owing to their popularity. With such a large-scale use, the data generated from these interactions would also be immense. This untapped goldmine of data can be used to further research and development of Natural Language Processing and can be used in many industries such as healthcare, technology, business etc. Sentiment analysis, with the help of Natural Language Processing can help industries process these huge datasets faster and efficiently and can be used, for example, in the healthcare industry to diagnose patients and develop diagnostic models for detecting chronic disease in its early stages. Web 2.0 has allowed massive user data to be generated which can be tapped to extract valuable information for various reasons that an individual, policy maker(s), organizations and Governments might need.
**Keywords:** Natural Language Processing, Healthcare, Technology, Artificial Intelligence, Data.

## INTRODUCTION

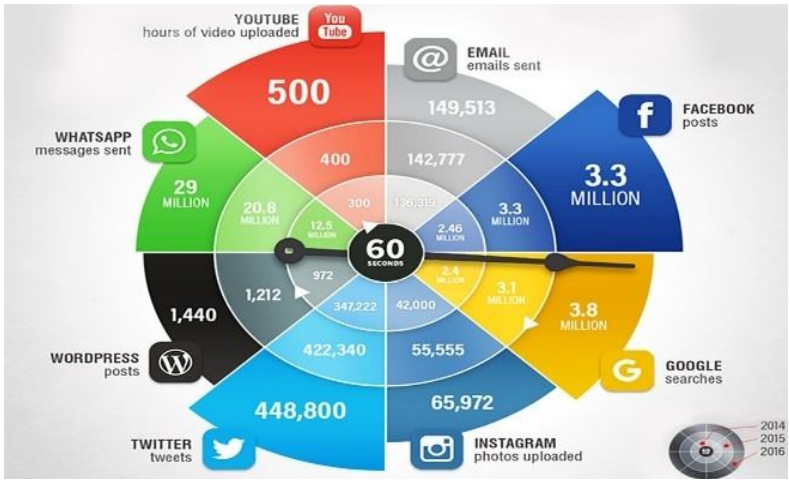
Natural language processing refers to a system's capacity to process sentences written in natural languages such as English rather than specialized artificial computer languages such as C, C++, JAVA, PROLOG, and so on. It is a branch of artificial intelligence that examines data in order to better comprehend natural human language. Natural language processing, or NLP, is a branch of machine learning that focuses on analyzing the nuances of how people communicate with one another so that it may replicate how we speak to and with one another without the constraints of the English language. It is impossible to overestimate the relevance of computer communication in ordinary language. Natural language processing (NLP) is undoubtedly the best example, with intelligent chatbots poised to revolutionize customer service and beyond. While computers have always been extremely useful for abstract tasks such as quantification, real humans have always presented a problematic interface. Despite the fact that computing technologies enable extremely rapid and precise communication channels, machines have never been satisfactory at understanding how and why we interact. NLP is focused to deciphering the cooperation between computers and machines through the use of language.

Today's consumers generate a vast amount of data through digital means. Because the data is created across multiple platforms, it is tough for businesses to keep track of it and respond effectively. Natural language processing (NLP) solutions give businesses a reliable way to keep track of a large amount of user-generated material across several digital channels. Companies can actively engage customers by replying to their queries or complaints using NLP techniques. NLP tools also assist businesses in installing various content filters, such as filtering undesired information or implementing spam filters on websites or social media pages. Businesses utilize NLP models to keep the quality of information on forums up to par. There will be no harmful backlinks or unsolicited advertising as a result of this. It can also auto-filter user-generated information that could be considered as hate speech by detecting “banned” words or vulgarism.NLP has become popular as a result of big data technologies, as this approach can now be used to handle enormous volumes of text data at unprecedented speeds using cloud/distributed computing. Machines can now interpret more language-based data than humans can, without becoming fatigued and in a consistent and unbiased manner. [1].

