Ritwik Raj Saxena[1]

[1]Graduate Teaching Assistant, Department of Physics and Astronomy; Research Advisee, Department of Computer Science, University of Minnesota-Duluth

April 09, 2024

Posted on 9 Apr 2024 — CC-BY 4.0 — https://doi.org/10.36227/techrxiv.171262869.98334885/v1 — e-Prints posted on TechRxiv are preliminary reports that are not peer reviewed. They should not b...

Abstract

Objective: The exponential growth of digital social platforms has not only connected individuals globally but has also provided a platform for users to freely express their experiences and viewpoints on topics spanning from consumer products and services to broader societal matters such as political issues. Within this expansive digital discourse, in the recent years, one notably discussed subject has been the SARS-CoV-2 (COVID-19) vaccines. In this article, our focus is on investigating the profound impact of neural networks in the analysis of sentiments expressed by people concerning the introduction and utilization of these vaccines. Background: Sentiment analysis, a critical facet of natural language processing (NLP), is replete with intricate associations in the linguistic landscape. Within its realms, many sophisticated methodologies, such as machine learning algorithms, including neural network architectures, are employed to decipher the intricate web of semantic relationships embedded in textual data, which include, but are not limited to, social media posts. From gathering business intelligence, to market research and competitor analysis, examining sentiments has found many practical uses. In the domain of COVID-19 vaccines, sentiment analysis has provided valuable insights into vaccine hesitancy, vaccine adoption rates, and public trust in the governmental setup and in the pharmaceutical industry. Methods: A systematic literature review (meta-analysis) was carried out to quarry scientific research on neural network-based analysis of sentiments about COVID-19 vaccines. Implementing a thorough search strategy, we isolated relevant articles and methodically examined them to discern key insights that contributed to our comprehension of the utility of neural networks in analyzing public opinion regarding COVID-19 vaccines. Conclusion: Our study provides insights affirming that neural networks have shown a surpassing capacity to discern intricate patterns within vast textual datasets. Their inherent ability to capture contextual nuances in language has enabled a nuanced understanding of diverse sentiments about COVID-19 vaccines. This has helped formulate strategies to alleviate negative sentiments about the vaccines leading to higher vaccine acceptance rates and management of the pandemic.

Examining Reactions about COVID-19 Vaccines: A Systematic Review of Studies Utilizing Deep Learning for Sentiment Analysis

Ritwik Raj Saxena[a]

a. Graduate Teaching Assistant, Department of Physics and Astronomy; Research Advisee, Department of Computer Science, University of Minnesota-Duluth.

Corresponding author: Ritwik Raj Saxena

Contents

Abstract

Objective: The exponential growth of digital social platforms has not only connected individuals globally but has also provided a platform for users to freely express their experiences and

viewpoints on topics spanning from consumer products and services to broader societal matters such as political issues. Within this expansive digital discourse, in the recent years, one notably discussed subject has been the SARS-CoV-2 (COVID-19) vaccines. In this article, our focus is on investigating the profound impact of neural networks in the analysis of sentiments expressed by people concerning the introduction and utilization of these vaccines.

Background: Sentiment analysis, a critical facet of natural language processing (NLP), is replete with intricate associations in the linguistic landscape. Within its realms, many sophisticated methodologies, such as machine learning algorithms, including neural network architectures, are employed to decipher the intricate web of semantic relationships embedded in textual data, which include, but are not limited to, social media posts. From gathering business intelligence, to market research and competitor analysis, examining sentiments has found many practical uses. In the domain of COVID-19 vaccines, sentiment analysis has provided valuable insights into vaccine hesitancy, vaccine adoption rates, and public trust in the governmental setup and in the pharmaceutical industry.

Methods: A systematic literature review (meta-analysis) was carried out to quarry scientific research on neural network-based analysis of sentiments about COVID-19 vaccines. Implementing a thorough search strategy, we isolated relevant articles and methodically examined them to discern key insights that contributed to our comprehension of the utility of neural networks in analyzing public opinion regarding COVID-19 vaccines.

Conclusion: Our study provides insights affirming that neural networks have shown a surpassing capacity to discern intricate patterns within vast textual datasets. Their inherent ability to capture contextual nuances in language has enabled a nuanced understanding of diverse sentiments about COVID-19 vaccines. This has helped formulate strategies to alleviate negative sentiments about the vaccines leading to higher vaccine acceptance rates and management of the pandemic.

Keywords: Sentiment Analysis, Natural Language Processing, Deep Learning, Machine Learning, Neural Networks, Support Vector Machines, Tokenization, Lemmatization, Segmentation, Stemming, Word2Vec, Bag of Words, SMOTE, COVID-19, Vaccine, Neutrosophic Set, Grey Data, Fuzzy Set Theory, Twitter, social media, GPT, Transformer, Recurrent Neural Networks, Logistic Regression, Stochastic Gradient

1. Introduction

“Words cut deeper than knives. A knife can be pulled out, words are embedded into our souls.” This quote, from William Chapman, underlines the importance of the impact of communication, that can be verbal, written, or virtual, over that of the impact which violent action has on bringing a possible point home. Sentiment analysis studies non-factual information, most often opinions and beliefs, that is inherent in communication.

Sentiment analysis, or opinion mining, implies the utilization of psychological, social, linguistic, computational, or semantic techniques to understand the feelings, viewpoints, or mindset of a person or a group of individuals towards a specific target entity, which could be a subject, a person, a place, an organization, a product, a service, a concept, a sociopolitical issue, or other similar

objects. Sentiment analysis is used to decide if a verbal or spoken excerpt voices positive, neutral, or negative sentiments. It is employed where the need is to mine propensity data from virtual content, including online messages, social media posts, blogs, news tidbits, and product and service reviews. Both government and private organizations then use this information to create services, designs, products, and policies that are more appealing to their target demographics, unleashing new opportunities to gain popularity and profit. It can also be used by political parties to field candidates that are popularly seen as more favorable in order to win polls. Sentiment analysis methods can be used as opinion polls and exit polls.

Sentiment analysis can be used to detect whether a piece of text, that is, a sentence, a paragraph, or a document, is subjective or objective. While subjective pieces of texts display viewpoints, objective pieces of texts revolve around factual statements. In this regard, it must be kept in mind that the implication of a word or a count is dependent on its usage. The more common use of sentiment analysis is to classify the positivity, negativity, or neutrality of sentiment of a piece of text. This is called polarity classification.

However, many pieces of text may be built out of a combination of negative and positive opinions. Under aspect-based sentiment analysis, text is classified by aspects to ascertain the sentiment of each aspect (Chakraborty *et al.*, 2023). The opinions of a person on a composite entity is commonly built of opinions, each of its own polarity, towards each aspect of that entity. The weighted sum of the sentiment of each aspect or feature gives us the overall sentiment of a piece of text, usually a document, tweet, review, or post. Aspect-based sentiment analysis, among many of its uses, helps in learning the overall reaction of consumers to a product, service, or an idea that is a potential a product or a service. In the realm of sentiment detection, NLP involves application of statistical methods and computational linguistics for text analysis.

Many studies have underscored the increasing importance of sentiment analysis in the contemporary business landscape, fueled by the rise of social media platforms, blogs, and review sites. Sentiment analysis has been portrayed as a valuable tool for companies seeking to navigate the digital realm. The ease with which consumers can share their opinions across the web has elevated online sentiments to a valuable resource for businesses. Sentiment analysis emerges as a valuable tool in this context, offering a systematic approach to distill meaningful information from the sea of online opinions, whose immensity poses a significant challenge for companies trying to extract valuable insights from consumers' comments. By employing sentiment analysis, organizations can sift through the noise and gain a deeper understanding of consumer conversations, enabling them to take more effective and better-targeted actions.

Organizations can track social media mentions and web references to competitors, analyzing consumer sentiments to gain a competitive edge in the market. This is termed as competitive intelligence mining. Sentiment analysis can be used to develop effective communication strategies in the realm of public relations. Sentiment analysis can be used to sequester sales leads, that is, potential clients. Information of industry trends gained through sentiment analysis can help align a company’s policies towards those products and services that may prove to be the most revenue-churning in the coming future. Sentiment analysis also helps companies in discovering influencers

with positive sentiments toward their products. These influencers can prove to be valuable assets for strategic public relations campaigns.

There are challenges and complexities associated with the implementation of sentiment analysis methodologies. Sentiment analysis is often depicted as a panacea for extracting valuable insights from the vast pool of online opinions. However, in many instances, sentiment analysis algorithms are not foolproof and can struggle with context, sarcasm, and evolving linguistic nuances. An algorithm may misinterpret a sarcastic comment as a positive sentiment, leading to inaccurate conclusions. Rhetorical statements like "I love how this product destroys everything" might be incorrectly interpreted as a positive sentiment due to the presence of the word "love" without considering the subsequent negative context. There is inherent subjectivity and cultural nuances that make sentiment analysis a complex task. biases introduced by cultural and linguistic diversity. Sentiment analysis algorithms may struggle to accurately interpret sentiments across different languages and cultural contexts, potentially leading to skewed results.

The indiscriminate use of sentiment analysis can raise privacy concerns, as organizations may be analyzing and interpreting users' sentiments without their explicit consent. This necessitates a transparent and ethical framework for the implementation of sentiment analysis, including obtaining informed consent from users. There is potential for bias in sentiment analysis algorithms, especially when dealing with diverse language patterns and cultural contexts. Incorporating machine learning models that can adapt to changing linguistic patterns and context enhances accuracy of sentiment analysis tasks.

1.1. A Bird’s Eye View of Sentiment Analysis

Sentiment analysis encompasses various tasks, approaches, and analyses, requiring a holistic perspective beyond mere categorization.

Cambria *et al.* (2017) proposed a three-layer structure with fifteen NLP problems (recorded and summarized by Ligthart et al., 2021):

1. Syntactics layer: Involving microtext normalization, sentence boundary disambiguation, part-of-speech (POS) tagging, text chunking, and lemmatization.
2. Semantics layer: Encompassing word sense disambiguation, concept extraction, named entity recognition, anaphora resolution, and subjectivity detection.
3. Pragmatics layer: Covering personality recognition, sarcasm detection, metaphor understanding, aspect extraction, and polarity detection.

1.1.1. Tasks in Sentiment Analysis

A. Subjectivity Classification: This task determines the presence and extent of subjectivity in the data. To classify the subjectivity of a piece of text, we determine whether it expresses subjective or objective content. Machine learning models, such as Support Vector Machines (SVM) or deep learning models like Bidirectional Long Short-Term Memory (LSTMs), which are trained on annotated datasets, are applied for subjectivity classification.

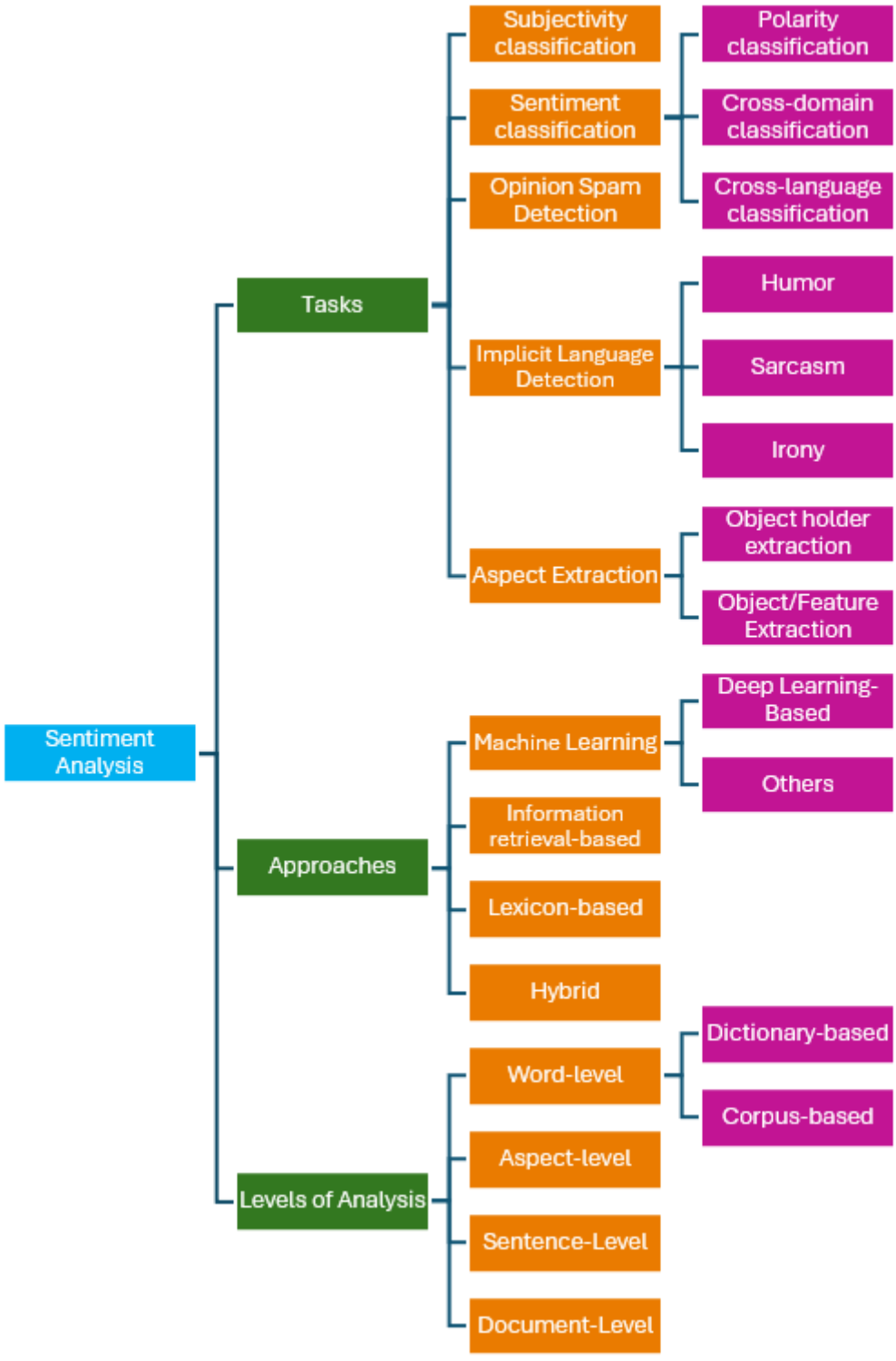


Figure 1: A summary of tasks, approaches, and levels of analysis in Sentiment analysis (borrowed/adapted from Ligthart *et al.*, 2021 and Kumar and Sebastian, 2012)

In the context of subjectivity classification and also otherwise, feature engineering usually involves selecting and crafting features that encapsulate syntactic patterns (which are

exemplified through syntactic layer tasks like n-gramming, POS tagging, microtext normalization, sentence boundary disambiguation, text chunking, lemmatization, etc.) and semantic patterns (which are exemplified through semantic layer tasks like semantic role labeling, word sense disambiguation, concept extraction, named entity recognition, anaphora resolution, creating word embeddings, and subjectivity detection, etc.) associated with subjective language.

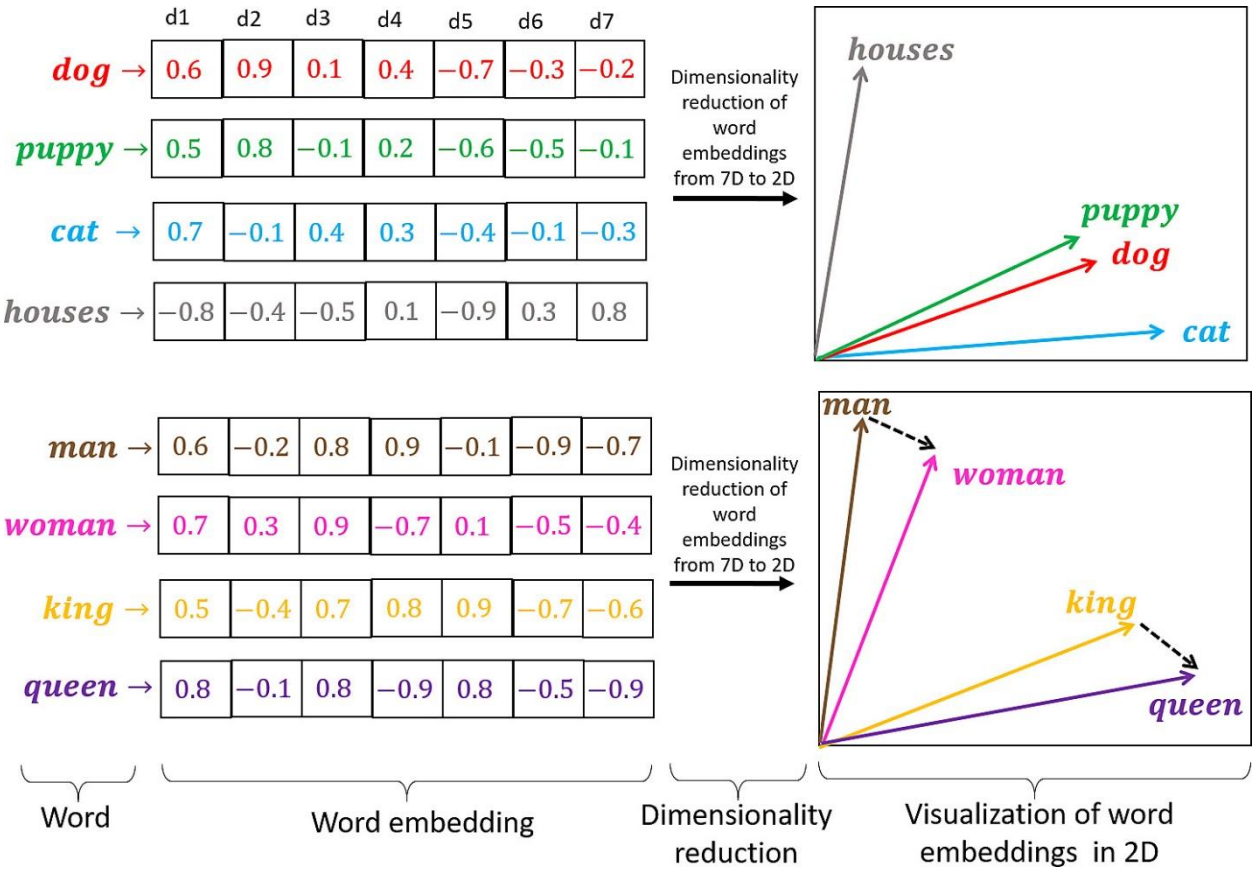


Figure 2: A diagrammatic representation of word embeddings. Here, d1, d2, and so on, represent the dimensions. Word embeddings, reduced to two-dimensional representations, are termed vectors (borrowed/adapted from Rozado, 2020).

B. Sentiment Classification: In this task, the sentiment expressed in a piece of text is identified. Advanced deep learning models, such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT), which are pre-trained on vast corpora of text and fine-tuned on sentiment-specific datasets, are applied. Attention mechanisms are used to capture nuances and dependencies within sentences. According to Ligthart *et al.*, polarity detection is a subtask under sentiment classification. Cross-domain and cross-language classification are specialized tasks within sentiment classification that focus on transferring knowledge from a data-rich source domain or language to a target domain or language where data and labels are scarce. Cross-domain classification involves detecting the sentiment of a target domain with limited data using a model that is trained on a source domain

which has an ample corpus of data on which the model can be trained. Cross-language classification, on the other hand, entails training a sentiment classification model on a dataset in the source language and applying it to a target (data-deficient) language. Both these subtasks involve knowledge transfer mechanisms (Ligthart *et al.*, 2021).

C. Opinion Spam Detection: Opinion spams are logically composed bogus reviews that aim to advertise or defame a product, a service, or a market player. They can be detected through methods like author profile analysis and detection of extreme sentiments. Examining the history of the author's reviews can reveal certain patterns, for example, frequent posting of similar opinions or sudden bursts of activity are usually anomalous. Further, opinion spamming accounts usually have incomplete and suspicious profiles. Opinions with overly positive or negative sentiments, without providing substantial details, can be also flagged as potential spam. Moreover, rapidly generated reviews or a high volume of opinions within a short time frame may indicate automated spamming. Machine learning models incorporating features like lexical diversity, sentiment consistency, and user behavior patterns are often explored to detect opinion spam. Ensemble models with various classifiers are also applied for enhanced performance.
D. Implicit Language Detection: Implicit language is characterized by the presence of elements like humor, sarcasm, irony, similes, and metaphors, and involves expressions that carry ambiguity in their meaning. These nuances can pose challenges in detection. Context-aware embeddings like Embeddings from Language Models (ELMo) and contextualized transformers like BERT (a pre-trained NLP model introduced by Google) are used to capture such subtle nuances. During studies involving implicit language detection, deep learning models are trained in a manner that they can identify implicit language patterns.
E. Aspect Extraction: Referred to as complimentary tasks by Kumar and Sebastian (2012), aspect extraction involves object holder extraction and object extraction or feature extraction. Object implies the sentiment, object holder implies the (target) entity holds the opinion that displays the mentioned object or sentiment, and feature implies the features of the opinion. Sequence labeling models like Conditional Random Fields (CRF) and Bidirectional Long Short-Term Memory with Conditional Random Fields (BiLSTM-CRF) are often implemented to identify and classify specific aspects within the text. Domain-specific embeddings are used to enhance performance on specialized corpora. Named Entity Recognition (NER) models with fine-tuning on datasets specific to object holder identification and co-reference resolution techniques which link pronouns to identified entities are applied for object holder extraction. Models with attention mechanisms are deployed to extract relevant features. Transfer learning on pre-trained language models are used to improve feature extraction.

