

How Responsible Is AI?

Identification of Key Public Concerns Using Sentiment Analysis and Topic Modeling

Dwijendra Nath Dwivedi, Krakow University of Economics, Poland*

Ghanashyama Mahanty, Utkal University, India

Anilkumar Vemareddy, University of Agricultural Sciences, India

ABSTRACT

Many businesses around the world are adopting AI with the hope of increasing their top-line and bottom-line numbers. The COVID-19 pandemic has further accelerated the journey. While AI technology promises to bring enormous benefits, the challenges come in similar proportions. In the current form, the requirements for transparency and trust are relatively low for AI systems. On the other hand, there is a lot of regulatory pressure for AI systems to be trustworthy and responsible. Challenges still exist both on the methods and theory side and how explanations are used in practice. The objective of this paper is to analyze Twitter data to extract sentiments and opinions in unstructured text. The authors attempted to use contextual text analytics to categorize the twitter data to understand the positive or negative sentiments and feelings for the AI ethical challenges and highlight the key concerns. Text clustering has also been performed on positive and negative sentiments to understand the key themes behind people's concerns.

KEYWORDS

AI Ethics, Data Bias, Responsible AI, Sentiment Analytics, Text Clustering, Topic Modeling, Twitter

1. INTRODUCTION

Humans are always in search of new tools and technologies to lead a better and productive life. Over the last three industrial revolutions, we have seen a massive shift from muscle power to mechanical power. The advancement in digitization, technology, and data analytics, focuses further on enhancing human capability by exploiting cognitive principles, otherwise known as Artificial Intelligence (AI), as a theme worldwide (Schwab, 2017). John McCarthy, the AI discipline founder, explains that Artificial Intelligence is the “science and engineering of making intelligent machines” (Walch, 2018).

Artificial Intelligence solutions are permeating into every walk of life. Some basic tasks like looking for information in the Google search engine, content writing, drawing to complex activities like seeking digital assistant and Robo advisor from a service provider are powered with AI algorithms. These algorithms are trained and validated based on available societal behavioral data with human ingenuity. While these automated decision-making systems bring enormous benefits to society, those can bring challenges, too, unless handling with utmost care.

There has been disagreement around the scientific definition of humans and their origin. However, the commonly accepted fact is that humans first appeared between 2-3 million years ago (Barras, 2016).

DOI: 10.4018/IJIRR.298646

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Despite evolving over millions of years, humans make irrational decisions and mistakes. Human decisions are colored by the amount of information they have, cognitive ability, the socio-economic condition they belong to, and many more possible dimensions. Similarly, the AI systems built by human endeavor and past societal behavior captured through data can be equally irrational and biased. These irrationalities or biases can appear in the form of infringement of privacy, discrimination, societal exclusion, accident, and rigging political systems (Cheatham, Javanmardian, and Samandari, 2019). After all, humans play a critical role in building these intelligent systems.

The global artificial intelligence market was valued at USD 62.35 billion in 2020 and is expected to grow at a compound annual growth rate (CAGR) of 40.2% from 2021 to 2028 (GrandViewResearch, 2021). AI is supposed to bring 21% of incremental impact in GDP in the United States of America by 2030 (Bughin *et al.*, 2018). Open databases have supported the rapid development of AI algorithms, which led to significant outcomes wherein different stakeholders have benefited to a greater extent. Ntoutsis *et al.* (2020) talk about the far-reaching AI impact on individuals and society, and their decisions might affect everyone, everywhere, and anytime, entailing concerns about potential human rights issues.

Artificial Intelligence systems can be a double-edged sword. While they bring substantial benefits into the decision-making process, any wrong decisions can lead to loss of life, reputational damage, revenue loss, societal unrest, regulatory backlash, criminal investigation, and diminished public trust (Cheatham, Javanmardian and Samandari, 2019). The problem gets magnified, especially when AI systems are built based on automated learning and deployed at scale principle. The AI mistakes can happen from the ideation of a problem through the design and deployment of the solution. Mistakes can be non-intentional to malicious intent, either exploit market conditions or defame and defeat economically and politically certain sections of society.

The series of recent AI mishaps have further ignited the debate. For example, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a case management and decision support tool used by the US courts estimated risk of re-offending was found to predict higher risk values for black defendants (and lower for white ones) than their actual risk (Angwin *et al.*, 2016). In another instance, the Google Advt tracking tool showed significantly more advertisements to high-paid men than their female counterparts (Datta, Tschantz, and Datta, 2015).

While humans are accountable for their decision-making activities, the moot question is who will be responsible when an AI system produces incorrect societal action. Are we penalizing the machines, the machines' creators, or the organization that promotes these machines? If stakeholders are many, how would the AI risk burden be shared? Since Artificial Intelligence is a relatively newer decision-making framework, we must discuss and debate possible consequences of an algorithm failure and its mitigation actions to minimize, if not eliminate, any unintended consequences of the system-based decisions.

