Reaction Towards The Emergence Of Artificial Intelligence (1358 words)

Muhammad Mujtaba Mir
Luddy School of Informatics, Computing, and Engineering
Indiana University - Bloomington

1. Introduction

Artificial intelligence (AI) refers to the simulation of human intelligence in machines. AI is characterized by its ability to rationalize and take actions without human intervention. Examples of AI-powered tools include Google's search algorithms, the Google Translator, and self-driving car technology. Currently, we have achieved applied AI, which performs narrow tasks such as facial recognition and natural language processing with accuracy levels that match or exceed human accuracy. While most AI applications are designed to benefit humanity, powerful tools can be used for harm when they fall into the wrong hands. Notable figures such as Stephen Hawking and Elon Musk have warned about the potential dangers of AI [1]. For example, AI could pose a threat to digital, physical, and political security through hacking, weaponization, and targeted disinformation campaigns [2].

2. Research Question

The purpose of this paper is to analyze the reaction to the emergence of Artificial Intelligence. The paper will describe the data collection and analysis process, present the conclusions drawn, and discuss areas for potential improvement.

3. Background & Literature

Previous research on the emergence of artificial intelligence (Miles Brundage et al., 2018) discusses the potential security threats from malicious uses of AI and proposes methods for forecasting, preventing, and mitigating these threats [3]. The paper analyzes the ways in which AI may impact the threat landscape in the digital, physical, and political domains and makes four high-level recommendations for AI researchers and other stakeholders.

In another paper (Menakshi Nadimpalli, 2017), the author discusses the benefits and risks of Artificial Intelligence [4]. Among the benefits of AI is its use in the finance and banking industry to monitor activities for suspicious behavior such as fraud. Systems put in place are designed to identify potential malpractices and prevent losses for companies.

4. Method

4.1. Data

The data was collected from Twitter using the SNScrape library in Python from June 1, 2021 to November 1, 2022. SNScrape is a scraper for social networking services that retrieves items such as user profiles, hashtags, or search results. Twitter was chosen as the source of data because it is an active platform where people share their opinions on a wide range of topics. The tweets were collected using the keywords ”benefits of artificial intelligence” and ”dangers of artificial intelligence” and stored in the ”ai1.csv” and ”ai2.csv” files, respectively. A total of 1000 and 750 tweets were collected for each keyword, respectively. Using two different and opposite keywords allowed us to create a balanced dataset by combining data from ai1.csv and ai2.csv, if an imbalance existed.

4.2. Analysis

In the first step, the tweets saved in the CSV files were stored in two separate pandas dataframes, ”df1” and ”df2,” for tweets containing the keywords ”benefits of artificial intelligence” and ”dangers of artificial intelligence,” respectively. Pandas was chosen for its convenient utility functions and methods for analyzing and manipulating data. The dataframes were then filtered based on the language of the tweets, and only tweets in the English language were retained. This was done using the ”lang” column, which stored the language of each tweet. After the filtering process, 982 tweets were left in ”df1” and 743 were left in ”df2.”

After that, the data was cleaned by removing HTML tags, special characters, multiple periods, and stopwords. Stopwords are common words that do not add significant information to the data and are usually filtered out (e.g., articles, prepositions, pronouns, conjunctions, etc.). The cleaning process helps improve the quality and clarity of the data.

For sentiment analysis, I used the Valence Aware Dictionary for Sentiment Reasoning (VADER) module in Python. VADER is specifically designed for sentiment analysis and has been shown to work well with social media data because it can handle slangs, capital letters, exclamation marks, emojis, etc. VADER outputs four values for a given sentence: positive, negative, neutral, and compound. The compound score was used to capture the sentiment because it is the sum of the positive, negative, and neutral scores, normalized between -1 (most extreme negative) and +1 (most extreme positive). Tweets were classified into three categories: positive, negative, and neutral based on the following thresholds: positive (1) if compound $\geq$ 0.03, negative (0) if compound ≤ -0.03, and neutral otherwise. The neutral tweets were discarded. After applying VADER and the above conditions to both dataframes, the number of positive tweets in ”df1” were 305, and the number of negative tweets in ”df1” were 513. Similarly, the number of positive tweets in ”df2” were 723, and the number of negative tweets in ”df2” were 19. There was a significant imbalance in the number of positive and negative tweets in ”df2,” so the negative tweets in ”df1” and the positive tweets in ”df2” were combined to create a fairly balanced dataset with 1236 tweets and stored in the ”df” dataframe.

The next step in the analysis was to train a machine learning algorithm for binary classification where the input was a tweet, and the output was either 1 (positive) or 0 (negative). Since machine learning algorithms only take numbers as input, it was necessary to convert the tweets into numerical representations. This was done using word embeddings, where dense vector representations of words capture their semantic meaning. In this case, I used the GloVe vector representation and converted the text into 50-dimensional vectors [5]. The reason for choosing 50 dimensions was that the amount of data was limited, and a 50-dimensional representation was appropriate for this data. I used the lazypredict library in Python to compare the performance of various machine learning algorithms across different metrics like accuracy, balanced accuracy, area under the curve, and F1-score. Based on the comparison, I chose the Logistic Regression model because of its simplicity and good performance. After making this choice, I used RandomizedSearchCV to find the best hyperparameter (C) value, which turned out to be 10.