Recent strides and research interest in the field of Natural Language Processing have made this area one of the most interesting, contemporary and important topics for study. The onset of big data and its collection, processing etc. have made it impossible for man to do all those by themselves. To put things in perspective, 1 Exabyte (10^18) of data is created on the internet every day, which is roughly the same as 250 million DVDs worth of data. Humankind generates the same amount of data in two days as it did from the origin of civilization until 2003. All these factors and points prove that NLP is going to guide and help us shape the future. The ongoing covid-19 pandemic has brought many challenges in various industries such as Healthcare, Technology, Retail etc. With these challenges comes opportunities to implement newer technologies to combat it. We shall explore these cases with special emphasis in the healthcare industry as a case study in this review research paper.

The objective of this study will be to explore how Natural Language Processing started being used in Sentiment Analysis. From simple chatbots to clinical diagnosis of patients suffering from psychological disorders, we shall see the flexibility of Natural Language Processing used in gauging emotions and sentiments. We shall also try to gauge the future and relevance of this topic while reviewing the work done towards it. We shall also see how mountains of user generated data can be sifted through easier via NLP to understand user review about a product or service by analyzing the sentiments in order to improve the same.




Reference: "What happens online in 60 seconds? | Smart Insights," *Smart Insights*, Mar. 30, 2017. https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/.

## INCEPTION

Syntactic Structures, written by Noam Chomsky, was published in 1957. He revolutionized existing linguistic principles in it, stating that the sentence structure must be changed for a computer to grasp a language. Chomsky developed a type of grammar known as Phase-Structure Grammar to achieve this objective, which painstakingly transformed natural English phrases into a structure that computers could understand. The overarching goal was to construct an artificial intelligence (AI) machine capable of thinking and communicating like a human brain.[2]

Natural Language Processing was given a fresh lease on life when the idea and need for Artificial Intelligence surfaced in the 1960s. LUNAR was founded in 1978 by W.A. Woods with the objective of studying, comparing, and interpreting chemical data on lunar rock and soil composition that had accumulated as a result of the Apollo moon landings, as well as answering the associated question. In the 1980s, computational grammar became a hot topic of study, and it was linked to the science of reasoning for meaning and taking into consideration the user's preferences and intents. In the 1990s, the rate of expansion of NLP quickened with the availability of parsers, grammars, tools, and practical materials related to NLP.[3].

In the 1980s, computational grammar became a hot topic of study, and it was linked to the science of reasoning for meaning and taking into consideration the user's preferences and intents. In the 1990s, the rate of expansion of NLP/MT quickened. Parsers, grammars, tools, and practical resources related to NLP/MT become available as a result of the parsers.

The history of natural language processing would be incomplete without including ELIZA, a chatbot programbuilt at MIT's Artificial Intelligence Laboratory between 1964 and1966.Joseph Weizenbaum was the one who came up with the idea. The analysis of key words is the first step in Eliza's input processing. Key words are significant, and scripts provide rankings to them. If key words are present, they are sorted by rank, and the input sentence is processed according to the rule associated with the key word with the highest rank.

Eliza uses a transformation rule that consists of two parts: breakdown and reassembly, after recognizing important words. The decomposition rule divides the input into chunks based on key words and applies the appropriate transformations. Then, when reassembling the sentence, the decomposition rule creates reassembly rules to follow. This rule rearranges the pieces of the input sentence formed by the decomposition rule in a suitable format. Overall, Eliza employs a set of rules to generate a response from user input.[4]. Because sentiment analysis is a relatively young topic of study, various terms have been invented to characterize it. In sentiment analysis, the words opinion mining, subjectivity analysis, review mining, and assessment extraction are all interchangeable. The purpose of subjectivity analysis, also known as subjectivity categorization, is to figure out if a statement or document expresses the author's thoughts or sentiments or only delivers facts. The majority of articles that use the phrase "sentiment analysis" focus on how to classify evaluations favorably or negatively.

