IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 January 2023, accepted 27 February 2023, date of publication 8 March 2023, date of current version 16 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3254503

RESEARCH ARTICLE

1000 Days of COVID-19: A Gender-Based Long-Term Investigation Into Attitudes With Regards to Vaccination

ERIK-ROBERT KOVACS[1], LIVIU-ADRIAN COTFAS[1], CAMELIA DELCEA[1], AND MARGARETA-STELA FLORESCU[2]

[1]Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, 010552 Bucharest, Romania
[2]Department of Administration and Public Management, Bucharest University of Economic Studies, 010552 Bucharest, Romania

Corresponding author: Erik-Robert Kovacs (erik.kovacs@csie.ase.ro)

The funding was provided by the Ministry of Research, Innovation and Digitization, through the project titled "Increasing Institutional Performance Through the Development of the Infrastructure and Research Ecosystem of Transdisciplinary Excellence in the Socio-Economic Field," under Grant 42PFE/30.12.2021.

ABSTRACT The coronavirus pandemic has undoubtedly been one of the major recent events that have affected our society at the global level. During this period, unprecedented measures have been imposed worldwide by authorities in an effort to contain the spread of the disease. These measures have led to a worldwide debate among the public, occurring not least within the forum of social media, tapping into pre-existing trends of skepticism, such as vaccine hesitancy. At the same time, it has become apparent that the pandemic affected women and men differently. With these two themes in view, the paper aims to analyze using a data-driven approach the evolution of opinions with regards to vaccination against COVID-19 throughout the entire duration of the pandemic from the point of view of gender. For this analysis, approximately 1,500,000 short user-contributed texts have been retrieved from the popular microblogging platform Twitter, posted between 30 January 2020 and 30 November 2022. Using a machine learning approach, several classifiers have been trained to identify the likely gender (*female* or *male*) of the author, as well as the stance of the specific post towards the COVID-19 vaccines (*neutral*, *in favor*, or *against*), achieving 85.69% and 93.64% weighted accuracy measures for each problem, respectively. Based on this analysis, it can be observed that most tweets exhibit a *neutral* stance, while the number of tweets *in favor* of vaccination is greater than the number of tweets opposing vaccination, with the distribution varying across time in response to specific events. The subject matter of the tweets varied more between stances than between genders, suggesting that there is no significant difference between the contents of tweets posted by females and males. We also find that while the overall engagement on Twitter with the topic of vaccination against COVID-19 is on the wane, there has been a rise in the number of *against* tweets continuing into the present.

INDEX TERMS Opinion mining, gender identification, stance detection, social media, covid-19, vaccine hesitancy.

I. INTRODUCTION

It is likely that few recent events elicit the same combination of feelings of dread, as well as a slight sense of unreality, as the COVID-19 pandemic. Starting as an unknown[1] outbreak in Wuhan, central China [1], it quickly spread worldwide [2], causing authorities to impose unprecedented measures in an effort to contain the spread of the

The associate editor coordinating the review of this manuscript and approving it for publication was Vijay Mago.

[1]Indeed, even as of December 2022, there is no generally-accepted theory as to the origins of the SARS-CoV-2 pathogen [1].

disease. For better or worse, medical terminology entered the vernacular, and the many waves of infections, the lockdowns, mask mandates, and the massive vaccination campaign that followed are now shared by billions across the globe as truly global common formative experiences.

At the same time, there is a relative dearth of literature discussing the impact of gender on this experience, especially when considering research which uses computational methods as they apply to social media. Social media appears to be an ideal source of data because, when confronted with lockdowns, travel restrictions, social distancing or working from home during the COVID-19 crisis, many people kept informed about the situation and in touch with their peer groups on social media, whose popularity increased significantly during this period [3].

This highlights the role of social media as an important mediator between people, businesses, institutions, and government, as it can represent a catalyst for healthy discussion and a source of reliable and high-quality information [4]. For instance Twitter, the fourth most popular social media platform in the US as of 2022 [5], continuously gained new users during the entire duration of the pandemic [3], and was specifically recognized in both the news media and the research literature as a significant resource that affected how the experience of the pandemic played out across the world [6], [7], [8].