1.1.2. Approaches Utilized for Sentiment Analysis

A. Machine Learning Approaches: As to traditional machine learning approaches, Random Forests and SVM with carefully engineered features have been frequently used for various NLP tasks with promising results. Hyperparameter tuning and feature importance analysis is commonly employed for model refinement. Deep learning or neural networks-based approaches have usually utilized transformer-based architectures (BERT, GPT) for tasks demanding contextual understanding. Pre-trained models are fine-tuned on task-specific

datasets, adjusting learning rates and optimization strategies. In this context, approaches in NLP that are not machine learning-based are called rule-based approaches.

B. Information Retrieval-Based Approaches: Such approaches are primarily centered around retrieving relevant information from large datasets or corpora. The goal is to find documents and passages that are most relevant to the query. Web scraping is used to retrieve information from the web. N-grams are used in search engines to improve the relevance of search results by considering sequences of words.

C. Lexicon-Based Approaches or Knowledge-based approaches: Lexicons, which are tailored to specific domains and tasks, are developed using word embeddings and statistical methods. Such approaches involve encoding explicit knowledge about a domain, often in the form of ontologies, taxonomies, or structured databases like lexicons. Lexicon-based models examine documents for words that express emotions. Lexicons are often integrated into rule-based systems for efficient analysis.

D. Hybrid Approaches: They primarily combine machine learning-based approaches with lexicon-based approaches. They incorporate pre-trained language models with rule-based systems. They are able to leverage the strengths of both approaches. They often utilize ensemble methods with models from different paradigms for robust predictions.

1.1.3. Levels of Analysis:

1. Word-Level analysis**:** To classify words based on their sentiment orientation of words and phrases is one of the primary tasks of sentiment analysis. This is so because most sentiment analysis studies employ the prior polarity of words and phrases for sentiment classification at higher levels. Various linguistic, morphological, and, like tokenization, lemmatization, stemming, and segmentation are utilized for word-level sentiment analysis. Word embedding models like Word2Vec, ELMo (Embeddings from Language Model) and Global Vectors for Word Representation (GloVe) are deployed for semantic understanding, while POS tagging and syntactic parsing are applied for fine-grained analysis. LLM-based models (like BERT, GPT and LLaMA) are also available for generating word embeddings, especially in conjunctions with frameworks like Langchain, which streamline the renderings of LLM applications.

    1.1. Dictionary-Based analysis: In the dictionary-based approach for sentiment analysis, we begin with a limited group of words that express different sentiments. These words are collectively referred to as a seed list. As we analyze more text, we keep adding similar and opposite words to the list from existing dictionaries like WordNet and SentiWordNet (Ligthart *et al.*, 2021; Kumar and Sebastian, 2012). Thus, a sentiment dictionary is forged. Statistical methods and domain-specific expertise are demanded in this task. Sentiment dictionaries are incorporated into the NLP pipeline for efficient word-level analysis.

    1.2. Corpus-Based analysis: Corpus-based methods in sentiment analysis depend on syntactic or statistical techniques. These techniques involve looking at how often a

word appears alongside another word whose sentiment or polarity is already known. By analyzing these co-occurrences in a large collection of texts (corpus), the method tries to understand the sentiment of words based on their associations with other words (Kumar and Sebastian, 2012). Statistical methods like TF-IDF (term frequency–inverse document frequency) are applied in a corpus-based analysis. Co-occurrence matrices, each of whose rows and columns correspond to words and each of whose cells contain the score representing how often the words occur together, are employed for semantic analysis and feature (information) extraction. Topic modeling techniques, for example Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), are used for uncovering latent patterns within large corpora.

2. Aspect-Level Analysis: This involves examining sentiments related to specific objects and their features within a text. Its scope traverses beyond gauging the sentiment of an entire document. Also termed entity-level or feature-level analysis, it acknowledges that while a document may have an overall positive or negative sentiment, the opinion holder might express diverse opinions about specific aspects of an entity. Identifying these aspects is crucial for accurately measuring aspect-level opinions (Ligthart *et al*., 2021; Tun Thura Thet *et al*., 2010; Kumar and Sebastian, 2012). Researchers often implement attention mechanisms in models to focus on specific aspects within the text. Aspect-specific classifiers are trained by computer scientists for detailed analysis.

3. Sentence-Level: Sentence-level analysis focuses on individual sentences within a document or a piece of text and is particularly employed for subjectivity classification. This approach operates under the fairly reliable assumption that text documents generally comprise sentences that either express opinions or do not, that is, they are either subjective or objective (Ligthart *et al*., 2021). Typically, pre-trained language models, which consider contextual dependencies, are used for sentence-level tasks. Hierarchical models and attention mechanisms are also explored for capturing subjectivity of and relationships between sentences.

4. Document-Level Analysis: In this, the entire document is taken as the basic unit whose sentiment orientation is to be computed. Document embeddings like Doc2Vec and BERT's context-aware embeddings are applied for holistic document understanding. Doc2Vec, also known as Paragraph Vectors, is an unsupervised algorithm that learns embeddings, in the form of fixed-size feature representations for variable-length pieces of text, such as sentences, paragraphs, or documents. Topic modeling and clustering techniques are employed to organize and summarize large document collections.

In data collection and processing, it is crucial that participants' opinions are captured as much as possible, and if there is any ambiguity in their response, it should be addressed. For example, fuzzy set theory, grey data, and neutrosophic set are helpful in modeling and performing computations on uncertain data.

Fuzzy set theory deals with sets in which elements have degrees of membership rather than a strict binary classification (belongs or does not belong). Membership values range between 0 and 1,

indicating the degree of belongingness of an element to the set. An element's membership degrees are expressed in terms of its degree of membership (μ) and its degree of non-membership ($\overline{\mu}$). The equation representing the relationship between these two components is:

$$\mu + \overline{\mu} = 1$$

In this manner, it is observed that sentiments in a fuzzy set are not strictly positive or negative.

Grey data refers to data that is partially known. This leads to ambiguity. Insufficient information makes it challenging to determine precise values. Sarcasm, irony, metaphors, and neutral sentiments may fall in this category.

Neutrosophic set is a generalization of the fuzzy set. It introduces a third component, indeterminacy, in addition to truth and falsity. Each element in a neutrosophic set is associated with three membership degrees: the degree of truth (T), the degree of indeterminacy (I), and the degree of falsity (F).

Mathematically, the degrees satisfy the condition (when normalized):

$$T + I + F = 1$$

Therefore, in case of a neutrosophic set, a sentiment can have a degree of positivity, a certain degree of negativity and a degree of greyness, and all these degrees sum up to one.

1.2. History of Sentiment Analysis

The importance of estimating opinions and sentiments for one's own benefit was discovered to be a powerful tool almost when Homo sapiens sapiens, the subspecies of Homo sapiens which is made of the only surviving members of genus Homo, was born. With Dunbar's (1998) powerful views on evolutionary psychology as the basis, it can be argued that human intelligence did not primarily evolve to help humans endure in adverse ecological scenarios like diseases and natural disasters, but to help the subspecies survive by engendering a rate of reproduction which was higher than their rate of death due to various causes, including hostilities. Possibly the first task a human accomplished by gauging opinions of someone else was finding and clinching a suitable mate. Mäntylä *et al.* (2018) theorize that sentiment analysis could be as old as the beginning of verbal communication. Ancient Greeks tried to ascertain opinions to find out their chances of winning a battle or those of a spy being loyal or disloyal (Thorley, 2004; Mäntylä *et al.*, 2018).

Darwin's 1872 publication 'The Expression of the Emotions in Man and Animals' is considered to be one of the forerunners of emotional research (Gendron and Barrett, 2009). Attempts to describe public opinion by estimating and quantizing it using questionnaires were first observed in 1900s (Drogba, 1931). In the late 1930s, the first scientific journal on public opinion was instituted (Public Opinion Quarterly; Mäntylä *et al.*, 2018).

Shannon, in his 1948 memorandum, 'A Mathematical Theory of Communication', introduced the concept of information entropy, which is defined as the measure of the meaningful information content in a message, and is dependent on the uncertainty reduced by that message (Shannon, 1948). The concept of Information entropy laid the basis of information theory. Shannon, through

his 1951 article ‘Prediction and Entropy of Printed English’, made a significant contribution to NLP and computational linguistics by providing a statistical foundation to language analysis (Sannon, 1951). This helped in the development of the first significant statistical language model in 1980s (Rosenfeld, 2000), giving context to statistical methods like n-gram model for sentiment analysis.

An important landmark in NLP, and, thereby, in sentiment analysis was Noam Chomsky’s work on transformational-generative grammar, a concept whose foundations were laid in 1950s (Singleton, 1974). His work influenced the development of computational linguistics by providing a theoretical framework for understanding the structure of sentences and the rules governing their transformation.

IBM collaborated with Georgetown University to develop a Russian-English machine translation setup which was publicly demonstrated in 1954. It generated hope for sophisticated automated rule-based systems, even though it was itself an experiment involving no more than 250 words and six grammar rules (Hutchins, 2004). It generated aspirations for systems capable of multiple artificial intelligence and NLP tasks, including in-context learning and sentiment analysis, highlighting the early recognition of broader applications for rule-based AI systems.

The SMART (System for the Mechanical Analysis and Retrieval of Text) Information Retrieval System, initiated at Harvard by Gerard Salton and developed at Cornell University by his group, incorporated rule-based approaches for document classification. The system focused on keyword-based retrieval and classification, relying on predefined rules to determine document relevance (Buckley *et al.*, 1993) Gerard Salton adopted the Vector Space Model for the SMART system (Salton *et al.*, 1975; Wong *et al.*, 1985), which he introduced alongside TF-IDF, which measures the importance of a word to a document and formed the basis of search engines and other systems. Some early work in information retrieval and document analysis may have involved attempts to mine opinions and subjective information from text. Possibly the earliest projects known to have explored techniques for identifying subjective language and opinions expressed in documents, thereby strengthening the groundwork for sentiment analysis concepts, were Turney’s and Pang’s, discussed later in this paper.

In the late 1970's, Charles Forgy developed the Rete algorithm at Carnegie Mellon University (Forgy, 1982). It formed the basis for many early AI tools, including NASA's CLIPS (Scott, 1994). It compared collections of patterns to collections of objects. While not specific to text classification, it became a widely used algorithm for rule-based systems. It efficiently matched rules against data patterns, including those relevant to text categorization. It facilitated faster and more scalable rule-based processing. Another project, the Cyc, initiated by Douglas Lenat, aimed to build a comprehensive knowledge base with common sense reasoning. While not strictly focused on document classification, Cyc involved the development of rules for understanding and categorizing information (Lenat, 1986). It provided insights into rule-based systems for knowledge representation.

Such rule-based systems gave an understanding into feature extraction, rule crafting, and the challenges of automating the categorization of textual data, setting the stage for future efforts into

document categorization and text classification. They laid the groundwork for many NLP tasks, including sentiment analysis. Furthermore, lexical works like William Wang's 1966 Dictionary on Computer, the first electronic database of Chinese dialects (Streeter, 1972), also served as an impetus in the field.

However, the chief trigger that paved the way for a systematic study of sentiment analysis was the explosion in the area known as psycholinguistics. Psycholinguistics was inspired by the works of Thorndike and Bartlett and came into popular use as a term after the publication of Pronko's 1946 article 'Language and Psycholinguistics: A Review' (Pronko, 1946). Psycholinguistics studies how language and psychological elements are related to each other. It is also called psychology of language (Jodai, 2011).

What augmented the role that psycholinguistics played in inspiring the development of the field of sentiment analysis is the appraisal approach to emotion. Appraisal approaches to emotion rely on the idea that emotions are not automatic physiological reactions to external stimuli but result from an individual's meaningful evaluation of those stimuli, thereby getting the name appraisal theories. These approaches had been in existence since the times of ancient Greek philosophers like Plato and Aristotle, but they entered limelight during 1960s due to the seminal work of scholars like Magda Arnold (1960) and Richard Lazarus (1980s and 1990s), who inspired exponential growth in this field (Roseman, 1996).

Appraisal theory to emotion studies the cause behind various people reacting dissimilarly to the same stimulus. The cognitive evaluation process is considered a central factor in the experience of emotions. The theory deliberates that emotions are subjective and arise when an individual attributes personal significance to an object, situation, or any entity. Emotions are contemplated as being dynamic and context-dependent, and individual differences in emotional responses are acknowledged by appraisal theorists. Evaluation of emotions involves a cognitive and reflective process where individuals assess the implications of a stimulus for their well-being, in line with their goals and values. Since, in accordance with this theory, a person's reaction to the same stimulus is independent of any other person's reaction to it, the study of unique sentiments of individuals became important with the advent of this theory. Therefore, researching individual variations of emotional experiences assumed significance within the scientific discourse.

Possibly influenced by Magda Arnold's work on appraisal theory, Weizenbaum developed ELIZA, a chatbot, at MIT in 1966 (Weizenbaum, 1966). ELIZA is one of the most well-known chatbots. It was designed to simulate a Rogerian psychotherapist and engaged users in text-based conversations. ELIZA used pattern matching and simple rule-based transformations to respond to user inputs. It could recognize keywords and generate responses based on predefined rules. ELIZA attempted to mimic human speech to return a statement into a question (Influentialfuture, n.d.). While ELIZA could not identify the polarity of a sentiment, it attempted to recognize the subjectivity of a statement, becoming the first attempt at rudimentary computational sentiment analysis.

English mathematician and computer scientist Alan Turing came out with the Turing Test in 1950. Through this test, Alan Turing raised the issue as to whether a computer program could converse

with a group of people without letting them realize that the entity they were talking to was a machine, and not a human (Adamopoulou and Moussiades, 2020). Kenneth Colby tried to answer Turing's question by creating PARRY, a chatbot which was crafted at Stanford University. PARRY simulated a person with paranoid schizophrenia. Like ELIZA, it engaged users in text-based conversations that touched upon emotional and psychological aspects. In PARRY's case, the focus was on simulating a specific mental health condition rather than general conversations. While not explicitly analyzing sentiment, PARRY's interactions delved into emotional and psychological aspects of conversation, providing a unique perspective on early attempts to model human behavior in language. PARRY became the first chatbot to pass the Turing Test (Botsplash, 2022).

Christian M.I.M. Matthiessen, William C. Mann and Sandra A. Thompson, together developed the Rhetorical Structure Theory (RST) in 1989. RST is described as a discourse analytical framework (Matthiessen *et al.*, 2018). RST is a model in discourse analysis that focuses on the hierarchical structure of texts and how different parts of a text are related to each other (Mann *et al.*, 1989). RST can be applied to analyze the rhetorical and discourse structures of textual content to better understand the sentiment expressed. RST can help identify the overall structure of a document or text. Sentiment in a document may vary across different sections, and understanding the structural organization can provide insights into how sentiments are distributed. It can also assist in identifying the key arguments, claims, and propositions within a text. Understanding these key elements is crucial in sentiment analysis, as sentiments are often expressed in relation to specific statements or claims.

The Defense Advanced Research Projects Agency (DARPA) initiated the TIPSTER Text program in 1991. It sought to improve HLT (Human Language Technology, n.d.) for managing multilingual corpora used in the intelligence process (Métais, 2013). In Japan, the first phase of ATR (Advanced Telecommunications Research) Interpreting Telephony Research project, also called ATR multilingual speech-to-speech translation (S2ST) system, started in 1986 and terminated in March 1993. Its third phase began in 2000. It focused on basic research for machine translation and speech recognition. It involved collaborations between American and Japanese researchers and furthered advancements in automatic language processing (Morimoto, 1993; Nakamura, 2004). The Penn Treebank Project contributed to the development of annotated corpora for training and evaluating NLP systems (Taylor, 2003).

Emotion classification the way we can classify an emotion based on its intensity and polarity finds a strong pedestal in emotion research and in affective science. The concept of affective computing emerged in 1990s. It emphasized recognizing and responding to human emotions by machines as a necessary need for development of truly intelligent artificial systems. It was pioneered by researchers, most notably Rosalind Picard at MIT (Picard, 2000). By treating emotion or sentiment recognition as a pattern recognition problem, and emotion expression as pattern synthesis, the concept of affective computing deeply contributed to the understanding of emotions in human-computer interaction. Developments like the concept of affective computing served to furnish a computational direction to the field of sentiment analysis.

WordNet was initiated by George A. Miller, a cognitive psychologist, in 1985 at Princeton University. The motivation was to create a comprehensive and structured lexical database that

reflects the organization of human knowledge about words and their meanings. The project received funding from the National Science Foundation (NSF) and collaboration from researchers at Princeton University, including Christiane Fellbaum, who played a significant role in WordNet's development. The first major release of WordNet, known as WordNet 1.0, was made publicly available in 1991 (Miller, 1995). It contained approximately 120,000 words grouped into synonym sets called synsets, with each set representing a specific lexical concept (Hane, 1999; Michael Hotchkiss, 2012) The chief relation between words in the WordNet was synonymy, as between the words locked and fastened. The most often encoded relation among synsets in the WordNet was the super-subordinate relation, which is also referred to as hyperonymy, hyponymy or ISA relation. This connects "more general synsets like {furniture, piece_of_furniture} to increasingly specific ones such as {bed} and {bunkbed}" (Princeton University, 2019).

SentiWordNet was developed by Esuli and Sebastiani (Esuli and Sebastiani, 2006). They used preexisting WordNet synsets for the purpose. SentiWordNet was described as "a lexical resource produced by asking an automated classifier Φ to associate to each synset s of WordNet (version 2.0) a triplet of scores $\Phi(s, p)$ (for $p \in P = \{\text{Positive, Negative, Objective}\}$) describing how strongly the terms contained in s enjoy each of the three properties." (Esuli and Sebastiani, 2008). SentiWordNet marked the initiation of the classification of sentiment in WordNet synsets, triggering an explosion in the growth of the field of sentiment analysis.

Late 1990s and early 2000s saw attempts to develop subjectivity lexicons (Hatzivassiloglou and McKeown, 1997; Wiebe, 2000; Wilson *et al.*, 2005). Peter Turney presented an unsupervised learning algorithm for classifying reviews as recommended (thumbs up) or not recommended (thumbs down). The detection of the polarity of a review was based on the average semantic orientation of the phrases in the appraised review which incorporated adjectives or adverbs (Turney, 2002). Pang *et al.* classified documents based on their overall sentiment. They used movie reviews as their data and determined whether a review was positive or negative. They found that machine learning techniques comprehensively outperformed human-produced baselines (Pang *et al.*, 2002). Liu *et al.*, in 2004, proposed a sentiment analysis approach known as Opinion Observer, which aimed to identify and track opinions in online news articles. Their work involved the development of a system that automatically extracted opinions and tracked how opinions evolved over time (Liu *et al.*, 2004).

During the incipient days of sentiment analysis, the technique was largely used on manuscripts or written paper documents. Now it is also used on a wide array of digital sources, including social media posts, online reviews, emails, news articles, and other forms of electronic text, reflecting the diverse and expansive nature of digital communication in contemporary society. It is applied to speech and audio, video content, images (including photographs, memes, and other related graphical representations), emoticons and emojis (Zou & Xiang, 2022), customer support interactions. surveys and feedback forms, product and service reviews, user comments in online forums, and in sensor data and internet of things (IOT) design and testing studies (to understand user sentiments related to smart devices and connected systems).

After Turney's, Pang's, and Liu *et al.*'s pioneering work related to development of sentiment lexicons and effective machine learning models for sentiment analysis, the evolution of sentiment

analysis paced on the shoulders of the viral growth in the popularity of social media. The explosion of social media platforms such as Twitter and Facebook in the early and late 2000s led to a surge in research on sentiment analysis of user-generated content. Twitter sentiment analysis, in particular, gained attention due to its real-time and short-text nature (Kumar and Jaiswal, 2019; Abdukhamidov, 2022).

Semantic analysis lies at the intersection of the fields of applied linguistics and computer science. It constitutes a synthesis of structured analysis of meaning of words, that is, semantic analysis, and computational components. In this context, SemEval constitutes a continuous series of evaluations (undertaken in the form of workshops) aimed at reviewing computational systems focused on semantic analysis and dedicated to advancing the forefront of NLP research. Its overarching mission is to push the boundaries of the current state of the art in semantic analysis while contributing to the development of top-tier annotated datasets. This initiative aims to address a spectrum of progressively complex challenges within natural language semantics. The primary goal of SemEval is to foster advancements in the development and evaluation of computational models that excel in understanding and interpreting the nuances of meaning in natural language. (In this context, *polysemy* implies a phenomenon where a word can have different meanings in different contexts.) SemEval became a regular annual feature since 2012. Each SemEval workshop tests its participants with a series of tasks within the field of NLP. For example, the SemEval-2021 Shared Task NLPContributionGraph, also known as the 'NCG task,' challenged SemEval 2021 participants to create automated systems capable of organizing contributions from scholarly NLP articles written in English (D'Souza *et al.*, 2021).

The adoption of deep learning techniques, particularly recurrent neural networks (RNNs), long short-term memory models, and transformer-based models like BERT, improved sentiment analysis performance. These models captured complex contextual information and achieved state-of-the-art results, increasing global interest in sentiment analysis. The focus shifted from document-level sentiment analysis to aspect-based sentiment analysis, where systems aim to identify sentiments toward specific aspects or entities within a document or sentence. This became crucial for fine-grained analysis in reviews and opinions. With the increasing use of multimedia content, researchers started exploring sentiment analysis in a multimodal context, combining text, images, audio, and video for a more comprehensive understanding of sentiment. Transfer learning techniques, especially pre-trained language models like GPT and BERT, have been adapted for sentiment analysis. Fine-tuning these models on sentiment-specific tasks has shown remarkable performance improvements (Park *et al.*, 2020; Kokab *et al.*, 2022; Jia *et al.*, 2023; Ahmed, S. F. *et al.*, 2023).