## MORDERN DAY USE AND APPLICATIONS

Nowadays, everyone wants the machine to communicate, and Natural Language Processing is the only means for a computer to do so (NLP). Take, for example, Amazon's Alexa, a talking assistant. It receives an inquiry through the medium of voice and can respond through the same media, i.e., voice. It can be used to ask questions, search for information, play music, and even book a cab. The technique for Natural Language Processing (NLP) in Alexa is demonstrated in this basic figure. Alexa is only one example; these talking computers, also known as Chatbots, can handle even the most complex interactions and operations connected to streamlined business using simply NLP. Previously, chatbots were only used for customer interaction and had limited conversational capabilities due to their rule-based nature. However, with the advent of Natural Language Processing and its integration with Machine Learning and Deep Learning, chatbots can now handle a wide range of tasks, including Human Resources and Health. [5][6]

Understanding customers has long been a priority for companies. However, all too frequently, these businesses fall into the trap of making product selections based on archaic marketing strategies like the consumer mood index. The consumer sentiment index provides useful information about macroeconomic conditions, but it does not assist in addressing consumer demand. Furthermore, conventional wisdom holds that offering a diverse range of products both online and in-store is the best strategy for improving the chances that customers will find ones they enjoy. Unfortunately, this strategy does not always work because many customers are perplexed by all of the options. To obtain information about customer demand, sentiment analysis uses natural language processing (NLP), text mining, and data mining capabilities. NLP is a branch of machine learning and artificial intelligence that enables machines to read and understand human language. It develops a means to evaluate millions of postings that are uploaded every day to social media sites like Facebook and Twitter because it is done by machines. Sentiment analysis examines the text for positive, negative, or neutral sentiment. It can also consider the meaning of the words as well as the context in which they were spoken.[7]

Customers are increasingly expressing their needs, ideas, preferences, and complaints online, which is why sentiment analysis is becoming popular among retailers and technology companies. Businesses can use sentiment analysis and natural language processing to learn what their consumers are saying and how they feel about the products they're selling. Retailers and IT businesses may now use social media to understand their consumers' mood in real time, learning how they feel about things on store shelves, store layouts, and ads. Furthermore, sentiment analysis may inform organizations about how strongly customers feel about a product and which aspects are responsible for those feelings. Technology companies may be able to turn their fortunes around in the future by incorporating sentiment analysis into their product releases. With real-time data, the merchandising team could make adjustments to rekindle public interest in the company's products.

## FUTURE AND APPLICATIONS IN WEB 2.0

Future opinion-mining systems will necessitate a larger and more diverse knowledge base of commonplace and commonsense data. More comprehensive knowledge must be linked to reasoning techniques impacted more deeply by human psychology and thought. Researchers will be able to better understand natural language opinions as a result of this, and the gap between (unstructured) multimodal data and (structured) machine-processable data will be bridged. Combining scientific theories of emotion with the practical engineering goals of analyzing sentiments in natural language text will result in more bio inspired approaches to the design of intelligent opinion-mining systems capable of dealing with semantic knowledge, making analogies, learning new affective knowledge, and detecting, perceiving, and "feeling" emotions. The capacity to tap into public emotion on social media is becoming more widely recognized as a valuable tool for market research, consumer segmentation, and stock price forecasting in strategic marketing planning and maneuvering. The healthy growth of the big data framework has fueled this evolution of technology adoption, causing applications based on Sentiment Analysis (SA) in big data to become commonplace for enterprises. The term "Web 2.0" is commonly used to denote a new way of thinking about data and data interactions on the internet. This evolution affects both technologies and practices.

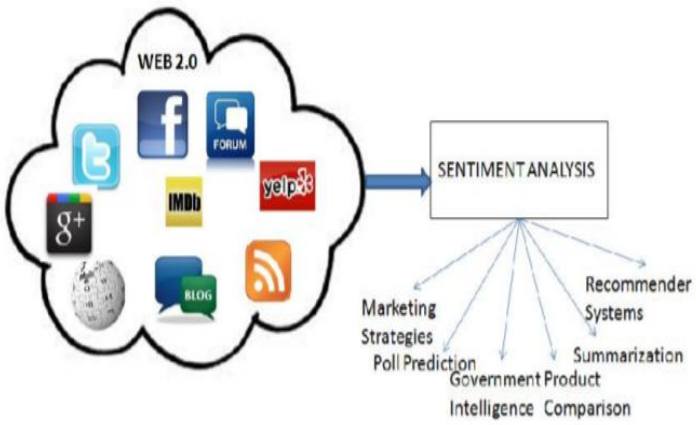
 

This progress may be seen in the way users engage with one another by acting directly on the website's content (for example, by adding tags to ease future search/retrieval/classification). As a result, Web 2.0 exemplifies community, dynamicity, and web interactions. As a result, websites are more reliant on tools such as wikis and blogs. [8][9]

