Utrecht University

Department of Information and Computing Science

Applied Data Science master thesis

Topic modeling and sentiment analysis on Artificial Intelligence tweets

First examiner:

Dennis Nguyen

Second examiner:

Jing Zeng

Candidate:

Shoufan Liao

Student Number:

9088725

July 4, 2023

Abstract

Since the advent of Artificial Intelligence, it has frequently been a subject of public discourse. This study will examine and analyse the content of tweets on Twitter about opinions and comments on Artificial Intelligence. This study examined through the usage of Latent Dirichlet Allocation (LDA) and Correlated Topic Model (CTM) to extract insightful topics and evaluate their performance, as well as conducted sentiment analysis using Vader and identifying top 10 opinion leaders in AI discourse on Twitter. In the data preprocessing, methods of the Spacy and Natural Language Toolkit (NLTK) library in Python have been conducted. The LDA model extracted 19 ideal topics while the CTM model retrieved 9 insightful topics from document collections, but these two models produced similar topics in general. The CTM model showed superior performance compared to the LDA model when evaluated on the coherence score where the CTM model coherence score is higher than that of LDA model. Sentiment analysis, by using Vader sentiment analysis, demonstrated a predominantly positive sentiment towards AI. Moreover, the top 10 opinion leaders, identified by three dimensions of Activity, Popularity, and Influence, all of those opinion leaders expressed positive sentiment towards AI.

Contents

1. Introduction

The COVID-19 epidemic began to spread over the world around the end of 2019, affecting people's lifestyles, causing changing people's lifestyles, preventing economic growth, and rising mortality rates, which has attracted public attention. According to Vaishya et al. (2020), Artificial intelligence (AI) has been applied in the COVID-19 pandemic event for seven aspects, including detection and identification of the infection, follow-up care, locating people's contacts and so on. During the COVID-19 period, the company has conducted AI based on big data to tackle uncertainties to minimize supply chain issues (Sheng et al., 2021). In the big data era, applications of artificial intelligence (AI) can be found in a variety of fields, including economics, health, autonomous driving and so on. Artificial intelligence (AI) applications can be seen in many industries such as economics, linguistics, health, autonomous driving, engineering and so on. Artificial intelligence (AI) has been conducted in the drug industry implementing deep learning and relevant modelling techniques to offer efficient and safe solutions (Zhu, 2020). In the marketing field, Artificial Intelligence (AI) also has been employed to make optimal strategies in order to tackle opportunities and challenges (Sachs,2016). Under Artificial Intelligence processing and techniques, big data can be managed better, which may make the computers perform more likely in human patterns and behaviour (Allam & Dhunny, 2019; Tecuci, 2012). Artificial intelligence (AI) can be helpful in dealing with the enormous amount of data derived from many sources such as the Internet of thing, mobile phone, etc. in many industries from governmental institutions and business corporations. As a result, Artificial Intelligence has attracted a lot of attention from a variety of industries.

As we can see, some applications of AI has achieved remarkable success like robots, natural language processing, and image processing. Thanks to AI, benefits from AI are increased physician performance levels, optimization operation efficiency in the logistics domain, prevention of fraud happening in the finance and banking industry, and standardized products (Nadimpalli, 2007). However, there are some risks and drawbacks to AI. Since Artificial intelligence (AI) has greatly changed and will continue to change human lifestyles, Huang et al. (2022) considered the ethics of AI including discrimination, privacy leakage, unemployment, and security risks and so on which has made these groups of individuals, organizations and society concerned about it. Thus, this study will explore the opinions and comments towards AI. And it is obvious that Twitter, a sort of social media, is a useful forum for people to share their thoughts. Tweets posted on Twitter contain their comments or opinions and sentiments towards events or products.

Therefore, this study explores the Twitter users' discussion about AI and their sentiment towards AI and identifying users who are the opinion leaders in AI discourse on Twitter. To address these questions, this study will use topic modeling to retrieve topics discussed by Twitter users, sentiment analysis to analyse tweets sentiment and identifying opinion leaders to detect the top 10 opinion leaders in AI discourse on Twitter.

This study has two contributions. First, it uses two different topic modeling methods to extract latent topics among document collections and evaluates their performance. Second, identifying opinion leaders is measured by combining three dimensions, which makes the result more reliable.

2. Literature Review

The volume of tweets presents significant challenges. There are hundreds of thousands of tweets talking about AI on Twitter, which is impossible by manually studying the AI discourse because of their scale and complexity. Thus, the need for computational methods has emerged. It is easy to analyze the text data like tweets using text mining, which is a powerful analysis tool for harnessing the potential of unstructured textual data through analysis to retrieve knowledge and uncover crucial patterns and relationships that are concealed in the data (Hassani et al., 2020). For instance, topic modelling methods can be used to retrieve the latent topics in AI discussions on Twitter, and sentiment analysis can provide insights into the overall sentiment of tweets about AI, giving us a sense of twitter users attitudes towards AI.

2.1 Artificial intelligence (AI) and big data

There is no common agreement on the definition of Artificial intelligence (AI). According to Duan et al. (2019), since its initial introduction in the 1950s, Artificial intelligence (AI) has once again become a hot topic due to the development of big data and supercomputing technology. Artificial intelligence (AI) also refers to the deployment of computers that can perform intelligent tasks that typically required human intelligence (Huang et al., 2019). Similarly, there is no common agreement on the definition of big data. Laney (2001) defined big data with three characteristics volume, velocity and variety respectively, which is also called the 3vs model. In the 3vs model, Velocity indicates the timeliness of big data, which must be carried out quickly and in a timely manner; Volume indicates a large amount of data; Variety indicates various types of data formats (i.e. text, documents, images, video etc.) (Chen et al., 2014). Apart from the 3Vs model, Nizam and Hassan (2017) have defined big data in 5Vs model, Volume, Variety, Velocity, Value, and Veracity, where Value refers to the evaluation of data utility that determines the identification of undiscovered values from the gathered data; Veracity refers to focusing on clearing and cleaning up incoming data to make data analysis easier.

2.2 Topic modelling

Topic modelling is an unsupervised method for detecting the main themes from a large number of unstructured document collections (Blei, 2012). Topic modelling postulates documents as mixtures of topics and topics as distributions over words within the document, which is a deeper understanding of human language learning and processing (Griffiths, 2007). Topic modelling is a type of generative probabilistic modelling and has been broadly used in the text mining domain (Liu et al., 2016). Topic modelling can be performed in various methods. In this study, Latent Dirichlet Allocation (LDA) and Correlated Topic Model (CTM) will be exploited.

2.2.1 Latent Dirichlet Allocation (LDA)

In topic modelling field, Latent Dirichlet Allocation (LDA) is one of the most popular techniques. The Latent Dirichlet Allocation (LDA) is proposed by Blei et al. (2003), which aims to identify topics from document collections, LDA assumes each document has different topics and each topic is modelled as a distribution over words. Weng et al. (2010) have already performed the LDA algorithm to extract the latent topics from enormous document collections on Twitter. LDA model can analyze and discover events through wavelet analysis, which technique can be used for identifying events from tweets (Cordeiro,2012). Jelodar et al. (2019) provide a summary of implementing LDA in different disciplines, and they classified these applications into seven different filed, such as social network and microblogs, linguistic science, political science, medical and biomedical, geographical and locations, crime prediction/evaluation and software engineering. In the health-related domain, Prier et al. (2011) conducted LDA algorithm to detect latent topics on a dataset including a significant quantity of tweets, and the result shows this method is pretty good for identifying topics in public health domain. In the politics-related domain, to detect the track of political event trend change on Twitter, the LDA algorithm is used for tracking the trend of topics from tweets posted during the period of presidential election in South Korea (Song et al., 2014). In the linguistic-related domain, LDA has been conducted to, without any manual adjustment, extract the latent topics from 762 published works where the study abstracts consist of terms "sustainability" and "social media", and also the LDA has been used for identifying the hot topic and cold topic by combining with time series analysis (Lee et al., 2021). To evaluate the companies' technology competitiveness, the LDA has been used to extract technological topics and then grouped technologies into specific categories, which is more efficient compared to traditional patent classifications while assessing an organization's technological competitiveness (Wang et al., 2020). In the sport-related domain, people usually talk about sports like football, LDA has been introduced to extract the topic from the tweets about football news in Bahasa Indonesia, and the result shows this model can provide several insightful topics (Hidayatullah et al., 2018). When it comes to sports topics in Indonesian, LDA performed better than that LSI in topic modelling (Negara et al., 2019).

2.2.2 Correlated Topic Model (CTM)

A limitation of LDA, however, is that it assumes topics are independent in document collections, which is not realistic (Blei & Lafferty,2006). An advanced approach has been proposed. The Correlated Topic Model (CTM) has been proposed by Blei & Lafferty (2006), which is a model relaxing the independent assumption of LDA and allows for a more in-depth understanding of the connections between topics where topic proportions use logistic normal distribution instead of a Dirichlet. As the CTM is an extension of LDA, the CTM performed better than LDA when it is introduced to a collection of OCRed articles from the journal Science and also offered a method to visualize and explore unstructured data (Blei & Lafferty,2006). The Correlated Topic Model (CTM) was performed to analyse document collections in the field of higher education, which provides insights into the research landscape of higher education (Daenekindt & Huisman, 2020). Correlated Topic Model (CTM) has been also conducted to detect hidden topics from Lithuanian news media articles to classify topics into three types (Rabitz et al., 2021). The CTM also has been employed in the web service domain to identify the latent factors from web services that performs better matching consumer queries when compared to LDA, PLSA and TextSearch methods (Aznag et al.,2013). The CTM is used for finding latent topics from customer experience reviews in the hospitality industry, and it is beneficial for business with regard to the improvement of customer experience and reputation management (Nave et al., 2018). Dybowski & Adämmer (2018) used the CTM to identify eight salient topics of tax policy news from a collection of presidential transcripts and then explored the relationship between consumer sentiment and economic activity. What's more, in order to deal with the problem that topics are related to one another, the CTM can be used in the facial expression recognition field because it is really impractical to assume each topic is uncorrelated to others, especially for facial expression analysis (Sang & Chan, 2015).

2.3 Sentiment Analysis

Users can post up-to-date opinions on everything and pose their own new idea on social platforms like Twitter. Sentiment analysis is a method that extracts or categorizes sentiment from reviews using natural language processing and text analysis, which has been used in a variety of industries, including marketing, social media, and so on (Hussein, 2018).Three levels of sentiment analysis can be distinguished: document-level, sentence-level, and aspect-level, document-level refers to aiming to identify positive or negative sentiment within the entire document, while sentence-level focuses on the emotion expression in each sentence, and aspect-level seeks to determine the sentiment polarity with relation to a certain aspect of entities

(Medhat et al., 2014). Additionally, Liu (2012) proposed that the difference between document and sentence level is not distinct due to sentence as a short document.
The study of Twitter sentiment analysis has lately caught the attention of many domains. Giachanou & Crestani (2017) did a review of sentiment analysis methods that have been proposed for processing Twitter data, this article gives a discussion about emotion detection, tweet sentiment quantification, and track of sentiment over time. Tweets have been subjected to sentiment analysis to categorize into positive, negative, and neutral (Agarwal et al., 2011). Wang & Fikis (2019) used sentiment analysis to classify the Twitter users' sentiment towards the hashtags #CommonCore and #CCSS between 2014 and 2015. And sentiment analysis was performed to detect sentiment and emotion from tweet content posted by users and comments, then created an emotion network that can be used for identifying the influential person's sentiment (Sailunaz & Alhajj, 2019). In addition, sentiment analysis was conducted to detect sentiment polarity towards the COVID-19 vaccine-related discussion textual data on social platform during the period between December 2020 and May 2021 (Melton et al., 2021). In this study, the tweets are performed sentiment analysis.

2.4 Identifying Opinion Leader

Opinion leaders are those people who occupy the central nodes of the network, having an unequal impact on the mindset and actions of other individuals (Rogers, 2003). According to Morone & Makse (2015), Twitter users who are the node enabling to diffuse most of the information to the whole network are regarded as the influential node. For example, in the policy-making field, the government are regarded as having more significant influence than the non-government (Song & Miskel, 2005). To explore the opinion leader on Twitter, communication network analysis has been used for detecting opinion leaders through five centralities(i.e. Indegree, Outdegree, In-Bonacich Power, Out-Bonacich Power, and betweenness centrality) (Wang & Fikis, 2019). According to Bamakan et al. (2019),

when it is demonstrated how effectively opinion leaders can steer a political stream, sell products, influence economic and marketing trends, or raise public awareness of environmental or public health problems, the significance of recognizing them may become more clear, so it indicates identifying opinion leader can be applied in various industries (e.g. sociology and psychology, education, etc.).
According to Li et al. (2013), opinion leaders are important in social networks because of their ability to influence others' beliefs and opinions through their superior impact. In the public health-related field, Swedish child health promotion has used the snowball method to identify potential opinion leaders (Guldbrandsson et al., 2012). In business and marketing-related fields, it is possible to use opinion leaders to promote goods and services because they can play multiple roles like experts, celebrities, and early adopters (Lin et al., 2018). Additionally, opinion leaders can be utilized to address cold start concerns in the recommender system, and the results suggest that doing so can improve the recommender system's accuracy by giving new users the right recommendations (Mohammadi & Andalib, 2017). In politics related field, to identify the opinion leader from the dataset corresponding to tweets of the Brazilian President political protests in 2015, Rocha et al. (2016) propose a methodology incorporating the detection of prominent users and sentiment analysis.