After training the model, topic modeling was also performed to see which topics were present in the most positive (1) and negative (0) tweets. To do this, I first separated the positive and negative tweets into two separate dataframes, ”pos” and ”neg” respectively. Then, I used BERTopic (a topic modeling technique) to perform the topic modeling on the ”pos” and ”neg” dataframes separately.

5. Results

The results of training a Logistic Regression model on the tweet data were promising, with an accuracy score of 0.98 on the training data and 0.96 on the test data. The effectiveness of the model was further demonstrated through the use of Principal Component Analysis to reduce the 50-dimensional input vectors to two dimensions for visual representation. The resulting plots, with the true and predicted labels, showed few instances of miss-classification.

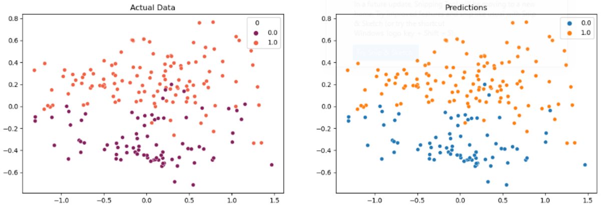

Figure 1. Scatterplot of real and predicted categories

The confusion matrix of test data is as follows:

Labels	0	1
0	96	7
1	2	143

The confusion matrix of train data is as follows:

Labels	0	1
0	323	5
1	5	457

Figure 2 shows the plot of the average sentiment on a daily basis, with the sentiment oscillating around a mean point and mostly negative. However, towards the end there is a hint of a positive trend in the mean sentiment, suggesting that the future trend may not be against Artificial Intelligence.

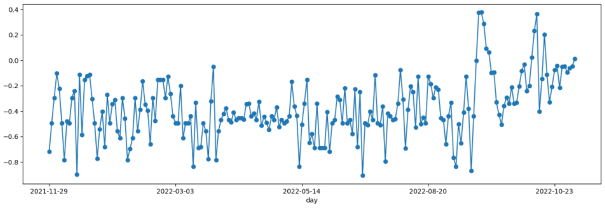

Figure 2. Mean sentiment

The results of topic modeling show that the positive tweets largely discuss the benefits of AI in the field of healthcare, medical diagnosis, impact in business, marketing etc., while the negative tweets focus on the dangers of AI, including mentions of Elon Musk’s concerns about the technology. The plots in Figures 3 and 4 provide visual representations of these findings by taking 1 topic from both ”pos” and ”neg” dataframes.

6. Limitations

The sentiment scores of the tweets were computed using VADER. However, it is important to note that VADER’s analysis may overlook important words or usage due to misspellings and grammatical mistakes, which are common in most, if not all, tweets. Additionally, VADER may not recognize discriminating jargon, nomenclature, memes, etc.,

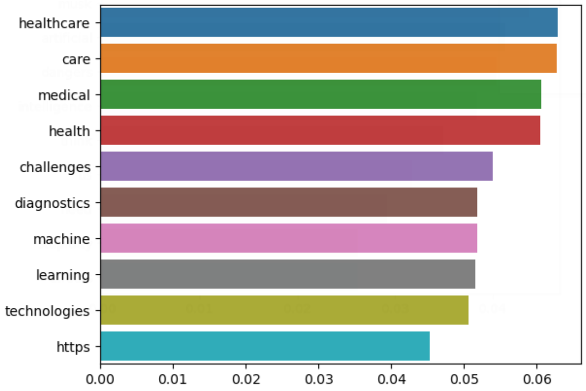

Figure 3. A topic from pos dataframe

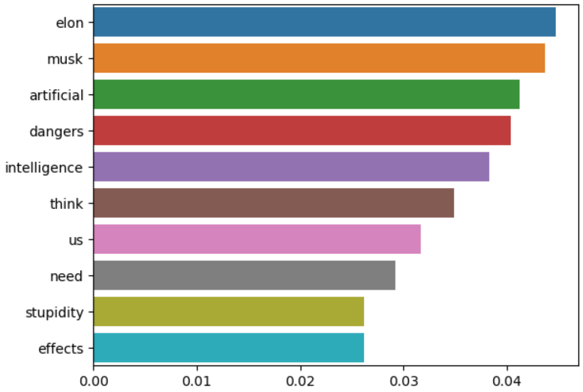

Figure 4. A topic from pos dataframe

which are commonly found in many tweets.

It is also important to note that this analysis only used 50-dimensional vectors to represent each tweet. Higher-dimensional representations, which capture the semantic meaning of words in more detail, exist. However, due to the limited amount of available data, a lower-dimensional representation was used to prevent overfitting by the machine learning algorithm.

To improve the analysis, we need to address the above mentioned limitations which will help improve the accuracy and reliability of the analysis.

References

[1] "Is Artificial Intelligence Dangerous? 6 AI Risks Everyone Should Know About". Bernard Marr Co., https://bernardmarr.com/is-artificial-intelligence-dangerous-6-ai-risks-everyone-should-know-about/

[2] Thomas, Mike. "7 Dangerous Risks of Artificial Intelligence". BuiltIn, https://builtin.com/artificial-intelligence/risks-of-artificial-intelligence.

[3] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., et al. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. https://doi.org/10.17863/CAM.22520

[4] Nadimpalli, Meenakshi. (2017). Artificial Intelligence Risks and Benefits. 6.

[5] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation.