N.B. The purpose of this paper is to present students with an example of how to write up the results of their CL study.

Sentiment Analysis of 'AI tweets'

Calbert Graham

University of Cambridge
crg29@cam.ac.uk

Abstract:
We conducted a sentiment analysis on recent Twitter data to investigate people's opinion of AI technology. We employed two different sentiment analysis techniques: (1) VADER, a rule-based approach that is optimised for analysis of social media data and (2) Flair, a state-of-the-art PyTorch machine-learning technique. The results of our experiments show that most Twitter users have an optimistic and positive view of AI technology, with a sizeable minority appearing to express concerns about the potential harms AI technology can cause. We created word clouds to examine the most frequent words in the tweets as well as the most frequent negative words. We end with a general discussion of the need to take public opinion, both positive and negative, into account in the development of AI technology.

Key words: *AI, FLAIR, VADER, EDI, sentiment, polarity*

1. Introduction

1.1. Background

Sentiment analysis or opinion mining is a natural language processing technique used to identify, extract and characterise subjective information in a text (Beigi, 2016). Sentiment analysis is a useful tool for determining the attitudes, opinions and emotions of a speaker or writer with respect to some topics. In sentiment analysis, historical data are used to make future predictions or to estimate public opinion. Social media sites (e.g., Facebook, Twitter, etc) are a treasure trove of data to gauge the opinions and sentiments of their users (Pak & Paroubek, 2010). Sentiment analysis has become one of the most popular applications of natural language processing. Social media data are used by companies in market analysis and product promotion. Sentiment analysis is also a popular technique in data mining, text mining and information retrieval. Being able to harness data on people's opinions on social, political, economic, health, and other issues has applications in various areas. In sentiment analysis, the texts that people produce are processed and analysed using various classification algorithms. One common approach is to classify texts on whether they are negative, positive or neutral based on the words used to discuss a topic. Working out the overall contextual polarity of a document gives us a sense of what a person's opinion is on the topic in question. People's opinions are by no means fixed, so access to historical data also allows us to track changes in people’s sentiments around certain topics, which may be useful to advertisers, policy developers and decision makers (Kavitha et al., 2022).

It is by no means an easy task to use social media information to determine people's opinions. Social media data can be messy and various preprocessing procedures are required to clean the data and extract only information that is relevant to the study being investigated.

This short study aims to better understand people’s opinion on artificial intelligence as a technology and its relation to ethics. We conducted sentiment analysis of Twitter data using Data Analytics and machine learning techniques.

1.2. Related work

Nausheen & Begum (2018) conducted a sentiment analysis to predict the 2016 US President election results. Whilst the paper was published after the results, making it difficult to validate the accuracy of the predictions, the overall analysis suggested that Hilary Clinton was viewed more favourably than Donald Trump. This is interesting because, although Clinton was defeated in the electoral college vote that determines the US presidency, she managed to win the popular vote. Akcora et al. (2010) tested an approach to identify breaking points in public opinion, proposing the use of microblogging sites to capture public opinion more accurately than traditional polling. Mandal et al. (2016) used a lexicon-based text classification model to analyse and predict sentiments from online reviews. Their model is different from other lexicon-based models in that, in addition to each negative or positive sentiment word, they provided a descriptor to capture the intensity of the sentiment. Mahadzir & Omar (2018) used opinion data about certain products and services from public users to create a sentence-visualisation tool for property investment. Similarly, data has been used to build models to forecast stock prices based on news articles (Kim et al., 2019; Shah et al., 2018). Bon-Itt & Skunkan (2020) tracked major trends in themes of concerns about the Covid-19 pandemic expressed by Twitter users. Various machine learning approaches have been used in sentiment analysis (see Zhang et al., 2018 and Fatima et al., 2022 for overviews of deep learning models, and Mohammad et al. 2013 for support vector machine (SVM)-based classifiers).