The area of sentiment analysis study has been rapidly expanding as a result of the huge and diverse data given by Web 2.0 applications. The data used in sentiment analysis comes from blogs, review sites, forums, microblogging sites, wikis, and social media sites. A review site is a website where users can post unfavorable comments about people, businesses, products, and services. The majority of sentiment analysis is done on movie and product review sites. Because the purpose of a review is to evaluate a single object, it is a single domain problem. Sentiment research on review sites can help manufacturers and future purchasers of a product. Manufacturers can use reviews to determine how well a product is received. They can deduce which characteristics the reviewers liked and disliked.[6]. The process of extracting meaningful conclusions is a tricky and long process. Data collection is the first step. One of the most crucial elements in the sentiment analysis process is this. Everything will depend on the quality of the data gathered and how it has been annotated or labeled from here on out. Live APIs for social media can be used to upload data. A news API can help you gather data from a variety of news sources, whereas a Facebook API can provide you access to all of the platform's publicly available data. Open-source repositories such as Kaggle, Amazon reviews, and Yelp can also be used. If you have data that you already have from a CRM application, you can manually upload that onto the sentiment analysis API as a .csv file. Customer relationship management (CRM) is a software program that helps you to keep track of all of your company's customer and prospect contacts and interactions. The second step is to analyze the data. The type of information in the data will determine how it is processed: text, image, video, or audio.

Audio transcription - To ensure that any video or audio file (e.g., podcast) in the data is not disregarded, the audio from the video data is transcribed using speech to text software.
Caption overlay - Any subtitles that appear in the video are evaluated for any appearing entities, aspects, or subjects that you have flagged as important.
Image overlay - Using OCR, the software can recognize and collect any images in a video or text data (optical character recognition)

Text extraction- During the sentiment analysis process, all text is recognized and extracted in the same way. Emojis and hashtags are also included, as they are an important aspect of social media sentiment research. Emojis must never be excluded from data processing because this could result in false positives or negatives. The third step would be to analyze the dataThere are several subtasks that need to be done for this stage of the sentiment analysis process. The dataset that will be used to train the model must be pre-processed and manually labeled before it can be used. By comparing the correctly categorized data with the wrongly classified data, this labelled data will be utilized to train the model. This will aid in the development of a brand's bespoke model. Custom Tags - During this stage of the process, custom tags for aspects and themes such as brand mentions, product names, and so on will be created for the data. Once the model has been trained, it will automatically partition text depending on these specially constructed tags. Topic Classification - A text's theme is assigned by the topic classifier. This sentence, for example, will be labeled with the topic "clothing" since "the dresses were wonderful, and I discovered some pretty excellent scarves as well." Sentiment Analysis - At this point, the platform isolates each aspect and theme, which is subsequently analyzed for sentiment. The range of sentiment scores is from -1 to +1. Zero can be used to describe a neutral statement.




Reference: A. Kumar and T. M. Sebastian, “Sentiment analysis: A perspective on its past, present and future,” *Int. j. intell. syst. appl.*, vol. 4, no. 10, pp. 1–14, 2012.