1.3. Techniques Utilized for Gathering Data for and Conducting Sentiment Analysis

Psychological and psycho-computational research, which includes sentiment analysis, can be classified into two main types: primary and secondary. Primary research involves firsthand data collection through direct interactions with users, and employs methodologies like interviews, surveys, and usability studies. On the other hand, secondary research utilizes information compiled by others, often sourced from books, articles, or journals. Another way to categorize research is based on the data type: qualitative and quantitative. Qualitative research relies on observations and

conversations, focusing on understanding users' needs and addressing questions related to factors behind actions, occurrences, phenomena, processes, and so on. In contrast, quantitative research involves calculable data gathered through numerical estimation, typically from large-scale surveys, to answer quantitative questions. Yet another way to label sentiment data mining research is classifying it based on whether the research setting is moderated or unmoderated. A facilitator is present in a moderated emotion data collection setting, while an unmoderated setting is not monitored.

Affect measures, also known as measures of affect or emotion, form a significant part of human behavioral research, and play a pivotal role in investigating human affect, encompassing emotions and moods (Ekkekakis, 2013). These measures involve gathering data through self-report studies where participants quantify their present emotional states or provide an average assessment of their feelings over an extended period (Cloos *et al.*, 2023). While certain affect measures include variations enabling the evaluation of inherent tendencies to experience specific emotions, assessments targeting such enduring traits are typically categorized as personality tests. This multifaceted approach aids researchers in delving into the dynamic and nuanced aspects of human affect, contributing valuable insights to the fields of psychology and personality assessment.

To conduct sentiment analysis, an investigator must obtain information through interaction with the persons whose sentiment is being studied, or must acquire, through lawful and ethical methods, publicly or privately available information about them for analysis. Privacy laws like Health Insurance Portability and Accountability Act (HIPAA) in USA, General Data Protection Regulation (GDPR) in European Union (EU) and Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada and privacy principles often apply to such data collection. Many tools have been discussed and found to be effective for measuring human emotions. They fall into various categories. The first category is self-report methods (Liu *et al.*, 2022).

Self-report methods are techniques that prompt respondents to express their emotions directly, either verbally or non-verbally. These methods include surveys, questionnaires, interviews, focus groups, ratings, rankings, Likert scales, emoticons, and so on. Self-report methods are useful for getting user feedback, opinions, and preferences, as well as for exploring the reasons and triggers behind user emotions. However, self-report methods also have some limitations, such as relying on user recall and honesty, being influenced by social desirability and expectations. Such issues are called biases, and some of the chief biases associated with self-reported methods are social desirability bias and recall bias (Althubaiti, 2016). Self-reporting methods are also unable to capture subtle, unconscious emotions. For example, the Hawthorne effect, also known as the observer effect, is a phenomenon in which individuals modify or improve their behavior simply because they are aware that they are being observed. It is related to desirability bias.

Surveys include open-ended (subjective, descriptive questions where a respondent can elaborate on their feelings with regards to the questions, and provide detailed and unrestrained responses, allowing for rich qualitative insights) or closed-ended questions (objective questions with a fixed number of options to select from) where participants express their opinions, providing valuable text data for sentiment analysis. In a comprehensive survey, various types of questions are strategically employed to gather diverse and nuanced responses from participants. Pairwise

comparison questions prompt individuals to choose between two options, offering a comparative perspective.

A forced-choice questionnaire is an instrument where respondents are compelled to select an answer from a predefined set of options for each question (Xiao *et al.*, 2022). A forced-choice question is a closed-ended question which a respondent cannot skip answering. As the respondents are forced to express their opinion, it eliminates the possibility of missing data, which requires techniques like imputation, additive smoothing, and interpolation to be addressed. This aids in data mining and data preprocessing. However, as the survey participants are forced to choose one of the alternatives, annoyance and confusion might be caused. Befuddlement might result if none of the given alternatives aligns with the answer a respondent would instinctively give to the question. And this often causes respondents' answers to not indicate their authentic feedback, leading to occurrence of false truths in the questionnaire and inaccuracies in an assessment – including sentiment analysis – of the answers to the questionnaire. In case a question is ambiguous, the respondents are still forced to answer it, leading to more frustration among survey-takers.

Figure 3: An example of a survey with a series of pairwise comparison scale questions. When paired comparison questions are combined with a scale, it is often referred to as pairwise comparison with a (rating) scale. In this type of assessment, respondents are presented with pairs of items, and in addition to choosing between the two, they are asked to provide a rating or score for the option they choose. The scale allows respondents to express the intensity of their preference or judgment for each item in the pair.

Nominal (multiple choice) questions are closed-ended questions which present a list of options, requiring participants to select the most fitting choice. Dichotomous questions, on the other hand, provide binary options for respondents to choose from. Rating scale questions involve assigning a numerical rating to statements and attributes. They introduce a graded scale, enabling participants to express their agreement or disagreement on a spectrum. Likert scales are specific types of rating scales which include a range of response options with clearly labeled anchor points. Usually, when answering a Likert scale question, respondents indicate their level of agreement or disagreement with a statement by selecting a point on the scale, usually ranging from "Strongly Disagree" to "Strongly Agree."

Contextual follow-up questions delve deeper into participants' responses by seeking clarifications and additional information. Matrix questions organize multiple items in a grid format, streamlining the collection of responses on related topics. Drop-down questions present a list of options in a dropdown menu format, enhancing survey aesthetics. Ranking questions require participants to prioritize items or preferences based on their significance. Checkbox questions offer multiple answer choices, permitting respondents to select all applicable options, fostering inclusivity in responses. Employing this diverse array of survey question types ensures a comprehensive exploration of participants' perspectives and preferences.

The item response theories (IRTs), also known as the latent response theories, including the Rasch Measurement Theory, involve mathematical models which aid in elucidating the link between unobservable (latent) traits and their observable manifestations, such as responses and performance outcomes (Wind and Hua, 2021; Fulcher and Harding, 2022). IRTs provide frameworks for understanding how individuals with different latent trait levels respond to items in assessments. The application of the IRTs extends beyond traditional educational assessments; IRTs can also be harnessed for sentiment analysis. By modeling the relationship between latent sentiments and observable responses, IRTs enable a more nuanced understanding of how individuals express sentiments across various contexts.

Interviews can capture in-depth, qualitative insights into individuals' sentiments. Open-ended interview questions can yield rich textual data for sentiment analysis. Like regular interviews, 1:1 (one-to-one) interviews provide a personalized setting for individuals to express their sentiments. Focus groups involve a facilitated discussion among a group of participants guided by a moderator. This method is used to gather collective insights, opinions, and perspectives on a specific topic. Analyzing the transcripts and notes from focus group discussions can provide insights into shared sentiments.

In-situ self-reports are data collection methods which record users’ current or recent experiences. These methods involve collecting data from individuals within their natural environment or context. In situ self-report studies involve participants reporting their experiences in real-time. The data can be analyzed for sentiment, especially if participants express their emotions or feelings. A habit-tracking app would be one of the examples of such user data recoding tools.

Ecological Momentary Assessment (EMA) is an in-situ self-reporting method which involves collecting real-time data in the participants' natural environment. Participants receive signals to report on their experiences, feelings, and behaviors at specific moments. For example, using a mobile app, participants receive prompts and notifications several times a day to report their current mood, stress level, or dietary choices. Experience Sampling Method (ESM) is another in-situ self-reporting method which captures participants' experiences by prompting them to provide information at random intervals throughout the day. It is a structured diary technique (Verhagen, 2016), that is, the data recorded through ESM has the same format across all respondents, paving way for an easier systematic analysis. ESM helps researchers understand the variability in experiences of respondents. An example of ESM would be a setting where participants carry a device that beeps at random times, prompting them to report their current activities, emotions, or social interactions.

A diary study is also considered to be an instance of ESM. Participants are asked to keep a daily diary where they document their experiences, emotions, and activities (Zhang, 2016). The participants of a diary study write entries at predetermined times throughout the day or whenever they feel compelled to record their thoughts. The diary could be physical or digital, depending on the study design. Through a diary study, researchers gain insights into participants' subjective experiences, reactions to events, and the context surrounding their behaviors over an extended period. Prominent examples of journaling experiences in the form of diary entries can be observed among space scientists aboard the International Space Station (ISS). Similarly, researchers conducting studies in polar regions, where conditions are extreme and isolation is prolonged, often maintain diary entries. These entries serve as a personal record of the researchers' encounters with the harsh polar environment, detailing scientific observations, weather conditions, and the psychological aspects of enduring extended stays in remote and challenging locations.

EMA and ESM are related terms and are often used interchangeably. Whereas some scholars make distinctions between the two, others consider them synonymous. Some distinctions between the two can be discussed. For instance, EMA involves real-time data collection triggered by external prompts at specific moments during the day. It involves scheduled assessments or random prompts (Shiffman, 2008). ESM is typically used to describe a broader category that includes both scheduled and random assessments. It can be argued that ESM captures data more continuously and incorporates both scheduled and spontaneous assessments. About EMA, it can be said that it involves fixed-time intervals or event-contingent assessments triggered by specific events or behaviors, while ESM typically involves random or semi-random time intervals, prompting participants to report their experiences without a fixed schedule.

Ambient EMA involves continuous data collection without explicit participant input. Sensors capture information about the environment and behavior automatically (Kim *et al.*, 2019). Using smartphone sensors, or wearable/ambulatory technologies [referred to as Mobile health (mHealth) care platforms by Kim *et al.*, 2019] such as fitness trackers and smartwatches including instruments such as Empatica E4 which record physiological and physical activity data like heart rate variability (also called HRV, and which is indicative of one’s emotional state) and number of steps taken in a specific time interval, researchers collect details on participants' location, movement, and ambient noise to infer contextual information without active participant reporting.

There are two kinds of self-reporting methods, in-situ and retrospective. While in-situ methods record real time and dynamic data, retrospective methods rely on memory (recall) of the participants (Wu *et al.*, 2020). Day Reconstruction Method (DRM) is an example of a retrospective self-reporting method, and it requires participants to reconstruct their daily activities and experiences retrospectively (Kahneman *et al.*, 2004). It involves recalling and reporting details about specific episodes during a given period.

Ethnography is a qualitative research method that involves prolonged engagement with a cultural community to understand their behaviors, beliefs, and practices (Morgan-Trimmer *et al.*, 2016). Ethnographers immerse themselves in the community to gain a holistic understanding. Understanding the cultural context through ethnography can help refine sentiment analysis models to better capture nuances in language and expression. While ethnography involves immersive,

long-term study of a community or culture, mobile ethnography involves participants documenting their experiences through multimedia (photos, videos, audio recordings, notes and so on) using their mobile devices. In a mobile ethnography study, participants use their smartphones to capture moments, thoughts, feelings, and experiences and provide context through annotations or voice recordings (Loh *et al.*, 2023). Mobile ethnography is also called digital ethnography or autoethnography, and it is another in-situ method.

The growing use of social media offers a substantial and novel reservoir of user-generated ecological data, often referred to as digital traces, that can be systematically gathered for sentiment data gathering efforts and computational research on respondents' emotions (Settanni *et al.*, 2018). Analyzing digital traces from social media platforms provides data on users' behaviors, sentiments, and interactions. Researchers strive to analyze tweets or Facebook posts to decipher trends in users' expressions, opinions, or reactions to events.

The Self-Assessment Manikin (SAM) is an affect measuring standard questionnaire tool designed to gauge individuals' emotional responses to various stimuli (Bradley *et al.*, 1994). It provides a quantifiable assessment of affective experiences. In modern applications, SAM is widely used in user experience research, particularly in the field of digital design and marketing. For instance, when testing a new mobile app, researchers might employ SAM to gauge users' emotional reactions to different interface designs, ensuring that the app elicits positive and engaging emotional responses. SAM's simplicity and versatility make it a valuable tool in understanding how individuals emotionally engage with contemporary digital experiences.

The Positive and Negative Affect Schedule (PANAS) is another psychological standard questionnaire tool used to assess an individual's positive and negative affective states (Diaz-Garcia, 2020). As an example, PANAS often finds application in workplace well-being initiatives. Companies implementing employee well-being programs usually employ PANAS to measure the impact of interventions on employees' positive and negative emotions. PANAS assessments can provide concrete insights into the effectiveness of initiatives such as mindfulness programs or flexible work arrangements, aiding organizations in crafting strategies that enhance overall job satisfaction and emotional well-being in the workplace.

Behavioral methods and empathy tools are also categorizations of methods that are used to measure participant emotion. Behavioral methods, also called observational methods, detect and analyze user actions and reactions, either online or offline. They are called observational methods because they are non-experimental in nature, that is, they do not seek to control or manipulate conditions in which observations are recorded. They record behavior in natural settings, and therefore, the possibility to record causal results using such methods is negated (Jhangiani *et al.*, 2019). Individual as well as group behavior depends on the situation (context), and that can give an understanding of feelings with limited accuracy.

Behavioral methods involve direct observation (where a researcher does not need to depend on reported observation conducted by someone else), and include eye tracking, mouse and click dynamics tracking, clickstream analysis (where the pages a user visits and user behavior while using an app or visiting a webpage), and keyboard dynamics tracking. They can also involve

recording facial expressions, gestures, body language, voice tone, and physiological responses. Eye-tracking studies involve measuring either where the user's gaze is concentrated or the movement of the eye, as an individual browses an app or a website. In the context of eye-tracking, heat maps represent where the visitor focused their eyes and how long they stared at a specific location, while saccade pathways record the eye's motion from between specified points of focus.

Behavioral methods are useful for measuring user attention, engagement, interest, frustration, or arousal, as well as for identifying user pain points and opportunities. However, behavioral methods also have some challenges, such as requiring dedicated personnel and specialized equipment and software, being intrusive and even invasive, and being difficult to interpret and contextualize. Figuring out what specific actions imply in specific cultural and psychological settings can be tough, as they might have different interpretations in diverse settings, leading to misunderstandings.

For example, maintaining eye contact during a conversation is often considered a sign of attentiveness and sincerity in Western cultures, while in some East Asian cultures, prolonged eye contact may be seen as impolite or confrontational. Similarly, the concept of personal space varies across cultures; what may be deemed as an acceptable distance for conversation in one culture might be considered too close or distant in another. Additionally, the interpretation of silence can differ significantly – some cultures may view it as a sign of contemplation or respect, while others might perceive it as awkward or indicative of disinterest. Further, putting together different types of behavioral data is complicated. Moreover, observing people has an amplified potential of causing privacy concerns.

Empathy tools are instruments designed to help researchers understand and empathize with users' emotions and experiences. These tools include techniques for gathering emotional data, such as user journey mapping and emotional mapping. Emotion mapping is a tool that visualizes user emotions along a journey or a process. It includes emotion curves, emotion wheels, emotion grids, and emotion cards. Emotion mapping is useful for understanding user emotional states, transitions, and patterns, as well as for evaluating user satisfaction and loyalty. However, emotion mapping also has some limitations, such as being subjective and variable, being influenced by external factors and biases, and being incapable of capturing the depth and richness of user emotions.

While user journey mapping is a strategic tool employed in user experience design to visualize and comprehend the entirety of a user's interactions, including emotional experiences, with a product or service, emotional mapping is a specialized approach within user experience design that focuses on understanding and visualizing the emotional responses users have during their interactions with a product or service. Personas play a crucial role in the context of user journey mapping, offering a human-centered approach to understanding and designing for different user segments. A persona represents a fictional but realistic character that embodies the traits, needs, and behaviors of a specific user group. Integrating personas into the user journey mapping process provides a more personalized and empathetic understanding of the user experience.

The concepts of user journey, customer journey, buyer journey, and user flow collectively represent the dynamic pathways individuals traverse while interacting with products or services. The user

journey encompasses the entire spectrum of user interactions, emphasizing the holistic experience from discovery to engagement and retention. The customer journey focuses on the broader relationship between a customer and a brand, considering touchpoints across multiple channels. The buyer journey specifically tracks the steps leading to a purchase decision. Meanwhile, user flow details the step-by-step sequence of actions users take within a system or interface. Together, these concepts provide a comprehensive understanding of the diverse pathways users and customers follow, guiding designers and marketers in optimizing every stage for a seamless, satisfying, and purpose-driven experience.

Physiological measures for measuring emotion, which include assessing Electrodermal Activity (EDA), HRV, Electrocardiogram (ECG) etc., using instruments that capture physiological parameters, in natural settings, also come under behavioral methods. Neuroscientific measures like Electroencephalography (EEG) and Magnetic Resonance Imaging (MRI) conducted in non-controlled settings are also prominent examples.

Participant observation is another method to gather emotion data. A type of non-experimental study where behavior is methodically observed to be recorded and analyzed, participant observation is a qualitative method which involves researchers actively engaging in the environment they are studying (Jhangiani *et al.*, 2019). The researcher, usually an anthropologist, sociologist, or psychologist, actively participates in the social phenomenon, social group, or social context being probed. While it may not be a typical method for sentiment analysis, observing participants in real-world settings is something that can provide insights into the context in which sentiment is expressed.

In-the-lab observation involves studying participants in a controlled laboratory setting. Researchers observe and record behaviors, reactions, and interactions under controlled conditions to gather data for analysis. Similar to participant observation, in-the-lab observation may not be a direct method for sentiment analysis. However, it could be used to observe participants' behaviors and expressions in a controlled setting, offering context for sentiment analysis.

Observation methodologies conducted in the lab setting aim to provide researchers with an understanding of users and their behaviors within their natural context. The goal of "deep hanging out" and active participation methods employed in in-the-lab observation method is to immerse oneself in the vicinity of the subjects, actively participating in the activities they engage in, fostering an experience of membership in the broader context, culture, or subculture under investigation. This approach goes beyond mere observation, emphasizing the formation of connections and empathy with the individuals and elements that hold significance for them. It is particularly valuable when the aim is to uncover authentic insights into users' experiences and behaviors in a more dynamic and realistic setting.

On the other hand, naturalistic or "Fly-on-the-Wall" observation serves a distinct purpose, focusing on gaining a profound understanding of how people behave in specific locations. This method is employed when researchers seek to study individuals unobtrusively, minimizing potential biases introduced by the Hawthorne effect (McCambridge *et al.*, 2014). The Hawthorne effect manifests when people act differently when they are aware of the fact that they are being observed. In the

fly-on-the-wall approach, researchers position themselves as passive observers, going to a designated location and discreetly observing ongoing activities without direct interaction or conversation with the subjects. By adopting the role of a "fly on the wall," researchers can capture genuine and unaltered behaviors, providing valuable insights into the dynamics of human behavior within a given environment.

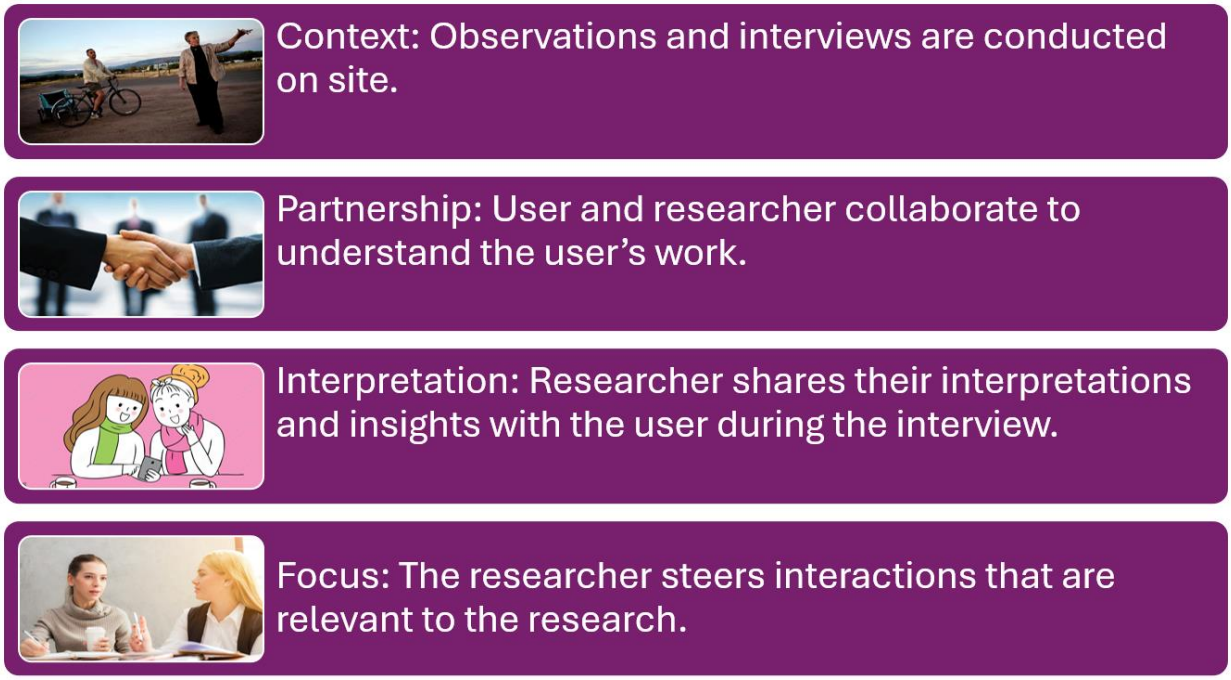


Figure 4: The Four Principles of Contextual Inquiry [adapted from Salazar (Nielsen Norman Group), 2020]

Contextual inquiry is a user-centered design method where researchers observe and interact with users in their real-world environment to understand how they work, think, and use products or services. This method provides insights into user needs and behaviors. Contextual inquiry involves studying users in their real-world context. Combining observation and self-report, contextual inquiry, a semi-structured interview method, seeks to obtain information about the context of use. In it, users are first asked a set of standard questions and then observed and questioned while they work in their own environments. It can provide valuable insights into how sentiment is expressed in natural language within specific situations, thereby contributing to the improvement of sentiment analysis models.