At the same time, social media has been known to facilitate the propagation of rumors [9], fake news [10], misinformation and disinformation campaigns [11], conspiracy theories [12], etc. and promote unhealthy conversations [13], [14] that endanger the public's ability to discern the facts and which sources of information are trustworthy. Indeed, during the COVID-19 crisis, a veritable "infodemic" swept social media, in which genuine concerns regarding civil liberties and personal rights were mixed in with potentially harmful falsehoods concerning COVID-19, its origins, symptoms, the availability of treatments, and especially with regards to the topic of vaccines and vaccination [2], [8], [15]. As one of the major measures taken by governments in order to curtail the pandemic, besides lockdowns [16], was to invest in vaccines, which were developed and deployed in record time, of special interest in this area is the subject of vaccine hesitancy, specifically how the opinion of the public evolved over time in relation with the major events of the pandemic. Since gender has also been widely cited as an important factor in the developments during the pandemic [8], [17], [18], [19], the question of how attitudes with regards to vaccination differ between women and men, and which issues concern which gender the most, are also relevant.

In view of these themes, we aim to contribute towards the literature on this subject with the current paper, by answering the following questions: 1) using social media data, can we develop computational tools that are able to automatically identify social media users as female or male based on their public information? 2) can we develop computational tools that can identify with satisfactory accuracy a text's stance towards vaccination against COVID-19? 3) using these tools, can we distinguish differences between users identified as female and those identified as male with regards to their stance towards vaccination? and 4) can we identify which topics are the most relevant for each stance and gender, respectively?

To answer these questions, we gathered approximately 1,500,000 short texts from the microblogging platform Twitter, called tweets, dating from between 30 January 2020[2] and 30 November 2022, together with the name of their authors, using the Twitter API. We then trained a linear support vector machine (SVM) able to identify the gender of a given name with over 85% accuracy and identified the gender of the authors. We manually annotated each tweet from a subset of the data as *neutral*, *in favor*, or *against* vaccination, and trained a series of machine learning classifiers on this data, achieving a 93.64% accuracy with the RoBERTa classifier. Using this model, we classified the remainder of the data and grouped it by gender and stance, analyzing the length of the tweets, determining commonly-occurring n-grams and identifying topics of concern for each stance-gender combination using latent Dirichlet allocation (LDA). The evolution of the topics by stance and gender has been put in relation to the main events that have occurred during the pandemic. The results show that there is no significant link between gender and stance, and that only small differences between genders exist when it comes to the contents of the discourse. Additionally, it was found that changes in the public's expression of their stance towards vaccination against COVID-19 usually come about as a result of events reported in the media. Finally, we discuss our findings, acknowledge the limitations of such a study, and give recommendations for future research. We also release our training data for further use in research.

II. LITERATURE REVIEW

In the following, the two main sub-problems of the proposed research process, namely gender identification and stance detection, are presented in terms of the current approaches found in the research literature. The natural language processing (NLP) task of stance detection is then introduced and discussed. The issue of vaccine hesitancy and gender differences regarding medical issues is highlighted while surveying applied research in the same vein as the present paper.

A. GENDER IDENTIFICATION

Gender identification in the context of NLP is a subtask of the author profiling (AP) NLP task, which aims to use text-based data in order to determine personal details about the author, such as gender, age, native language, etc. [21], [22].

[2]The day when the World Health Organization declared the outbreak a public health emergency of international concern, which we considered the start of the COVID crisis on a global level [20].

This task can be more formally defined as the task of finding the tuple $< a, g >$ given a sample x_i (a text in a text expressed in natural language), in which a is the text's author and g is the author's gender, $g \in$ *{female, male}*. A wide spectrum of approaches has emerged, grouped into two main categories: one we term intrinsic gender detection, in which nothing is known about the author except for the text x_i, while the other we term metadata-based, in which there is some other information available about the author besides the text, such as their name, occupation, preferred pronouns, etc. While it could be argued that only the first approach constitutes a legitimate NLP task, in practice the two form a continuum, as often the text itself can give away information about the author's gender, which the classifier that also considers metadata can extract and leverage.

From a technical standpoint, the traditional approach to this problem is lexicon- or dictionary-based [23], wherein the gender of the author is evaluated based on pattern-matching for specific words that are inherently gendered in their texts. The more recent approach is to use machine learning [21], in which the gender is determined by a classifier based on a number of labeled examples, or texts that have been manually annotated as having been composed by an author of the specified gender. The two approaches can be combined, for instance by using counts of gendered words as features when representing documents for a machine learning approach [24].