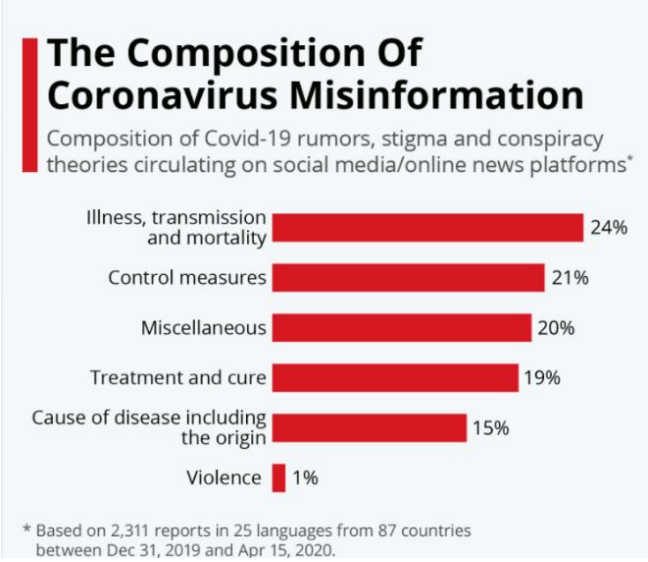
 

This polarity assignment is significant because, while the platform awards individual scores to different characteristics such as convenience, speed, cleanliness, functionality, drinks, ambience, and so on, it is the aggregate score that is used to determine the audience's sentiment toward the brand. So, if three of the seven elements are rated poorly (-.65) and four are rated positively (+.5), the sentiment analyzer will assign an average score to the brand's overall sentiment. The fourth and final step would be to visualize the data. The insights are immediately translated into actionable reports in the form of graphs and charts when all of the steps in the sentiment analysis process have been completed. These reports can then be distributed among teams. These visual reports are extremely significant since they allow the organization to see granular, aspect-based outcomes. When the organization get an average score for their brand or service, they can use the sentiment analysis dashboard to filter the data to see which features received high scores and which received low scores. This will help them determine which areas require more attention than others. As a result, one gains actionable information at this point of the sentiment analysis steps, which they can utilize to determine the best course of action for a companies or service's growth ambitions.[10][11][12]

## CASE STUDY: COVID -19 AND NATURAL LANGUAGE PROCESSING

The COVID-19 pandemic has had a tremendous impact on society, both as a result of COVID-19's severe health impacts and as a result of public health measures taken to slow its spread. Many of these issues stem from information needs, and attempts to fulfill these needs have resulted in an information overload for both scholars and the general public. Many of the information demands made urgent by the COVID-19 epidemic can be addressed using natural language processing (NLP), a branch of artificial intelligence that translates human language.

The elderly, those with impairments, and those with numerous comorbidities are among the susceptible categories who have been identified as being at a greater risk of getting more severe and life-threatening COVID-19 infections. The COVID-19 pandemic led healthcare providers and systems around the world to curtail or limit less-urgent treatments, such as rehabilitation for persons with acute and chronic diseases and disorders. For some patients, the treatment delay is annoying but not life-threatening. A delay or interruption in treatment can have a substantial impact on recovery and efficacy for other patients. Deferring rehabilitation therapy is undesirable since it leads to poorer physical and psychological consequences for patients, as well as increasing the strain on the healthcare system in the future to deal with the mounting backlog. Rehabilitation has become increasingly important during the COVID-19 pandemic. Rehabilitation is needed to meet the needs of people with acute and chronic illnesses, as well as to help people recover from severe COVID-19 infections that require long-term intensive care and respiratory support. Rehabilitation for post-COVID patients has been demonstrated to be costly to healthcare systems, with the average cost of post-COVID patients' rehabilitation services being nearly twice that of non-COVID patients' rehabilitation services.




Reference: "The Composition of Coronavirus Misinformation," *Statista Infographics*. https://www.statista.com/chart/22527/composition-of-covid-19-misinformation/.

Artificial intelligence (AI) and machine learning (ML) approaches can be used to support healthcare professionals and healthcare delivery at this moment, when healthcare resources are being stressed because to the epidemic. In-person meetings with doctors, occupational therapists, and physiotherapists can be supplemented with these AI/ML technologies.