Our study therefore builds on the work of other researchers in the field by using information-retrieval tools to obtain Twitter data, and in applying sentiment analysis techniques to investigate people’s opinion of a topic of huge significance: artificial intelligence. Our motivation was to get a ‘snapshot view’ of public sentiment on AI and to ascertain the level of interest in the benefits vs. the risks of AI. We explore the following specific research questions:

1. What is the overall polarity ranking of the data? i.e., which of positive, negative, and neutral are the most- and least-frequent labels?
2. What words are most frequently associated with AI?

3. What words are most associated with negative AI tweets?
4. How frequently do keywords on EDI and ethics appear?

2. Method

Below, we briefly describe the procedures for information retrieval, preprocessing, analysis and visualisation. For details about packages and the Python script used, please consult the Github repository: https://github.com/crgraham/AItweets.

2.1. Information retrieval

At the time of writing, Twitter allows the retrieval of tweets by using their API and third-party applications. We used a third-party, open-source, Python-based tool - Twitter Intelligence Tool (Twint, 2022) - to obtain users' English tweets with the hashtag #AI. Tweets were then filtered and classified based on their sentiment: negative, positive, neutral. Stop words (i.e., terms that were manually excluded) included 'https'. There were 19,883 tweets in the dataset, extracted from across accounts worldwide. The tweets were all posted on users' Twitter accounts between 30 August 2022 - 1 September 2022.

2.2. Preprocessing

Twitter data is unstructured and may contain URLs, hashtags, white spaces, emoticons, and non-ASCII and other non-English symbols mixed in with the English text. We used Python libraries to remove the non-relevant part of the text data. Whilst this process cannot filter out every non-conventional English form, it significantly improves the quality of the text and is an essential first step.

2.3. Analysis

We used two approaches to analyse and extract sentiment scores: Valence Aware Dictionary for sEntiment REasoning (VADER) and Flair, a PyTorch NLP framework.

2.3.1. VADER

VADER is optimised for analysing the sentiment of social media posts. It is a lexicon- and rule-based approach which uses a list of lexical features that are labelled as negative, positive or neutral according to the semantic orientation of the word to calculate the overall text sentiment (Hutto & Gilbert., 2014). VADER returns a classification for each tweet along with a valence score for each input sentence that falls into a given sentiment category. The overall classification is based on an average of these valence scores.

The compound score generated by the model is the most commonly used method for sentiment analysis (Hutto & Gilbert, 2014). According to the developers, 'it is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalised to be between -1 (most extreme negative) and +1 (most extreme positive)'.

We followed the typical threshold values suggested by the developers for classifying the sentiment of each tweet:

1. Positive sentiment has compound score >= 0.05
2. Neutral sentiment has compound score > -0.05 and <0.05.
3. Negative sentiment has compound score <= -0.05

This shows that the higher the score, the more positive the tweet.

2.3.2. FLAIR

Flair is a state-of-the-art machine learning approach in Natural Language Processing with various functionalities including Named Entity Recognition (NER), text embeddings, and models for sentiment analysis. The model was originally trained on IMDb movie review data which has the one disadvantage that, unlike VADER, it is not specifically optimised for analysis of social media data. However, it has the advantage over rule-based models such as VADER of being able to make predictions based on out-of-corpus data, including misspelt words.

2.4. Data summarisation and visualisation

The final step was to use the filtered tweets to generate a word cloud, a bar chart and tables to summarise/visualise the results.

3. Results

The results of both analyses revealed that in this experiment most of the tweets with "AI" were positive, followed by neutral and negative. However, the number of tweets in each category varied between the two systems as shown below:

3.1. *In relation to RQ1: the polarity of tweets*

3.1.1. VADER vs FLAIR

Table 1. Evaluation of Polarity and Valence Measures (VADER)

Sentiment	Number of tweets	Compound score
Positive	10,237	0.5437
Negative	2,193	-0.4485
Neutral	7,453	0.0001

Table 2. Evaluation of Polarity and Probability (Flair)

Sentiment	Number of tweets	Probability
Positive	8,887	0.888
Negative	5,330	0.922
Neutral	5,666	0.000

The polarity results are compared side-by-side in Figure 1.