The paper's main objective is to summarize public concerns about responsible AI theme discussion in the Twitter forum. The study uses the Naïve Bayes principle to extract subjective emotions and feelings covering both Twitter messages' positive and negative sentiments. Businesses often use it to detect social data sentiment, gauge brand reputation, and understand customers. Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc.), urgency (urgent, not urgent), and even intentions (interested v. not interested). The paper aims to understand and analyze the public's sentiments regarding AI ethics and trustworthy challenges. The idea is to know if most of the tweets have positive or negative feelings about the theme. At the same time, it would be essential to understand the key clusters and segments of positive and negative tweets.

The present study aims to address the following research questions:

- a) What is the distribution of positive and negative sentiments emerging from public tweets in the context of AI Ethics and regulatory challenges?

b) What are the key segmentations and clusters in the positive and negative sentiments?

2. METHODOLOGY

2.1 Data Source

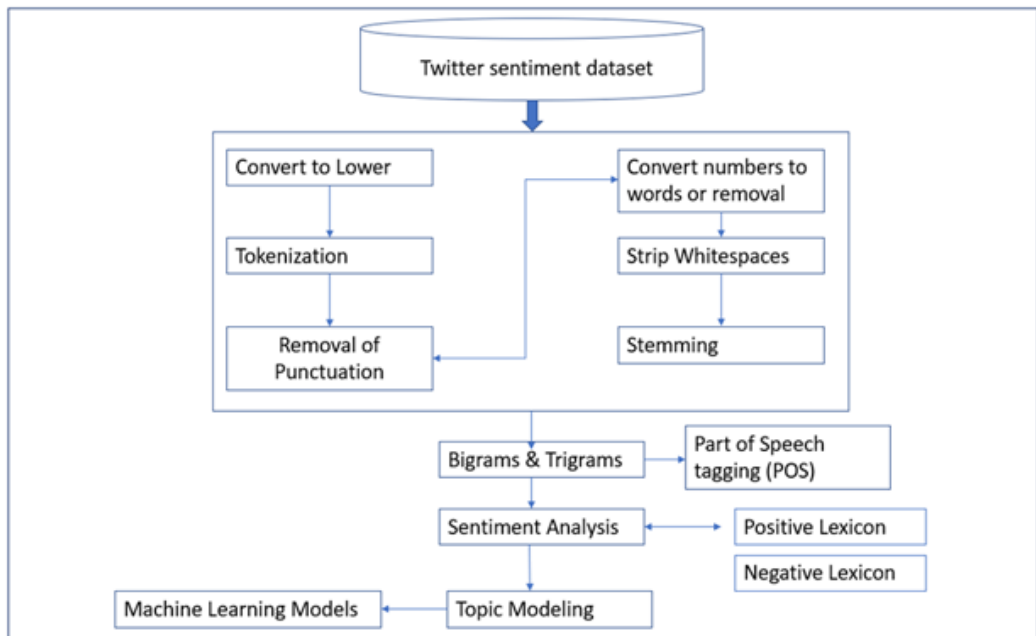
For the study, the data was extracted from Twitter, as it is one among the open data sources containing rich information on the large set of topics discussed by the public. Therefore, the people's tweets regarding AI ethics, challenges, trustworthiness, associated risks, and related issues from February 2021 to March 2021 were extracted. In addition, the 18,182 tweets from Twitter covering comments/tweets from different parts of the World were extracted. The relevant tweets pertaining to AI were extracted with the most care by using the python program.

2.2 Analytical Procedure

The present study was carried out by adopting novel natural language processing techniques, namely 1) Sentiment analysis and 2) Topic modeling. The detailed methodology followed as under

Firstly, the **sentiment analysis refers to computational classification/categorization of sentiments in the piece of text, especially refers to people's attitude on particular aspects of positive or negative or neutral**. In the present context, the sentiment analysis was employed to classify people's tweets are positive or negative or neutral in response to AI ethics, trustworthiness, risk in usage, applicability, accuracy, etc.

Figure 1. Flow chart of pre-processing steps of Twitter text data.



The accuracy and adaptability of the sentiment analysis outcomes always depended on the form or structure of text data used for sentiment analysis. The text data should be extensively cleaned before the sentiment polarity calculations. This step is needed for the text analysis to transform

human language into a machine-readable format for further processing and analysis. The selection and pre-processing steps were majorly dependent on the form or structure of text data. The detailed pre-processing steps followed in the present study were depicted in the flow chart below

In the pre-text processing of twitter text data, initially, duplicate rows were removed to avoid unbiased results. Subsequently, converted all the text into lower cases to prevent multiple copies of the same word. For example, (“Ai Ethics” ai ethics” will be considered as two different words). The punctuations are removed as they might add extra information or reduce the size of the training dataset while handling text. Also, eliminated stop words that are frequently occurring words in the text by using the text blob library in python. The tweets with many spelling mistakes or short words were observed in the Twitter data. Hence spelling correction steps are performed with the help of the text blob library.