Gender detection based on the knowledge of the author's personal name constitutes a type of metadata-based gender identification. This task is made possible by a feature of many human cultures, namely that they encode gender through language not only through markers of grammatical gender, such as personal and possessive pronouns in English, but also through the choice and use of anthroponyms, or personal names [25]. For instance, in European contexts, the name of a person usually has to conform to certain rules, possibly including being explicitly gendered [26]; in some contexts these are even enforced by a government agency [26], [27]. Not all languages and cultures exhibit this property; for instance in Mandarin Chinese, first names might not always be attributable to a specific gender [28].

Other languages and cultures, such as Russian, go as far as to assign a patronymic component in personal names that explicitly marks the bearer as belonging to a certain gender by suffixing the name of their father with the grammatical components *-ovna,-evna* (female) / *-ovich*,*-evich* (male) [29]. And while in English such components are not usually the norm, many European (including English) first (or given) names are of common Latin descent, allowing them to be identified by their internal phonetic structure – such as ending in the sound / ə / (female), or a consonant sound (male) [25]. This remains the case despite the fact that English common nouns do not carry grammatical gender information anymore [30].

This hypothesis was tested for the case of German and Romanian by Năstase and Popescu [30]. The authors obtain evidence that grammatical gender can be predicted with high accuracy from the form of the nouns even for common nouns (that is, in languages with grammatical gender), and they use exotic hand-built features based on the phonetics of the words in order to train a kernel ridge regression (KRR) classifier [30]. Moorthy et al. [31] also obtain similar results using a random forest (RF) classifier and hand-built features in order to establish the "gendered-ness" of certain marketing terms, such as the names of popular brands and products.

But developing complex, hand-built feature sets can be difficult and time-consuming. The regularities mentioned above enable the development of automated character-level tokenizers that can construct features representing these sub-word structures within personal names. The resulting vectors can then be learned by a suitable machine learning algorithm, allowing for the development of high-accuracy classifiers. Such a technique was presented by Malmasi and Dras [32], who successfully inferred gender and ethnicity from a dataset of personal names. The authors present a linear support vector machine (SVM) classifier able to achieve a 81.30% accuracy for the gender identification task using character-level trigram features [32].

These studies show that given the appropriate linguistic context (that is, that the target language indeed carries gender information within anthroponyms relevant to the actual gender of the bearer), the presence of training data of suitable size and quality, and the construction of the appropriate feature set, the gender of names can be determined with high accuracy.

B. STANCE DETECTION

Stance detection is a natural language processing task related to opinion mining and sentiment analysis. It aims to identify the value judgement related to a certain target within a text [33], or whether the opinion expressed in the text is neutral towards a target, supports it, or is opposed to it. The target can be a concept, opinion, person, product, company, political party, or any other entity. More formally, it can be defined as the task of finding tuple $<t, s>$ given sample x_i, where t is the target entity and s represents the stance inferred from the content of text x_i, $s \in$ *{in favor*, *neutral*, *against}* [2].

Stance detection should not be confused with the closely-related task of polarity detection, which involves finding the general sentiment behind a text – *neutral*, *positive*, or *negative*. While they share the goal of gaining insight into the subjective experience of the author through automatically evaluating the meaning of a text, stance analysis is different in that it aims to identify a judgement and not the general sentiment, as texts with a negative sentiment can express a positive view of a certain subject and vice versa [34], and there need not be any correlation between the two [33]. In addition, sentiment analysis and polarity analysis can be performed in the absence of a target, while stance analysis presupposes the existence of a target which is judged [33].

As with gender identification, sentiment detection, polarity detection, and stance detection techniques are closely related and fall into two main categories: lexicon- or dictionary-based, which attempt to quantify stance based on the occurrence of certain words, and machine learning methods, in which machine learning algorithms are used to infer the stance [2]. As in the case of gender identification, hybrid methods which combine the two approaches can also be applied. Lexicon-based approaches cited in the literature include VADER [35] and NCRLex [36]. The advantage of such an approach is that it is easy to implement and interpret, it is fast, and has modest computational requirements. For generic tasks it is possible to obtain good results with this approach [37]; however, it is usually not sufficiently domain-specific to generalize well to arbitrary targets.

The machine learning approach has seen an increased interest in the recent literature [38], especially with the advent of large transfer learning models, such as BERT and its variants [33]. Two of the most popular topics where sentiment analysis, polarity analysis and stance detection have been widely studied in the literature are COVID-19-related social phenomena, such as the role of China [39], vaccinations [2], [7], [40], [41], and lockdowns [42], and finance-related research, either related to the stock market [42], [43] or cryptocurrencies [44], [45].

In this area, Melton et al. [7] use data gathered from Reddit, another popular social media platform, to perform lexicon-based sentiment analysis on the data, and extract topics using latent Dirichlet allocation (LDA). The authors detect a generally positive sentiment, but they note that this suspiciously remains constant with time, contradicting observations from other research on the subject. The researchers suggest that the rules and structures specific to each social media platform, such as the upvote/downvote mechanism and the strict community guidelines and moderation that are enforced on Reddit, add a significant bias to such observations [7].

Liu and Takikawa [39] used deep learning to assess the trends regarding anti-China discourse in Japanese news media, performing stance analysis with the People's Republic of China as the target. The authors collect comments from Japanese news web sites containing keywords related to China and COVID-19, and fine-tune BERT and RoBERTa on an annotated dataset obtained from news headlines [39]. The study finds that the publishing of anti-China articles is correlated with a change in stances among the public, measurable in the volume of comments expressing anti-China views [39].

The results obtained in these studies show that the approach is mature enough to warrant use in industry as well as applied empirical research, provided a domain-specific dataset of good quality and sufficient size is available.

C. STANCE DETECTION WITH REGARDS TO VACCINATION USING SOCIAL MEDIA DATA AND A FOCUS ON GENDER

Since the start of the COVID-19 crisis, gender was identified as a potentially important factor in its handling and communication [8], [46], [47], [48]. Lockdowns and other measures also affected women and men differently [19], reflecting in part gender differences in occupation and income [49]. The issue of COVID-19 vaccine hesitancy in relation to gender has been widely addressed in traditional empirical research [46], [47], [48]. However, some of these studies have very small sample sizes with participants restricted to a narrow geographical location and a specific time period, thus being unable to capture the evolution of aggregate stances across the world over longer time periods. For this purpose, the computational methods mentioned above appear to be more suitable.

In this area of research, Cascini et al. [8], in their comprehensive review of the literature, only mention two studies which have focused on gender that employ methods similar to our own. This aspect was studied from the standpoint of polarity analysis in social media texts by Zhang et al. [41], who identify gender among other personal details as relevant to users' opinion regarding vaccination in the course of their research on Weibo, a Twitter-like microblogging platform popular in China. The study finds that males are more positive towards vaccination than females [41].

Similarly, Ansari and Khan [50] discuss gender as a relevant factor influencing vaccination-related polarity, and use a multinomial naïve Bayes (MNB) classifier to identify the gender of the tweets' authors, though they do not report performance metrics, nor a comprehensive methodology with regards to gender identification, as this was not the focus of their paper. The authors report no significant differences between the genders [50].

As such, to the best of our knowledge, we are the first to apply computational methods to the problem of the impact of gender on stances towards vaccination with a view towards identifying relevant differences emerging from long-term trends.

III. DATA AND METHODS

The steps taken to perform the proposed analysis can be seen in Figure 1.

A. TRAINING DATA ACQUISITION

The first step is to access the Twitter REST API retrieving a dataset consisting of English-language tweets relevant to the vaccination debate. While the tweets must be in English, they can be from all over the world where English is spoken or otherwise used – that includes the US, Canada, the UK, Ireland, Australia, New Zealand, India, Pakistan and Bangladesh, Israel, and several African countries. We queried the full archive search endpoint using the keywords described in Table 1, themselves inspired by the earlier research on vaccine hesitancy by Cotfas et al. [2]. We used the "relevancy" ranking criterion to obtain the best-quality tweets only [51]. We also retrieved the authors' names. We ran queries for each day between 30 January 2020 and 30 November 2022, and 500 tweets per day were retrieved obtaining approximately 500,000 tweets in total; then, using two random 2-hour

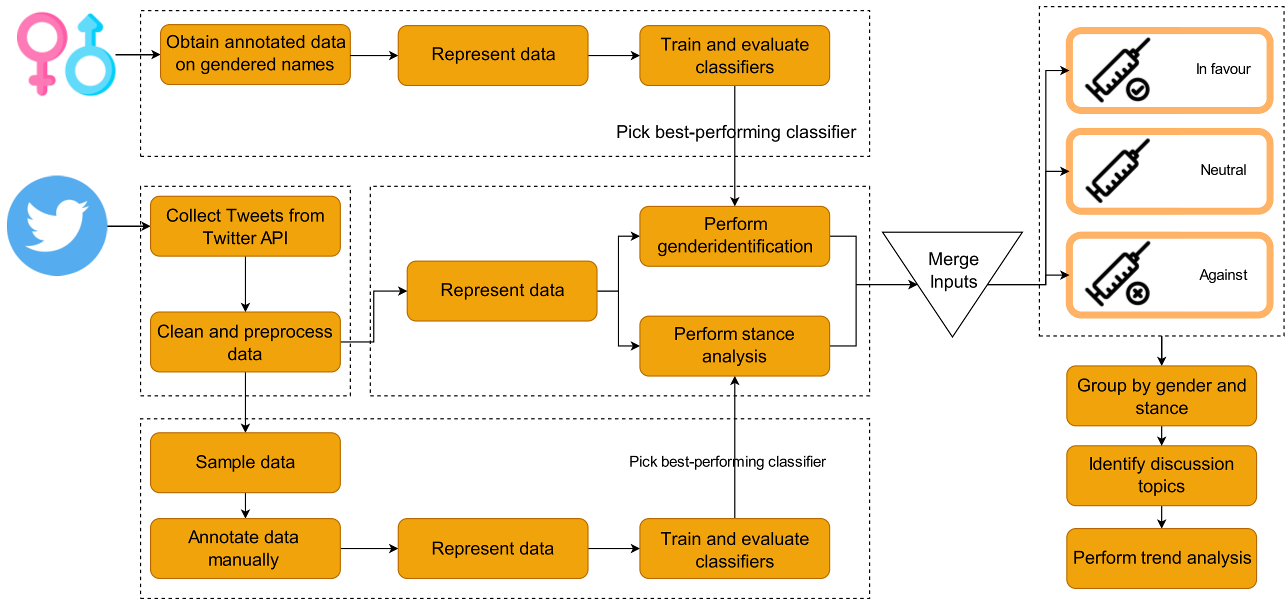


FIGURE 1. **The methodology used.**

TABLE 1. **Keywords used for data retrieval.**

Topic	Keywords
COVID-19	covid, covid19, covid-19, coronavirus, corona outbreak, coronavirus pandemic, wuhanvirus, 2019nCov
Vaccination	vaccine, vaccination, vaccinate, vaccinating, vaccinated

windows for each day, a further maximum of 1,500 tweets were retrieved for each day. In total, 1,440,567 tweets were collected, ensuring a representative distribution across the entire period of the pandemic until the present.

The next step was to gather data for the gender identification sub-problem. The user's name, handle, and bio were considered initially as potential inputs for gender identification; in addition to this, the identification of gender from the tweets themselves was also briefly considered as an intrinsic gender identification problem [24], [52]. However, due to the higher difficulty of these tasks, as well as the lack of suitable training data, the name was ultimately deemed the best candidate for building a reliable classifier, due to the surprisingly large number of users in our data who had identifiable personal names set as their user name (this is a free text field, so arbitrary input is possible).

As such, the next step was acquiring training data for the preparation of a classifier able to predict the gender of personal names. The prediction of gender based on personal names is, as discussed above, a well-researched problem in the field. Manual annotation is unnecessary in this case, as many suitable data sets have been made available by previous research or government organizations. One of the most cited is the Baby Names from Social Security Card Applications dataset made available by the US government, composed of 35,556 personal names and the associated biological sex of the persons bearing those names, gathered from Social Security records in the US between 1880 and 2021 [53]. This dataset has the advantage of containing data gathered over a long time period, but it is limited in that it contains data relevant to only one culture; yet in the case of personal names it is clear that their associated gender is culturally determined [25], [28], [30], [32].

Another suitable dataset we identified was the Gender by Name Data Set available in the University of California Irvine Machine Learning Repository, containing 147,270 personal names and the associated biological sex of the persons bearing those names, dating from between 1880 and 2019 and gathered from a worldwide population (US, UK, Canada, and Australia) [54]. The higher diversity of the data reflects this worldwide distribution; however, the overall quality of the data is lower due to errors such as typos, obviously meaningless names, etc. We opted to use the latter dataset because, in addition to containing most of the data found in the former, it is larger and has more diverse contents, reflecting the observations that while Twitter posts tend to lean heavily towards a North American user base, it is nevertheless a global platform [2], [40].

B. CLEANING, PREPROCESSING AND ANNOTATION

Starting from the Gender by Name Data Set we removed any duplicates or names that appeared to be incorrect/gibberish (composed of only symbols, repeated letters, only one letter, etc.). The final dataset contained 133,970 names, with 62.3% being female and 37.7% male. The names, while of different ethnic and linguistic backgrounds, were already latinized, but some exhibited accents, diacritics, or other special characters;

TABLE 2. Number of tweets by author type as identified via named entity recognition.

Author type	Tweet Count	Percentage
ORG[a]	346,292	29.27%
PERSON	419,833	35.48%
UNK[b]	417,030	35.25%
Total	**1,183,155**	**100.00%**

[a]organization
[b]unknown, mostly pseudonymous users

these were left untouched. The names were converted to lowercase, characters other than letters were substituted with the token "*", and the special token "$" was appended to the start as well as the end of the name so information concerning first and last characters can be captured during vectorization.

As for the quality of the name data within the data gathered from Twitter, it was quite variable; in addition to pseudonymous or corporate users, many users used different name orders, and some of them had artifacts such as emoji, exotic fonts, and titles such as "Ph.D." or "Dr." interspersed with their otherwise valid personal name. Nevertheless, manual inspection suggested many users did have identifiable name information available.

In order to parse the personal name out of the user name field, several approaches were considered, including pattern-matching heuristics, training a separate classifier, and using a pre-trained parser. Ultimately, we decided to use the named entity recognition (NER) capabilities of the spaCy[3] Python package to identify and extract personal names in the data.

We ran the spaCy tokenizer on all names and saved those that could be identified as human names. Those identified as organizations were assigned the dummy gender "ORG", as the property of gender does not apply to them. Finally, the names that contained no known entities were considered "UNK" (unknown) users. Upon manual inspection of samples, these consisted mainly of pseudonymous users. The data was quite evenly distributed between the three categories. The proportions in the data can be seen in Table 2.

In the case of the tweets, we converted the entire corpus to lowercase and removed all punctuation, URLs, HTML entities, and other non-character symbols. We left in hashtags and @mentions as these can provide some information and context with regards to the stance. For the same reason we changed emojis to their textual representation. We removed duplicate tweets, resulting in a final set of 1,183,155 tweets.

A 0.9% stratified random sample was drawn from the tweets, resulting in a sample size of 10,697 texts. These tweets were manually annotated by three human annotators with one of the labels *neutral*, *in favour*, or *against* based on the stance towards vaccination expressed by the tweet. The annotators were unaware of the name of the author or their predicted gender. Conflicts were solved by discussing the relevant tweet and obtaining a consensus on their stance. After annotation, the class distribution was highly imbalanced. With such a small dataset, class imbalance significantly reduces the performance of the classifiers, so a balanced sub-sample was drawn via undersampling of the majority class (*neutral*), resulting in a final, perfectly balanced training set of 2,739 tweets (913 tweets in each class). This subset is used to train a series of machine learning models, of which the best-performing one is then used to infer the stance towards vaccination for the entire dataset. The balanced annotated dataset is available at the following link: https://github.com/erkovacs/1000-days-covid-19-a-gender-based-long-term-investigation-attitudes-with-regards-to-vaccination.

C. DATA REPRESENTATION

We performed experiments with multiple data representations for both sub-problems. For the gender identification sub-problem we ran experiments with character-level n-gram representations with or without TF-IDF, while for the stance detection sub-problem we tried document-term matrix, TF-IDF, spaCy word vectors, spaCy sentence vectors, and the native embeddings for the BERT and RoBERTa models.

1) DOCUMENT-TERM MATRIX

A document, or individual sequence of text from a corpus of related texts, can be thought of as a sequence of tokens (n-grams; words or groups of words). If we take the co-occurrence of these n-grams as the fundamental feature of the document, we can represent any document as an $n \times m$ document-term matrix M in a vector space described by the vocabulary of the corpus. In the matrix M, the columns represent the features (tokens), the rows represent one document per row, and each element M_{ij} represents the corresponding token's frequency (also known as the count, term frequency, or TF – see Equation 1), or how many times that specific token occurs within the document.

$$TF(t, d) = \frac{\text{number of occurences of } t \text{ in } d}{\text{number of words in } d} \tag{1}$$

where t is the n-gram or "term", and d is the current document.

This representation is suitable for any sequence-based data, and as described below, can be generalized to be used at the character-level as well.

Within our Python notebooks, we used the CountVectorizer class included in the scikit-learn[4] library to build the document-term matrix.

2) TF-IDF

The TF-IDF-enhanced document-term matrix is an iteration on the token-count document-term matrix which seeks to encode a measure of a token's importance by weighting its simple count (here called the term frequency/TF, see

[3]https://spacy.io/

[4]https://scikit-learn.org/

Equation 1) by a measure of its overall frequency in the whole corpus (here called the inverse document frequency, IDF, see Equation 2) [55].

$$IDF(t,c) = ln \frac{\text{number of documents in } c}{\text{number of documents in } c \text{ containing } t} \quad (2)$$

where t is the n-gram or term, and c is the corpus containing all documents.

This technique is meant to better capture the relative importance of the token within the document, as for instance certain words, especially function words such as articles (''the''), prepositions (''in'', ''at'', ''on''), conjunctions (''and'', ''or''), and certain adverbs (''here'', ''there'') are extremely common both within and between any set of documents, yet are imbued with very little semantic content, as they primarily express relations between other words. Thus, TF-IDF (Equation 3) seeks to capture the fact that relatively rare words within the corpus that are common within a few documents might have special significance [55].

$$TF - IDF(t,d,c) = TF(t,d) \cdot IDF(t,c) \quad (3)$$

While such a reasoning is sound in the case of sentences or entire documents, it is arguable whether it applies to sub-word structures, as these are much less heterogenous (there are theoretically 18,278 distinct letters, bigrams, and trigrams that can be formed using the English alphabet, and of these, only a fraction actually occur in practice; this is orders of magnitude fewer than how many distinct words there are in English, not to mention word-level bigrams and trigrams). In our experiments, using TF-IDF at an n-gram level led to a slight performance decrease for the classical ML models and a modest performance increase in the case of the neural networks, as can be seen in Table 3.

For all models we used the TfidfVectorizer class from the scikit-learn package in our Python notebooks.

3) CHARACTER-LEVEL N-GRAM ENCODING

In English, as well as in many other languages across the world, individual words, including anthroponyms, are written using sequences of individual characters. This enables us to use a similar encoding scheme as in the case of the document-term matrix, only at the character instead of the token level. As discussed above, these sequences can be represented as vectors of size m, in this case in a vector space described by the alphabet used, plus special tokens such as the ''*'' token denoting a non-alphabetical character or the ''$'' delimiter we used to mark the beginning and the end of the sequence. To capture sub-word multi-character structures, we used character-level n-gram features of $n \in$ [1], [3], inspired by [32]. We avoided using higher-order n-grams in order not to start treating entire names as features. We obtained the best results when we used the plain n-gram frequencies for bigrams and trigrams without TF-IDF, with an alphabet limited to 1024 features.

In our Python notebooks, we used the CountVectorizer and TfidfVectorizer classes from scikit-learn with the analyzer parameter set to ''char''.

4) SPACY WORD AND SENTENCE VECTORS

Much like the seminal structuralist thesis stating that meaning is predicated on the difference between signs [56], the meaning of words can be represented in a vector space of abstract meaning, where their similarity in meaning would be represented as a proximity in space. The features describing this vector space can be automatically learned from large corpora of text; many word vector models (usually called ''word embeddings'' in the literature) are available, such as GloVe [57], word2vec [58], as well as specialized embeddings for use with large Transformer-based models [59], [60]. We opted to use the vector representations included in the spaCy package available for the Python language, which are based on GloVe. We used the large English language model, which can represent words as 300-dimensional vectors. Using the same approach, spaCy is also capable of representing entire documents as 300-dimensional vectors by averaging the vectors of the words they contain. Both approaches were tried in experiments, but sentence vectors presented better results and faster training time, as well as a more manageable memory footprint.

Word vectors depend on context, and spaCy's implementation cannot represent words outside its language model's vocabulary. Thus, personal names not in the vocabulary would have been represented as null vectors. This makes this representation unsuitable for the gender identification task.

5) TRANSFORMER WORD EMBEDDINGS

Because of the unique way in which they have been trained, the Transformer-based models require their own native tokenizers to obtain pre-trained word embeddings as well as their associated attention mask. BERT uses a WordPiece encoder with a vocabulary size of 30,000 [61], while RoBERTa uses byte-pair encoding tokenizer derived from GPT-2 which uses a vocabulary of size 50,000 sub-words [60].

There are certain known issues with how personal names are represented by these tokenizers [62]. Thus, this representation was used only for the stance detection sub-problem.

We used the AutoTokenizer class from the Huggingface[5] provider with the PyTorch[6] package as a tensor library and an interface with the GPU in our Python notebooks.