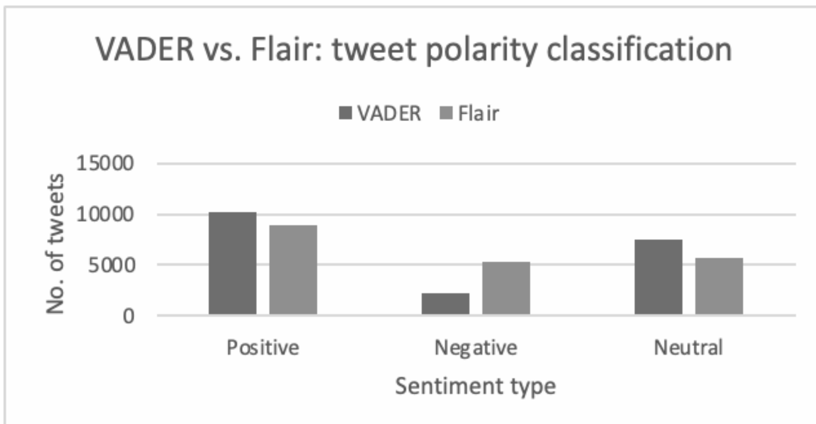

Figure 3: VADER vs. Flair Tweet Sentiment Polarity

3.1.2. Data samples

Samples of tweets are shown in Table 3. Examples are chosen from the dataset where both VADER and Flair were in agreement on the classification.

Table 3 Samples of tweets classified into the three sentiments.

Sentiment	Original tweets (before pre-processing)
Positive	Bytesview's advanced #textanalysis tools let you compile and analyze large volumes of unstructured text data and extract actionable #Insights to boost your business processes. Visit Us @ https://t.co/4s2wJKAZzA #sentimentanalysis #dataanalysis #MachineLearning #AI https://t.co/gMEi7l9VzS
	The Tintra Team are in the Middle East this and next week working on the exciting next step in our roadmap. Keep connected for updates during the trip #tnt #tintra #qatar #ai https://t.co/3FNltqFYQW
Negative	The debate around #AI generated #art is heating up. Here's what a leading #IntellectualProperty lawyer has to say about where @craiyonAI might be treading in unchartered and difficult waters. #MachineLearning #midjourneyAi

	#copyright #GDPR #ethicaltech https://t.co/4vuJaRIyae
	AI offers customers more personalised products and services, but AI bots can still contain human favouritism and the work from the @UniofOxford and @VodafoneGroup is so important to ensure ethical #AI. https://t.co/c6Co1C91Jv https://t.co/M98CvwZRdr
Neutral	#AIArtwork generated September 01, 2022 at 03:01AM. Source: https://t.co/jq73SDshph Hashtags: #AbstractArt #AI #Art #Artist #Artwork #GAN #GenerativeAdversarialNetwork #Painting https://t.co/xnigFgEnUX
	Made it. We have a full house! Our workshop consists of three parts: ü§ñ #ML Explainability and Fairness üë§ Human centred #AI #Design üôåüèΩ Hands-on implementation session #DN22 https://t.co/8xQPkjfjx7

3.2. *In relation to RQ2: the most frequent words*

The most frequent words appearing with "AI" are shown in the Word Cloud in Figure 2.

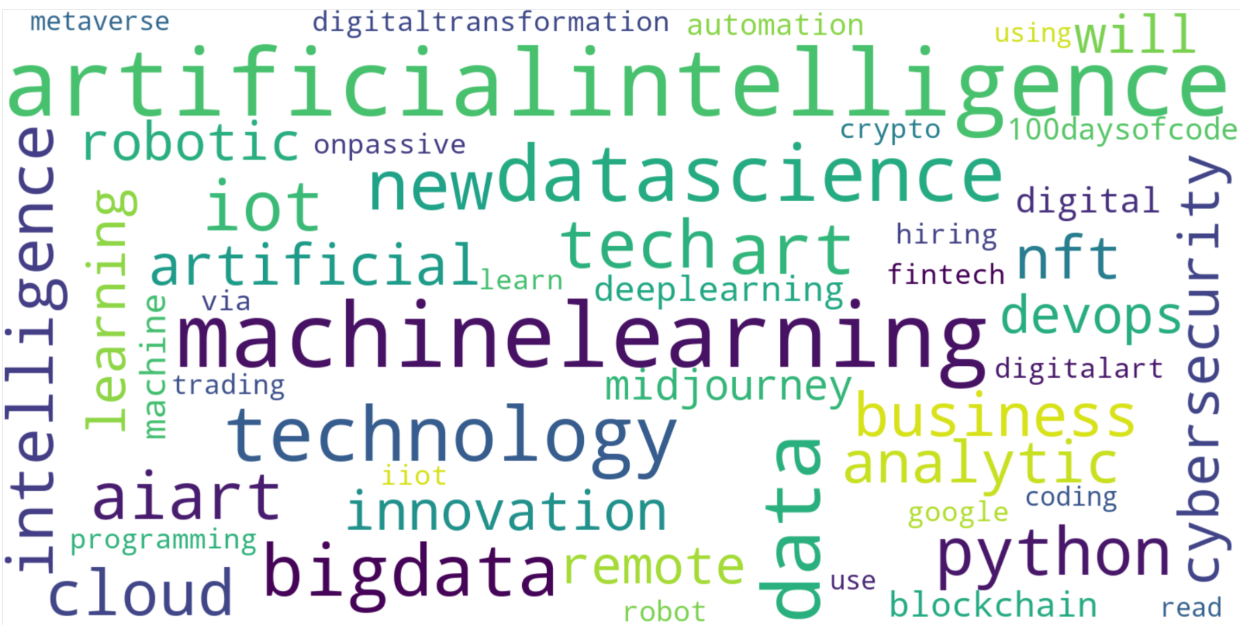

Figure 2: Word cloud for tweets with "AI" in full dataset

3.3. *In relation to RQ3: the most frequent words in negative tweets*

The next step was to get a sense of the most frequent words appearing in negative tweets. We therefore filtered out tweets that were rated as either positive or neutral and created a smaller dataset of tweets that are classified as "negative" by both VADER and Flair systems (total count of 1,933). We then created a Word Cloud shown in Figure 3.

Figure 3: Word cloud for tweets with “AI” in sub-dataset of tweets classified as negative.

3.4. *In relation to RQ4: keywords in EDI*

In a final step we searched the entire original dataset for key words that are associated with Equality, Diversity, and Inclusion (EDI), AI ethics/bias and online safety, as shown in Figure 4.