2.5 Research Question

This study will explore what are Twitter user's discussions and sentiments towards Artificial intelligence (AI). In this regard, this study proposes the following research questions:
Research Question 1: What are the topics Twitter users talking about on Artificial intelligence?
Research Question 2: Which topic modelling method performs better in this context when compared to LDA and CTM?
Research Question 3: What are the most tweets' sentiments about Artificial intelligence?

Research Question 4: Who are the opinion leaders and what are those opinion leaders' sentiment towards Artificial intelligence?

3. Methodology

3.1 Data

3.1.1 Data collection

The dataset, in this study, is provided by the project team with Twitter Application Programming Interface (Twitter API). Twitter API is a set of tools that allows retrieving tweets messages, user information, geographic location, timestamp and other actions on Twitter (Twitter,2023). The data was collected by the English words or phrases or # hashtags (e.g.#ai, #bigdata, #iot, #deeplearning, #artificialintelligence, etc.) related to the Artificial Intelligence (AI) discourse from January 2010 to September 2022. When a tweet matching any of the related AI discourse was collected and then stored in this dataset. There are about 1.1 million tweets in the dataset. In this dataset, it contains the following variables (Twitter,2023), as shown in Table 1.

Table 1. Dataset columns description

Variable	Description
author id	the unique code of the author posted tweet
username	the screen name of the user, is unique but subject to change.
created_at	the time of tweet created
geo	the geographic information of the tweet
id	the unique identifier of tweet
lang	the language of tweet
like_count	the number of likes of tweet
quote_count	the number of quote of tweet
reply_count	the number of responses to the tweet
retweet_count	the number of times the tweet has been retweeted
source	the posted platform of the tweet
tweet	the text content of tweet
in_reply_to	the tweet was in reply to another user
users	the tweet mentioned/tagged other users
followers	the number of followers of the user
tweet count	the number of tweets posted by the user

3.1.2 Ethical and Legal consideration

The data is collected by TwitterAPI, under this tool, only publicly published tweets information can be collected by TwitterAPI. According to Twitter (2023), TwitterAPI allows users to control their own non-public Twitter information and only provide this information to developers using TwitterAPI if they have given permission to access it. Twitter has also informed users how their information data would be accessed and used through TwitterAPI, which contains three components, transparency, control, and approved uses, respectively; transparency refers to providing users with a clear understanding of how and which their data may be accessed and used; control means users can protect tweets(information will not share through API), manage applications(authorized which applications can access to your account) and delete content(not available anymore if deleting); approved uses indicates if a developer wants to infer or derive from sensitive information which is prohibited by Twitter (Johnson,2018). In addition, even if some sensitive tweets are included, it would not be a problem because these data will be aggregated together in topic modelling and sentiment analysis.

3.1.3 Data preparation

The original dataset should be processed before exploring insightful patterns from the data. First, due to just focusing on English tweets, data is needed to check and removed non-English tweets according to the "lang" attribute from the dataset, and the "created_at" attribute indicates the date and changes it into this format (year-month-day). Second, a new column needs to be created called "involve_count", referring to the number of tweets related to AI by aggregation function (sum) of each user, which represents their involvement in AI discourse. Third, duplicated tweets need to be removed like retweeted tweets according to the "tweet" column. Next, we just choose the columns relevant to our study for further preprocessing.
As for more in-depth preprocessing, the text data need to be preprocessed including tokenization, removing stop words, lemmatization, etc. In this

study, Natural Language Toolkit (NLTK) and Spacy library have been used to do data preprocessing, which are packages in Python. Step 1(text cleaning): remove URLs, stand-alone numbers, punctuations, hashtags(#), and mention(@) from the tweets. Step 2(Filter Nouns): only select Nouns using Spacy tool for the first round. Step 3(Lowercasing and lemmatization): convert all words and characters into lowercase in case of sensitivity of the case, then reduce words to their root form. Step 4(tokenization and removing stop words): tokenize the tweet into individual words or tokens, and remove common and insignificant words as well as punctuations since they do not make any sense to semantic meaning, English stop words were used from the NLTK library and also added personalized stop words. Step 5(removing emojis and emoticons): remove the emojis and emoticons in tweets. Step 6(noisy character removal and second Nouns filtering): remove noisy characters like 'rt', 'gt', apostrophe and others, then use the part-of-speech(pos) method from NLTK to only select Nouns for the second times in order to ensure clean data better. When performing data preprocessing in sentiment analysis, there is a little difference, only the Step 1(text cleaning) will be conducted in the in-depth preprocessing. The emoticons, emojis and punctuation will be included because Vader use these characters to represent intensity and polarity (Hutto & Gilbert,2014; Na et al.,2021).

3.2 Topic Modeling

3.2.1 Latent Dirichlet Allocation (LDA)

LDA is an approach to identifying latent topics from large document collections without labelled data since it is an unsupervised method (Blei et al.,2003). In this study, LDA model used the LdaModel algorithm from the Gensim library built in Python (Řehůřek & Sojka,2010). Before implementing LdaModel, several steps need to be done. First, although it has already removed stop words and punctuations, there are still some noisy words needed be filtered out that can be done through the filter_extremes method. This method has two parameters named no_below and no_above, where no_below indicates removing tokens that appear in less than the minimum times in documents, and no_above refers to removing tokens that appear in more than the threshold fraction of the total number of documents, where the fraction ranges from 0 to 1. After implementing many experiments, the no_below is set to 70, and the no_above is set to 1. Second, two hyperparameters, the number of topics (k) and the number of iterations, can significantly influence topic performance. Based on the examination of the number of topics from 2 to 24, this study has selected the optimal number of topics as 19 (coherence value: 0.5748) with the optimal coherence value, as shown in Figure 1. With regard to the number of iterations, this study has implemented the iterations range of 50 to 1000 by manual, and the result indicates the optimal number of iterations is 200.

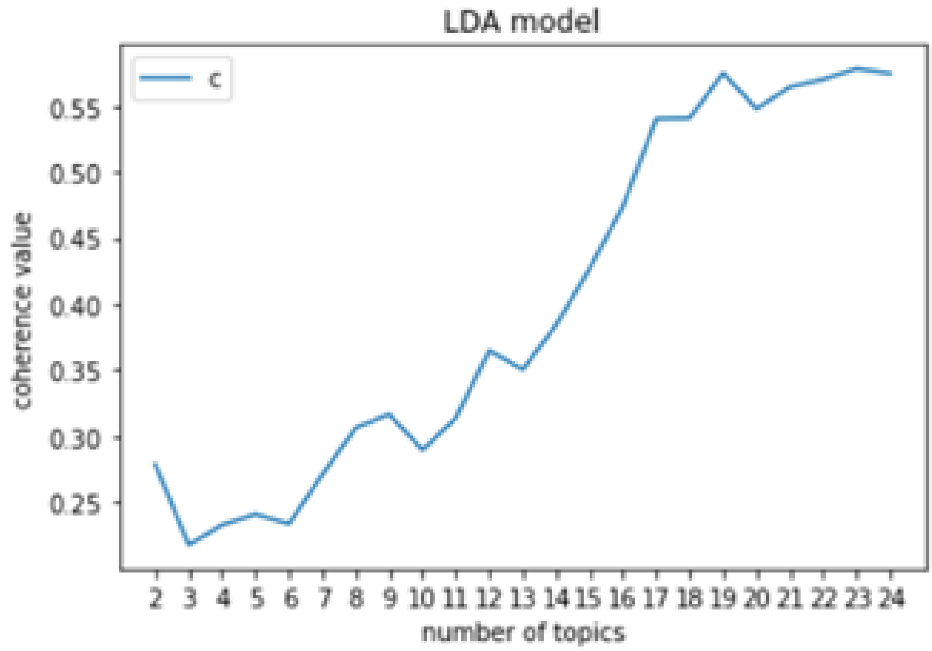


Figure 1. Detecting optimal number of topics using coherence value in LDA

3.2.2 Correlated Topic Model (CTM)

However, there is a limitation of LDA that LDA assumes topics are independent with collection documents (Blei & Lafferty,2006). To address this independence issue, Correlated Topic Model (CTM) has been used. The most benefit of CTM is that it can identify the dependencies and relation-

ships existing among topics within a collection of documents.
This study uses CTMmodel method of tomotopy package from Python (Lee, 2022).For CTMmodel, three hyperparameters also needed to be set. The first one parameter is the number of topics (k). The second parameter is the minimum document frequency of tokens(min_df). The last parameter is the number of top words to be removed(rm_top). When considering min_df, this study min_df is set to 70 and the rm_top is set to 0, to maintain the corresponding to LDA model parameters, which indicates any token appears in fewer than 70 documents will be removed. As for the number of topics k, this CTM model is trained for each number of topics within the range from 2 to 24, selecting the optimal number of topics with the optimal coherence value, which finally is set to 9(coherence value: 0.6557), as shown in Figure 2. In CTM model, the model is trained many times of iterations. To select the best iterations, in this study, iteration has been set to 50 after performing multiple tests.

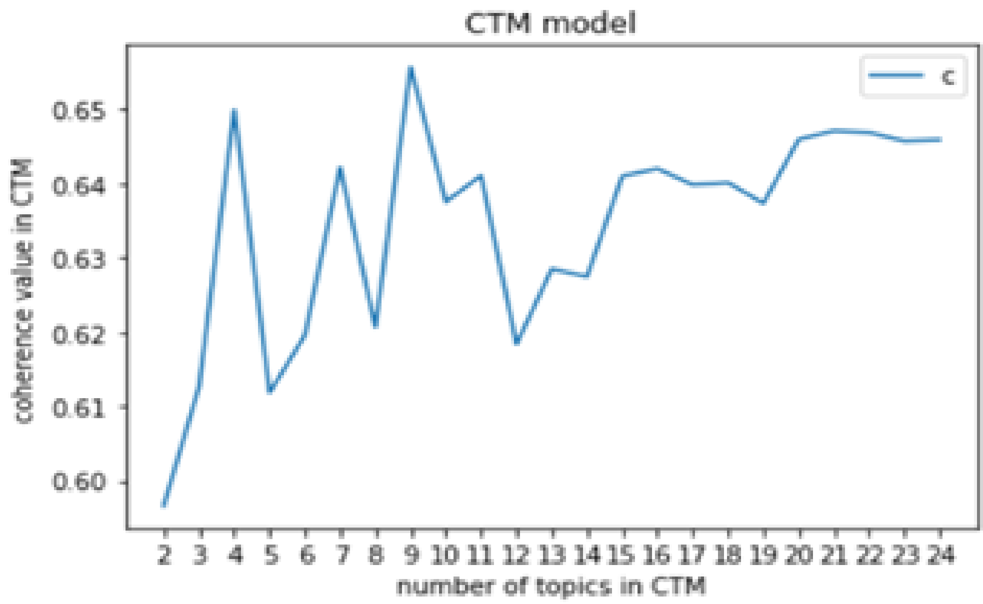


Figure 2. Detecting optimal number of topics using coherence value in CTM

3.2.3 Evaluation of topic modeling

To find the optimal number of topics, there are three types of evaluation of topic models, qualitative (one is keeping track of a topic's top n words, the other is word and topic intrusion), quantitative (Perplexity and Coherence) and mixed approach (Giri, 2022). A lower perplexity score indicates greater performance because it quantifies how well a model predicts unobserved data (Blei et al.,2003; Giri, 2022). The research, however, has shown that perplexity negatively correlates to human interpretability (Chang et al.,2009). Qualitative measures like word intrusion and topic intrusion are also good way to evaluate the performance of topic modelling while it is time consuming and complexity. Therefore, this study will not consider the word intrusion and topic intrusion measures. Thus, this study will only use coherence to evaluate the performance of LDA and CTM models. The coherence quantifies the degree of semantic relatedness between topics, where the higher coherence score indicates the more interpretable topics (Mimno et al.,2011). And the coherence has been conducted to find the optimal number of topics in LDA and CTM (Na et al.,2021; Melton et al., 2021).

In this study, therefore, to find the optimal number of topics, the coherence score is used in LDA and CTM models. In addition, the coherence score is also used for evaluating the performance of LDA and CTM models. The coherence score has four popular metrics and has been built in Gensim Library in Python (Řehůřek & Sojka,2010), u_mass, c_uci, c_npmi, and c_v, especially the c_v metric of coherence had been regarded as performing better, more detail explanation can be seen in Röder et al.(2015) . This study will use the 'c_v' metric to measure the coherence score in LDA and CTM, which is the common metric implemented by many research studies. In brief, the coherence score is a meaningful and useful measure.

To validate the results of LDA and CTM models, the metric of coherence value and manual inspection will be used to evaluate the quality of LDA and CTM models, where manual inspection refers to inspecting the most dominant words in each topic if they make sense and related, and coherence value indicates the higher the value, the better the model.