After the above steps, tokenization was done to divide the text into a sequence of words or sentences, transforming our tweets into a blob and converting them into a series of words. Followed by Stemming refers to the removal of suffices, like “ing,” “ly,” “s,” etc., by a simple rule-based approach by using Porter Stemmer from the NLTK library of python. Some may use Lemmatization as a more practical option than stemming because it converts the word into its root word, rather than just stripping the suffices. It makes use of the vocabulary and does a morphological analysis to obtain the root word. Therefore, researchers usually prefer lemmatization over stemming.

After basic preprocessing steps of cleaning the text, extracted the features using the following natural language techniques. N-Grams identify the combination of multiple words used together. We have used N-grams, bigrams, and trigrams. Unigrams have not captured much information as compared to bigrams and trigrams. Hence, bigrams or trigrams were used to capture the language structure, like what letter or word was likely to follow the given one. Further, part-of-speech tagging mainly assigns speeches to each text word based on its context and definition (nouns, verbs, adjectives, and others).

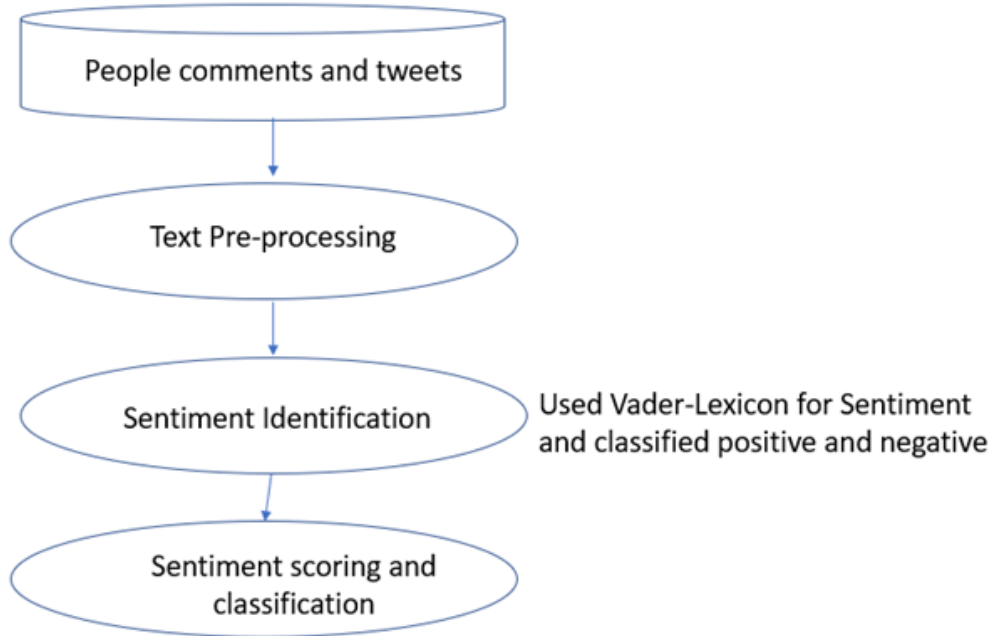
Secondly, **topic modeling** is the process of extracting or obtaining required features from the bag of words. This is an important technique since each word present in the corpus is considered a natural language processing feature. This feature reduction will help us focus on the right content instead of going through the entire text in the training data. There are many methods used for topic modeling. Latent Dirichlet Allocation (LDA) is one method used to analyze the topic modeling in the present study.

LDA is a statistical and graphical model used to obtain relationships between multiple documents in a corpus. It is developed using the variation exception maximization (VEM) algorithm to get the maximum likelihood estimate from the whole text corpus. Traditionally, this can be solved by picking out the top few words in the bag of words. However, this completely lacks the semantics in the sentence. This model follows the concept that the probabilistic distribution of topics can describe each document, and words can explain each topic. Thus, it helps to get a much clearer vision of how the topics are connected. It considers all corpus of entire documents in the data. After preprocessing of the corpus, each bag of words consists of common words. Using the LDA model, the topics related to each document have been derived and can group all corpus into a particular group for further usage. The flow chart below details the process of topic modeling

3. CONTRIBUTIONS AND VALUE ADD

AI risk is an emerging topic. The study uses open-source, available Twitter data to understand the key emerging themes in the AI benefits and cost space to the society using a text mining framework. The study aims to bring the critical issues to the forefront and create awareness among the practitioners and regular users to address the key concerns while embarking on the AI implementation journey. Also, it contributes to text mining literature by providing a framework for analyzing public sentiments.

Figure 2. Process flow for topic modeling



4. LITERATURE REVIEW

The literature has been reviewed under three broad themes: AI risk and ethical challenges, sentiment analysis using twitter data, and text clustering methods.