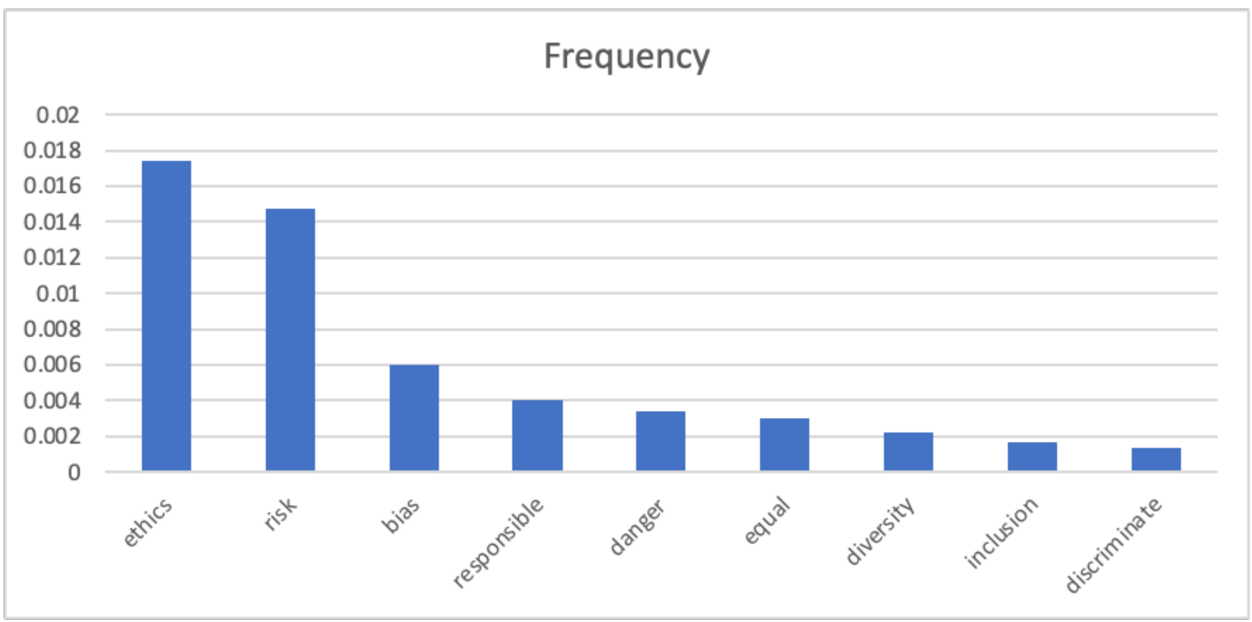

Figure 4: Frequency of key words in the full dataset

4. Discussion and conclusion

We conducted a sentiment analysis of tweets to measure public opinion on artificial intelligence. The results of the experiment showed that most tweets were classified as positive, which we interpret as suggesting that users had a generally positive opinion of AI. The Word Clouds revealed a highly optimistic view of AI with words like ‘innovation’, ‘big data’, ‘coding’, ‘learning’, ‘programming’, ‘technology’, etc. The positive view of AI seems consistent with a model in which humans and AI are working together cooperatively. This idea of an enhanced AI-human partnership is usually referred to as augmented intelligence. (cf. Felice et al., 2022 for a detailed discussion of this.) In a joint policy document, the Royal Society and the American Academy for Arts and Sciences state that: “In this context, interdisciplinary discussions about the impact of AI on areas of societal interest are important in contributing to an environment in which the benefits of AI are brought into being safely and rapidly, and where these benefits are shared across society equitably and inclusively.” (Royal Society, 2018). We share this view.

Whilst there were far fewer negatively classified tweets than positive ones, among this category prominent words included ‘crime’, ‘fraud’, ‘embezzlement', ‘ponzi’, etc. It might be speculated that among users who are concerned about the negative impacts of AI, safety and online security are important issues. These are among 39 issues Stahl (2021) identified as causing significant public concern about ethics in AI. Addressing the potential risks that people perceive AI technology may pose to the wider community of target users is essential for the development of good, ethical practice in AI. More research is therefore needed to better understand people’s opinion about specific AI technologies and their impact on their daily lives. Developers of AI technology must be partners in creating an inclusive culture defined by ethical practice in AI that seeks to understand the needs of all user groups.

Taken together, the sentiment analysis reveals an intriguing pattern of optimistic and positive language in tweets on artificial intelligence. The valence measure of the rule-based VADER model suggests an equally polarised level of intensity between tweets classified as positive and those classified as negative. That is to say, when tweets were positive, they had a level of intensity that was roughly matched in the opposite direction as when they were negative.