Contextual inquiry aims to understand how the surrounding environment impacts interactions. This approach is most effective when biases introduced by observation and discussion are not a concern, and researchers are particularly interested in specific tasks. Unlike participant observation where the researcher actively engages in the task, contextual inquiry involves defining tasks in advance and conducting sessions in the natural environment where the tasks typically occur. Participants are asked to accomplish a series of tasks with minimal interruptions while verbalizing their thought process. The back-and-forth interaction between the researcher and participant allows for a comprehensive understanding of the task, with the participant explaining their actions, the

researcher offering interpretations, and the participant either agreeing or correcting. This methodology's strength lies in its ability to unveil unexpected details, habitual behaviors, and invisible processes that might otherwise go unnoticed, providing valuable insights into the intricacies of user behavior.

Web scraping refers to mining or harvesting data from online sources. It can be performed manually, but it is usually performed by automation tools like web crawlers or bots. Employing such tools is cost-intensive but saves time. The information is extracted before being exported and stored in a way which is more convenient for a user, like a spreadsheet or a database. Data scraped from the web can be used for sentiment analysis.

Online data sources, such as social media, reviews, ratings, and forums, are rich data sources for sentiment analysis. Examples include data obtained from Facebook, YouTube, Twitter, product review web pages, and online forums like Reddit. Analyzing text from these sources allows for real-time tracking of sentiments expressed by individuals or communities. The textual data collected through these probes can be analyzed for sentiment. The exploration of online sources delves into the dynamics of social interactions conducted on digital platforms, shedding light on the ways message boards, social media, and other tools facilitate social engagement and information dissemination. By scrutinizing online content and activities, researchers gain a deeper understanding of how individuals interact, communicate, form communities, and learn in the digital landscape.

This exploration extends to unraveling the development and propagation of ideas, providing valuable insights into information dissemination patterns. Online studies are marked by lack of control on participants, and a lack of detailed observation opportunities can afford challenges to the data collection. Online studies must consider questions of privacy and online consent. Cultural probes are a design research method involving the use of various creative tools and activities to elicit responses from participants about their daily lives, preferences, and experiences. It aims to uncover deep insights into cultural aspects.

Emotion recognition is a tool that uses artificial intelligence and computer vision to detect and identify user emotions from facial expressions or body movements. It can include webcams, other kinds of cameras, sensors, or wearables. Emotion recognition is useful for measuring user affect, mood, and personality, as well as for providing user feedback and personalization. However, emotion recognition also has some issues, such as being prone to errors and inaccuracies, being sensitive to privacy and ethics, and being unable to capture the context and meaning of user emotions.

Emotion synthesis is a tool that uses artificial intelligence and computer graphics to generate and display user emotions on virtual or augmented reality devices. It can include avatars, characters, or scenarios. Emotion synthesis is useful for creating user empathy, immersion, and presence, as well as for testing user reactions and responses. However, emotion synthesis also has some challenges, such as being complex and costly, being unrealistic or uncanny, and being unable to capture the authenticity and diversity of user emotions.

Sentiment analysis is a tool that uses NLP and machine learning to extract and classify user emotions from text or speech. It can include reviews, comments, feedback, social media posts, chat messages, or transcripts. Sentiment analysis is useful for gauging user sentiment, polarity, intensity, and valence, as well as for discovering user insights and trends. However, sentiment analysis also has some drawbacks, such as being dependent on the quality and quantity of data, being affected by sarcasm, irony, or ambiguity, and being unable to capture complex or mixed emotions. Detecting sarcasm and irony in text is particularly challenging. While a statement may appear positive in literal terms, its intended sentiment could be negative, requiring a deep understanding of context and cultural nuances. Discerning sarcasm and accurately interpreting sentiment can be difficult even for humans, not to mention the added complexity for machine learning models like neural networks.

2. Background

The successful curtailment of an epidemic is intricately linked to widespread vaccine acceptance within the population. This is especially true about an epidemic associated with a highly contagious virus like SARS-CoV-2. Vaccination, particularly in case of viral diseases, has proved to be a powerful tool in the public health arsenal, for it not only provides individual protection but also contributes to achieving herd immunity. Achieving herd immunity is a key objective in controlling the spread of infectious diseases. Herd immunity occurs when a sufficiently large proportion of the population has achieved immunity against the causative agent of the epidemic, either through vaccination or after surviving a previous infection. The spread of herd immunity reduces the overall transmission of the causative agent, lending protection to even those who are not immune.

Vaccination plays a crucial role in breaking the chain of transmission. When a significant portion of the population is vaccinated, the spreading disease vector encounters fewer susceptible individuals. This makes it tough for the vector to spread efficiently. Limiting its spread helps in slowing down or, in some cases, causing complete standstill to the transmission dynamics of the epidemic. Vaccinated individuals who do contract the virus are more likely to experience milder forms of the disease. This not only reduces the burden on healthcare systems but also contributes to overall community well-being. A lower severity of illness is found to be particularly decisive in preventing overwhelming surges in hospitalizations and fatalities during an epidemic.

High vaccine acceptance is vital for protecting vulnerable populations, such as the elderly, individuals with underlying health conditions, and those who cannot receive the vaccine due to medical reasons. These groups are at higher risk of severe outcomes, and widespread vaccination helps create a protective shield around them. Furthermore, vaccination can play a complimentary role to other public health measures implemented to control epidemics. While measures like social distancing, mask-wearing, and testing are important, vaccination provides a sustainable, long-term solution. It allows for the gradual easing of stringent control measures and assists in restoring normalcy to society.

A high level of vaccine acceptance reduces the likelihood of resurgences of the causative agents and the emergence of new variants. The virus is less likely to mutate and evolve when its transmission is significantly curtailed through vaccination, preventing the potential setbacks

associated with new variants (Saxena *et al.*, 2017). Epidemic control is not limited to individual countries. Achieving global vaccine acceptance is crucial for preventing the international spread of the virus. As the world is interconnected, the control of an epidemic requires coordinated efforts and high vaccine acceptance on a global scale. There was reported anticipation for a singular vaccine providing long-lasting immunity. However, as new variants emerged and data indicated potential waning immunity over time, the need for booster shots and yearly renewal of doses became a topic of discussion, eroding public trust in the efficacy of vaccination in context of COVID-19 (Lin *et al.*, 2023).

The curbing of an epidemic is a collective effort, and vaccine acceptance is a lynchpin in this endeavor. Sufficient vaccination coverage within the population contributes to the interruption of transmission, protects vulnerable individuals, and sets the stage for the eventual control and, ideally, the eradication of the epidemic. Public health campaigns and communication strategies that foster trust, dispel myths, and promote vaccine acceptance are essential components in achieving these goals.

In case of COVID-19 pandemic, the vaccination drives and the public acceptance of the vaccines that were rolled out encountered unprecedented hurdles, ranging from vaccine development timelines to the spread of misinformation. Some researchers argue that the fast-tracked development and approval of SARS-CoV-2 vaccines, such as the Pfizer-BioNTech and Moderna vaccines, raised concerns among the public about the safety and efficacy of these vaccines. The unprecedented speed of development led to skepticism, despite clinical trials being held (Chavda *et al.*, 2022). According to Chavda *et al.*, the emergency vaccination discovery approach led to the omission of some initial preclinical steps in vaccine development, which could have been a cause of concern for people. Research studies on the safety and efficacy of these vaccines were pivotal in addressing such causes of concern and building trust.

Studies have indicated that the lowered counteracting capacity of the vaccines against some coronavirus variants created doubts in productive vaccine manufacturing. The emergence of SARS-CoV-2 variants raised questions about vaccine efficacy, prompting the need for variant-specific vaccines. Further, challenges in cold chain management for vaccine distribution remained critical. A proposed solution to this was to revise the vaccine sequences using methodologies like directed evolution, in line with the chief mutations in the SARS-CoV-2 virus that caused the appearance of new variants (Saxena *et al.*, 2017; Harvey *et al.*, 2021; Chavda *et al.*, 2022). The expedited vaccine development faced ethical considerations (Jalilian *et al.*, 2023), and issues related to vaccine accessibility, safety, and efficacy persisted. Most vaccines were emergency use-approved, and not fully approved, which led to more skepticism about COVID-19 vaccination. Furthermore, the rapid development and rollout of COVID-19 vaccines and mass vaccination drives across the world was itself one of the causes of emergence of new, vaccine-resistant variants (Rouzine *et al.*, 2023), leading to potentially enhanced distrust about the vaccines.

Infodemic, or information epidemic, which implies rapid spread of misinformation, has played a significant role in vaccine hesitancy, or the widespread hesitation to receive COVID-19 vaccine. Research studies analyzing social media trends identified the widespread dissemination of false information regarding vaccine contents, side effects, and conspiracy theories. In some instances,

unfounded conspiracy theories emerged, such as the false claim that the introduction of the COVID-19 vaccine involved implanting microchips or nano-chips into the human body (Ulah *et al.*, 2021; Skafle *et al.*, 2022). The conspiracy theorists further alleged that 5G networks would then be utilized to transmit signals to these chips (Islam *et al.*, 2021), purportedly enabling control over humanity by a shadowy 'world elite'. Theories like these have purportedly led to many people, in 2024, still celebrating not taking the vaccine. Such conspiracy theories have sufficed to generate unwarranted mistrust, and this fact underscores the significance of critically evaluating sentiment information in the age of digital misinformation spreads in the form of an infodemic.

Public apprehension about the long-term effects of the vaccines was also a hurdle with regards to successful COVID-19 vaccination. Research-oriented studies examining the long-term safety and efficacy of the vaccines became essential in addressing such concerns. Providing evidence-based information on the long-term benefits and safety of the vaccines was crucial in influencing public perception (Nuwarda *et al.*, 2022; Majid *et al.*, 2022; Mir and Mir, 2024).

Alliances between governments, non-government organizations and the private sector, aimed at tackling the pandemic, were common. In this context, measures that were considered coercive in nature, for example, government-mandated vaccination and getting COVID-19 vaccines, often multiple doses and boosters of it, being made prerequisites for getting or continuing employment by multiple establishments (Papastephanou, 2021; Lange and Shullenberger, 2024), sufficed to increase public cynicism about COVID-19 vaccines.

The situation in which countries prioritize securing early access to stocks of vaccines, sometimes to the extent of stockpiling essential components for local vaccine production, is termed vaccine nationalism (Hafner *et al.*, 2022). The global disparity in vaccine distribution and instances of vaccine nationalism fueled public skepticism. Some studies highlighted the challenges of ensuring equitable access to vaccines and the potential consequences of unequal distribution. These studies underscored the importance of global collaboration and fair distribution strategies for building trust. Issues of vaccine access and health disparities influenced public acceptance. Research-oriented studies on vaccine distribution strategies and their impact on vulnerable populations highlighted the need for targeted interventions. Addressing barriers to access and ensuring equitable vaccination opportunities were essential for overcoming hesitancy.

Reports of rare adverse events associated with COVID-19 vaccination raised concerns among the public. Rare cases of life-threatening blood clots (thrombi), anaphylaxis (hypersensitivity or allergic reaction), myocarditis (inflammation of myocardium), pericarditis (inflammation of pericardium), tinnitus (ringing in ears), and hearing changes (Deb *et al.*, 2020; Yaamika *et al.*, 2023), emerged, causing panic. Research studies investigating the frequency and risk factors of these events made an effort to provide critical insights. Clear communication based on scientific evidence attempted to contextualize the rarity of such events and the overall safety of vaccination but, reportedly, there are people that continue to feel that they are fortunate for not getting COVID-19 vaccines.

Political influence and its impact on public trust were significant factors in guiding public sentiment about COVID-19 vaccines (Bolsen and Palm, 2021). Research studies exploring the role

of political communication in shaping public opinion on vaccination highlighted the importance of transparent and science-based messaging. Instances where political agendas influenced vaccination campaigns underscored the need for unbiased, evidence-driven communication. Further, the use of fetal cell lines in some COVID-19 vaccines ate away at public confidence among individuals who prioritized certain moral and religious beliefs (Zimmerman, 2021), especially people who believed the fetal cell lines were associated with abortion while considering abortion as something sinful. The disclosure of such practices sparked a phenomenon in which a subset of the population was left questioning the ethical considerations behind vaccine development.

The barriers to public acceptance of SARS-CoV-2 vaccines were, and continue to be, multifaceted, and require research-driven strategies to address concerns, combat misinformation, and build trust. However, the first step in this process requires the quantification of public distrust, and for that purpose, sentiment analysis becomes an invaluable tool. By conducting sentiment analysis, areas where false information is gaining traction can be pinpointed, which can enable authorities to counteract these narratives with targeted communication strategies.

Negative sentiments often arise when individuals perceive a lack of transparency and, therefore, trustworthiness, in the information provided by health authorities. Sentiment analysis helps in identifying patterns of distrust or skepticism, allowing for the development of communication strategies that prioritize transparency, clarity, and the dissemination of accurate information. By directly addressing concerns voiced through sentiment analysis, health organizations can foster a more positive perception and trust in the COVID-19 vaccine.

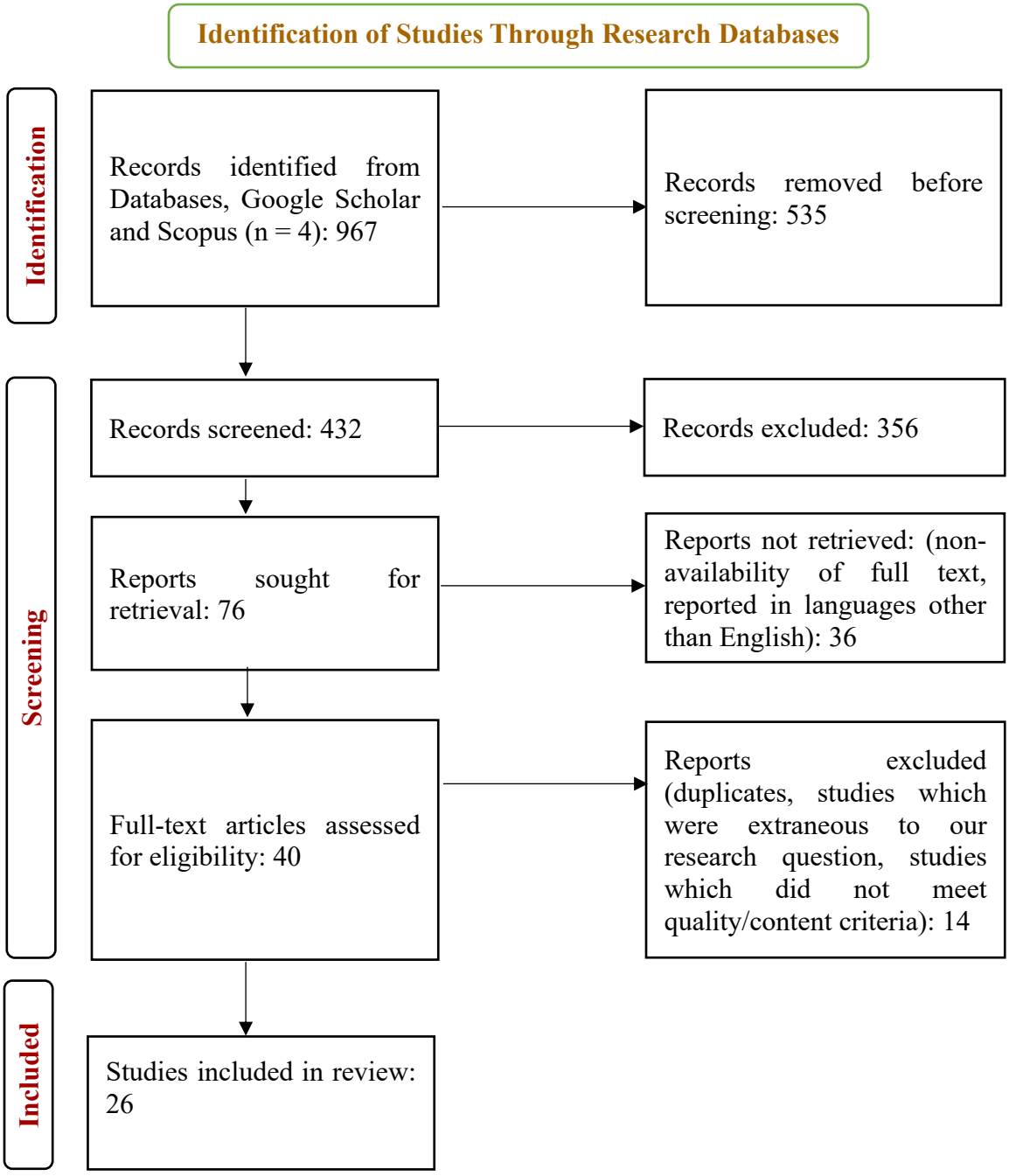
Moreover, sentiment analysis aids in monitoring the evolution of sentiments over time. It allows authorities to identify emerging trends and narratives, enabling them to stay ahead of potential issues related to misinformation. Real-time monitoring of sentiment helps in swiftly countering false claims or rumors before they gain widespread traction. By understanding sentiment dynamics, public health campaigns can be adjusted to not only provide accurate information but also to address evolving concerns and mitigate the impact of misinformation on public perception.

3. Methods

The purpose of this paper was to review the studies which employ deep learning algorithms for sentiment analysis associated with COVID-19 vaccination. For this, we performed a secondary study. A meta-analysis was carried out by employing the PRISMA technique. The databases searched were Scopus, PubMed Central, IEEE Xplore, and ScienceDirect, while the search engine used was Google Scholar. A total of 26 studies were finally identified to be reviewed. The decision was based on criteria discussed in the PRISMA flowchart, apart from the number of citations and the publication date, so that we chose only those articles which had not been sufficiently assessed yet.

The keywords (including phrases) used in the final search query were sentiment, sentiment analysis, opinion mining, deep learning, neural networks, COVID-19, SARS-CoV-2, vaccine, vaccines, and vaccination. Operations like ‘AND’ and ‘OR’ were deployed at relevant locations in the search query so that the most relevant results were revealed. The phrase ‘opinion mining’

contributed the least to our search query. The review table encapsulates the gist of the studies we reviewed.



Flowchart 1: PRISMA flowchart for our systematic review. The flowchart summarizes our identification, screening, and exclusion procedures.

S. No.	Title of the Publication	Authors	Year	Model Conceptualized
1.	Sentiment Analysis For COVID-19 Vaccine Popularity	Saeed *et al.*	2023	Unnamed (proposed) DNN model
2.	Social Media Text Analysis on Public's Sentiments of Covid-19 Booster Vaccines	Kristian *et al.*	2023	Used combinations of existing tools
3.	Social Media Sentiment Analysis on Third Booster Dosage For COVID-19 Vaccination: A Holistic Machine Learning Approach.	Ghosh *et al.*	2023	Compared preexisting models
4.	Sentiment Analysis Using Machine Learning and Deep Learning on Covid 19 Vaccine Twitter Data With Hadoop Mapreduce.	Kul and Sayar	2023	Compared preexisting models
5.	Temporal Analysis and Opinion Dynamics Of COVID-19 Vaccination Tweets Using Diverse Feature Engineering Techniques	Ahmed, S. *et al.*	2023	Proposed an unnamed framework
6.	Sentiment Analysis On COVID-19 Vaccine Tweets Using Machine Learning And Deep Learning Algorithms.	Jain *et al.*	2023	Compared preexisting models
7.	Sentiment Analysis Of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches	Ainapure *et al.*	2023	Compared preexisting models
8.	DFM: Deep Fusion Model For COVID-19 Vaccine Sentiment Analysis.	Rani and Jain	2023	A deep fusion model (DFM)
9.	Users' Reactions to Announced Vaccines Against COVID-19 Before Marketing In France: Analysis Of Twitter Posts	Dupuy-Zini *et al.*	2023	Built a classifier using a preexisting model
10.	COVID-19 Vaccine Hesitancy: A Global Public Health and Risk Modelling Framework Using an Environmental Deep Neural Network, Sentiment Classification with Text Mining and Emotional Reactions From COVID-19 Vaccination Tweets	Qorib *et al.*	2023	Compared preexisting models
11.	COVID-19 Vaccine Rejection Causes Based on Twitter People's Opinions Analysis Using Deep Learning.	Alotaibi *et al.*	2023	Compared preexisting models
12.	Deep Learning Analysis Of COVID-19 Vaccine Hesitancy and Confidence Expressed on Twitter In 6 High-Income Countries: Longitudinal Observational Study.	Zhou *et al.*	2023	Finetuned a preexisting model
13.	A Novel TCNN–Bi-LSTM Deep Learning Model for Predicting Sentiments of Tweets About COVID-19 Vaccines.	Aslan	2022	TCNN–Bi-LSTM deep learning model
14.	Predicting The Sentiment of South Korean Twitter Users Toward Vaccination After the Emergence Of COVID-19 Omicron Variant Using Deep Learning-Based Natural Language Processing.	Eom *et al.*	2022	Created their own (unnamed) model using preexisting models
15.	Sentiment Analysis of Covid-19 Vaccine with Deep Learning	Nuser *et al*.	2022	Proposed an unnamed hybrid framework
16.	Sentiment Analysis System For COVID-19 Vaccinations Using Data of Twitter	Khalid *et al.*	2022	An unnamed sentiment analysis system
17.	Covid-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset.	Reshi *et al.*	2022	An ensemble model LSTM-GRNN
18.	Competitive Capsule Network Based Sentiment Analysis on Twitter COVID'19 Vaccines	Prabha and Rathipriya	2022	Competitive Capsule Network (CompCapNets)
19.	Covid-19 Vaccine Hesitancy: A Social Media Analysis Using Deep Learning	Nyawa *et al.*	2022	Compared preexisting models
20.	Sentiment Analysis to Extract Public Feelings on COVID-19 Vaccination	Almurtadha *et al.*	2022	An unnamed sentiment analysis system

21.	Enhanced Sentiment Analysis Regarding COVID-19 News from Global Channels.	Ahmad, W. *et al.*	2022	Cov-Att-BiLSTM
22.	KEAHT: A Knowledge-Enriched Attention-Based Hybrid Transformer Model for Social Sentiment Analysis.	Tiwari and Nagpal	2022	KEAHT
23.	Deep Learning-Based Sentiment Analysis Of COVID-19 Vaccination Responses from Twitter Data.	Alam, K.N. *et al*	2021	Compared preexisting models
24.	Sentiment Analysis Of COVID-19 Vaccine Tweets in Indonesia Using Recurrent Neural Network (RNN) Approach	Sandag *et al.*	2021	Compared preexisting models
25.	Machine Learning and Lexical Semantic-Based Sentiment Analysis for Determining The Impacts Of The COVID-19 Vaccine	Alam, S. *et al.*	2021	Used Combinations of Existing Tools
26.	Systematic Delineation Of Media Polarity On COVID-19 Vaccines In Africa: Computational Linguistic Modeling Study	Gbashi *et al.*	2021	Used Combinations of Existing Tools

Table 1: Systematic Review Table – The Twenty-Six Studies Finally Chosen to be Reviewed.