4.1 AI Risk and Ethical Challenges

Dwijendra and Mahanty (2021), by leveraging the AI incident database, identified some key AI risk areas from known risk incidents from recent years. Hagendorff (2020) analyzed and compared 22 ethical guidelines, highlighting overlaps but also omissions. Also, the study provides some interesting recommendations on how the effectiveness in the demands of AI ethics can be improved. Maas, (2018) suggests that large-scale, cascading errors in AI systems are inevitable. Box and Snaveley, (2019) discussed the human bias in machine learning. Martinho, Kroesen and Chorus, (2020) integrated both theoretical and empirical lines of thought to address the matters of moral reasoning in AI Systems. Tamboli (2019), shared challenges that change in data over time brings and hence affects statically programmed (or assumed) relationships, more commonly known as “concept drift”. Thomas(2021), shared concerns and the technical issues raised when humans are replaced by artificial intelligence (AI). Holzinger, Haibe-Kains and Jurisica (2019), explains sufficient, and quality data is needed to solve critical challenges in medicine. He mentions effectively and efficiently integrating diverse clinical, imaging, and molecular profile data is necessary to understand complex diseases.

4.2 Sentiment Analytics Using Twitter

Asghar *et al.*(2018), analyzed consumers’ views for major automobile Brands Using Twitter Data. They found that Audi has 87% of the positive tweets compared to 74% for BMW, 84% for Honda, 70% for

Toyota, and 81% for Mercedes. Godea et al.(2015), explored the general sentiments and information dissemination concerning electronic cigarettes or e-cigs using Twitter and found that Twitter users are mainly concerned with sharing information (33%) and promoting e-cigs(22%). Olorunnimbe and Viktor (2015) study captured individuals' sentiments as they evolve while exploring political sentiments on Twitter for opinion mining. Dinkić et al.(2016), used Twitter data to determine the popularity of city locations of interest and public spaces in general. Esiyok and Albayrak (2015), compared the performance of the Naïve Bayes and Maximum Entropy classification methods for predicting marketing trends. Choi and Kim (2013), performed sentiment analysis for tracking breaking events and found that his study offers diverse evidence to prove that Twitter has valuable information for monitoring breaking news worldwide. Rao and Srivastava(2014), evaluated public opinion tweets in driving investment decisions. Steede et al.(2018), analyzed the sentiment and content analysis of Twitter content regarding antibiotics' use in livestock.

4.3 Text Clustering Methods

There have been challenges around managing the explosion of electronic document archives. Sharper and new tools and techniques are required to organize, search, index, and review extensive data collections in a time-efficient manner (Alghamdi and Alfalqi, 2015). As a generalization, there are two broad approaches to process text- natural language processing (NLP) and statistical-based programs like topic modeling (Hofmann, 2001). Unlike NLP methods that tag parts-of-speech and grammatical structure, statistical-based models like topic models are based mainly on the 'bag-of-words (BoW) assumption. In BoW models, the collection of text documents is quantified into a document-term matrix (DTM) that counts the occurrence of each word (columns) for each document (rows). In most topic models like LDA, the DTM is one of two model inputs and the number of topics(Wesslen, 2018). Deerwester et al.(1990), presented one of the first topic models using latent semantic analysis (LSA) and singular value decomposition (SVD), in which a large DTM is decomposed into a set of about 100 orthogonal factors from which the original matrix can be approximated by linear combination. They assumed some underlying latent semantic structure and used statistical techniques to estimate this latent structure.

Asmussen and Møller (2019), presented a framework to leverage the topic modeling technique for performing an exploratory literature review on an extensive collection of papers. The framework they propose enables a large volume of documents to be reviewed in a transparent, efficient, and reproducible way using the LDA method. In general, there are two methods for automatically processing documents- supervised learning and unsupervised learning. Supervised learning includes manually coding a collection of documents before conducting an analysis, which involves a high amount of time to achieve the result. On the other hand, unsupervised learning methods, such as topic modeling, do not have the pre-requisite to manually code the documents, saving a lot of time for an exploratory review of the extensive collection of papers. Gottipati, Shankararaman and Lin(2018), leveraged topic modeling and data visualization methods to analyze student feedback comments from seven undergraduate courses taught at Singapore Management University. They assessed rule-based methods and statistical classifiers for extracting the topics. Al-Obeidat, Kafeza and Spencer(2017), further proposed an opinions sandbox for topic extraction, sentiment analysis for pulling issues and their associated sentiments from a database. They used LDA for topic extraction and the "bag-of-words" sentiment analysis algorithm, where polarity is determined based on the frequency of positive/negative words in a document. Benedetto and Tedeschi, (2016), highlight the standard approaches of sentiment analysis in social media streams and related cloud computing issues. Big data is divided into four features, namely four V's of big data- volume, velocity, variety, and Veracity. Volume is the most considerable amount of data that should be stored and processed. Velocity is the frequency of the incoming data. At the same time, variety describes different data types, whereas veracity refers to the trustworthiness and accuracy of the data available.

5. RESULTS & DISCUSSIONS

The present study can be viewed broadly under two headings, sentiment analysis, and topic modeling. Firstly, sentiment analysis classifies twitter text data into positive, positive, negative, and neutral sentiments. In the process, the text blog library of python was used to assess a sentence's polarity, which uses naive Bayes (probabilistic algorithms that use Bayes's Theorem to predict a text's category). As a result, it generates a score ranging between -1 to +1, which is strongly negative to strongly positive sentiments.

Secondly, topic modeling, the key topics were extracted, which discovers keywords in the sentiments that capture the recurring text theme and is widely used to analyze the large sets of sentiments to identify the most common topics more quickly and efficiently. Finally, we have applied latent semantic analysis and singular value decomposition for text clustering in terms of methodology. In the present context of text mining, clustering divides tweets into various groups based on the presence of similar themes, where it describes which topic of AI is in most debate/dialog.

5.1 Sentiment Analysis

Sentiment analysis for free-flowing text like Twitter data can effectively combine natural language processing (NLP) and machine learning algorithms to reduce sentiment scores to a sentence or phrase. Sentiment analysis can give an idea about public opinion, brand reputation, key concerns, customer experience, customer perception, and an overall index of how optimistic the common public is about a topic of interest, the people's attitude towards AI. Figure 3 represents the word cloud extracted from the Twitter data before sentiment analysis and topic modeling. It may not be possible to identify the positive /negatives /neutral opinions from this word cloud. However, it shows the popular/ repeated words used in the discussion in relation to the topic of interest.

Figure 3. Word cloud before sentiment analysis and topic modeling



As discussed in the methodology, four layers of sentiment analysis are currently followed in the study of the Twitter text data. The first is to break down the raw Twitter text into a list of different components sometimes referred to as tokens that reflect the person's emotions towards AI ethics, trustworthiness, credibility, risk association, and accuracy of AI models. Secondly, isolated the tokens with sentiments and ignored the rest. In the third step, assigned polarity scores to each component of the tokens. Finally, in fourth assigned sentiments such as very positive, positive, neutral, very negative, or negative based on the polarity scores. Table 1 depicts the distribution of sentiments on AI ethics and associated parameters considered with respect to the AI. The results found that around 44% (8018 tweets) were neutral, followed by 39% positive (positive 25% & very positive 14%, in total 7018 tweets), and the rest showed negative sentiments on AI. The positive tweets of 39% on AI infers that these segments of populations have more trust and favor AI. On the other hand, around 17% may disagree with AI, and the rest of the population neither supported nor favored AI.

Table 1. Sentiment segmentations of Twitter data

Sentiments	Counts	Distribution (%)
Neutral	8,018	44
Positive	4,554	25
Very positive	2,464	14
Negative	2,176	12
Very negative	970	5
Total	18,182	

The positive and negative sentiments derived from the sentiment analysis were shown in fig. 4 & 5 wherein most frequent/key positive/negative words are shown in bigger. As the frequency decreases, words will be shown in smaller sizes.

5.2 Topic Modeling

Topic models play an essential role in exploring text data, especially with a large volume of text data, to understand the structures and groups of interest. The LDA was used for the topic modeling wherein the key topics were extracted from the bag of words. As the model follows the concept of the probabilistic distribution of topics that describes each document, the probabilistic distribution of words can explain each topic to get a clearer vision of how the topics are connected.

The resulted cluster of LDA were depicted in the table 2 & 3, the table 2 represents the negative sentiments towards AI which means these segments of population are not in favour of AI. The key negative sentiments expressed are AI is risk to human intelligence, not accurately project the probabilities and not trustworthy to be used as solution. However, AI is developing rapidly and unpredictable, extremely irresponsible and shows the risk of negative consequences. The public discusses on implementation of the artificial intelligence models in government sector. The concerns raised in politics on information derived from AI etc. and reflect the ethical concerns of the future.