3.3 Sentiment Analysis

Vader sentiment analysis method has been used for sentiment analysis from tweets about the 2016 US election by Elbagir & Yang (2019) and from tweets content about covid-19 by Abdulaziz et al. (2021). Hutto and Gilbert (2014) proposed the Valence Aware Dictionary and sEntiment Reasoner (VADER) tool, which is a rule-based and lexicon sentiment analysis tool with a focus on social media, and it demonstrated that Vader outperformed more established sentiment lexicons like LIWC in terms of benefits. The Vader sentiment analysis can be used for classifying polarity (positive, negative or neutral) and also calculating the intensity of emotion.
In this study, therefore, sentiment analysis is implemented through the Vader sentiment analysis tool from Natural Language Toolkit (NLTK) package built in Python. According to Hutto and Gilbert (2014), the Valence score varies from -4 to +4, where -4 denotes the most negative sentiment and +4 denotes the most positive, and the compound score is determined by adding up each word's value score, and then normalized to fall between -1 (the most extreme negative) and +1 (the most extreme positive), indicating that the compound score is the most useful metric to measure a particular sentiment. In this study, it will adopt the typical threshold compound value (Hutto & Gilbert,2014) to classify tweets polarity into positive, negative and neutral. The threshold value is set as following:
(i) Positive, if compound value >= 0.05
(ii) Negative, if compound value <= -0.05
(iii) Neutral, Otherwise
To verify the results generated by Vader, 100 tweets were randomly selected from the dataset, manually labeled by the author, and then compared with the sentiments labeled by Vader. As a result, the accuracy rate is over 0.7, indicating that Vader worked correctly. While there are some limitations of Vader such as sarcasm and irony performed misinterpreted, and lack of large context understanding, it is still the optimal choice in social media sentiment analysis (DeLancey,2020; Wang et al.,2020).

3.4 Identifying Opinion leader

To detect the opinion leaders from Twitter whose tweets are posted related to political activism, three variable dimensions have been conducted: social connectivity, user involvement and user identity (Xu et al., 2014). Sailunaz & Alhajj (2019) identified opinion leaders from Twitter users text calculating influence score through five dimensions: the number of followers, the number of likes, the number of retweets, the number of agreed comments and the number of disagreed comments. In terms of identifying opinion leaders on Twitter, Riquelme & González-Cantergiani (2016) classified the measures into three categories: Activity (measured by how active the user is), Popularity (measured by how well-known the user is) and Influence criteria (measured user's action influence on other users in the Twitter network). Opinion leader was detected by three dimensions: popularity, activity and authority from tweets related to sports events in 2016, examining the proxies for popularity, activity, and authority, including the number of followers, the number of tweets, and PageRank, all of which show that a single indicator is insufficient to identify opinion leaders (Lamirán-Palomares et al., 2019).

In terms of Activity measure, the number of tweets has been proposed to rank users on Twitter (Nagmoti et al., 2010). Riquelme & González-Cantergiani (2016) proposed the General activity measured by the number of original tweets posted, number of replies posted, number of retweets accomplished by the user and number of tweets of other users marked as likes by the author. In addition, Pal & Counts (2011) proposed Topical signal (TS) to assess the extent of user involvement in a specific topic where TS is calculated by adding up the user's total number of original tweets, replies, and retweets, then dividing by the total number of tweets.

With regard to the Popularity measure, it has been suggested to calculate the ratio of followers and the summation of followers and followees (Nagmoti et al., 2010). Aleahmad et al. (2016) used the exponential of the number of followers as shown in equation (1), where λ is a constant set as 1 by default. As for influence criteria, Retweet Impact (RI) is one of the metrics to evaluate the user's text content's impact by calculating the number of retweeted tweets, which was proposed by Pal & Counts (2011), where RI is calculated by the number of original tweets by the user and retweeted by other users, then multiplied by the logarithm of how many users have retweeted the author's tweets in total. Retweet and like in tweets can be regarded as two powerful actions (Riquelme & González-Cantergiani 2016).

Because of the various of metrics for detecting opinion leaders, there is no common agreement equation or gold standard to detect the opinion leaders on Twitter. Based on the dataset, in this study, Activity, Popularity and Influence will be used for identified opinion leaders, and the Popularity measure used as equation (1), and the Activity measure used as equation (2) where the number of tweets related to AI indicates how many tweets the user posted involved in AI event on Twitter. As for Influence measure, this study will adopt the number of posted tweets excluding retweeted tweets marked as 'n_tweets', then multiplied by the logarithm of the summation of 'like_count' and 'retweet_count', marked as "lr_times"(lr_times= "like_count"+ "retweet_count"), thus the influence score shows in equation (3). Then, normalizing the Popularity, Activity and Influence score is used by the Min-max normalization to scale value in the range from 0 to 1.

However, to avoid some user accounts by artificially buying followers, likes counts and retweet counts, or those users having the most followers but not involved in the AI discourse. There is a correlation between the quantity of followers and "lr_times", as shown in Figure 3. There are some outlier users that should be removed. Interquartile Range (IQR) is a good measure to detect outliers, setting Q1 minus 1.5 *IQR and Q3 add 1.5 *IQR as the boundary where Q1 is the 25th percentile and Q3 is the 75th percentile in the dataset (Thomas,2022).

After removing outliers, a new variable, Indicator finally is generated by adding Popularity and Activity and Influence scores together. The calculation of this Indicator is shown in equation (4). This Indicator variable is used for detecting opinion leaders, where the higher the value, the more likely the user is an opinion leader.

$$\textbf{Popularity} = 1 - e^{-\lambda * \#\text{followers}} \tag{1}$$

$$\textbf{Activity} = \frac{\text{number of tweets related to AI}}{\text{total number of tweets}} \tag{2}$$

$$\textbf{Influence} = \text{n\_tweets} * \log(\text{lr\_times}) \tag{3}$$

$$\textbf{Indicator} = \text{Popularity} + \text{Activity} + \text{Influence} \tag{4}$$

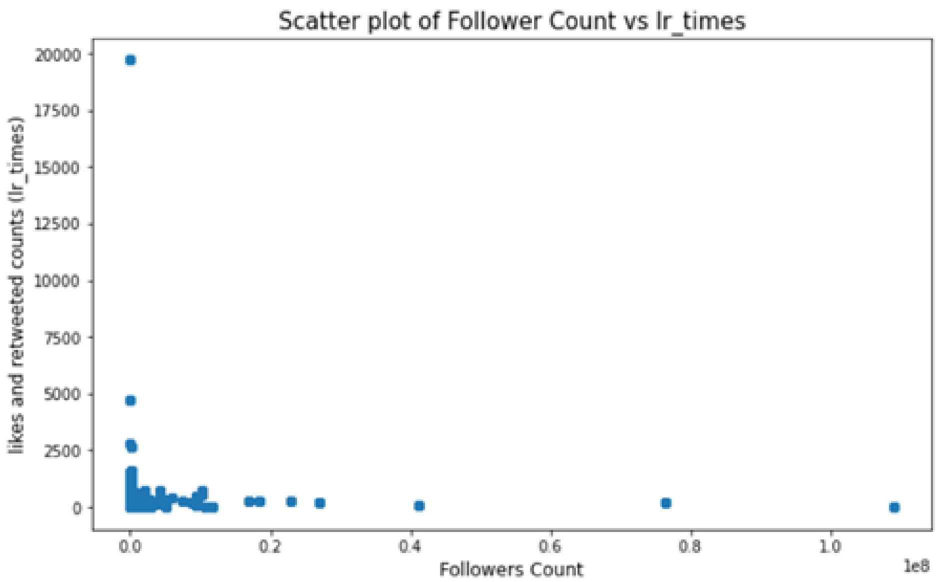


Figure 3. Removing outlier opinion leaders

4. Results and Analysis

In this section, the result will be represented. There are five parts of results, including the topics generated by the LDA and CTM models, the performance result of LDA and CTM models, sentiment analysis and identifying opinion leader.

4.1 LDA topic model

After performing the LDA model, it generated 19 ideal topics, which is shown in Table 2, and each topic was manually annotated topic description. The annotation of the topic is chosen by observing the top terms for each topic with their probability and randomly checking if the tweets are most likely belonging to the topic. Proportion in Table 2 shows the average prominence of each topic across all documents, where a higher proportion value means that the more likely an average document would focus on that topic. In addition, each topic's top 10 words are also visualized by using word cloud, as shown in Figure 4. In word cloud, the much larger size of a word, the more important weight is in a topic.

Although LDA model generates a slight difference in topics when running each time, topics each time extracted are similar in general. There are 19 topics extracted from the document collections by LDA model, these topics are related to job employment, education, architecture, video game, robots, finance, global market, energy and health. The 19 topics extracted can represent an overall view of discussion of AI discourse on Twitter. Topics can be categorized into three types, Finance, Business & job, and Entertainment & Society. With regard to Finance, topics 4 and 14 can be grouped into this classification, talking about the implementation of AI and its influence on the Finance industry. As for Business &job, topics 1, 3, 6, 9, 10, 11, 12, 13, 16, 17 and 18 can be categorized together. These topics are mainly talking about the AI affecting global business and employment, which indicates traditional business will be replaced by digital business and traditional ways of product and design will also be impacted, as well as the job employment is going to receive changes. Topics 2, 5, 7, 8, 15 and 19 are classified into Entertainment & Society, which indicates AI will change human entertainment style and make unrealistic dream become realistic in future and also influence future education and healthcare. In general, the majority of topics are concerned with AI influence in Business and the job market. In addition, considering the average prevalence of each topic across the document collections, Topic 1 ("Intelligence Technology in Design and Architecture"), which is characterized by a higher proportion value of approximately 36%, emerges as the most prominent. The topic 1 maintains the highest proportion during the period and it reached peak of over 40% in 2016, more detail of which can be seen in the appendix B in the figure 18.

Table 2. Topics generated in LDA model

Topic	Words	Topic Description	Proportion
1	intelligence technology system art decision program issue detail line home	Intelligence Technology in Design and Architecture	35.88%
2	video game conversation trading eye journey man artist war core	Video Game	4.01%
3	model book reason place plan type music law ability choice	Legal Aspects in Creative Industry	2.84%
4	application growth control view engine bank room color press passion	Application in Financial industry	2.72%
5	process piece word page employee cancer training heart email song	Health Awareness in workplace	3.6%
6	business edge content country internet history environment member production train	Digital Business and Production	3.1%
7	team life risk person number fan shop battle shot reaction	Digital Sport	3.61%
8	idea care knowledge class humanity episode practice picture dream career	Personal Ambition and Human Care	3.05%
9	world article event student performance end change energy city demand	Global Events and Changes	5.25%
10	design security hand quality help stage need show treatment emotion	Design Quality and Security	2.5%
11	problem management order friend moment communication tv limit guide attack	Problem Solving and Management	2.32%
12	image question story car study search drug answer discovery kid	Automatic Drive	3.31%
13	work network code source brain skill interview money vision fun	Career and Job Market	3.51%
14	machine service fact body possibility stuff amount tax fear ground	Automatic Service in Financial Transaction	3.8%
15	topic baby school position road supply meeting hype house file	Education and Family Life	2.03,
16	way language level prediction report term group worker version assistant	Technology impact in Multilingual Work Environment	4.46%
17	project future job power impact paper reality point woman kind	AI Impact on Employment	5.22%
18	company course voice developer child ticket traffic opinion girl weather	Voice Technology Business	3.91%
19	robot case computer value mind action food rule folk movie	Implications of Robotics	4.88%