A case-study summary of our work on an AI/ML and Natural Language Processing (NLP) system for a telephone-based Rehabilitation Advice Line will also be presented as an example of such a system. These technologies can also provide continuity of care when in-person appointments are too risky due to subsequent waves of the COVID-19 epidemic. Furthermore, beyond the pandemic's urgent needs, the implementation of these devices will continue to be beneficial for providing care to isolated and rural populations.[13] A misinformation epidemic accompanied the COVID-19 health disaster. Due to the lack of knowledge about the virus's nature and origins, conspiracy theories arose, which were widely disseminated on YouTube and other online social media platforms. Efforts to detect and filter relevant falsehoods were expedited by YouTube, but the results were mixed. [14]. People have expressed a variety of emotions on social media sites such as Twitter throughout the surge of COVID-19 instances and tougher lockdowns. During COVID-19, social media played a crucial role in motivating academics to use NLP and machine learning technologies to analyze data. Deep learning sentiment analysis revealed that World Health Organization Tweets failed to provide public guidance. Sentiment analysis was used to investigate the impact of a nationwide lockdown in India due to the COVID-19 outbreak, and it was discovered that people in India viewed the fight against COVID19 positively, with the majority agreeing with the government's decision to announce the lockdown in order to flatten the curve. The sheer number of tweets enable researchers and policy makers to observe the impact and result of steps taken to curb the virus. [15]

## REFERENCES


1. "The basics of NLP and real time sentiment analysis with open-source tools," *www.linkedin.com*. https://www.linkedin.com/pulse/basics-nlp-real-time-sentiment-analysis-open-source-tools-%C3%B6zg%C3%BCr-gen%C3%A7 (accessed Sep. 23, 2021).
2. "NLP - overview," *cs.stanford.edu*. https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html (accessed Sep. 23, 2021).
3. Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/247926251_The_Lunar_Science_Natural_Language_Information_System_Final_Report. [Accessed: 23-Sep-2021].
4. Researchgate.net. [Online]. Available: https://www.researchgate.net/publication/221152425_A_Multi-modal_Eliza_Using_Natural_Language_Processing_and_Emotion_Recognition. [Accessed: 25-Sep-2021].
5. B. Lutkevich and E. Burns, "What is natural language processing? An introduction to NLP," Techtarget.com, 02-Mar-2021. [Online]. Available: https://searchenterpriseai.techtarget.com/definition/natural-language-processing-NLP. [Accessed: 27-Sep-2021].
6. X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, Jun. 2015, Doi: 10.1186/s40537-015-0015-2.
7. "(PDF) Sentiment analysis using product review data," ResearchGate. https://www.researchgate.net/publication/316666447_Sentiment_analysis_using_product_review_data.
8. P. Durville and F. Gandon, "Filling the Gap between Web 2.0 Technologies and Natural Language Processing Pipelines with Semantic Web," IEEE Xplore, Oct. 01, 2009. https://ieeexplore.ieee.org/document/5291527 (accessed Oct. 13, 2021).
9. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?" Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02, 2002, Doi: 10.3115/1118693.1118704.
10. Á. García-Crespo, R. Colomo-Palacios, M. Mencke, and J. M. Gómez-Berbís, "CUSENT: Social Sentiment Analysis Using Semantics for Customer Feedback," Social Web Evolution: Integrating Semantic Applications and Web 2.0 Technologies, 2009. https://www.igi-global.com/chapter/social-web-evolution/29290 (accessed Oct. 14, 2021).
11. J. A. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, and E. Herrera-Viedma, "Carrying out consensual Group Decision Making processes under social networks using sentiment analysis over comparative expressions," Knowledge-Based Systems, vol. 165, pp. 335–345, Feb. 2019, Doi: 10.1016/j.knosys.2018.12.006.
12. K. Unnamalai, "Sentiment Analysis of Products Using Web," Procedia Engineering, vol. 38, pp. 2257–2262, 2012, Doi: 10.1016/j.proeng.2012.06.272.
13. J. Carriere *et al.*, "Case Report: Utilizing AI and NLP to Assist with Healthcare and Rehabilitation During the COVID-19 Pandemic," *Frontiers in Artificial Intelligence*, vol. 4, Feb. 2021, Doi: 10.3389/frai.2021.613637.
14. H. O.-Y. Li, A. Bailey, D. Huynh, and J. Chan, "YouTube as a source of information on COVID-19: a pandemic of misinformation?" *BMJ Global Health*, vol. 5, no. 5, p. e002604, May 2020, Doi: 10.1136/bmjgh-2020-002604.
15. A.H. Alamoodi et al., "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," Expert Systems with Applications, Oct. 2020, Doi: 10.1016/j.eswa.2020.114155.