Figure 4. Word cloud for positive sentiments



Figure 5. Word cloud for negative sentiments

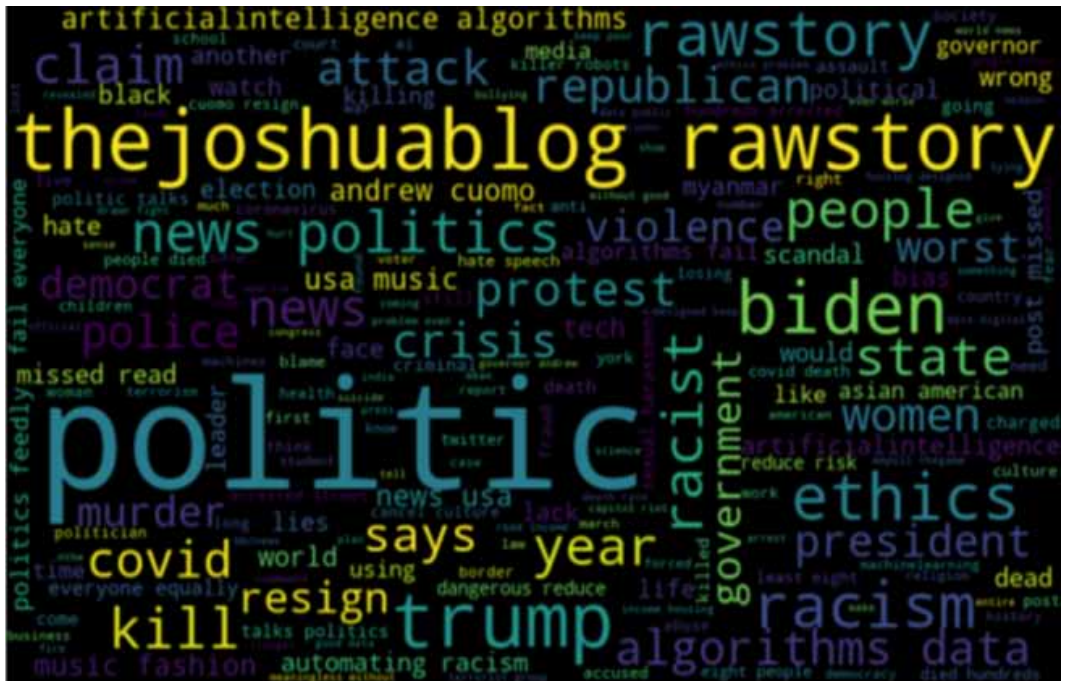


Table 2. Cluster- Negative Sentiments Segmentations

Segments	Topic Modeling: Negative Sentiments
0	“show_risk,+yldixdrui,+model,+artificialintelligence,+human”,
1	“germany,+humble,+beginning,+sinister,+scandal”,
2	“politic,+common,+upside,+digi,+many”,
3	“insufficient,+tool_completely,+expert,+facebook,+bias”,
4	“artificialintelligence,+algorithm,+datum,+digital,+watch”,
5	“risk_amplify,+data,+discrimination,+forward,+bias”,
6	“automating_racism,+fail_everyone,+equally,+jrcwbrvbdw,+politic”,
7	“ethic,+alg,+critical,+wilson,+access”,
8	“artificialintelligence,+bias,+behind_assistant,+accent,+algorithm”,
9	“politic,+news,+harm,+biden,+warn”

Model Performance

The model performance was checked using perplexity. The measure traditionally used for topic models is the $\text{perplexity}(\text{test set } w) = \exp\{-L(w)/\text{count of tokens}\}$

which is a decreasing function of the log-likelihood $L(w)$ of the unseen documents w ; the lower the perplexity, the better the model. We have got a very low confusion and Coherence score.

Perplexity: -11.535942740878582

Coherence Score: 0.502168380482148

Machine Learning and Artificial Intelligence systems are shaping up to our society – for better and worse (Helbing *et al.*, 2019). One of the dominant themes that emerge from the analysis is these tools have the ability to decide the way to think and act and got the potential to destabilize democracy. In 2016, the Data Science firm Cambridge Analytica used the tool to manipulated American voters and influenced their political decision-making process. This is a classic example of sinister motives that created a massive scandal all over the World. Racism and sexism are other areas of concerns for people around the World. In 2015, a study by the University of Washington showed that a simple image search in the USA for “CEO” shows only 11 per cent female, whereas, in reality, 27 per cent of the CEOs are female in the country. With improvements in computer vision, both digital and physical surveillance is becoming a key concern for society.

5.2.1 Key Negative Topics Derived From LDA

An artificial intelligence model shows more risk than human intelligence and insufficient expertise and completely biased tools. Risk amplification & data discrimination, bias algorithms may lead negative impacts hence raises the question of trustworthiness. Political news automatic racism harms the people.

The table 2 represents the positive sentiment segmentation wherein these segment of population are in favor of AI. The topics identified from the analysis will help to improve the AI ethics, emphasize in understanding and inclusiveness of AI. The analysis also showed the public sectors and political systems are excited to use the AI system as these will help in taking the good decisions and to create policies in the government sector. AI is responsible for developing better machine learning models globally with legal and trustworthy.

5.2.2 Key Positive Topics Derived From LDA

Artificial Intelligence implementation is growing lot faster than many of us anticipated. Numerous positive themes are emerging from the public discourse. One of the significant potentials of AI technology is to build automation of repetitive work. Identification of defective assets having an automated virtual chatbot to respond to customer queries are kinds of stuff delegated to Artificial Intelligence agents(Thomas, 2021). Health is another area where AI systems have started playing a critical role.