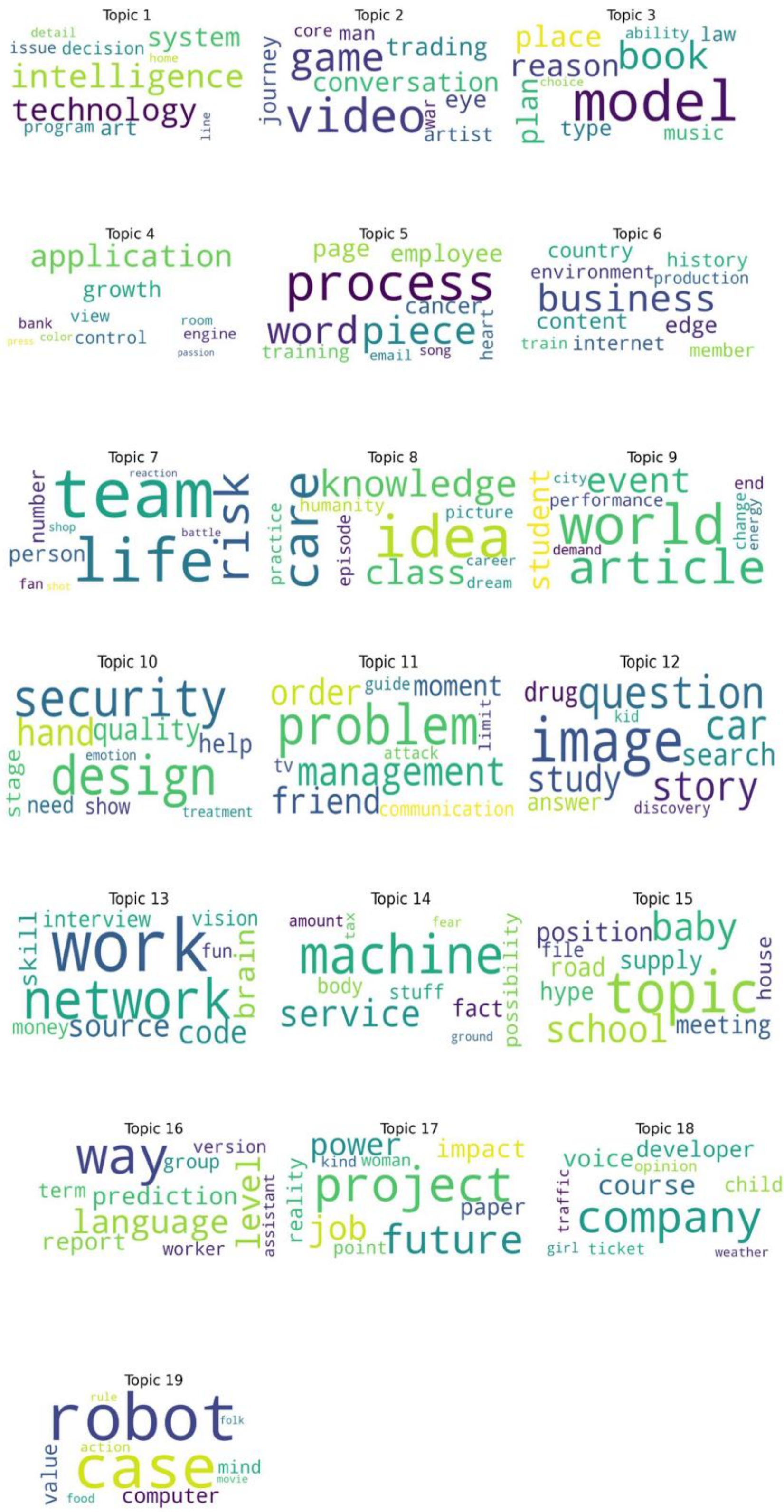


Figure 4. Topic generated in LDA model using word cloud

4.2 CTM topic model

After implementing the CTM model, the topic distribution among the document collections, generated 9 interpretable topics, which is shown in Table 3, and each topic was manually labelled topic description. The explanation of the topic is chosen by examining the top terms for each topic with their probability and randomly checking if the tweets are most likely belonging to the topic. Proportion in Table 3 shows the average prominence of each topic across all documents, where a higher proportion value means that the more likely an average document would focus on that topic. In addition, each topic's top 10 words are also visualized by using word cloud, as shown in Figure 5. The much larger size of a word, the more important weight is in a topic.
When compared to the LDA model, the CTM model generated a smaller number of topics. These 9 topics can be grouped into two categories, Business, and Media & Entertainment respectively. The first category is Media & Entertainment including topics 1,3,4,5,6,7 and 9, the prominent words for this category, for instance, are movie, music, tv, computer, and game. By contrast, the other two topics are related to the Business domain, the most probable words are business, project, world, service customer, etc. In brief, the topics generated by CTM and LDA models are similar in general because they are mainly related to business and entertainment industries. In addition, considering the average prevalence of each topic across the document collections, topic 1 ("Perception and Discussion of Movie Industry"), with a higher proportion value of about 28%, emerges as the most prominent. The topic 1 maintains the highest proportion between 2010 and 2022 and it reached peak of over 30% in 2011 while topic 3 nearly reached its peak of about 20% in 2010, more detail of which can be seen in the appendix B in the figure 19.
In this study, the CTM model depicts the topics relationships using the Network library in Python, as shown in Figure 6. It only plots the top-tenth percentile correlation between topics to find the most significant relationship. According to Figure 6, topic 3,4,5,6 and 7 shows significant correlation,

where the correlation value of topic 3 and 4 is 0.1978, the correlation value of topic 4 and 5 is 0.1838, the correlation value of topic 5 and 6 is 0.1814, the correlation value of topic 6 and 7 is 0.1384. In topics 3 and 4 with the highest correlation score of 0.1978, the most prominent words in topic 3 are: "intelligence, song, job, judge, show, winner, class, singer, etc.", and topic 4 are: "machine, voice, story, human, image, word, point, interview, etc.". These words can be interpreted as the discussion topic is mainly about AI impact in the media context, these two topics are significantly correlated which makes sense. The least significant correlation is topics 6 and 7 with a 0.1384 correlation score. The most probable words are topic 6(life, article, work, video, model, car, player, book, program, software, etc.), and topic 7(game, computer, practice, problem, case, process, challenge, level, value, audition, etc.), these topics indicate the impact of AI in media and video games, which will influence our lifestyles.

Table 3. Topics generated in CTM model

Topic	Words	Topic Description	Proportion
1	industry brain research movie mind event market kid scientist science	Perception and Discussion of Movie Industry	27.69%
2	world project solution service power platform result development opportunity reality	Powerful Solutions for Global Market	12.4%
3	intelligence song job judge show man winner class woman singer	Music Judgement	13.23%
4	machine voice story human image word point interview student risk	Risks and Opportunities in Human and Machine Interactions in Education	9.42%
5	company girl team tool application tv boy impact expert release	Impact of TV and Application on Youth	8.55%
6	life article work video model car player book program software	Media and Software in Lifestyle	8.12%
7	game computer practice problem case process challenge level value audition	Computer Game	7.12%
8	way robot business experience customer idea agent reason face innovation	Robot for Business	6.58%
9	technology future system performance question fan network music art step	The Future of Entertainment in Arts	6.89%

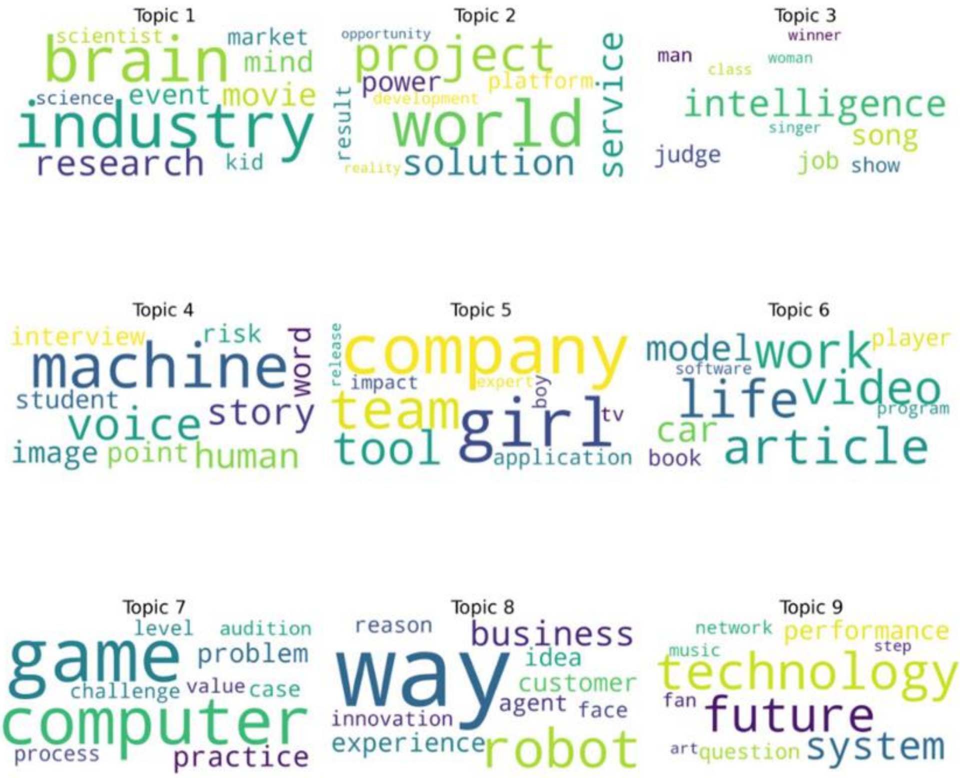


Figure 5. Topic generated in CTM model using word cloud

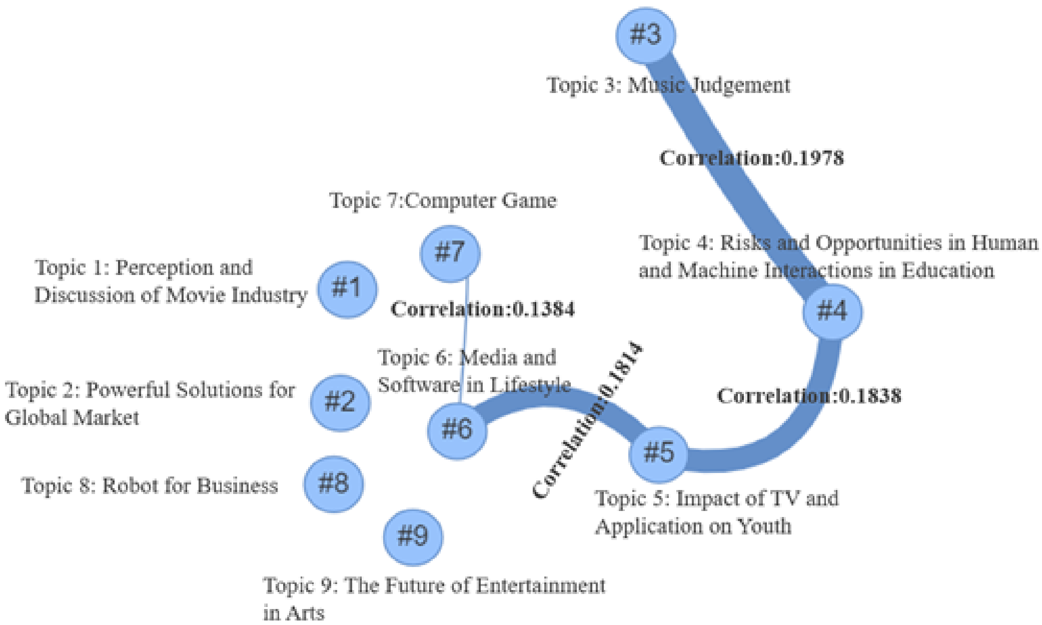


Figure 6. Topics significant correlation in CTM model

4.3 Evaluation of LDA and CTM

In this subsection, this study will evaluate the performance of LDA and CTM models implementation on the dataset. Perplexity and coherence of quantitative measures can be used for evaluating topic quality. However, perplexity is considered not a perfect metric and even gives a negative correlation with human interpretability when evaluating topic quality (Chang et al.,2009; Newman et al.,2010; Ray et al.,2019). Therefore, this study will use the coherence score to evaluate the performance of the LDA and CTM models. In the LDA model, it selected 19 topics as the optimal number of topics with a coherence value of 0.5748. After selecting the 19 topics in LDA model, each topic is validated by observing the top words which is also make sense and easily interpretable, indicating that the result generated by LDA model is meaningful. By contrast, the ideal number of topics has been chosen as 9 in the CTM model, where the coherence value stands at 0.6557. Similarly, when selecting the 9 topics in CTM model, each topic is validated by observing the top words and they are interpretable to human.

In addition, this study changes the number of topics from 2 to 24 and calculates the coherence score in LDA and CTM models, as shown in Figure 7. In Figure 7 the CTM model coherence score is always higher than that of LDA model, which indicates CTM model performs better than that of LDA model. It is also worth noting that the coherence experience fluctuation in the LDA model starting at around 0.28 then grows slowly and finally reached its highest point at about 0.57, while the coherence in CTM remains stable where it starts at near 0.6 and reached the highest value with 0.65 or so. This study has also computed the coherence score using ′u_mass′, ′c_uci′ and ′c_npmi′ measures and the mean coherence score of these metrics over all the topics ranging from 2 to 24 showing CTM model coherence score is a little higher than that in LDA model, more detail of which can be seen in appendix B in the figure 17. As a result, it indicates the performance of CTM model is better than of LDA model based on the coherence score as the evaluation metric in this study.

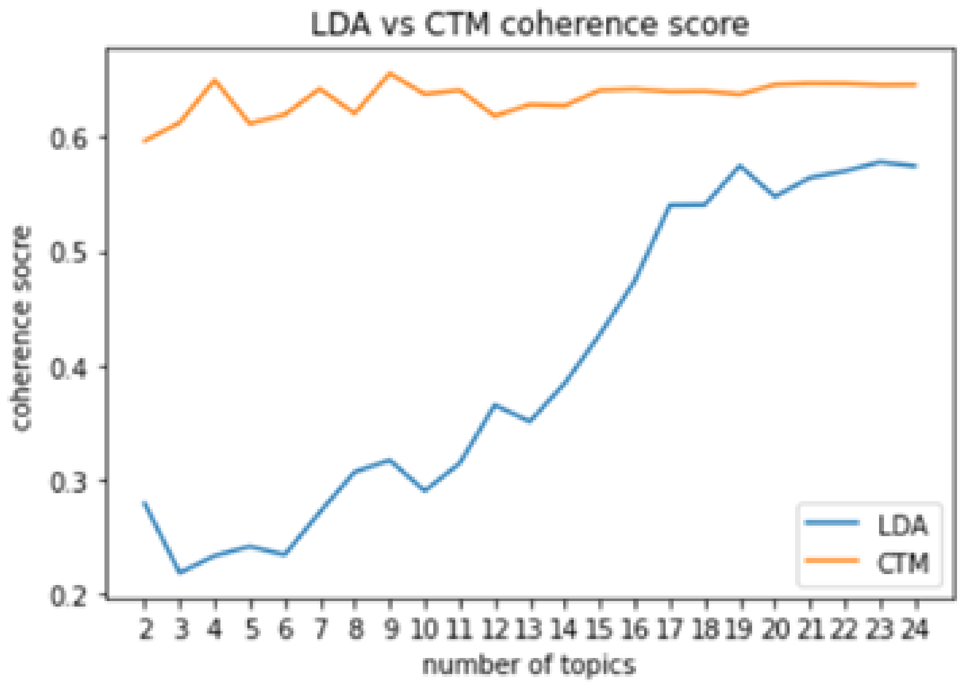


Figure 7. Coherence score in LDA and CTM model

4.4 Sentiment Analysis

In this study, each tweet is analysed by using the Vader sentiment analysis tool in Python. Then each tweet is classified into three types of polarity, 'Positive', 'Negative', and 'Neutral'. It shows the proportion of tweets that are classified as 'Positive', 'Negative', and 'Neutral', where half proportion of tweets toward the AI are Positive sentiment while only about 16% of tweets are Negative, as shown in Figure 8.

Figure 9 shows the number of tweets in different polarities sentiments per year. The number of tweets in Positive sentiment is always higher than that in Negative sentiment. The number of tweets talking about AI shows an overall decreasing trend. In the year 2011, however, both Positive and Negative sentiment graph reached their peaks respectively, indicating AI was the most popular topic in this year when compared to other years.

In terms of intensity sentiment score, this study analyses the total intensity sentiment score per year and each tweet's average intensity sentiment score per year, respectively. In Figure 10, the total intensity sentiment score of each year is shown, and it reflects that the absolute total intensity senti-

ment score of Positive sentiment far outweighs the absolute value of Negative sentiment between 2010 and 2022. Interestingly, in the year 2011, each graph reached its peak respectively where Positive intensity sentiment score is over 4000 and Negative intensity sentiment score is below -1500, which is also corresponding to the polarity sentiment distribution in 2011. However, the difference between the total intensity sentiment score of Positive and Negative is becoming smaller with time flying by, which indicates users are more familiar with AI and are no longer as resistant or crazy about it.
When it comes to each tweet's average intensity sentiment score, this study computes the average intensity sentiment score by using the total intensity sentiment score of each polarity divided by its number of tweets, as shown in Figures 11 and 12. The average intensity sentiment of Positive and Negative both show fluctuation. Similarly, the Positive average intensity sentiment reached the maximum value at about 0.56 in the year 2011 and the Negative average intensity sentiment got a minimum value at about -0.47 in 2011, which implies the discussion of AI is the most intensive in this year. Nevertheless, in the year 2016, the Positive average intensity sentiment score dramatically dropped to its lowest point of about 0.49, which indicates the tweets towards AI become less positive in this year.

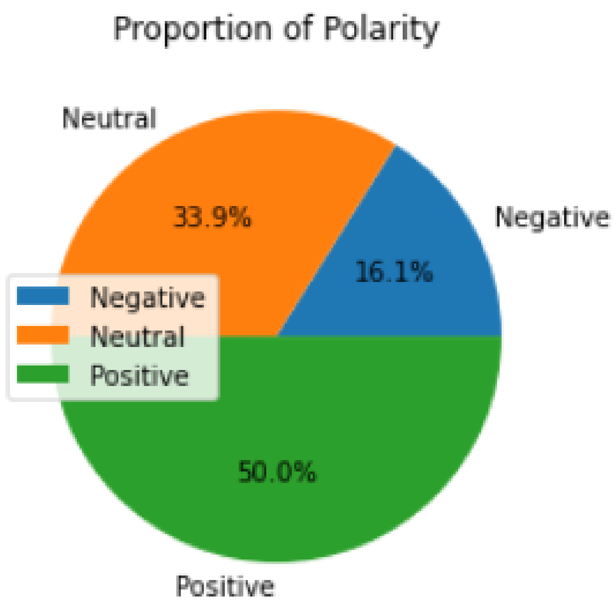


Figure 8. The proportion of polarity

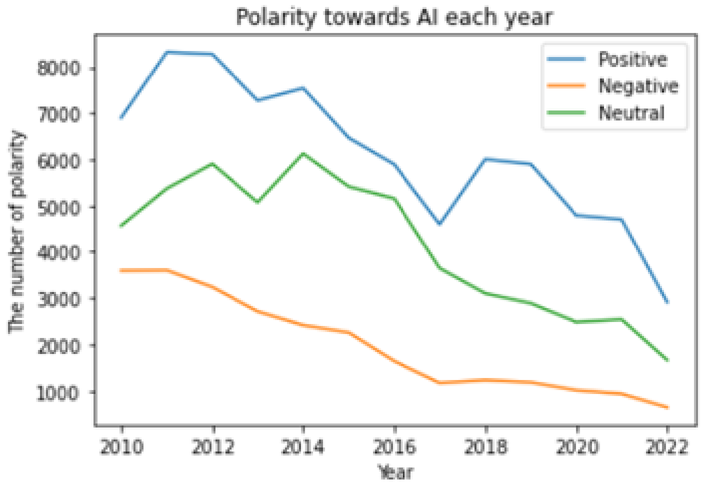


Figure 9. The polarity per year

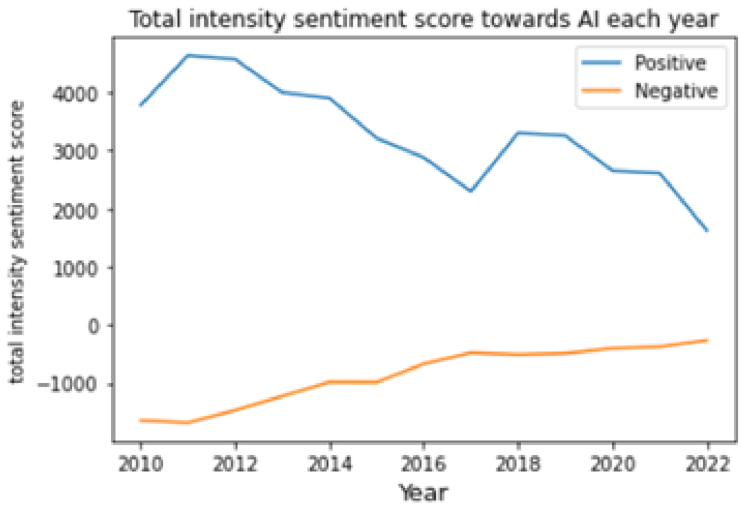


Figure 10. The total intensity sentiment score per year

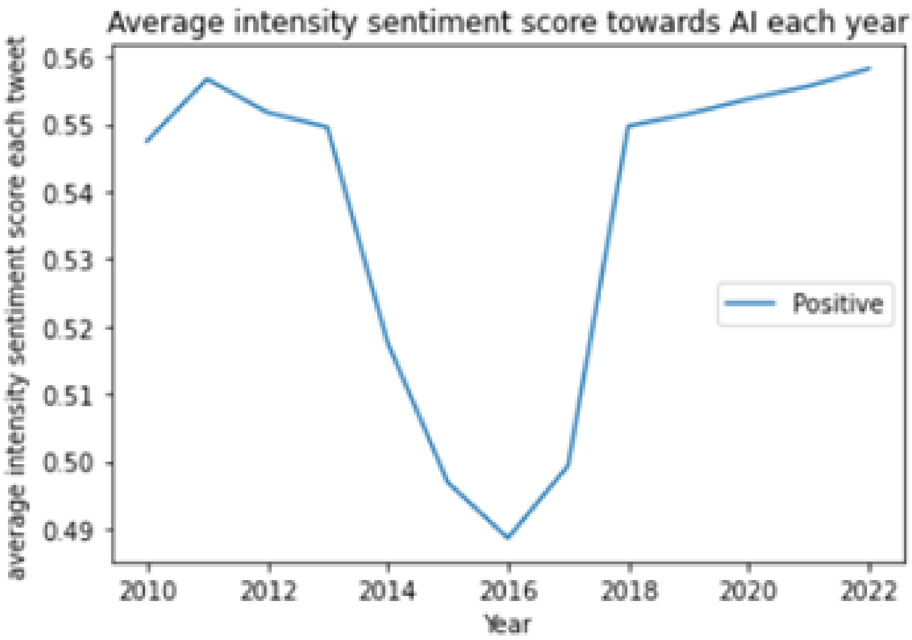


Figure 11. The average intensity sentiment score in Positive per year

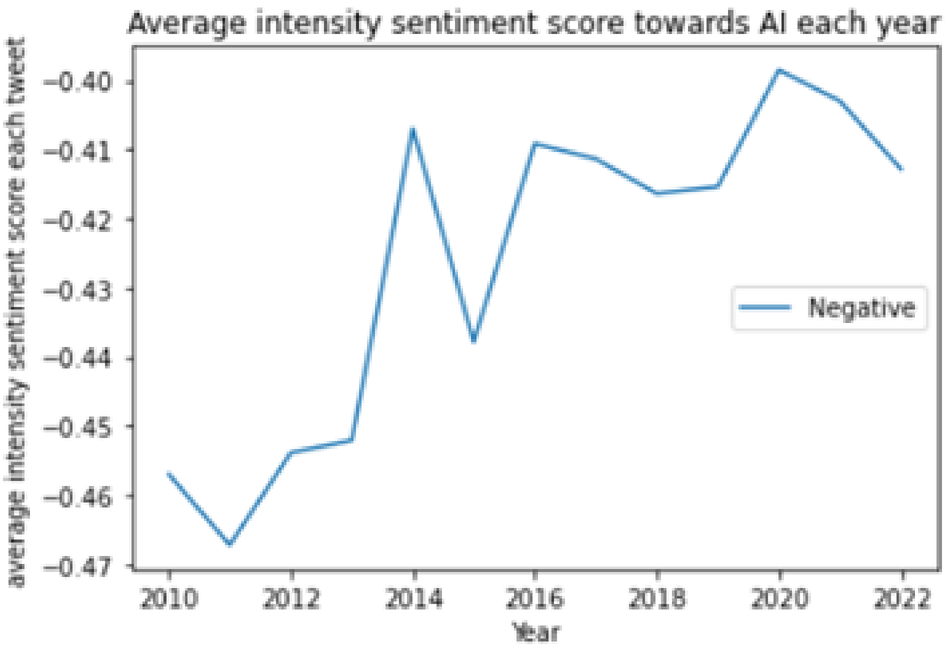


Figure 12. The average intensity sentiment score in Negative per year

4.5 Identifying opinion leader

To detect the opinion leader in Twitter social communication, this study uses three features, Activity, Popularity and Influence. This study identifies the top 10 opinion leaders from the dataset, the result is represented using word cloud in Figure 13.
First, this study explores what is the polarity of those opinion leaders towards AI, and the result is represented in Table 4. In this study, the sentiments of the top 10 opinion leaders are classified by calculating the net difference between the quantity of "Positive" and "Negative" tweets posted by each of these opinion leaders. If the count of "Positive" tweets surpasses that of "Negative" tweets, the opinion leader is classified as "Positive" shown in the "Label" column of Table 4, and vice versa. The result reveals a prevailing positive sentiment towards AI among these opinion leaders.
Second, the intensity sentiment score of those top 10 opinion leaders is also explored and analysed, as shown in Figures 14 and 15. When the total intensity sentiment score of each opinion leader is higher than 0 then it will be labelled as 'Positive', otherwise 'Negative'. This result shows each of these opinion leaders is 'Positive', which is also identical to the outcome as mentioned earlier in Table 4. The top 10 opinion leaders are all regarded as 'Positive' sentiment towards AI. With regard to username 'Zayy7_', this username has the highest value in total intensity sentiment score of nearly 60 and also the largest average intensity sentiment score was over 0.35.
In brief, the top 10 opinion leader identified by this study, expressed 'Positive' sentiment towards AI, and most of them are the stakeholders in AI discourse.
In addition, 19 topics are generated by LDA model, and the topic1 labelled with "Intelligence Technology in Design and Architecture" is most engaged topic by all these top 10 opinion leaders. And the username "Zayy7_" shows the significant involvement in the topics 2(labelled with "Video Game") and 8(labelled with "Personal Ambition and Human Care") while the username "ilanawaber" also has a lot of engagement in the topic 8, more details

of which can be seen in the appendix B in the figure 20.
By contrast, there are 9 ideal topics generated by CTM model. Overall, all these top 10 opinion leaders were deeply involved in the discussion of topic 1 (labelled with “Perception and Discussion of Movie Industry”). Nevertheless, in the topic 3(labelled with “Music Judgement”), the username “Zayy7_” and “franborrell” heavily involved in the topic 3 while the username “Zayy7_” and “ilanawaber” are also more associated with the topic 6(labelled with “Media and Software in Lifestyle”), more details of which can be seen in the appendix B in the figure 21.
In general, most of opinion leaders participated in the discussion of topic 1 generated from both LDA and CTM models. Moreover, in addition to the engagement on topic 1, some opinion leaders (e.g. "Zayy7_", "ilanawaber") have more engagement on other topics compared to other opinion leaders.