4. Discussion

When it comes to sentiment analysis using neural networks, the machine learning process for studying the sentiments of pieces of text assumes importance. The sentiment analysis pipeline involves a series of machine learning steps, integrating NLP techniques for robust sentiment understanding. It begins with raw training data collection. For instance, a dataset containing product reviews with associated sentiment labels belonging to the set {Positive, Negative, Neutral} can serve as training data. To train its model to successfully recommend a movie or a show to a user, sites like Netflix and Amazon Prime, which utilize recommender systems, the pipeline depends on collecting vast amounts of user data, including viewing history, search queries, and user ratings.

Sampling is the main technique employed for data selection. It is ensured that using a sample would work almost as well as using the entire data, which implies that the sample is representative, and all classes have sufficient data points. Techniques like undersampling the majority classes or SMOTE (Synthetic Minority Over-sampling Technique) are applied to balance the class distribution and handle imbalanced classes in the training dataset. A sample is representative if it has approximately the same property (of interest) as the original set of data. A diverse, holistic, and inclusive dataset is crucial for robust training and model generalization.

Once the data is collected, sentiment annotation is accomplished. This is done through a process where either a team of human annotators label each text sample, or clustering algorithms (unsupervised learning) perform the task, creating a labeled dataset. In text preprocessing and cleansing, the focus is on preparing the data for subsequent analysis. Techniques such as removing stop words, handling punctuation, and stemming and lemmatization are applied. In a study on Twitter sentiment analysis, this step involves handling user mentions, hashtags, and URLs peculiar to the platform.

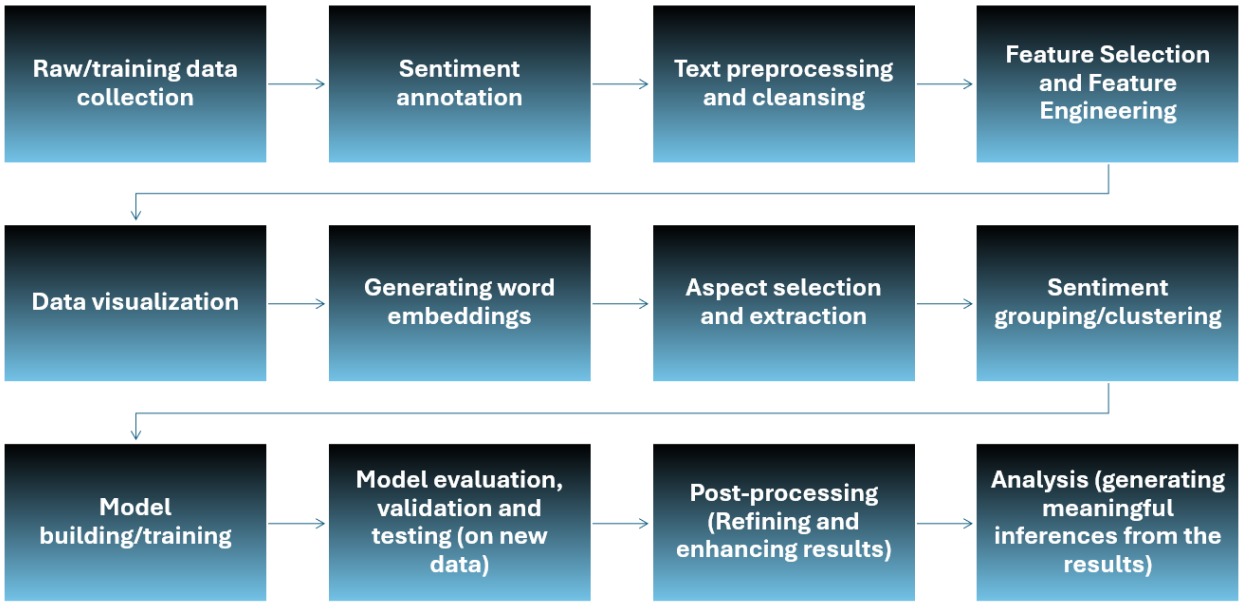


Figure 5: Machine Learning Process for Sentiment Classification and/or other Sentiment Analysis Tasks

Feature selection and engineering involve crafting features that encapsulate syntactic and semantic patterns. Feature selection process reduces the set of features so that it contains only the most informative and useful subset of features from the dataset. The feature engineering process, on the other hand, creates new features and/or modifies existing ones to make them more suitable to the model (Feature Engineering Vs Feature Selection, n.d.). Correlation analysis, chi-square tests, and information gain are filtering methods which help in performing feature selection (Gupta, 2020), and feature normalization, one-hot encoding and creating polynomial features are techniques for feature engineering (Müller and Guido, 2016).

In a study analyzing political discourse, features may include syntactic structures reflecting the complexity of statements. Using n-grams to capture context in a sequence of words or leveraging part-of-speech tagging for syntactic insights can constitute a feature selection and engineering method. Word embeddings, generated in the next step, capture semantic relationships. Word embeddings may be generated using an individual model, called encoder. This is distinguished from the model generated further downstream in the pipeline, the decoder, which decodes the embeddings to generate consequential, workable information (Saxena and Saxena, 2024). Embeddings like Word2Vec or GloVe are employed. In an NLP system, embeddings reveal nuanced relationships between words. In a financial sentiment analysis project, for example, embeddings may serve to reflect market sentiment.

Once the relevant data and features have been selected, the data can be visualized using scatterplots, histograms, heat maps and so on, to see how representative the data is and how balanced the features are. This is crucial in exploratory data analysis. The reasoning behind visualizing data lies in uncovering patterns, anomalies, or imbalances that would affect the performance of the machine learning model that would be created later in the pipeline.

Aspect selection and extraction are crucial for understanding sentiments about specific entities. In a restaurant review analysis, aspects like "service," "food quality," and "ambiance" become key elements. Sentiment grouping and clustering, illustrated through unsupervised techniques like k-means, helps uncover patterns in large datasets. Clustering is done based on data visualization already performed.

Sentiment classification involves training models on the annotated data. Techniques such as RNNs or transformer-based models like BERT and GPT are employed. In an e-commerce context, the classification model can discern sentiments in customer feedback. Model evaluation ensures reliability and robustness. To ensure an effective evaluation, metrics like precision, recall, and F1 score are employed. Validation on a part of training data that was not used in training the model helps us test the consistence of training process. Testing of the model on an entirely new dataset, possibly a set of latest reviews, ensures the model's adaptability to evolving sentiments.

Post-processing fine-tunes results. The task of handling minority classes is performed. Even if the balance is addressed in the training data, imbalanced classes can still impact the evaluation of the model's performance. In imbalanced datasets, accuracy alone might not provide an accurate picture, and F1 score, precision, recall and area under ROC curve become meaningful. For instance, if a binary classification model is trained on a dataset where 90% of instances belong to class A and 10% to class B, a model predicting all instances as class A would still have high accuracy (90%). However, it would perform poorly in identifying instances of class B. Such imbalance is achieved at post-processing stage,

Sentiment thresholds are also adjusted. Post-processing adjustment of thresholds can be crucial in dealing with imbalanced sentiment classes. It allows for optimizing the model's performance based on the specific needs and priorities related to imbalanced class distribution. Sentiment thresholds define the score (confidence level) at which a sentiment label is assigned. In a binary classification, a threshold of 0.5 usually indicates that a score above 0.5 is classified as positive, and below 0.5 as negative. Post-processing involves fine-tuning such thresholds based on the desired trade-off between precision and recall. Setting a higher threshold increases precision (fewer false positives) but may lead to missing some true positives. Lowering the threshold increases recall (capturing more true positives) but may result in more false positives.

In a study on social media sentiment, this step could involve addressing challenges like sarcasm detection. Data visualization can again be carried out at this stage to see how well the model performed. The analysis stage delves into meaningful insights. Advanced visualization tools help understand sentiment trends over time and across different segments. In a healthcare sentiment analysis project, for instance, the analysis reveals evolving sentiments about specific medical treatments.

Sentiment analysis is essentially a data mining process. It helps us mine sentiment data from the target data.

In the case of sentiment analysis, postprocessing and analysis involves assessing the significance of sentiment predictions along with iterative improvement of the model and the depiction of its results based on new data and user feedback. Statistical methods are employed to analyze the

relevance of sentiments in the dataset. Here, sentiments that have a higher impact on decision-making and user experience are identified. Sentiment insights are presented in a visually understandable format. Sentiment trends, patterns, or anomalies are visualized. Sentiment distributions and changes over time may be depicted.

Figure 6: Summary of a Data Mining Process

As iterated before, during postprocessing and analysis, sentiment data is made actionable for users and decision-makers. Insights which can guide decision-making are derived from the results of the sentiment analysis task performed by the model. Sentiments that may require immediate attention and areas where user satisfaction is particularly high or low are highlighted. The presentation of sentiment insights is tailored to the demands of the end user. The needs of the audience are considered, and sentiment information is presented in a format that is most meaningful to them. For instance, summarized reports and visual dashboards are provided that cater to specific user preferences.

While machine learning has automated and, therefore, eased the process of sentiment analysis, and made it more efficacious, deep learning has shown even more promise than other machine learning techniques in this field. Sentiment analysis researchers leverage the ability of neural networks to autonomously extract intricate patterns and representations from complex data. In NLP, deep architectures like convolutional neural networks (CNNs) and RNNs have demonstrated enhanced value in learning hierarchical features and capturing subtle relationships within raw text. This hierarchical understanding is particularly advantageous when dealing with sentiment analysis, where the sentiment expressed in a sentence is often influenced by the sentiments of individual words and their combinations.

Pretrained deep learning models like BERT and GPT-4, which have been repeatedly trained over huge corpora of textual data, have surpassing capacities in sentiment detection, faring well even when faced with sarcasm, context-dependent remarks (for example, saying “Take the leap!” to someone needing encouragement to found a business presents contrasting emotions when uttered before a person standing on the edge of a rooftop), pieces of text with multiple contrasting opinions (for example, she made it but missed the entire show), pieces of text which seem objective but are implicitly subjective (for example, he entered the theater when the show was just about to get over.), and pieces of text containing negation (for example, their trip was bereft of jollity and all that is needed to make an outing memorable).

Neural networks, especially deep learning architectures like Graph Neural Networks (GNNs) and RNNs, excel at learning hierarchical representations of data (Saxena and Saxena, 2024). In sentiment analysis, this allows them to automatically extract meaningful features from raw text or sequences of data. Deep learning models like Transformers, RNNs and LSTM are more efficient

at capturing contextual dependencies within text, considering the sequential nature of language. This is crucial for sentiment analysis, as the sentiment expressed in a sentence often depends on the surrounding words and phrases.

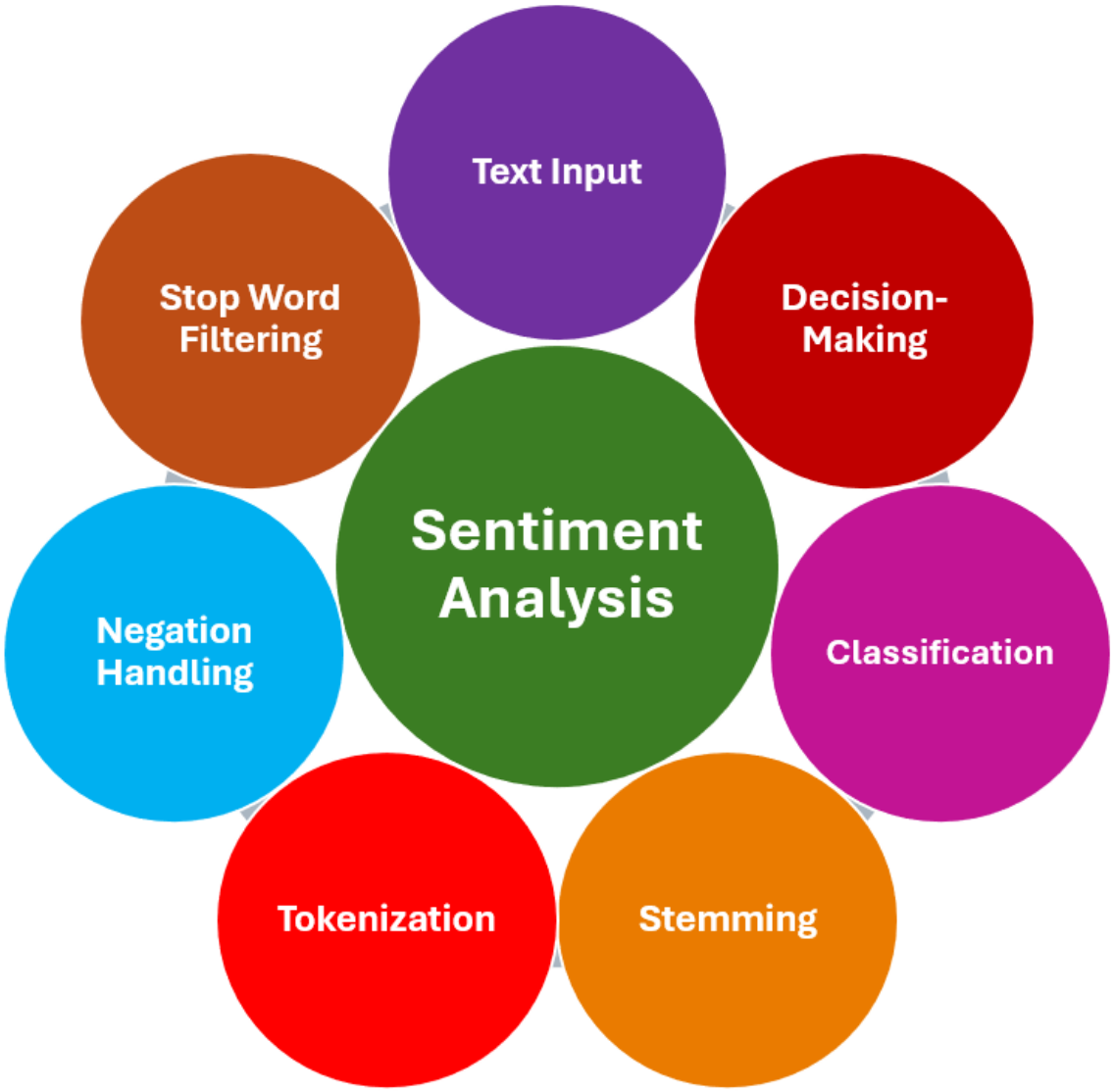


Figure 7: Components of Sentiment Analysis

Pre-trained word embedding generators, such as Word2Vec (a two-layer model) and GloVe, are used to generate vector representations of words in a piece of text. Further, embeddings from transformer models like BERT provide semantically rich representations of words. BERT employs a bidirectional approach to consider the entire context of a word, leading to more accurate representations and improved sentiment analysis outcomes. Generated by the encoder model in a particular task, these embeddings enhance the decoder's understanding of the relationships between words, improving sentiment analysis accuracy.

Sentences and documents can vary in length, and deep learning models, particularly recurrent architectures, can handle variable-length input sequences efficiently. This flexibility is essential for processing text data in sentiment analysis. Non-neural network machine learning models often require manual feature engineering. Neural networks, on the other hand, can automatically learn relevant features, reducing the need for extensive feature engineering efforts. Sentiment analysis tasks can involve complex relationships and subtle expressions. Deep learning models can adapt to such intricacies. They can encapsulate delicate patterns and sentiment variations that are usually challenging for rule-based and traditional machine learning approaches to capture.

Furthermore, the concept of transfer learning has been instrumental in enhancing sentiment analysis performance. Transfer learning allows for leveraging knowledge gained from one task to boost performance on another, even with limited task-specific data. Pre-trained models, such as LLaMA, trained on large corpora for general language understanding, can be fine-tuned for specific tasks with limited labeled data. Transfer learning demonstrates exceptional performance in various NLP tasks, including sentiment analysis. The transfer of knowledge from a pre-trained model to a task-specific model has proven effective in scenarios in which obtaining large, labeled datasets for sentiment analysis might be challenging.

One of the notable advancements in recent years is the incorporation of attention mechanisms into deep learning models. Attention mechanisms enable models to dynamically focus on relevant parts of input sequences, allowing for a more nuanced understanding of contextual information. This attention to specific words or phrases enhances the models' ability to discern sentiment subtleties and contributes to improved accuracy in sentiment classification.

In the context of sentiment analysis, the ability of deep learning models to handle variable-length sequences is a crucial advantage. Unlike traditional machine learning approaches that often struggle with varying text lengths, deep learning models, notably RNNs, can effectively process sequences of different lengths. In doing so, they can accommodate the inherent variability in natural language expressions. Deep learning models also enable end-to-end learning, where the entire sentiment analysis process, from raw input data to final predictions, is learned directly from the data. This minimizes the need for manual intervention in different stages of the pipeline. Some of the latest deep learning approaches to conduct sentiment analysis with regards to COVID-19 vaccines are discussed in this paper.

4.1. Determining Relative Popularity of Various COVID-19 Vaccines

Saeed *et al.*, in their 2023 article, have explored sentiment analysis on social media data related to COVID-19 vaccine. They have focused on exploring Twitter data. Their study aims to gauge public sentiments and emotions regarding different vaccines and their popularity. Their sentiment analysis process involves the collection of tweets through web scraping, while focusing on various machine learning and deep learning models. The researchers have gathered tweets related to COVID-19 vaccines using the Tweepy API. They have used naïve bayes, SVM, decision trees, K-Nearest Neighbors (KNN) algorithm, and a deep neural network (DNN) for sentiment analysis. They claim that the DNN outperformed other models with 97.87% accuracy (Saeed *et al.*, 2023).

The models were trained on 70% of the dataset they generated, and 30% were reserved for evaluation. Standard metrics like accuracy and precision were employed for performance evaluation. The SVM model achieved an accuracy of 94.87%, while Naïve Bayes achieved 83.23%, and the Decision Tree model achieved 76.54%. The study provided options for daily, weekly, and monthly popularity detection of COVID-19 vaccines based on sentiment analysis. Future research suggestions include exploring sentiment analysis in other languages, incorporating hybrid models, and leveraging transfer learning for improved results.

4.2. Sentiment Analysis on COVID-19 Vaccine Boosters in Indonesia

Kristian *et al.*, in their 2023 study, have analyzed social media text related to public's sentiments about COVID-19 booster vaccines in Indonesia. The authors contend that as the COVID-19 pandemic progresses, countries must grapple with waning vaccine effectiveness. This realization has led to booster vaccine recommendations. Indonesia, in response, initiated booster vaccinations, and this study delves into public sentiments on this matter through social media platforms, particularly Twitter and YouTube (Kristian *et al.*, 2023).

Their sentiment analysis process involves data mining, pre-processing, and the application of various machine learning algorithms, including CatBoost Classifier, SVM, Random Forest, Logistic Regression (LR), Gaussian Naïve Bayes (GaussianNB), Bernoulli Naïve Bayes (BernoulliNB), and BiGRU (Bidirectional Gated Recurrent Unit, a deep learning model, which uses gating mechanism on RNNs) for NLP. Imbalanced data scenarios are addressed using SMOTE, Random Over Sampling (ROS), and Random Under Sampling (RUS). The CatBoost Classifier, with ROS, emerged as the top performer, achieving 88% accuracy, with BiGRU also exhibiting strong results at 83% accuracy.

They utilized data crawling to harvest data from YouTube and Twitter. They have employed steps like case folding, removing numbers, punctuation, and special text, spell checking, stemming, and word embedding, for data preprocessing. To tackle imbalanced data, Random Oversampling (ROS), Random Undersampling (RUS) and SMOTE have been used. For sentiment classification, algorithms like CatBoost Classifier, SVM, Random Forest, LR, GaussianNB, BernoulliNB, and BiGRU are applied. For performance evaluation, data is split into 80% training and 20% testing. Evaluation metrics include accuracy, precision, recall, and F1-score. Based on these metrics, they chose the best model.