Table 3. Cluster 2- Positive sentiments segmentations

Segments	Topic Modeling: Positive Sentiments
0	“importance,+make,+lnbfch,+understandable_theshiftfi, + inclusive”,
1	“excited,+politic, +biden,+trump,+project”,
2	“discuss,+collaboration,+codedbiaspb,+bannister,+outstanding”,
3	“principles_responsible,+accountability_inclusiveness,+reliability_safety,+fairness_transpare,+integrity”,
4	“artificialintelligence,+algorithm,+datum,+digital,+tech”,
5	“responsibleai,+do,+machinelearne,+algorithmic,+global”,
6	“legal,+trust,+framework_europe,+cozibdy,+poll”,
7	“techethic,+come,+month,+march,+wrap”,
8	“think,+transparency,+lead,+discrimination,+class”,
9	“film,+surgery,+investigate,+netflix,+intelligence”

Model Performance Checking Using Perplexity:

Perplexity: -12.63675039019591

Coherence Score: 0.5301629075500442

The health care industry is one of the most challenging industry and score for AI is increasing. Image analysis to drug discoveries is performed by AI on a regular basis (Delijewski and Haneczok, 2021). Forest and agriculture are other areas where AI started showing tremendous benefits. Be it checking moisture quality, the content of the soil, pest management, weather, to auto watering plants are handled through AI algorithms(Columbo et al., 2021). Using AI for managing cybersecurity, fraud and credit management are other fertile ground for AI implementation.

AI systems have to be more explainable, transparent, interpretable, built on unbiased data. It should be designed on human-first and common-sense principles. Governments have to play an active role in creating awareness of AI technologies, legal aspects around them. Since AI is growing drastically public sector, governments, business sectors should bring the code of ethics into the AI discussions. From the above positive topics, general thinking about the AI principles should be responsible, accountability, reliability, safety, fairness, transparency, and integrity. Also, the Netflix uses the AI to recommend movies to users based on the history and top pics from overall users and highest-ranked movies. So it's a clear indication of how accurate AI recommendations are, and its shows how ethical or responsible these AI systems are. In recent years, greater AI transparency into all working models will help mitigate fairness, discrimination, and trust. All these have increased the attention.

6. CONCLUSION

Artificial Intelligence technology started dominating every aspect of life. Social media, health, airline, oil exploration, banking and payment, cyber surveillance and many other industries are rapidly adopting these new technologies. With the rise in digitization, better storage and computing power, autonomous algorithmic decisions are expressive of scope for AI in the future as these bear enormous speed and are cost-efficient.

These autonomous decision-making systems are built based on data captured through our day-to-day behavior and reflect on societal actions to a large extent. But, of course, not every action taken in the past is unbiased. These unjust or biased data points, along with any defect in the AI design and delivery process, will have the potency of accentuating and magnifying the problem.

Our in-depth analysis of public sentiments from the Twitter feeds certainly not only looking at the merits of Artificial Intelligence implication but cluster out some of the key issues and challenges these new technologies bring to the society. Given the rapid pace of AI implementations, these challenges may get magnified and rise to social tension unless handled with the proper framework and policies. Therefore, we must build adequate social coherence and the ethical principle so that AI technology will be an enabler of society and not a disabler.

CONFLICT OF INTEREST:

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

FUNDING AGENCY

Publisher has waived the Open Access publishing fee.

REFERENCES

- Al-Obeidat, F., Kafeza, E., & Spencer, B. (2017). Opinions Sandbox: Turning emotions on topics into actionable analytics. *International Conference on Emerging Technologies for Developing Countries*, 110–119.
- Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.*, 6(1).
- Angwin, J. (2016). Machine bias. *ProPublica*, 23(2016), 139–159.
- Asghar, Z. (2018). Sentiment analysis on automobile brands using Twitter data. *International Conference on Intelligent Technologies and Applications*, 76–85.
- Assmussen, C. B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, 6(1), 1–18.
- Barras, C. (2016). We don't know which species should be classed as human. *BBC Earth*, 11.
- Benedetto, F., & Tedeschi, A. (2016). Big data sentiment analysis for brand monitoring in social media streams by cloud computing. In *Sentiment Analysis and Ontology Engineering* (pp. 341–377). Springer. doi:10.1007/978-3-319-30319-2_14
- Box, J., & Snively, E. (2019). Do You Know What Your Model is Doing? How Human Bias Influences Machine Learning. The Clinical Data Science Conference- PHUSE EU Connect.
- Bughin, J. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy*. McKinsey Global Institute.
- Cheatham, B., Javanmardian, K., & Samandari, H. (2019, Apr.). Confronting the risks of artificial intelligence. *McKinsey Quarterly*.
- Choi, D., & Kim, P. (2013). Sentiment analysis for tracking breaking events: a case study on twitter. *Asian conference on intelligent information and database systems*, 285–294. doi:10.1007/978-3-642-36543-0_30
- Columbo, J. A. (2021). Association of Adoption of Transcatheter Aortic Valve Replacement With Center-Level Perioperative Outcomes. *JAMA Network Open*, 4(2), e2037885-e2037885.
- Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proc. Priv. Enhancing Technol.*, 2015(1), 92–112. doi:10.1515/popets-2015-0007
- Deerwester, S. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391–407.
- Delijewski, M., & Haneczok, J. (2021). AI drug discovery screening for COVID-19 reveals zafirlukast as a repurposing candidate. *Medicine in Drug Discovery*, 9.
- Dinkić, N. (2016). Using sentiment analysis of Twitter data for determining popularity of city locations. *International Conference on ICT Innovations* 156–164.
- Dwijendra, D. N., & Mahanty, G. (2021). A text mining-based approach for accessing AI risk incidents. *International Conference on Artificial Intelligence*.
- Esiyok, C., & Albayrak, S. (2015). Twitter sentiment tracking for predicting marketing trends. In *Smart Information Systems* (pp. 47–74). Springer. doi:10.1007/978-3-319-14178-7_2
- Godea, A. K. (2015). An analysis of twitter data on e-cigarette sentiments and promotion. *Conference on Artificial Intelligence in Medicine in Europe*, 205–215. doi:10.1007/978-3-319-19551-3_27
- Gottipati, S., Shankaraman, V., & Lin, J. R. (2018). Text analytics approach to extract course improvement suggestions from students' feedback. *Research and Practice in Technology Enhanced Learning*, 13(1), 1–19.
- GrandViewResearch. (2021). *Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution, By Technology (Deep Learning, Machine Learning, Natural Language Processing, Machine Vision), By End Use, By Region, And Segment Forecasts, 2021 - 2028*. Available at: <https://www.grandviewresearch.com>
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*. Springer, 30(1), 99–120. doi:10.1007/s11023-020-09517-8