A keyword search reveals that some users are indeed concerned about the issue of ethics and bias in AI, and representation around EDI. To be clear, we are not making any association between the occurrence of these pre-selected words and any level of support in tweets. After all, tweets are short and do not always provide sufficient context to allow for objective interpretation of results. Rather, we merely suggest that this gives us a sense of how frequently these words appear in Twitter discourse on AI.

Future work could present more refined search terms that aim to determine public opinion on specific AI technologies (e.g., ASR technology), and would give us more insight into the polarity of AI tweets. It could also attempt to address in more detail people's opinions on AI and its impacts on their way of life. Finally, it would be useful for future work on sentiment analysis to track changes in attitudes over time and inform discussions about the responsible development and deployment of AI.

5. References

Akcora, C., Bayir, M., Demirbas, M., et al. (2010). Identifying breakpoints in public opinion. In Proceedings of the First Workshop on Social Media Analytics (SOMA '10). Association for Computing Machinery, New York, NY, USA, 62–66. https://doi.org/10.1145/1964858.196486

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief. Sentiment Analysis and Ontology Engineering.

Boon-Itt, S., & Skunkan, Y. (2020). Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. JMIR public health and surveillance, 6(4), e21978. https://doi.org/10.2196/21978

Fatima, N., Imran, A., Kastrati, Z., et al. (2022). "A systematic literature review on text generation using deep neural network models," IEEE Access, 2022.

Felice, F., Petrillo, A., Luca, C., et al. (2022). Artificial Intelligence or Augmented Intelligence? Impact on our lives, rights and ethics. Procedia Computer science 200, pp. 1846-1856.

Hutto, C., Gilbert, E. (2014). "Vader: A parsimonious rule-based model for sentiment analysis of social media text." Eighth International AAAI Conference on Weblogs and Social Media. 2014.

Kavitha, M., Naib, B., Mallikarjuna, B., et al. (2022). "Sentiment Analysis using NLP and Machine Learning Techniques on Social Media Data," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 112-115, doi: 10.1109/ICACITE53722.2022.9823708.

Kim, J., Seo, J., Lee, M., et al. (2019). "Stock price prediction through the sentimental analysis of news articles," in 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), 2019, pp. 700–702.

Mahadzir, N., & Omar, M. (2018). "A sentiment analysis visualization system for the property industry," *International Journal of Technology*, vol. 9, no. 8, pp.1609–1617.

Mandal, S., & Gupta, S. (2016). "A Lexicon-based text classification model to analyse and predict sentiments from online reviews," 2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2016, pp. 1-7, doi: 10.1109/ICCECE.2016.8009549.

Mohammad, S., Kiritchenko, S., & Xiaodan, Z. (2013). NRC-Canada: Building the

State-of-the-Art in Sentiment Analysis of Tweets. In Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pages 321–327, Atlanta, Georgia, USA. Association for Computational Linguistics.

Nausheen, F., & Begum, S. (2018). "Sentiment analysis to predict election results using Python." 2nd International Conference on Inventive Systems and Control (ICISC), pp. 1259-1262, doi: 10.1109/ICISC.2018.8399007.

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).

Royal Society (2018). AI, Society and Good: Note of discussions at a Royal Society and American Academy workshop. Centre for Advanced Study in the Behavioural Sciences, Stanford University, 2018.

Shah, D., Isah, H., & Zulkernine, F. (2018). “Predicting the effects of news sentiments on the stock market,” in 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4705–4708.

Stahl B. C. (2021). Ethical Issues of AI. Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies, 35–53. https://doi.org/10.1007/978-3-030-69978-9_4

Twint. (2022). https://github.com/twintproject/twint . Accessed in September 2022.

Zhang, L, Wang, S, Liu, B. (2018). Deep learning for sentiment analysis: A survey. WIREs Data Mining Knowl Discov. 2018; 8:e1253. https://doi.org/10.1002/widm.1253

6. Appendices

Packages and Python script codes and the data used in this report can be found in the Github repository for this report: https://github.com/crgraham/AItweets.