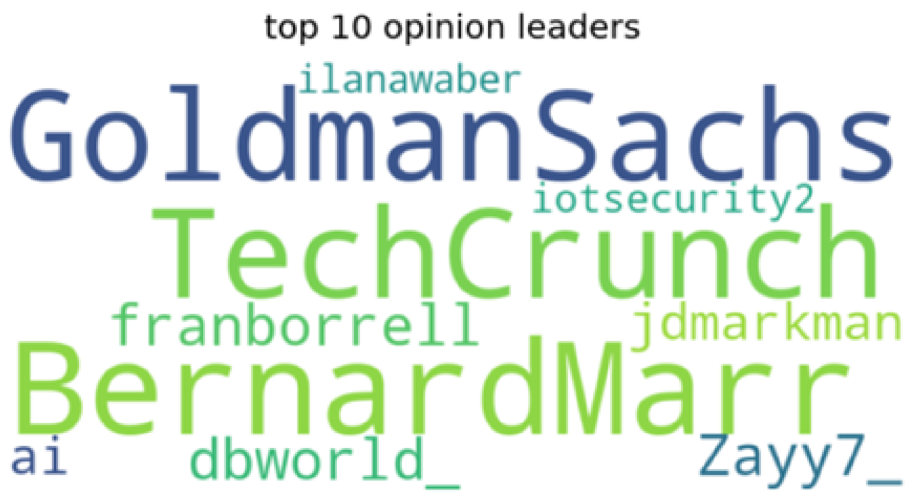


Figure 13. The top 10 opinion leader visualization in word cloud

Table 4. The top 10 opinion leaders sentiment

Username	Label
ilanawaber	Positive
iotsecurity2	Positive
ai	Positive
jdmarkman	Positive
franborrell	Positive
dbworld_	Positive
Zayy7_	Positive
TechCrunch	Positive
GoldmanSachs	Positive
BernardMarr	Positive

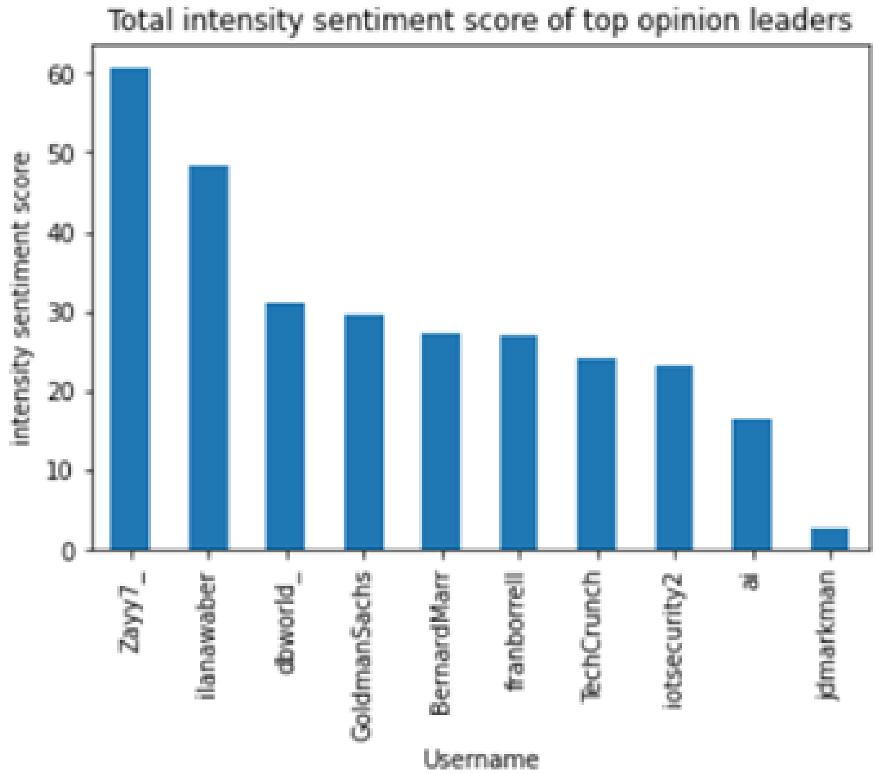


Figure 14. The top 10 opinion leader total intensity sentiment score

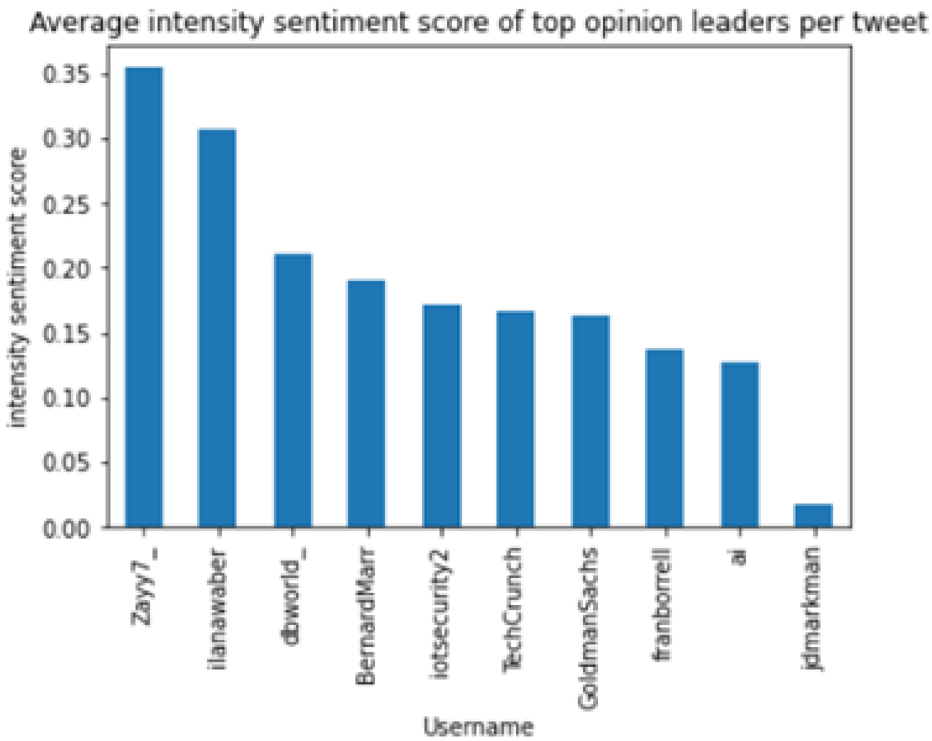


Figure 15. The top 10 opinion leader average intensity sentiment score

5. Conclusion and Discussion

In this section, the answer to the research question, conclusion, limitations and future work will be represented.

5.1 Answers for the research question

The answer to Research Question 1: What are the topics Twitter users talking about on Artificial intelligence? The result has been represented in Tables 2 and 3. On the one hand, from the LDA model, the topics are about job employment, education, architecture, video game, robots, finance, the global market, energy and health. On the other hand, topics generated by the CTM model, topics are related to movies, music, tv, computer game, business, and world customer service. Briefly, these topics are mainly related to business and entertainment industries.
To answer Research Question 2: Which topic modelling method performs better in this context when compared to LDA and CTM? This study used the coherence score as the evaluation metric, and the result is shown in Figure 7. It can conclude that the performance of the CTM model is better than that of the LDA model because the CTM model has a higher coherence value when compared to the LDA model.
To answer Research Question 3: What are the most tweet sentiments about Artificial intelligence? This study has implemented Vader sentiment analysis to analyse polarity and intensity sentiment scores respectively. First, the proportion of polarity in Figure 8, suggests the half of Twitter tweets are positive sentiment to AI while only about 16% of tweets are negative and nearly 34% of tweets are Neutral. Then, when classifying tweets sentiment into 'Positive' and 'Negative' categories, computing the total intensity sentiment score of each category is shown in Figure 10, indicating intensity sentiment score in positive category is higher than that in negative category.

In general, the majority of tweets show positive sentiments towards Artificial intelligence.
To answer Research Question 4: Who are the opinion leaders and what are those opinion leaders' sentiment towards Artificial intelligence? This study utilizes three dimensions to detect opinion leaders. The top 10 opinion leaders are username with "ilanawaber", "iotsecurity2", "ai"," "jdmarkman"," "franborrell", "dbworld_", "Zayy7_", "TechCrunch", "GoldmanSachs" and "BernardMarr". And all of these top 10 opinion leaders are positive sentiment towards Artificial intelligence, as shown in Table 4.

5.2 Conclusion

This study aims to explore the implementation of the topic modeling method of Latent Dirichlet Allocation (LDA) and Correlated topic model (CTM) to retrieve insightful topics and evaluate the performance of LDA and CTM models, and analyse the tweets sentiment as well as identify opinion leaders in the Artificial Intelligence discourse on Twitter.
This study was conducted by several procedures. In the data preprocessing step, non-English tweets, duplicated tweets, hashtags, punctuations, stand-alone numbers and stop words are removed, and two times selections for Nouns have been implemented. In the topic modeling procedure, to get an optimal model, noisy words is filtering out through removing tokens that appear less than or more than threshold value in documents. Then LDA and CTM models are optimized through parameters of the number of topics and iterations. Both the optimal number of topics in LDA and CTM models is selected based on the coherence value. The topics extracted from the LDA and CTM models are similar in general, but the number of topics generated from CTM model is less than that of LDA model. In addition, when only considering the coherence score as the evaluation metric, the performance of the CTM model is better than that of LDA model. In the sentiment analysis, Vader sentiment analysis has been performed on tweets after only removing links, stand-alone numbers, hashtags and mentions, but keeping

the punctuations and emojis. And the result of sentiment analysis shows the majority of tweets represent positive sentiment towards Artificial Intelligence and the total intensity sentiment score also shows the strength of positive sentiment is larger than negative sentiment. To identify the top 10 opinion leaders, three features are conducted together, Activity, Popularity and Influence, and all of these opinion leaders expressed positive sentiments towards Artificial Intelligence.

5.3 Limitation and Future Work

5.3.1 Limitation

There are several limitations in this study. The first limitation is the evaluation metric of performance and choosing the optimal number of topics in the LDA and CTM models. This study only uses the coherence score to select the optimal number of topics and assess the performance of the LDA and CTM models. Although getting the highest coherence score, some topics are not good or easy for human interpretability. The quality of data also affects the model performance, when performing the topic modeling, only tweet content is considered without including metadata such as tweets author, date of post, etc. Moreover, the coherence score is one metric for assessing performance, but it is not the best metric or the only way to do so. For example, qualitative measures should be considered together as one of evaluation metrics such as word intrusion and topic intrusion (Chang et al.,2009; Giri, 2022). Next, the second limitation is sentiment analysis. Vader sentiment analysis has been conducted on tweets in this study, where Vader could not identify the sarcasm tweet well leading to wrongly classifying tweets' sentiment polarity (Wang et al.,2020). And since the dataset is not including labelled tweets, it is not available to assess the accuracy of labelling tweets' sentiment by using Vader. Although a subset of data had been randomly sampled and labelled by author, it introduces bias and inconsistency because different people may have different interpretations of the sentiment on the same content. The third limitation is identifying opinion leaders, in this study, it does not combine the retweet impact with mention impact to identify opinion leaders (Pal & Counts, 2011).

5.3.2 Future work

In future work, topic modelling should not only include tweets content, but it also should consider combining the metadata of tweets with tweets content to extract topics, which can be achieved by using Structural Topic Model (STM) (Roberts et al., 2013). In terms of evaluation metrics of topic model performance, comprehensive metrics should be adopted such as a combination of qualitative and quantitative measures. In sentiment analysis, various sentiment analysis methods should be used rather than only Vader sentiment analysis tool. Lastly, more dimensions or features should be considered and combined when identifying opinion leaders.

Bibliography

Abdulaziz, M., Alotaibi, A., Alsolamy, M., & Alabbas, A. (2021). Topic based Sentiment Analysis for COVID-19 Tweets. International Journal of Advanced Computer Science and Applications, 12(1). https://doi.org/10.14569/IJACSA.2021.0120172

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011). Sentiment analysis of twitter data. In Proceedings of the workshop on language in social media (LSM 2011) (pp. 30-38).

Aleahmad, A., Karisani, P., Rahgozar, M., & Oroumchian, F. (2016). OLFinder: Finding opinion leaders in online social networks. Journal of Information Science, 42(5), 659–674. https://doi.org/10.1177/0165551515605217

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. Cities, 89, 80–91. https://doi.org/10.1016/j.cities.2019.01.032

Aznag, M., Quafafou, M., & Jarir, Z. (2013). Correlated topic model for web services ranking. International Journal of Advanced Computer Science and Applications, 4(6).

Bamakan, S. M. H., Nurgaliev, I., & Qu, Q. (2019). Opinion leader detection: A methodological review. Expert Systems with Applications, 115, 200–222. https://doi.org/10.1016/j.eswa.2018.07.069

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Blei, D., & Lafferty, J. (2006). Correlated topic models. Advances in neural information processing systems, 18, 147.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. Advances in neural information processing systems, 22.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. Mobile Net-

works and Applications, 19(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0
Cordeiro, M. (2012). Twitter event detection: combining wavelet analysis and topic inference summarization. In Doctoral symposium on informatics engineering (Vol. 1, pp. 11-16).
Daenekindt, S., & Huisman, J. (2020). Mapping the scattered field of research on higher education. A correlated topic model of 17,000 articles, 1991–2018. Higher Education, 80(3), 571–587. https://doi.org/10.1007/s10734-020-00500-x
DeLancey J.(2020).Pros and Cons of NLTK Sentiment Analysis with VADER. Code Project. Retrieved from https://www.codeproject.com/Articles/5269447/Pros-and-Cons-of-NLTK-Sentiment-Analysis-with-VADE
Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management, 48, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021
Dybowski, T. P., & Adämmer, P. (2018). The economic effects of U.S. presidential tax communication: Evidence from a correlated topic model. European Journal of Political Economy, 55, 511–525. https://doi.org/10.1016/j.ejpoleco.2018.05.001
Elbagir, S., & Yang, J. (2019). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. Hong Kong.
Giachanou, A., & Crestani, F. (2017). Like It or Not: A Survey of Twitter Sentiment Analysis Methods. ACM Computing Surveys, 49(2), 1–41. https://doi.org/10.1145/2938640
Giri. (2022, August 19). Topic Model Evaluation. High Demand Skills. Retrieved from https://highdemandskills.com/topic-model-evaluation/?utm_content=cmp-true
Griffiths, M. S., Tom. (2007). Probabilistic Topic Models. In Handbook of Latent Semantic Analysis. Psychology Press.
Guldbrandsson, K., Nordvik, M. K., & Bremberg, S. (2012). Identification of potential opinion leaders in child health promotion in Sweden using network analysis. BMC Research Notes, 5(1), 424. https://doi.org/10.1186/

1756-0500-5-424
Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text Mining in Big Data Analytics. Big Data and Cognitive Computing, 4(1), Article 1. https://doi.org/10.3390/bdcc4010001
Hidayatullah, A. F., Pembrani, E. C., Kurniawan, W., Akbar, G., & Pranata, R. (2018). Twitter Topic Modeling on Football News. 2018 3rd International Conference on Computer and Communication Systems (ICCCS), 467–471. https://doi.org/10.1109/CCOMS.2018.8463231
Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: A survey. Enterprise Information Systems, 13(1), 132–144. https://doi.org/10.1080/17517575.2018.1493145
Huang, C., Zhang, Z., Mao, B., & Yao, X. (2022). An Overview of Artificial Intelligence Ethics. IEEE Transactions on Artificial Intelligence, 1–21. https://doi.org/10.1109/TAI.2022.3194503
Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University - Engineering Sciences, 30(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002
Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. Multimedia Tools and Applications, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4
Johnson R. (2018, April 26). Twitter developer platform and user privacy. Twitter Developer Blog. Retrieved from https://blog.twitter.com/developer/en_us/topics/tools/2018/twitter-developer-platform-and-user-privacy
Lamirán-Palomares, J. M., Baviera, T., & Baviera-Puig, A. (2019). Identifying Opinion Leaders on Twitter during Sporting Events: Lessons from a Case Study. Social Sciences, 8(5), Article 5. https://doi.org/10.3390/socsci8050141

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), 1.
Lee, J. H., Wood, J., & Kim, J. (2021). Tracing the Trends in Sustainability and Social Media Research Using Topic Modeling. Sustainability, 13(3), Article 3. https://doi.org/10.3390/su13031269
Lee, M. (2022). bab2min/tomotopy: 0.12.3 (version 0.12.3) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6868418
Li, Y., Ma, S., Zhang, Y., Huang, R., & Kinshuk. (2013). An improved mix framework for opinion leader identification in online learning communities. Knowledge-Based Systems, 43, 43–51. https://doi.org/10.1016/j.knosys.2013.01.005
Lin, H.-C., Bruning, P. F., & Swarna, H. (2018). Using online opinion leaders to promote the hedonic and utilitarian value of products and services. Business Horizons, 61(3), 431–442. https://doi.org/10.1016/j.bushor.2018.01.010
Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. SpringerPlus, 5(1), 1608. https://doi.org/10.1186/s40064-016-3252-8
Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011
Melton, C. A., Olusanya, O. A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. Journal of Infection and Public Health, 14(10), 1505–1512. https://doi.org/10.1016/j.jiph.2021.08.010
Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In Proceedings of the 2011 conference on empirical methods in natural language processing (pp. 262-272).
Mohammadi, S. A., & Andalib, A. (2017). Using the opinion leaders in

social networks to improve the cold start challenge in recommender systems. 2017 3th International Conference on Web Research (ICWR), 62–66. https://doi.org/10.1109/ICWR.2017.7959306
Morone, F., & Makse, H. A. (2015). Influence maximization in complex networks through optimal percolation. Nature, 524(7563), Article 7563. https://doi.org/10.1038/nature14604
Na, T., Cheng, W., Li, D., Lu, W., & Li, H. (2021). Insight from NLP analysis: COVID-19 vaccines sentiments on social media. arXiv preprint arXiv: 2106.04081.
Nadimpalli, M. (2017). Artificial intelligence risks and benefits. International Journal of Innovative Research in Science, Engineering and Technology, 6(6).
Nagmoti, R., Teredesai, A., & De Cock, M. (2010). Ranking Approaches for Microblog Search. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 1, 153–157. https://doi.org/10.1109/WI-IAT.2010.170
Nave, M., Rita, P., & Guerreiro, J. (2018). A decision support system framework to track consumer sentiments in social media. Journal of Hospitality Marketing & Management, 27(6), 693–710. https://doi.org/10.1080/19368623.2018.1435327
Negara, E. S., Triadi, D., & Andryani, R. (2019). Topic Modelling Twitter Data with Latent Dirichlet Allocation Method. 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), 386–390. https://doi.org/10.1109/ICECOS47637.2019.8984523
Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). Automatic evaluation of topic coherence. In Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics (pp. 100-108).
Nizam, T., & Hassan, S. I. (2017). Big Data: A Survey Paper on Big Data Innovation and its Technology. International Journal of Advanced Research in Computer Science, 8(5).
Pal, A., & Counts, S. (2011). Identifying topical authorities in microblogs. Proceedings of the Fourth ACM International Conference on Web Search

and Data Mining, 45–54. https://doi.org/10.1145/1935826.1935843
Prier, K. W., Smith, M. S., Giraud-Carrier, C., & Hanson, C. L. (2011). Identifying Health-Related Topics on Twitter. In J. Salerno, S. J. Yang, D. Nau, & S.-K. Chai (Eds.), Social Computing, Behavioral-Cultural Modeling and Prediction (pp. 18–25). Springer. https://doi.org/10.1007/978-3-642-19656-0_4
Rabitz, F., Telešienė, A., & Zolubienė, E. (2021). Topic modelling the news media representation of climate change. Environmental Sociology, 7(3), 214–224. https://doi.org/10.1080/23251042.2020.1866281
Ray, S. K., Ahmad, A., & Kumar, C. A. (2019). Review and implementation of topic modeling in Hindi. Applied Artificial Intelligence, 33(11), 979-1007.
Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks (pp. 45-50). ELRA.
Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. Information Processing & Management, 52(5), 949–975. https://doi.org/10.1016/j.ipm.2016.04.003
Roberts, M. E., Tingley, D., Stewart, B. M., & Airoldi, E. M.(2013). The Structural Topic Model and Applied Social Science. Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation.
Rocha, L., Mourão, F., Vieira, R., Neves, A., Carvalho, D., Bandyopadhyay, B., Parthasarathy, S., & Ferreira, R. (2016). Connecting Opinions to Opinion-Leaders: A Case Study on Brazilian Political Protests. 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 716–725. https://doi.org/10.1109/DSAA.2016.77
Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In Proceedings of the eighth ACM international conference on Web search and data mining (pp. 399–408).
Rogers, E. (2003). Diffusion of innovation (5th ed.). New York, NY: The Free Press.
Sachs, G. (2016). Profiles in Innovation: Artificial Intelligence-AI. Machine

Learning and Data Fuel the Future of Productivity.
Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. Journal of Computational Science, 36, 101003. https://doi.org/10.1016/j.jocs.2019.05.009
Sang, R., & Chan, K. P. (2015). A Correlated Topic Modeling Approach for Facial Expression Recognition. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 989–993. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.149
Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2021). COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions. British Journal of Management, 32(4), 1164–1183. https://doi.org/10.1111/1467-8551.12441
Song, M., & Miskel, C. G. (2005). Who Are the Influentials? A Cross-State Social Network Analysis of the Reading Policy Domain. Educational Administration Quarterly, 41(1), 7–48. https://doi.org/10.1177/0013161X04269515
Song, M., Kim, M. C., & Jeong, Y. K. (2014). Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter. IEEE Intelligent Systems, 29(2), 18–26. https://doi.org/10.1109/MIS.2014.20
Tecuci, G. (2012). Artificial intelligence. WIREs Computational Statistics, 4(2), 168–180. https://doi.org/10.1002/wics.200
Thomas S. (2022, January 24). Calculate Outlier Formula: A Step-By-Step Guide. Outlier org. Retrieved from https://articles.outlier.org/calculate-outlier-formula
Twitter. (2023). Twitter API rules and policies. Twitter Help Center. Retrieved Retrieved May 18, 2023 from https://help.twitter.com/en/rules-and-policies/twitter-api
Twitter. (2023). Twitter API v2 data dictionary. Developer Platform. Retrieved Retrieved May 18, 2023 from https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/user
Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(4), 337–339. https://doi.org/

10.1016/j.dsx.2020.04.012
Wang, X., Yang, X., Wang, X., Xia, M., & Wang, J. (2020). Evaluating the competitiveness of enterprise's technology based on LDA topic model. Technology Analysis & Strategic Management, 32(2), 208–222. https://doi.org/10.1080/09537325.2019.1648789
Wang, Y., & Fikis, D. J. (2019). Common Core State Standards on Twitter: Public Sentiment and Opinion Leaders. Educational Policy, 33(4), 650–683. https://doi.org/10.1177/0895904817723739
Wang, Y., Shen, G., & Hu, L. (2020). Importance evaluation of movie aspects: Aspect-based sentiment analysis. In 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE) (pp. 2444-2448). IEEE.
Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). TwitterRank: Finding topic-sensitive influential twitterers. Proceedings of the Third ACM International Conference on Web Search and Data Mining, 261–270. https://doi.org/10.1145/1718487.1718520
Xu, W. W., Sang, Y., Blasiola, S., & Park, H. W. (2014). Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election. American Behavioral Scientist, 58(10), 1278–1293. https://doi.org/10.1177/0002764214527091
Yin, H., Song, X., Yang, S., & Li, J. (2022). Sentiment analysis and topic modeling for COVID-19 vaccine discussions. World Wide Web, 25(3), 1067-1083.
Zhu, H. (2020). Big Data and Artificial Intelligence Modeling for Drug Discovery. Annual Review of Pharmacology and Toxicology, 60(1), 573–589. https://doi.org/10.1146/annurev-pharmtox-010919-023324

Appendix

Appendix A

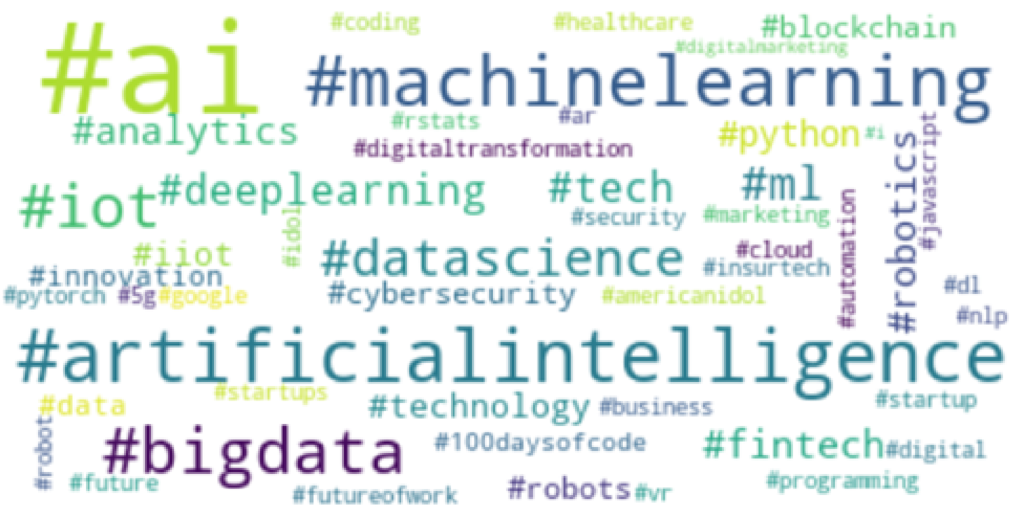
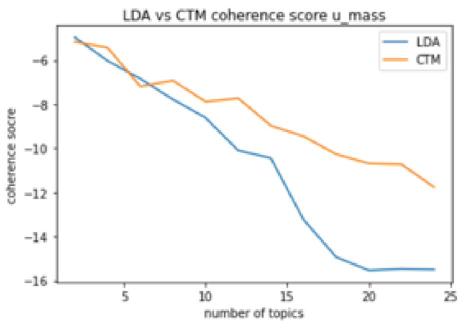


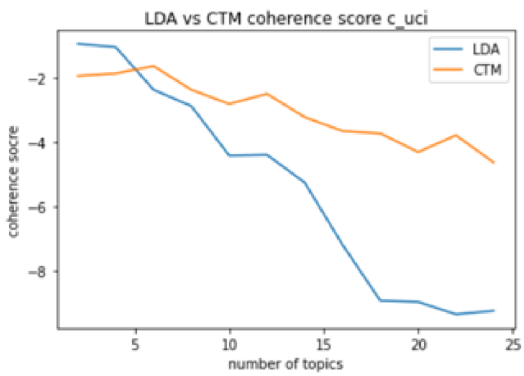
Figure 16. Top 50 hashtags in the dataset

Appendix B

metric: u_mass; average value : in LDA is -10.783702586012895; in CTM is -8.51240696134134



metric: c_uci; average value : in LDA is -5.41779498743603; in CTM is -3.025912946871203



metric: c_npmi; average value : in LDA is -0.20699093634345947; in CTM is -0.1532346817041983

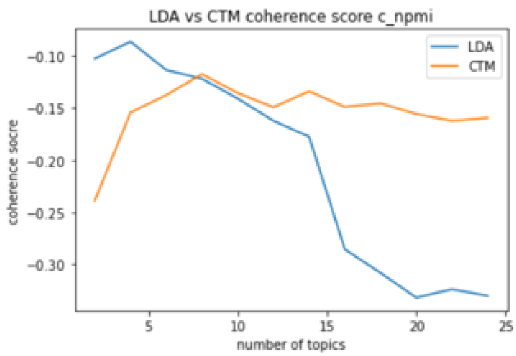


Figure 17. Coherence score in LDA and CTM models using different metrics

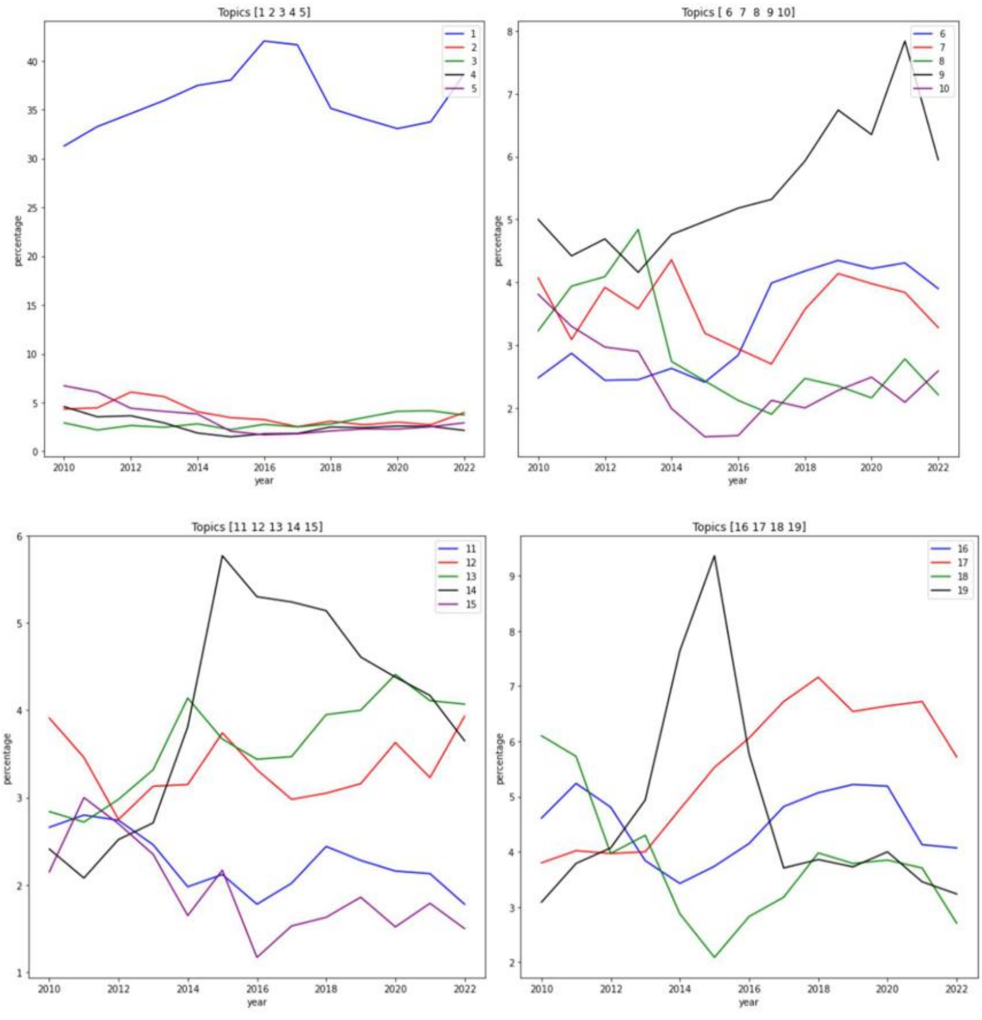


Figure 18. Topics distribution with time series in LDA model

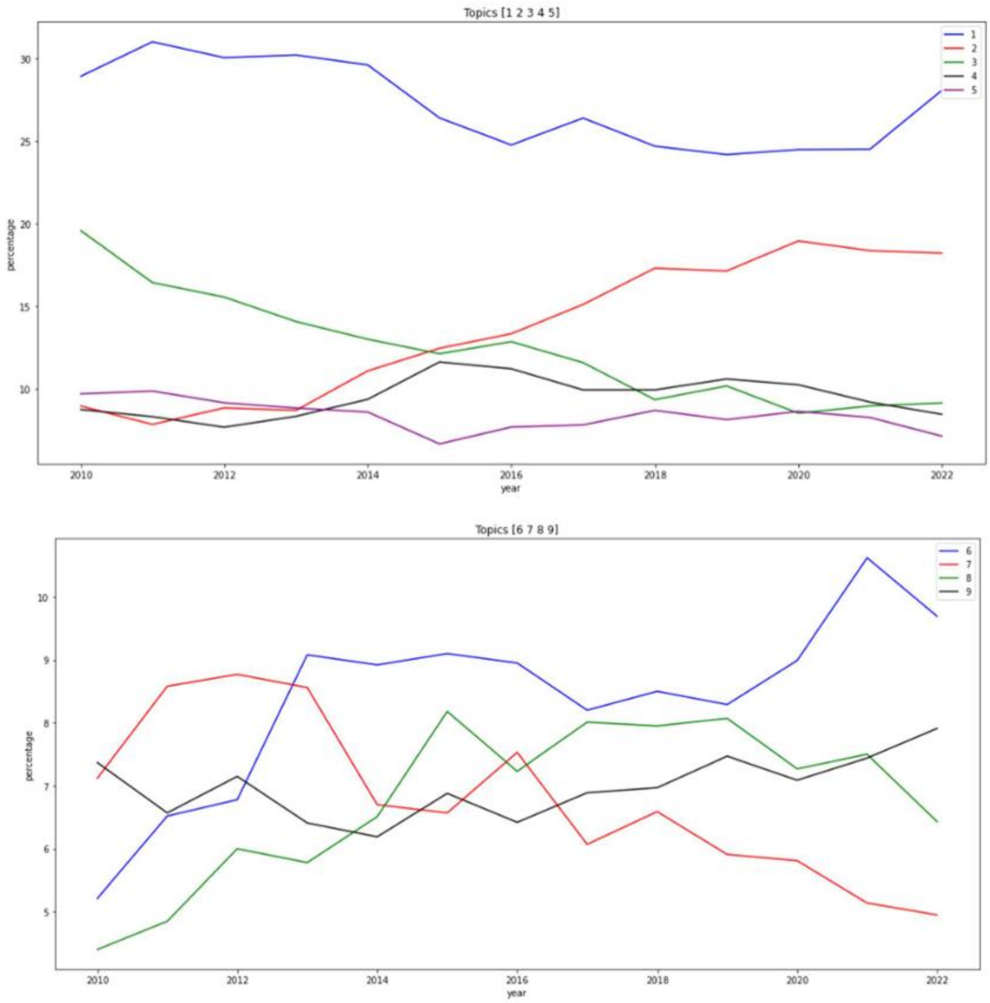


Figure 19. Topics distribution with time series in CTM model

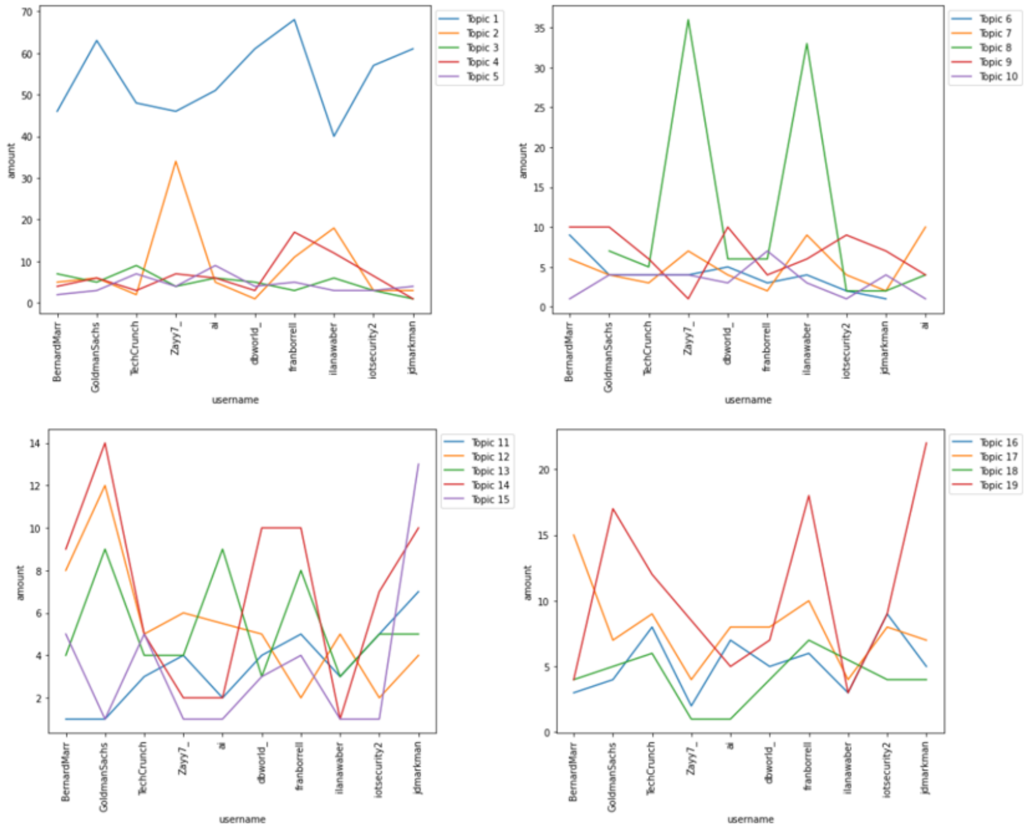


Figure 20. The distribution of top 10 opinion leaders engagement in each topic in LDA model

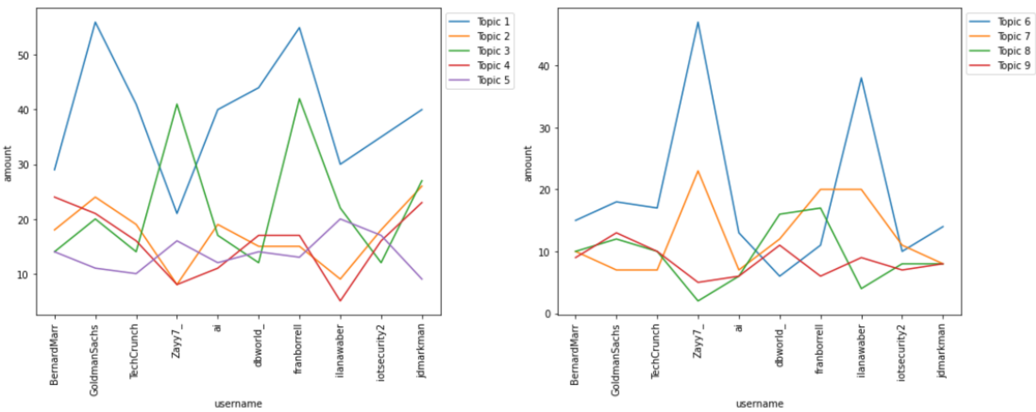


Figure 21. The distribution of top 10 opinion leaders engagement in each topic in CTM model

Appendix C

The code of this study can be seen through this link:
https://github.com/Faslio/AI_implication.git