Helbing, D. (2019). Will democracy survive big data and artificial intelligence? In *Towards digital enlightenment* (pp. 73–98). Springer. doi:10.1007/978-3-319-90869-4_7

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1), 177–196.

Holzinger, A., Haibe-Kains, B., & Jurisica, I. (2019). Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. *European Journal of Nuclear Medicine and Molecular Imaging*. Springer, 46(13), 2722–2730. doi:10.1007/s00259-019-04382-9 PMID:31203421

Maas, M. M. (2018). Regulating for 'Normal AI Accidents' Operational Lessons for the Responsible Governance of Artificial Intelligence Deployment. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 223–228. doi:10.1145/3278721.3278766

Martinho, A., Kroesen, M., & Chorus, C. (2020). An Empirical Approach to Capture Moral Uncertainty in AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 101. doi:10.1145/3375627.3375805

Ntoutsis, E. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1356.

Olorunnimbe, M. K., & Viktor, H. L. (2015). Tweets as a vote: Exploring political sentiments on twitter for opinion mining. *International Symposium on Methodologies for Intelligent Systems*, 180–185. doi:10.1007/978-3-319-25252-0_19

Rao, T., & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. In *State of the art applications of social network analysis* (pp. 227–247). Springer. doi:10.1007/978-3-319-05912-9_11

Schwab, K. (2017). *The Fourth Industrial Revolution*. Portfolio. Crown Publishing Group.

Steede, G. M. (2018). A sentiment and content analysis of Twitter content regarding the use of antibiotics in livestock. *Journal of Applied Communications, Agricultural Communicators in Education*, 102(4), 1B–1B.

Tamboli, A. (2019). Evaluating Risks of the AI Solution. In *Keeping Your AI Under Control* (pp. 31–42). Springer. doi:10.1007/978-1-4842-5467-7_4

Thomas, R. (2021). How AI Is Driving The New Industrial Revolution. *Forbes*. Available at: <https://www.forbes.com>

Walch, K. (2018). Artificial Intelligence Is Not A Technology. *Forbes*.

Wesslen, R. (2018). *Computer-assisted text analysis for social science: Topic models and beyond*. arXiv preprint arXiv:1803.11045.

Dwijendra Nath Dwivedi is M. Phil in development studies from Indira Gandhi institute of development research. He is currently perusing PhD. From Cracow University of economics Poland. He possesses over 17 years of experience in applying Artificial Intelligence and Advanced Analytics across different industries, e.g. BFSI, Government, Telco, and utilities in various functional areas, e.g. risk and marketing. He conducts AI Value seminars and workshops, for the executive audience and for power users. As a thought leader, he is bridging the gap between business needs and analytical enablers and to drive analytical thinking into successful business strategies. He has published many papers in high impact factor and peer reviewed Scopus indexed Journals.

Ghana Mahanty leads MENA Data science, UAE and Global Data Science Center of Excellent, India for Visa Inc. He is a data enthusiast with expertise in mobilizing the organization to deliver value out of data and specializing in business management, data transformation strategy, and machine learning and artificial intelligence solutions for Retail Banking and Payment industry. He has worked extensively across North America, Asia, Europe and Middle East regions. Currently pursuing PhD. in Economics from Department of Analytical and Applied Economics, Utkal University, Bhubaneswar, India. His research area of interests are machine learning, artificial intelligence, macro econometric and structural data modelling, panel modelling, money, banking and payments.