They found that the CatBoost Classifier with ROS achieved the highest accuracy, with strong metrics for neutral (78%), negative (91%), and positive (99%) sentiments. BiGRU performed well with an overall accuracy of 83%, exhibiting strength in neutral (88%) and negative (91%) sentiments but facing challenges with positive sentiment (72%). The authors acknowledge the diversity of public sentiments on COVID-19 vaccine boosters in Indonesia. They conclude that CatBoost Classifier with ROS stands out as the most accurate method. Data collection permits, dialects in comments, and fitting problems with BiGRU, were challenging according to them. They believe that future research should focus on expanding and cleaning datasets to address fitting issues and explore more advanced data cleaning techniques.

4.3. Sentiment Analysis of Tweet Responses About the Third Booster Dosage

Ghosh *et al.*'s 2023 study focuses on sentiment analysis of user tweet responses regarding the third booster dosage for COVID-19 vaccination. It employs various machine learning algorithms, including naïve bayes, KNN, RNN, and Valence Aware Dictionary for sEntiment Reasoner (VADER). They used a simple sentiment categorization based on a threshold sentiment polarity score, with values above the higher threshold considered positive, values below the lower threshold considered negative, and those in between categorized as neutral.

They discovered that India had the highest incidence of positive sentiments about the third booster and the lowest incidence for negative sentiments for the same. They also found that VADER sentiment analysis stood out with impressive results, achieving 97% accuracy, 92% precision, and 95% recall compared to other models. RNN, their only deep learning model, achieved an accuracy of 81%. They highlight the potential of VADER as a cost-effective, unsupervised learning approach for sentiment analysis in healthcare discussions. Their study suggests the applicability of their model to other infectious diseases in the future.

4.4. Identifying Twitter Opinions about COVID-19 Vaccine

Kul and Sayar, in 2022, have discussed a real-time implementation of a system that identifies Twitter opinions about the COVID-19 Vaccine using the big data tool Hadoop. They have divided all tweets into three categories: Positive, Neutral, and Negative. They conducted **s**entiment analysis using various machine learning techniques, including LR, Random Forest, DNN and CNN. They acquired Twitter data. The data was cleaned and transformed using Hadoop. Mentions, hashtags, links, and retweets were removed before stemming, punctuation removal, tokenization, removal of stopwords and multiple whitespaces. Hadoop MapReduce was used for distributed processing (Kul and Sayar, 2022).

The processed data was labeled using TextBlob. The results were stored in Hadoop Distributed File System (HDFS). They found that DNN performed better than CNN while LR performed better than RF. The highest accuracy their model achieved was 90.42%. They acknowledged the need for rapid data processing due to its accumulating nature. For potentially improved outcomes, they mentioned the use of alternative data vectorization and vector sizes, especially in conjunction with models like LSTM. Among their deep learning algorithms, compared to the 81% training accuracy of their CNN algorithm, their DNN algorithm gave better training accuracy at 0.84 or 84%. However, their best model proved to be their bigram-based random forest with an accuracy of 88%.

Aslan, in her 2022 article, describes a TCNN–Bi-LSTM deep learning model that was used for predicting sentiments of tweets about COVID-19 vaccines. To begin with, the raw dataset underwent preprocessing to enhance data quality. FastText word embedding was applied to extract strong features from the dataset, leveraging location-based and sub-word information features. A new dataset was collected from Twitter for experimental results. The study aims to facilitate the detection of inappropriate, incomplete, and erroneous information about vaccination. Posts

containing false information were tagged with content warnings to mitigate the impact of incorrect information (Aslan, 2022).

A hybrid CNN and Bi-LSTM model named TCNN–Bi-LSTM is proposed. A two-stage convolutional layer (TCNN) was introduced to extract more meaningful local features, followed by a Bi-LSTM model capturing long-distance dependencies. Experimental results exemplified how the proposed method demonstrates promising results compared to baseline deep learning and machine learning models. The TCNN–Bi-LSTM model with FastText outperformed all other baseline deep learning models. FastText word embedding technique achieved the best performance among other word embedding techniques (Glove and TF-IDF). The accuracy and loss curves of the proposed TCNN–Bi-LSTM model with FastText, Glove, and TF-IDF consistently beat other models. It demonstrated higher performance scores across all word vectors, with an accuracy of 98% for FastText.

In another paper, crafted by Eom *et al.* in 2022, Twitter data in South Korea was collected. The authors focused on keywords related to vaccines after the Omicron variant outbreak (27th November 2021 to 14th February 2022). They conducted network analysis to explore the relationship between potential keywords associated with vaccination post-Omicron on Twitter (Eom *et al.*, 2022). They developed a sentiment analysis model for speech regarding vaccination after the Omicron variant using five algorithms: SVM, RNNs, LSTMs, BERT, and KoBERT (Korean BERT). The authors conducted topic modeling on Twitter users regarding vaccination post-Omicron. They identified coexisting sentiments of expectation, distrust, and anxiety related to vaccine efficacy, with keywords like "Omicron," "vaccine," "vaccine inequality," and "breakthrough infection."

The authors employed CONCOR (CONvergence of iterated CORrelations) analysis to identify keyword factor types related to Omicron and vaccines. They discovered clusters such as Omicron and vaccination status, Infection and treatment, Vaccine effectiveness and the need for vaccine research, and increase in confirmed cases and deaths. They evaluated the performance of sentiment classification models with metrics like accuracy, precision, and F1-score. KoBERT beat other models, showing a 5% improvement over BERT, likely due to optimization for the Korean language. KoBERT demonstrated the best performance across accuracy (71%), precision, and F1 score metrics. KoBERT, optimized for the Korean language, demonstrated superior predictive performance in sentiment analysis.

Ahmed, S. *et al.* published a study in 2023 which aims to explore public opinion dynamics on COVID-19 vaccination through sentiment analysis of related tweets. They have proposed a sentiment analysis framework for COVID-19 vaccination-related tweets. They investigated TF-IDF, Bag of Words (BoW), Word2Vec, and a combination of TF-IDF and BoW. They employed various classifiers, including Random Forest, Gradient Boosting Machine (GBM), Extra Tree Classifier (ETC), LR, Naïve Bayes, SGD, Multilayer Perceptron (MLP), CNN, BERT, LSTM, and RNN. They utilized a large dataset of COVID-19 vaccination-related tweets and employed manual and TextBlob-based labeling for performance evaluation (Ahmed, S. *et al.*, 2023).

For feature engineering, they examined the influence of TF-IDF, BoW, Word2Vec, and feature union of TF-IDF and BoW on model accuracy. ETC with BoW features showed the highest accuracy (92%) and was identified as the most suitable approach for sentiment analysis of COVID-19 vaccine-related tweets. BoW was generally the most effective feature representation method across all classifiers. Deep learning models, including BERT, LSTM, and CNN, did not perform as well as non-deep learning models. This was due to a small dataset size, sparsity, and the need for extensive hyperparameter tuning.

A study conducted by Jain *et al.*, published in 2023, focuses on sentiment analysis of tweets using machine learning and deep learning techniques to categorize them into positive and negative sentiments. As a data source, they used a Kaggle dataset consisting of categorized tweets with positive and negative sentiments. As to vector representation techniques, they utilized BoW and TF-IDF to convert tweet reviews into vector representations. The machine learning algorithms they deployed include LR, Naïve Bayes, SVM, among others, while their deep learning algorithms are LSTM and BERT. They found that SVM achieved the highest accuracy among their traditional machine learning models at 88.7989%. They obtained 90.42% accuracy with BERT which was higher than that of LSTM. Their analysis showed that BERT outshone other models (Jain *et al.*, 2023).

The research of Alam, K.N. *et al.*, published in 2021, aims to analyze sentiments regarding various vaccines based on Twitter data. Their data collection period spanned seven months' worth of tweets on common vaccines worldwide. They utilized VADER for initial sentiment analysis and categorized sentiments into positive, negative, and neutral groups. For predictive modeling, they used RNN architectures (LSTM and Bi-LSTM) to assess sentiment performance. They used accuracy, precision, recall, F-1 score, and confusion matrix to validate the models. With VADER, 33.96 percent of tweets were found to be positive, 17.55 percent to be negative and 48.49 percent to be neutral. Their accuracy with LSTM was 90.59% and with Bi-LSTM was 90.83% (Alam K. N. *et al.*, 2021).

Nuser *et al.*'s 2022 paper aims to analyze user sentiment towards the COVID-19 vaccine using a hybrid deep learning model. The proposed model combines CNN for feature extraction and LSTM for monitoring long-term dependencies between words (Nuser *et al.*, 2022). The proposed network topology settings contribute to high performance in sentiment analysis of COVID-19 vaccine tweets. Their dataset constituted 13,190 tweets, which were used for extensive experiments. As an evaluation metric, accuracy is used for assessing model performance. It was found that the hybrid model achieved an accuracy of 83% and outperformed the accuracy of both CNN and LSTM algorithms individually.

Khalid *et al.* published a paper in 2022 proposed a system based on Bi-LSTM model. They utilized data extracted from Twitter spanning from January 1st to September 3rd, 2021, encompassing 131,268 tweets. The overall validation accuracy of their model was found to be 74.92% (Khalid *et al.*, 2022).

A study presented by Santag *et al*. in 2021 described how sentiment analysis was performed on COVID-19 vaccine-related issues in Indonesia using a variety of RNNs and traditional machine

learning methods. Data was collected through Twitter API crawling. Among traditional machine learning models, SVM demonstrated the highest accuracy, precision, and recall among traditional machine learning methods, with an RMSE value of 0.117 (Santag *et al*., 2021). The types of RNN algorithms tested using LSTM approach included simple RNN, Bi-LSTM, and GRU. All of them exhibited similar performance with an accuracy of 91%, a recall of 91% and a precision also of 91%, The RMSE was 0.085.

Ainapure *et al.*'s 2023 study investigated the sentiments of Indian citizens regarding the COVID-19 pandemic and vaccination drive. The study utilizes text messages from Twitter. It utilizes both deep learning (Bi-LSTM and GRU) and lexicon-based techniques for sentiment classification. VADER and NRCLex tools were used for polarity classification (Ainapure *et al.*, 2023). Bi-LSTM and GRU were employed for predictions. On the COVID-19 dataset, Bi-LSTM achieved an accuracy of 92.70% and GRU demonstrated an accuracy of 91.24%. On vaccination Tweets, Bi-LSTM's accuracy was 92.48% and GRU's accuracy was 93.03%.

Reshi *et al.*, in 2022, published a study which aimed to analyze global sentiments regarding COVID-19 vaccination through a large-scale Twitter dataset. The evaluation involves various lexicon-based methods and machine/deep learning models for sentiment analysis (Reshi *et al.*, 2022). They applied NLP and machine learning for sentiment analysis. They evaluated different lexicon-based methods, including TextBlob, VADER, and AFINN. They proposed an ensemble model named LSTM-GRNN for sentiment classification, which combines LSTM, GRU and RNN. Among lexicon-based methods, TextBlob demonstrated better results compared to VADER and AFINN for sentiment analysis. The ensemble LSTM-GRNN model accomplished an accuracy of 95%, outdoing both traditional machine learning and deep learning models studied in the research.

Rani and Jain in 2023 suggested a deep fusion model which amalgamated deep learning models with a meta-classifier. This was done to beat the challenge of contextual polarity ambiguity in user-generated data. It was able to identify sentiment polarity with accuracy and improved the categorization of COVID-19 vaccines and omicron variant tweets (Rani and Jain, 2023). The deep learning Models they used included LSTM, GRU, CNN, Bi-LSTM, Stacked LSTM, Stacked GRU, a combination of LSTM and CNN, and a combination of GRU and CNN. The accuracy LSTM achieved for vaccine-related discourse was 78.13% and for Omicron-related discourse was 79.21%. GRU achieved a similar performance to LSTM with lower training time. BiLSTM outperformed LSTM and GRU with an accuracy of 81.53% for vaccine-related discourse and 80.13% for Omicron-related discourse. CNN failed to show performance improvement. Stacked and combination models showed better performance than individual models. Their DFM bested all other models with an overall accuracy of up to 88%.

Prabha and Rathipriya published a study in 2022 examining tweets related to COVID-19 vaccination to extract sentiment information. They utilized traditional machine learning models and deep learning models. The traditional machine learning models they used were Random Forest, LR, SVC, and K-Neighbor Classifier. The deep learning models they used were CNN, LSTM, CNN-Static (Static Convolutional Neural Network), Bi-LSTM, Emdd + Conv (Embedding plus Convolution), CapsNets (Capsule Networks), and their proposed approach (CompCapNets or

Competitive Capsule Networks). With a mean accuracy of 0.9817 and a median accuracy of 0.98 over all datasets, CompCapNets outshone other models (Prabha and Rathipriya, 2022).

Guerdoux *et al.*, in 2022, published a report about a preliminary study comparing the inference times of a deep learning model for predicting Twitter users' opinions on COVID vaccines. They implemented the model using PyTorch and CamemBERT, a French version of BERT. The study aimed to evaluate the impact of using TorchServe, a library for serving PyTorch models, on inference times. In their related study, accomplished previously but published later than their 2022 study, they implemented a classifier based on CamemBERT for forecasting the sentiment of Twitter users about potential vaccines against COVID-19. They extracted nearly 350,000 French tweets related to these keywords. Their dataset contained a total of 16% positive tweets, 41% negative tweets, and 43% neutral tweets. The model achieved an F1-score of 0.75 (Dupuy-Zini *et al.*, 2023; Guerdoux *et al.*, 2022).

4.5. Sentiment Analysis to Gauge COVID-19 Vaccine Hesitancy and Acceptance

Qorib *et al.*'s study in 2023 focuses on sentiment analysis and emotion classification of public tweets related to COVID-19 vaccine hesitancy. The researchers conducted daily downloads of public tweets from Twitter using the Twitter API. The preprocessing steps included stemming, lemmatization, and sentiment labeling using the NRCLexicon (NRC Word-Emotion Association Lexicon) technique. Tweets were categorized into ten classes, encompassing positive sentiment, negative sentiment, and eight basic emotions. A statistical analysis, specifically a t-test, was employed to assess the statistical significance among basic emotions. The results indicated strong relationships between certain emotion pairs, such as joy–sadness, trust–disgust, fear–anger, surprise–anticipation, and negative–positive. The study utilized various deep neural network models, including 1DCNN (one dimensional CNN), LSTM, Multiple-Layer Perceptron (MLP), and BERT, for sentiment and emotion classification (Qorib *et al.*, 2023).

The achieved accuracies were 88.6% for 1DCNN, 89.93% for LSTM, and 96.71% for BERT in COVID-19 sentiment analysis. Emotion classification was extended beyond basic emotions to include complex ones like aggressiveness, contempt, remorse, disapproval, awe, submission, love, and optimism. The research provides insights into the evolving sentiment dynamics surrounding COVID-19 vaccines, inferring that people's sentiment has become more positive over time. The study concludes that artificial neural network techniques, particularly BERT, demonstrated superior computational performance in sentiment analysis and model validation compared to traditional supervised learning methods.

In another article, which was published in 2023, Alotaibi *et al.*, have presented a comprehensive analysis of public sentiments towards COVID-19 vaccines through multi-class sentiment analysis. Twitter data from January to December 2020 related to COVID-19 vaccines was collected from Kaggle. Preprocessing steps included removing non-English tweets, handling null values, removing duplicates, and standard text preprocessing techniques like stemming, lemmatization, and tokenization. Machine learning classifiers used included LR, SVM, KNN Decision Tree, Multinomial Naïve Bayes (MNB), and Random Forest. Stochastic Gradient Descent (SGD) and

Gradient Boosting were used to minimize loss. Deep learning models which were employed included RNN, LSTM, GRU, RNN-LSTM, and RNN-GRU (Alotaibi *et al*., 2023).

Negative tweets were analyzed to identify rejection causes using LDA. Five rejection causes the authors identified were, Lack of safety, Side effect, Production problem, Fake news and Misinformation, and Cost. Machine learning classifiers (LR, SVM, KNN, Decision Trees, MNB, Random Forests, XGBoost) and deep learning models (Simple RNN, LSTM, GRU, RNN-LSTM, RNN-GRU) were used to classify negative tweets based on rejection causes. The authors argue that Decision Tree and GRU yielded the best accuracy rates of 92.26% and 96.83%, respectively, for classifying public opinions on COVID-19 vaccines. LR and LSTM achieved accuracies of 89.97% and 91.66%, respectively.

A study conducted by Zhou *et al.* in 2023 endeavored to explain spatiotemporal distribution of COVID-19 vaccine hesitancy along with vaccine confidence expressed on Twitter across a number of major English-speaking countries throughout the period of the COVID-19 pandemic. The authors collected 5,257,385 English-language tweets in conjunction with COVID-19 vaccination from January 1, 2020, to June 30, 2022, in six countries: the United States, the United Kingdom, Australia, New Zealand, Canada, and Ireland. They developed transformer-based deep learning models which were trained to classify tweets into categories indicating intent to accept or reject COVID-19 vaccination and belief in the effectiveness and in the unsafety of the vaccine. They scrutinized sociodemographic factors linked to COVID-19 vaccine hesitancy as well as confidence in the United States employing bivariate and multivariable linear regressions (Zhou *et al.*, 2023).

They adopted the COVID-Twitter-BERT (CT-BERT) model, specially pretrained on COVID-19–related tweets, to improve language interpretation within the specific domain. CT-BERT was fine-tuned using a manually annotated dataset of 8,073 tweets on COVID-19 vaccines. The annotated tweets were divided into a training set (80%), a development set (10%), and a test set (10%). They used average deep learning prediction of an individual's tweets to measure vaccine hesitancy. Spatiotemporal tendencies were determined as the average of all individuals within a specific time period, place, or mentioning a specific vaccine manufacturer.

They concluded that the six countries whose tweets they studied experienced similar evolving trends of COVID-19 vaccine hesitancy and confidence. They found that the prevalence of intent to agree to COVID-19 vaccination decreased from 71.38% (March 2020) to 34.85% (June 2022) with some fluctuations. They further discovered that the belief in COVID-19 vaccines being unsafe continuously rose from 2.84% (March 2020) to 21.27% (June 2022). As to the specific vaccines, they found that Moderna and Pfizer reached higher acceptance rates (>50%) in Ireland, the United States, and Canada, AstraZeneca vaccine had the highest acceptance rate in the United Kingdom (50.48%), while Johnson & Johnson vaccine was most accepted in Ireland (60.04%).

A study conducted by Nyawa *et al.* in 2022 identifies tweets linked to vaccine hesitancy during the COVID-19 pandemic. The study compared various models, including traditional machine learning models (LR, Random Forests, Support Vector Classifier (SVC), KNN, Gradient Descent, Decision Trees, Gboost, AdaBoost), and deep learning models (LSTM, RNN, and a light version of LSTM). Deep learning models outperformed traditional machine learning models in detecting

vaccine-hesitant messages on social media. The Random Forests model achieved the highest accuracy among traditional models (83%), while LSTM models achieved an accuracy rate of 86%, highest among all models.

Almurtadha *et al.*, in 2022, published a study which used Twitter API to retrieve Arabic tweets. They then assessed public acceptance of COVID-19 vaccination against disease using NLP and deep learning methods. Their results depicted a generally positive opinion on COVID-19 vaccination.

4.6. Finding Effects of COVID-19 Vaccine

Alam, S. *et al.*, in their 2021 study, have provided an analysis of public sentiments regarding the COVID-19 vaccine using Twitter data obtained from Kaggle spanning from January to December 2020. The data underwent preprocessing, including cleaning, and was labeled based on textual sentiment using TextBlob and VADER lexical semantic methods. Various machine learning methods, such as RNN, CNN, LR, and a merged model (RNN+CNN), were employed to analyze public sentiments and visualize concerns related to COVID-19 vaccination throughout 2020 (Alam *et al.*, 2021).

The methodology involves pre-processing the Twitter data, labeling it using TextBlob and VADER, and training and testing machine learning methods including RNN, CNN, LR, and their merged model. The results from both TextBlob and VADER are compared to achieve the highest accuracy and better understand the reasons behind them. The study discusses the performance of the RNN model, LR, CNN, and the merged model on labeled tweet data by utilizing TextBlob and VADER. The accuracy rates of these classifiers were compared, with the RNN model achieving the highest accuracy rates of 99.34% for TextBlob and 98.20% for VADER, while LR yielded the lowest accuracy.

4.7. Sentiment Analysis regarding COVID-19 News and Vaccines

Ahmad *et al.* in 2022 presented a novel approach named Cov-Att-BiLSTM for sentiment analysis of COVID-19 news headlines using DNNs. The primary goal is to develop an effective model that can accurately capture public sentiment regarding COVID-19 news and vaccine. The proposed approach integrates semantic-level data labeling. The authors gather Twitter datasets from six global news channels (BBC, CGTN, CNN, DW, RT, France24) using the Twitter API. All tweets posted by these channels between January 01, 2020, and December 31, 2021, are retrieved, resulting in a total of 328,370 tweets. After filtering for COVID-19 related keywords, 73,138 tweets are identified for analysis (Ahmad *et al.*, 2022).

Data cleaning and preprocessing are performed using the NLTK library in Python. The text is cleaned by removing punctuation, URLs, and mentions. Hashtags are processed by removing the character representing the hashtag, and all text is converted to lowercase. The cleaned data is then used for sentiment analysis. Sentiments are labeled on the data using three pre-trained models: TextBlob, Vader, and ROBERTa-base. These models assign sentiment scores or labels to the tweets based on their content and context.

The proposed neural network architecture, Cov-Att-BiLSTM, incorporates attention mechanisms, Bidirectional Long Short-Term Memory (BiLSTM) layers, and an embedding layer. The attention mechanism is applied to each representation of the BiLSTM layer, determining the importance of each word in the context. The paper conducts a comparative analysis with various machine learning classifiers (SVM, LR, KNN, Bernoulli Naïve Bayes, Decision Trees) and the proposed Cov-Att-BiLSTM model. The classifiers are evaluated using different embedding methods, including Word2Vec, FastText, and GloVe. Performance metrics such as Accuracy, Precision, Recall, and F1 score are used to assess the models.

The Cov-Att-BiLSTM model outperforms other classifiers, achieving the highest accuracy, precision, recall, and F1 score on the testing dataset. The choice of Word2Vec embedding further improves the F1 score compared to GloVe. The sentiment analysis is extended to news related to different COVID-19 vaccines. Pfizer vaccine news has the highest positive rate, while Sinopharm vaccine news has the lowest negative rate and the highest neutral rate.

4.8. Attention-based Transformer for Sentiment Analysis

Tiwari and Nagpal's article, published in 2022, introduces an innovative model, the Knowledge-Enriched Attention-based Hybrid Transformer (KEAHT), for sentiment analysis. KEAHT overcomes limitations of DNNs and leverages explicit knowledge from LDA topic modeling and a lexicalized domain ontology (Tiwari and Nagpal, 2022). It utilizes pre-trained BERT for accurate text analysis. Benchmark datasets, "COVID-19-Vaccine-Labelled-Tweets" and "Indian-Farmer-Protest-Labelled-Tweets," are presented by the authors. COVID-19-Vaccine-Labelled-Tweets dataset had a total of 2988 data points, with 1465 positive, 1249 neutral and 274 negative data points, while the farmer protest dataset had 1239 positive, 1483 neutral and 621 negative data points. KEAHT model achieved 92.84% training and 90.63% testing accuracy for COVID-19 vaccine-related tweets, and 92.63% training and 81.49% testing accuracy for Indian farmer protest-related tweets.

4.9. COVID-19 Vaccine Sentiment Analysis in Africa

Gbashi *et al.*, in 2021, published a study which delineated their experiments to mine COVID-19 vaccine sentiment in Africa. Their study examined sentiments expressed in media communications, including Google News headlines and snippets and Twitter posts. 637 Twitter posts and 569 Google News headlines/snippets obtained between February 2 and May 5, 2020, were used as data. TextBlob, VADER, and Word2Vec combined with a Bi-LSTM were utilized for sentiment analysis. The authors found that sentiments expressed in Google News and Twitter during the specified period were predominantly passive or positive towards COVID-19 vaccines in Africa. Their model achieved an accuracy of 92%.

5. Results

Out of the twenty-six studies investigated, twelve studies have explicitly discussed proposing new (or their own) models for performing all or some of their tasks they intended to perform as part of their reviewed research. Some of these models have been given specific labels and epithets while others have not. However, since all studies involve at least some level of original (a few of the

articles may even be considered to contain seminal research) work, the unambiguous discussion as to proposing a new model becomes somewhat irrelevant.

Twenty of the appraised studies involve an element of comparison while six focus on deploying a single model. Among the twenty studies which show a direct comparison of different machine learning models, fourteen studies specifically pitted traditional machine learning models against deep learning models, while the remaining six only compared two or more deep learning models with one another.

In four (28.57%) out of the fourteen studies which involved comparison between the performance of traditional machine learning models and that of deep learning models, traditional machine learning models exhibited higher accuracy and/or other metrics when compared to deep learning models, while the rest (71.43%) displayed better statistics for deep learning models. This suggests a prevailing trend in favor of deep learning models. However, the choice between traditional and deep learning models may still depend on specific factors like the characteristics of the datasets.

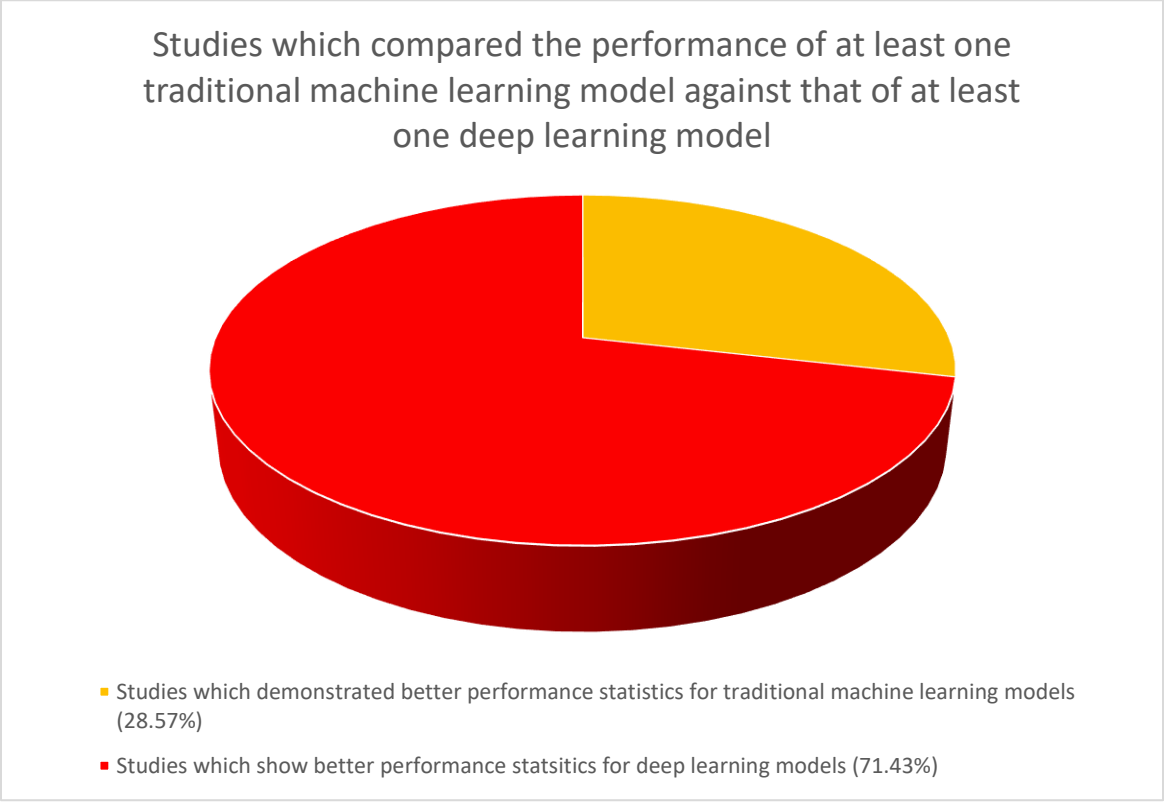


Figure 8: Evaluation of Model Effectiveness: Traditional vs. Deep Learning Approaches

6. Conclusion

The COVID-19 disease began in 2019 and turned into a pandemic in 2020. It came with a series of possible offshoots, including rare ones like its potential involvement in triggering autophagy or in promoting autoimmune disease (Li *et al.*, 2021; Peng *et al.*, 2023; Saxena and Saxena, 2023).

While the mortality from the disease was not the preponderant concern, its transmittivity, both pre-symptomatic and post-symptomatic, was surpassing. In view of its capacity to spread from person to person and infect communities in quick succession, the pharmaceutical industry as well as the governments accelerated their efforts to manufacture vaccines for achieving control on the spread of the disease. While the pandemic was itself a major trigger of panic, the rushed drug trials and rollout of the vaccines raised further concerns among the public. The safety and efficacy of the vaccines became the centerpieces of public discourse, and a plethora of opinions was expressed across the wide tapestry of the virtual world, specifically the social media. The discourse gained velocity in 2020 and matured in 2021, and it was around that time that studies on sentiment analysis on COVID-19 vaccine opinions started getting launched. Figure 9 shows the trend of how the number of published studies about COVID-19 vaccine sentiment increased since 2021.

Interestingly, the vaccines were later found to have some efficacy in assuaging COVID-19's rarer adverse byproducts, like fanning the flames of autoimmune disease (Peng *et al.*, 2023). Furthermore, some concerns about the vaccines were indeed genuine as the vaccines came with their own suite of potential side effects, including the risk of instigating Guillain-Barré syndrome and idiopathic thrombocytopenia purpura (Shah *et al.*, 2022; Yazdani *et al.*, 2023), both of which are themselves autoimmune disorders.

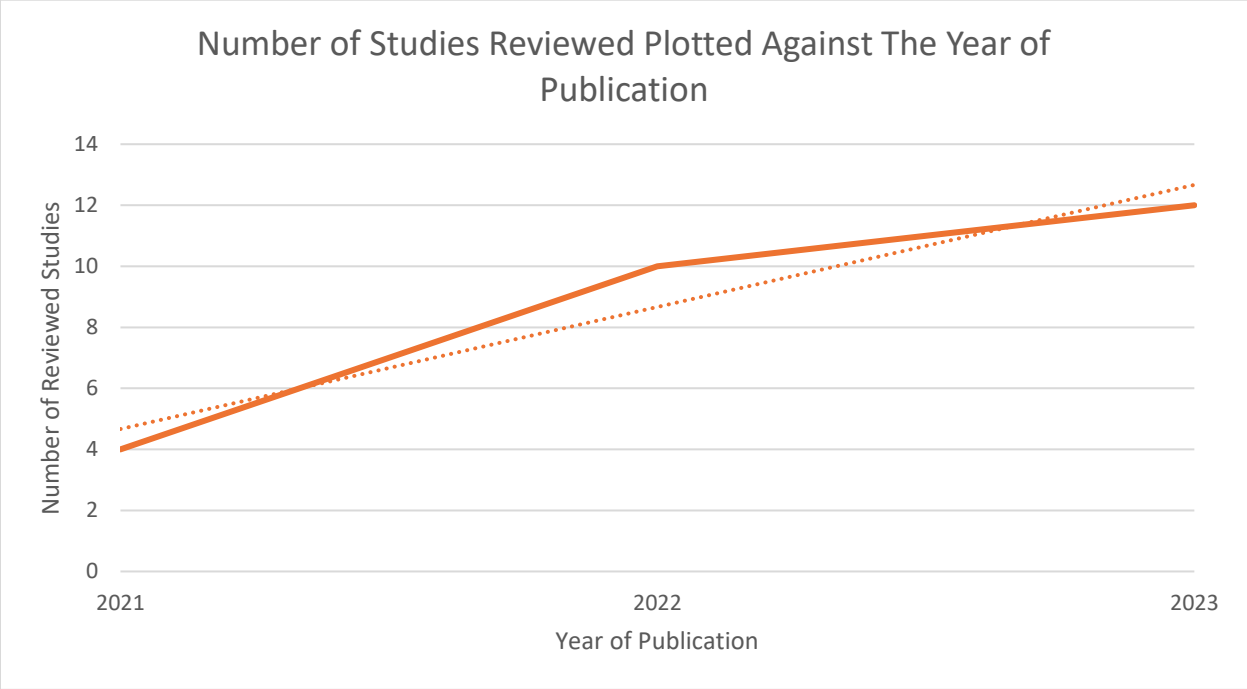


Figure 9: The Growth in The Number of Published Studies About COVID-19 Vaccine Sentiment Over the Past Three Years

We reviewed twenty-six publications which involved deep learning for analysis of sentiments about COVID-19 vaccines. We aimed to find out how deep learning had made the procedure of sentiment analysis related to COVID-19 vaccines better and more accurate, and we demonstrated the same using our in-depth secondary study. It was observed that in the recent trend in research studies about sentiment analysis regarding COVID-19 vaccines, there is an increasing incidence

of innovative implementation models which improve the sentiment analysis process. The cardinal goals of sentiment analysis about opinions in relation to COVID-19 vaccines have been to assist decision making, specifically with regards to allaying and mitigating concerns that have arisen among the common populace in the context of the vaccines. While the relevance of COVID-19 vaccine discourse has somewhat waned since the development of herd immunity, new variants of SARS-CoV-2 virus continue to be discovered to be circulating among the population (Schnirring, 2024; Katella, 2024). Although uncertain, it is still possible that new vaccines are needed if a variant of SARS-CoV-2 virus emerges which is resistant to existing vaccines and is truly a cause of concern. In this regard, the continued evolution of sentiment analysis models holds promise for informed decision-making and public engagement in the ongoing discourse surrounding COVID-19 vaccines.

Declaration of Funding

The author(s) did not receive any funding for this article.

Declaration of Conflict of Interest

The author(s) declare(s) no conflict of interest.

Acknowledgments

The authors are grateful to Dr. Naresh Macker for his aid and support during the drafting of this manuscript.

References

1. Chakraborty, A. K., & Das, S. (2023). Chapter 8 - A comparative study of a novel approach with baseline attributes leading to sentiment analysis of COVID-19 tweets (D. Das, A. K. Kolya, A. Basu, & S. Sarkar, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/pii/B9780323905350000136
2. Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic Reviews in Sentiment Analysis: A Tertiary Study. Artificial Intelligence Review, 54. https://doi.org/10.1007/s10462-021-09973-3
3. Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. International Journal of Intelligent Systems and Applications, 4(10). https://doi.org/10.5815/ijisa.2012.10.01
4. Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment Analysis Is a Big Suitcase. IEEE Intelligent Systems, 32(6), 74–80. https://doi.org/10.1109/mis.2017.4531228
5. Rozado, D. (2020). Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. PLOS ONE, 15(4), e0231189. https://doi.org/10.1371/journal.pone.0231189
6. Thet, T. T., Na, J.-C., & Khoo, C. S. G. (2010). Aspect-Based Sentiment Analysis of Movie Reviews on Discussion Boards. Journal of Information Science, 36(6), 823–848. https://doi.org/10.1177/0165551510388123

7. Dunbar, R. I. M. (1998). The Social Brain Hypothesis. Evolutionary Anthropology: Issues, News, and Reviews, 6(5), 178–190. https://doi.org/10.1002/(sici)1520-6505(1998)6:5%3C178::aid-evan5%3E3.0.co;2-8
8. Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers. Computer Science Review, 27, 16–32. https://doi.org/10.1016/j.cosrev.2017.10.002
9. Thorley, J. (2004). Athenian Democracy. In Google Books. Psychology Press. https://books.google.com/books/about/Athenian_Democracy.html?id=d0g9M8kM53sC
10. Gendron, M., & Barrett, L. F. (2009). Reconstructing the Past: A Century of Ideas About Emotion in Psychology. Emotion Review, 1(4), 316–339. https://doi.org/10.1177/1754073909338877
11. Droba, D. D. (1931). Methods Used for Measuring Public Opinion. American Journal of Sociology, 37(3), 410–423. https://doi.org/10.1086/215733
12. Public Opinion Quarterly. OUP Academic. https://academic.oup.com/poq
13. Shannon, C. E. (1948, July 15). A Mathematical Theory of Communication. Web.archive.org. https://web.archive.org/web/19980715013250/http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf
14. Shannon, C. E. (1951). Prediction and Entropy of Printed English. Bell System Technical Journal, 30(1), 50–64. https://doi.org/10.1002/j.1538-7305.1951.tb01366.x
15. Rosenfeld, R. (2000). Two decades of statistical language modeling: where do we go from here? Proceedings of the IEEE, 88(8), 1270–1278. https://doi.org/10.1109/5.880083
16. Singleton, J. (1974). The Explanatory Power of Chomsky's Transformational Generative Grammar. Mind, 83(331), 429–431. https://www.jstor.org/stable/2252745
17. Pronko, N. H. (1946). Language and Psycholinguistics: A Review APA PsycNet. Psycnet.apa.org. https://psycnet.apa.org/record/1946-03243-001
18. Hutchins, W. J. (2004). The Georgetown-IBM Experiment Demonstrated in January 1954. Machine Translation: From Real Users to Research, 102–114. https://doi.org/10.1007/978-3-540-30194-3_12
19. Buckley, C., Salton, G., & Allan, J. (1993). The SMART information retrieval project. HLT '93: Proceedings of the Workshop on Human Language Technology, March 1993, Pages 392. https://doi.org/10.3115/1075671.1075771
20. Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613–620. https://doi.org/10.1145/361219.361220
21. Wong, S. K. M., Ziarko, W., & Wong, P. C. N. (1985). Generalized vector spaces model in information retrieval. Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '85. https://doi.org/10.1145/253495.253506
22. Forgy, C. L. (1982). Rete: A fast algorithm for the many pattern/many object pattern match problem. Artificial Intelligence, 19(1), 17–37. https://doi.org/10.1016/0004-3702(82)90020-0
23. Scott, S. L. (1994). Optimal Pattern Distributions in Rete-based Production Systems. N95-19755. https://ntrs.nasa.gov/api/citations/19950013339/downloads/19950013339.pdf

24. Lenat, D., Prakash, M., & Shepherd , M. (1986). CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks. AI Magazine. https://doi.org/10.5555/13432.13435
25. Streeter, M. L. (1972). DOC, 1971: A Chinese dialect dictionary on computer. Computers and the Humanities, 6(5), 259–270. https://doi.org/10.1007/bf02404242
26. Jodai, H. (2011). An Introduction to Psycholinguistics. https://files.eric.ed.gov/fulltext/ED521774.pdf
27. Roseman, I. J. (1996). Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory. Cognition & Emotion, 10(3), 241–278. https://doi.org/10.1080/026999396380240
28. Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45. https://doi.org/10.1145/365153.365168
29. Chatbots: A story of evolution. (n.d.). Influentialfuture.com. Retrieved February 13, 2024, from https://influentialfuture.com/pages-65-chatbots-a-story-of-evolution
30. Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2(100006). Sciencedirect. https://doi.org/10.1016/j.mlwa.2020.100006
31. Botsplash. (2022, April 19). Chatbots: A Brief History Part I - 1960s to 1990s. Www.botsplash.com. https://www.botsplash.com/post/chatbots-a-brief-history
32. Matthiessen, C. M. I. M., Wang, B., Mwinlaaru, I. N., & Ma, Y. (2018). "The axial rethink" – making sense of language: an interview with Christian M.I.M. Matthiessen. Functional Linguistics, 5(1). https://doi.org/10.1186/s40554-018-0058-8
33. Mann, W. C., Matthiessen, C. M. I. M., & Thompson, S. A. (1989). Rhetorical Structure Theory and Text Analysis. Pragmatics & Beyond, 39. https://doi.org/10.1075/pbns.16.04man
34. Métais, E., Meziane, F., Sararee, M., Sugumaran, V., & Vadera, S. (2013). Natural Language Processing and Information Systems: 18th International Conference on Applications of Natural Language to Information Systems, NLDB 2013, Salford, UK, Proceedings. In Google Books. Springer Berlin Heidelberg. https://books.google.com/books/about/Natural_Language_Processing_and_Informat.html?id=0KXQnAEACAAJ
35. Morimoto, T. (1993). Automatic Speech Translation at ATR. https://aclanthology.org/1993.mtsummit-1.8.pdf
36. Nakamura, S. (2006). The ATR Multilingual Speech-to-Speech Translation System | IEEE Journals & Magazine | IEEE Xplore. Ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/1597243/
37. Taylor, A., Marcus, M., & Santorini, B. (2003). The Penn Treebank: An Overview. Treebanks, 5–22. https://doi.org/10.1007/978-94-010-0201-1_1
38. Picard, R. W. (2000). Affective Computing. MIT Press. https://mitpress.mit.edu/9780262661157/affective-computing/
39. Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39–41. https://doi.org/10.1145/219717.219748

40. Hane, P. J. (1999). Beyond Keyword Searching—Oingo and Simpli.com Introduce Meaning-Based Searching. Newsbreaks.infotoday.com. https://newsbreaks.infotoday.com/nbreader.asp?ArticleID=17858
41. Michael Hotchkiss. (2012). George Miller, Princeton psychology professor and cognitive pioneer, dies. Princeton University. https://www.princeton.edu/news/2012/07/26/george-miller-princeton-psychology-professor-and-cognitive-pioneer-dies?section=topstories
42. Princeton University. (2019). WordNet | A Lexical Database for English. Princeton.edu. https://wordnet.princeton.edu/
43. Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. Aclanthology.org. https://aclanthology.org/L06-1225/
44. Esuli, A., & Sebastiani, F. (2008). Automatic generation of lexical resources for opinion mining. ACM SIGIR Forum, 42(2), 105. https://doi.org/10.1145/1480506.1480528
45. Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting on Association for Computational Linguistics -. https://doi.org/10.3115/976909.979640
46. Wiebe, J. (2000). Learning Subjective Adjectives from Corpora. https://cdn.aaai.org/AAAI/2000/AAAI00-113.pdf
47. Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis (pp. 347–354). https://aclanthology.org/H05-1044.pdf
48. Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews (pp. 417–424). Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia. https://aclanthology.org/P02-1053.pdf
49. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. Cornell Bowers CIS Computer Science. www.cs.cornell.edu. https://www.cs.cornell.edu/home/llee/papers/sentiment.pdf
50. Liu, B., Hu, M., & Cheng, J. (2004). Opinion Observer: Analyzing and Comparing Opinions on the Web. https://www.cs.uic.edu/~liub/publications/www05-p536.pdf
51. Zou, H., & Xiang, K. (2022). Sentiment Classification Method Based on Blending of Emoticons and Short Texts. Entropy, 24(3), 398. https://doi.org/10.3390/e24030398
52. Kumar, A., & Jaiswal, A. (2019). Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency and Computation: Practice and Experience, e5107. https://doi.org/10.1002/cpe.5107
53. Abdukhamidov, E., Juraev, F., Abuhamad, M., El-Sappagh, S., & AbuHmed, T. (2022). Sentiment Analysis of Users' Reactions on Social Media during the Pandemic. Electronics, 11(10), 1648. https://doi.org/10.3390/electronics11101648
54. D'Souza, J., Auer, S., & Pedersen, T. (2021). SemEval-2021 Task 11: NLPContributionGraph - Structuring Scholarly NLP Contributions for a Research Knowledge Graph. ArXiv (Cornell University). https://doi.org/10.18653/v1/2021.semeval-1.44
55. Park, H., Song, M., & Shin, K.-S. (2020). Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories. Knowledge-Based Systems, 187, 104825. https://doi.org/10.1016/j.knosys.2019.06.033

56. Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. Array, 100157. https://doi.org/10.1016/j.array.2022.100157
57. Jia, J., Liang, W., & Liang, Y. (2023). A Review of Hybrid and Ensemble in Deep Learning for Natural Language Processing. https://doi.org/10.48550/arXiv.2312.05589
58. Ahmed, S. F., Bin, S., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Ali, A. B. M. S., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review. https://doi.org/10.1007/s10462-023-10466-8
59. Ekkekakis, P. (2013). The Measurement of Affect, Mood, and Emotion | Social psychology, in *www.cambridge.org*. Cambridge University Press. Retrieved February 24, 2024, from https://assets.cambridge.org/97811070/11007/frontmatter/9781107011007_frontmatter.pdf
60. Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. Psychological Assessment, 35(3), 189–204. https://doi.org/10.1037/pas0001200
61. Liu, T., Meyerhoff, J., Eichstaedt, J. C., Karr, C. J., Kaiser, S. M., Kording, K. P., Mohr, D. C., & Ungar, L. H. (2022). The relationship between text message sentiment and self-reported depression. Journal of Affective Disorders, 302, 7–14. https://doi.org/10.1016/j.jad.2021.12.048
62. Althubaiti, A. (2016). Information Bias in Health research: definition, pitfalls, and Adjustment Methods. Journal of Multidisciplinary Healthcare, 9(9), 211–217. https://doi.org/10.2147/JMDH.S104807
63. Xiao, Y., Liu, H., & Li, H. (2017). Integration of the Forced-Choice Questionnaire and the Likert Scale: A Simulation Study. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.00806
64. Wind, S., & Hua, C. (2021). Rasch Measurement Theory Analysis in R. In bookdown.org. CRC Press. https://bookdown.org/chua/new_rasch_demo2/many-facet-rasch-model.html
65. Fulcher, G., & Harding, L. (2022). The Routledge Handbook of Language Testing. Routledge & CRC Press. https://www.routledge.com/The-Routledge-Handbook-of-Language-Testing/Fulcher-Harding/p/book/9781138385436
66. Zhang, X., Pina, L. R., & Fogarty, J. (2016). Examining Unlock Journaling with Diaries and Reminders for In Situ Self-Report in Health and Wellness. Europe PMC (PubMed Central). https://doi.org/10.1145/2858036.2858360
67. Verhagen, S. J. W., Hasmi, L., Drukker, M., van Os, J., & Delespaul, P. A. E. G. (2016). Use of the experience sampling method in the context of clinical trials. Evidence-Based Mental Health, 19(3), 86–89. https://doi.org/10.1136/ebmental-2016-102418
68. Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. Annual Review of Clinical Psychology, 4(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415
69. Kim, J., Marcusson-Clavertz, D., Yoshiuchi, K., & Smyth, J. M. (2019). Potential benefits of integrating ecological momentary assessment data into mHealth care systems. BioPsychoSocial Medicine, 13(1). https://doi.org/10.1186/s13030-019-0160-5

70. Wu, Y.-H., Stangl, E., Chipara, O., Gudjonsdottir, A., Oleson, J., & Bentler, R. A. (2020). Comparison of In-Situ and Retrospective Self-Reports on Assessing Hearing Aid Outcomes. Journal of the American Academy of Audiology, 31(10), 746–762. https://doi.org/10.1055/s-0040-1719133
71. Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science, 306(5702), 1776–1780. https://doi.org/10.1126/science.1103572
72. Morgan-Trimmer, S., & Wood, F. (2016). Ethnographic methods for process evaluations of complex health behaviour interventions. Trials, 17(1). https://doi.org/10.1186/s13063-016-1340-2
73. Loh, C. E., Sun, B., & Lim, F. V. (2023). "Because I'm always moving": a mobile ethnography study of adolescent girls' everyday print and digital reading practices. Learning, Media and Technology, 1–20. https://doi.org/10.1080/17439884.2023.2209325
74. Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. Cyberpsychology, Behavior, and Social Networking, 21(4), 217–228. https://doi.org/10.1089/cyber.2017.0384
75. Jhangiani, R. S., Chiang, I-Chant. A., Cuttler, C., & Leighton, D. C. (2019). Observational Research. Pressbooks.pub; Pressbooks. https://kpu.pressbooks.pub/psychmethods4e/chapter/observational-research/
76. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9
77. Díaz-García, A., González-Robles, A., Mor, S., Mira, A., Quero, S., García-Palacios, A., Baños, R. M., & Botella, C. (2020). Positive and Negative Affect Schedule (PANAS): psychometric properties of the online Spanish version in a clinical sample with emotional disorders. BMC Psychiatry, 20(1). https://doi.org/10.1186/s12888-020-2472-1
78. McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic Review of the Hawthorne Effect: New Concepts are Needed to Study Research Participation Effects. Journal of Clinical Epidemiology, 67(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015
79. Salazar, K. (2020). Contextual Inquiry: Inspire Design by Observing and Interviewing Users in Their Context. Nielsen Norman Group. https://www.nngroup.com/articles/contextual-inquiry/
80. Saxena, R. R., Saxena, A., & Saxena, R. (2017). Vulnerability to a Bioterrorism Attack and the Potential of Directed Evolution as a Countermeasure. Ejbio.imedpub.com; Electronic Journal of Biology, 2017, Vol.13(2): 125-130. https://ejbio.imedpub.com/articles/vulnerability-to-a-bioterrorism-attack-and-the-potential-ofdirected-evolution-as-a-countermeasure.pdf
81. Lin, C., Bier, B., Tu, R., Paat, J. J., & Tu, P. (2023). Vaccinated Yet Booster-Hesitant: Perspectives from Boosted, Non-Boosted, and Unvaccinated Individuals. Vaccines, 11(3), 550. https://doi.org/10.3390/vaccines11030550
82. Chavda, V. P., Yao, Q., Vora, L. K., Apostolopoulos, V., Patel, C. A., Bezbaruah, R., Patel, A. B., & Chen, Z.-S. (2022). Fast-track development of vaccines for SARS-CoV-2: The shots that saved the world. Frontiers in Immunology, 13. https://doi.org/10.3389/fimmu.2022.961198

83. Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. Nature Reviews Microbiology, 19(7), 409–424. https://doi.org/10.1038/s41579-021-00573-0
84. Jalilian, H., Amraei, M., Javanshir, E., Jamebozorgi, K., & Faraji-Khiavi, F. (2023). Ethical considerations of the vaccine development process and vaccination: a scoping review. BMC Health Services Research, 23(1). https://doi.org/10.1186/s12913-023-09237-6
85. Rouzine, I. M., & Rozhnova, G. (2023). Evolutionary implications of SARS-CoV-2 vaccination for the future design of vaccination strategies. Communications Medicine, 3(1), 1–14. https://doi.org/10.1038/s43856-023-00320-x
86. Ullah, I., Khan, K. S., Tahir, M. J., Ahmed, A., & Harapan, H. (2021). Myths and conspiracy theories on vaccines and COVID-19: Potential effect on global vaccine refusals. Vacunas (English Edition), 22(2), 93–97. https://doi.org/10.1016/j.vacune.2021.01.009
87. Skafle, I., Nordahl-Hansen, A., Quintana, D. S., Wynn, R., & Gabarron, E. (2022). Misinformation about Covid-19 Vaccines on Social Media: Rapid Review. Journal of Medical Internet Research, 24(8). https://doi.org/10.2196/37367
88. Islam, M. S., Kamal, A.-H. M., Kabir, A., Southern, D. L., Khan, S. H., Hasan, S. M. M., Sarkar, T., Sharmin, S., Das, S., Roy, T., Harun, M. G. D., Chughtai, A. A., Homaira, N., & Seale, H. (2021). COVID-19 vaccine rumors and conspiracy theories: The need for cognitive inoculation against misinformation to improve vaccine adherence. PLOS ONE, 16(5), e0251605. https://doi.org/10.1371/journal.pone.0251605
89. Nuwarda, R. F., Ramzan, I., Weekes, L., & Kayser, V. (2022). Vaccine Hesitancy: Contemporary Issues and Historical Background. Vaccines, 10(10), 1595. https://doi.org/10.3390/vaccines10101595
90. Majid, U., Ahmad, M., Zain, S., Akande, A., & Ikhlaq, F. (2022). COVID-19 vaccine hesitancy and acceptance: a comprehensive scoping review of global literature. Health Promotion International, 37(3). https://doi.org/10.1093/heapro/daac078
91. Mir, S., & Mir, M. (2024). The mRNA vaccine, a swift warhead against a moving infectious disease target. Expert Review of Vaccines. https://doi.org/10.1080/14760584.2024.2320327
92. Papastephanou, M. (2021). Pandemic Totalitarianisms, Limit Situations and Forced Vaccinations. Philosophy International Journal, 4(4). https://doi.org/10.23880/phij-16000214 Hafner, M., Yerushalmi, E., Fays, C., Dufresne, E., & Van Stolk, C. (2022). COVID-19 and the Cost of Vaccine Nationalism. Rand Health Quarterly, 9(4), 1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9519117/
93. Lange, E. L., & Shullenberger, G. (2024). COVID-19 and the Left: The Tyranny of Fear. In Google Books. Taylor & Francis. https://books.google.com/books?hl=en&lr=&id=nRD3EAAAQBAJ&oi=fnd&pg=PA22&dq=totalitarianism+covid-19+vaccine+sentiment&ots=QMZWrlbK_0&sig=-oR7Z6eoClZI6LpIDbTbY3O139c#v=onepage&q&f=false
94. Deb, B., Shah, H., & Goel, S. (2020). Current global vaccine and drug efforts against COVID-19: Pros and cons of bypassing animal trials. Journal of Biosciences, 45(1). https://doi.org/10.1007/s12038-020-00053-2

95. Yaamika, H., Muralidas, D., & Elumalai, K. (2023). Review of adverse events associated with COVID-19 vaccines, highlighting their frequencies and reported cases. Journal of Taibah University Medical Sciences, 18(6), 1646–1661. https://doi.org/10.1016/j.jtumed.2023.08.004
96. Bolsen, T., & Palm, R. (2021). Politicization and COVID-19 vaccine resistance in the U.S. Progress in Molecular Biology and Translational Science, 188(1), 81–100. https://doi.org/10.1016/bs.pmbts.2021.10.002
97. Zimmerman, R. K. (2021). Helping Patients with Ethical Concerns about COVID-19 Vaccines in light of Fetal Cell Lines used in some COVID-19 Vaccines. Vaccine, 39(31). https://doi.org/10.1016/j.vaccine.2021.06.027
98. Feature Engineering Vs Feature Selection | Blog. (n.d.). www.playerzero.ai. Retrieved March 4, 2024, from https://www.playerzero.ai/advanced/product-builder-facts/feature-engineering-vs-feature-selection
99. Gupta, A. (2020, October 10). Feature Selection Techniques in Machine Learning. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/
100. Müller, A. C., & Guido, S. (2016). 4. Representing Data and Engineering Features - Introduction to Machine Learning with Python [Book]. Www.oreilly.com. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/ch04.html
101. Saxena, R. R., & Saxena, R. (2024). Applying Graph Neural Networks in Pharmacology. Authorea. TechRxiv, Advance Sagepub. https://www.techrxiv.org/doi/full/10.36227/techrxiv.170906927.71541956/
102. Saeed, M., Ahmed, N., Mehmood, A., Aftab, M., Amin, R., & Kamal, S. (2023). Sentiment Analysis for COVID-19 Vaccine Popularity. KSII Transactions on Internet and Information Systems, 17(5), 1377–1395. https://doi.org/10.3837/tiis.2023.05.004
103. Kristian, Y., Yesenia, A. V., Safina, S., Pravitasari, A. A., Sari, E. N., & Herawan, T. (2023). Social Media Text Analysis on Public's Sentiments of Covid-19 Booster Vaccines. Lecture Notes in Computer Science, 209–224. https://doi.org/10.1007/978-3-031-37105-9_15
104. Ghosh, P., Dutta, R., Agarwal, N., Chatterjee, S., & Mitra, S. (2023). Social Media Sentiment Analysis on Third Booster Dosage for COVID-19 Vaccination: A Holistic Machine Learning Approach. Lecture Notes in Electrical Engineering, 179–190. https://doi.org/10.1007/978-981-19-8477-8_14
105. Kul, S., & Sayar, A. (2022). Sentiment Analysis Using Machine Learning and Deep Learning on Covid 19 Vaccine Twitter Data with Hadoop MapReduce. Lecture Notes in Networks and Systems, 859–868. https://doi.org/10.1007/978-3-030-94191-8_69
106. Aslan, S. (2022). A Novel TCNN–Bi-LSTM Deep Learning Model for Predicting Sentiments Of Tweets About COVID-19 Vaccines. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.7387
107. Eom, G., Yun, S., & Byeon, H. (2022). Predicting the sentiment of South Korean Twitter users toward vaccination after the emergence of COVID-19 Omicron variant using deep learning-based natural language processing. Frontiers in Medicine, 9. https://doi.org/10.3389/fmed.2022.948917
108. Ahmed, S., Khan, D. M., Sadiq, S., Umer, M., Shahzad, F., Mahmood, K., Mohsen, H., & Ashraf, I. (2023). Temporal analysis and opinion dynamics of COVID-19 vaccination tweets

using diverse feature engineering techniques. PeerJ, 9, e1190–e1190. https://doi.org/10.7717/peerj-cs.1190

109. Jain, T., Verma, V. K., Sharma, A., Saini, B., Purohit, N., Bhavika, Mahdin, H., Ahmad, M., Darman, R., Haw, S.-C., Shaharudin, S. M., & Arshad, M. S. (2023). Sentiment Analysis on COVID-19 Vaccine Tweets using Machine Learning and Deep Learning Algorithms. International Journal of Advanced Computer Science and Applications, 14(5). https://doi.org/10.14569/ijacsa.2023.0140504
110. Alam, K. N., Khan, M. S., Dhruba, A. R., Khan, M. M., Al-Amri, J. F., Masud, M., & Rawashdeh, M. (2021). Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data. Computational and Mathematical Methods in Medicine, 2021, 1–15. https://doi.org/10.1155/2021/4321131
111. Nuser, M., Alsukhni, E., Saifan, A., Khasawneh, R., & Ukkaz, D. (2022). Sentiment Analysis of Covid-19 Vaccine with Deep Learning. Journal of Theoretical and Applied Information Technology, 30(12).
112. Khalid, E. T., Talal, E. B., Khamees, M. K., & Yassin, A. A. (2022). Sentiment Analysis System for COVID-19 Vaccinations Using Data of Twitter. Indonesian Journal of Electrical Engineering and Computer Science, 26(2), 1156. https://doi.org/10.11591/ijeecs.v26.i2.pp1156-1164
113. Sandag, G. A., Manueke , A. M., & Walean, M. (2021). Sentiment Analysis of COVID-19 Vaccine Tweets in Indonesia Using Recurrent Neural Network (RNN) Approach | IEEE Conference Publication | IEEE Xplore. ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/9649648
114. Ainapure, B. S., Pise, R. N., Reddy, P., Appasani, B., Srinivasulu, A., Khan, M. S., & Bizon, N. (2023). Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches. Sustainability, 15(3), 2573. https://doi.org/10.3390/su15032573
115. Reshi, A. A., Rustam, F., Aljedaani, W., Shafi, S., Alhossan, A., Alrabiah, Z., Ahmad, A., Alsuwailem, H., Almangour, T. A., Alshammari, M. A., Lee, E., & Ashraf, I. (2022). COVID-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset. Healthcare, 10(3), 411. https://doi.org/10.3390/healthcare10030411
116. Rani, S., & Jain, A. (2023). DFM: Deep Fusion Model for COVID-19 Vaccine Sentiment Analysis. Lecture Notes in Networks and Systems, 227–235. https://doi.org/10.1007/978-981-19-9228-5_20
117. Prabha, V. D., & Rathipriya, R. (2022). Competitive Capsule Network Based Sentiment Analysis on Twitter COVID'19 Vaccines. Journal of Web Engineering. https://doi.org/10.13052/jwe1540-9589.2159
118. Dupuy-Zini, A., Audeh, B., Gérardin, C., Duclos, C., Gagneux-Brunon, A., & Bousquet, C. (2023). Users' Reactions to Announced Vaccines Against COVID-19 Before Marketing in France: Analysis of Twitter Posts. Journal of Medical Internet Research, 25, e37237. https://doi.org/10.2196/37237
119. Guerdoux, G., Tiffet, T., & Bousquet, C. (2022). Inference Time of a CamemBERT Deep Learning Model for Sentiment Analysis of COVID Vaccines on Twitter. Studies in Health Technology and Informatics. https://doi.org/10.3233/shti220714

120. Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cotae, P. (2023). COVID-19 Vaccine Hesitancy: A Global Public Health and Risk Modelling Framework Using an Environmental Deep Neural Network, Sentiment Classification with Text Mining and Emotional Reactions from COVID-19 Vaccination Tweets. International Journal of Environmental Research and Public Health, 20(10), 5803–5803. https://doi.org/10.3390/ijerph20105803
121. Alotaibi, W., Alomary, F., & Mokni, R. (2023). COVID-19 vaccine rejection causes based on Twitter people's opinions analysis using deep learning. Social Network Analysis and Mining, 13(1). https://doi.org/10.1007/s13278-023-01059-y
122. Zhou, X., Song, S., Zhang, Y., & Hou, Z. (2023). Deep Learning Analysis of COVID-19 Vaccine Hesitancy and Confidence Expressed on Twitter in 6 High-Income Countries: Longitudinal Observational Study. Journal of Medical Internet Research, 25, e49753. https://doi.org/10.2196/49753
123. Nyawa, S., Tchuente, D., & Fosso-Wamba, S. (2022). COVID-19 vaccine hesitancy: a social media analysis using deep learning. Annals of Operations Research. https://doi.org/10.1007/s10479-022-04792-3
124. Almurtadha, Y., Ghaleb, M., & Mohammed, A. (2022). Sentiment Analysis to Extract Public Feelings on COVID-19 Vaccination. Lecture Notes in Networks and Systems, 639–648. https://doi.org/10.1007/978-3-031-16865-9_51
125. Alam, S., Shovon, S. D., & Joy, N. H. (2021). Machine learning and Lexical Semantic-based Sentiment Analysis for Determining the Impacts of the COVID-19 Vaccine | IEEE Conference Publication | IEEE Xplore. ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/9885671
126. Ahmad, W., Wang, B., Martin, P., Xu, M., & Xu, H. (2022). Enhanced sentiment analysis regarding COVID-19 news from global channels. Journal of Computational Social Science, 6(1), 19–57. https://doi.org/10.1007/s42001-022-00189-1
127. Tiwari, D., & Nagpal, B. (2022). KEAHT: A Knowledge-Enriched Attention-Based Hybrid Transformer Model for Social Sentiment Analysis. New Generation Computing. https://doi.org/10.1007/s00354-022-00182-2
128. Gbashi, S., Adebo, O. A., Doorsamy, W., & Njobeh, P. B. (2021). Systematic Delineation of Media Polarity on COVID-19 Vaccines in Africa: Computational Linguistic Modeling Study. JMIR Medical Informatics, 9(3), e22916. https://doi.org/10.2196/22916
129. Li, F., Li, J., Wang, P.-H., Yang, N., Huang, J., Ou, J., Xu, T., Zhao, X., Liu, T., Huang, X., Wang, Q., Li, M., Yang, L., Lin, Y., Cai, Y., Chen, H., & Zhang, Q. (2021). SARS-CoV-2 spike promotes inflammation and apoptosis through autophagy by ROS-suppressed PI3K/AKT/mTOR signaling. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1867(12), 166260. https://doi.org/10.1016/j.bbadis.2021.166260
130. Peng, K., Li, X., Yang, D., Chan, S. K. W., Zhou, J., Yuk, E., Sze, C., Tsz, F., King, C., Chan, E. W., Leung, W. K., Lau, C.-S., & Wong, I. C. K. (2023). Risk of autoimmune diseases following COVID-19 and the potential protective effect from vaccination: a population-based cohort study. EClinicalMedicine, 63, 102154–102154. https://doi.org/10.1016/j.eclinm.2023.102154

131. Saxena, R. R., & Saxena, R. (2023). Potential Of Medium to Long-Term Fasting to Trigger an Autoimmune Response Through Hyperaggressive Autophagy. IJRDO - Journal of Biological Science, 9(1), 1–11. https://doi.org/10.53555/bs.v1i1.5937
132. Shah, H., Kim, A. S., Sukumar, S., Mazepa, M. A., Kohli, R., Braunstein, E. M., Brodsky, R. A., Cataland, S. R., & Chaturvedi, S. (2022). SARS-CoV-2 vaccination and immune thrombotic thrombocytopenic purpura. Blood, 139(16), 2570–2573. https://doi.org/10.1182/blood.2022015545
133. Yazdani, A. N., DeMarco, N., Patel, P., Abdi, A., Velpuri, P., Agrawal, D. K., & Rai, V. (2023). Adverse Hematological Effects of COVID-19 Vaccination and Pathomechanisms of Low Acquired Immunity in Patients with Hematological Malignancies. Vaccines, 11(3), 662–662. https://doi.org/10.3390/vaccines11030662
134. Schnirring, L. (2024, March 4). BA.2.87.1 COVID variant detected in Southeast Asia | CIDRAP. www.cidrap.umn.edu. https://www.cidrap.umn.edu/covid-19/ba2871-covid-variant-detected-southeast-asia
135. Katella, K. (2024, January 31). 3 Things to Know About JN.1, the New Coronavirus Strain. Yale Medicine. https://www.yalemedicine.org/news/jn1-coronavirus-variant-covid