D. DATA AGGREGATION AND TOPIC ANALYSIS

We apply topic modelling using LDA to model the main discussion topics, as well as gain insight into how these differ across gender and stances. LDA is a generative model that can be used to reveal unobserved structures in text data (or abstract topics) [55]. Several hyperparameter values cited in the literature [63] were tried for each class-gender

[5]https://huggingface.co/

[6]https://pytorch.org/

combination. We optimized LDA by trial and error for each event-gender-stance combination and obtained our best results when modelling 4 topics. For this purpose, we used the Gensim[7] package. The criterion applied was how subjectively different the topics are between themselves, and their inter-topic distances as shown by the PYLDAvis[8] package. We used Bag-of-Words encoding for text representation, and we excluded terms appearing in fewer than 15 tweets or more than 90% of tweets, but we varied these parameters for some events where the topics appeared to be too homogenous. We also identified the most important n-grams for each period and each class using TF-IDF.

IV. CLASSIFIERS

Both sub-problems are multi-class classification problems, which we elected to solve using a machine learning approach. However, the types of classifiers used will differ between the sub-problems, as they differ in their scale: the gender identification sub-problem is a word-level problem, while the stance analysis sub-problem is a document-level problem. Furthermore, two types of machine learning algorithms will be used: classical machine learning algorithms, which use statistics, probability, and other mathematical modelling techniques, and deep learning algorithms, which use neural networks.

A. EVALUATION CRITERIA

In order to detect overfitting and get a good estimate of out-of-sample performance, all experiments have been run using 5-fold stratified cross-validation [2], [40], [64]. This is a process, widely used in the literature, which ensures all k "folds", which are equal-sized "slices" of the data containing $\frac{n}{k}$ training examples, are used in turn for training as well as evaluation, at the cost of k separate model fits. For each fold, k-1 folds are used for training, while the remaining fold is used for validation. If the problem is a multi-class classification problem, care must be taken to preserve class distribution within folds; for this purpose, we used the stratified k-fold data splitter included in the scikit-learn Python package.

Due to the different types of algorithms we used, there was a need to use common metrics to compare their performance. Because we considered both sub-problems as multi-class classification problems, we chose the average class accuracy, also known as the balanced accuracy score (Equation 5) [65], as the best measure of classifier performance. We also report the class-wise accuracy scores for each class, calculated as the recall (Equation 4) when considering only one class as "positive" [65]. This measure is expressed as a percentage or a number between 0 and 1 (higher is better). If the class accuracies are close to each other, it indicates that the classifier is not biased towards one or more classes to the detriment of the others.

$$recall = \frac{TP}{TP + FN} \tag{4}$$

$$accuracy_B = \frac{1}{k} \sum_{i=1}^{k} recall_i \tag{5}$$

where TP is the number of true positives (the positive prediction matches the label) and FN is the number of false negatives (the negative prediction does not match the label), k is the number of classes, i is the index of the class currently being considered, and the $recall_i$ is the recall calculated when considering only the i-th class as the "positive" class.

Finally, we report the F1 score (Equation 7), which is another commonly-reported evaluation metric [2], [52]. This is calculated as the harmonic mean of the precision (Equation 6) and the recall (Equation 4) [65].

$$precision = \frac{TP}{TP + FP} \tag{6}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{7}$$

where FP is the number of false positives (the positive prediction does not match the label).

The values reported have been calculated on the test data only and represent an average of the scores across all k folds.

B. MULTINOMIAL NAÏVE BAYES

The naïve Bayes classifier is a probability-based classical machine learning algorithm that applies Bayes' theorem to predict a sample being of a given class based on the likelihoods of each of its features being part of that class [65]. It is naïve in that it assumes that the features are independent, which in the case of document classification is not a correct assumption. Nevertheless the Bayesian classifier has proven useful in practical applications, and for establishing a baseline against which to compare more powerful models [2], [65].

We used multinomial naïve Bayes classifier for both sub-problems, since it is suitable for multinomially-distributed data, like token or character frequencies in text. Distinct feature sets have been used for the two sub-problems. For gender identification, the best results were obtained using 1024-dimensional vectors consisting of character-level n-gram features, $n \in$ [1], [3], while for the stance classification, 2-3000-dimensional word-level n-gram features of $n \in$ [1], [2] proved more useful for all models. For both sub-problems, we ran a grid search over several values of the smoothing parameter α, with the best results obtained for values 0.5 and 0.9 respectively.

In our Python notebooks, we used the MultinomialNB class included in the scikit-learn package.

C. RANDOM FOREST

The random forest (RF) classifier is an ensemble classical machine learning algorithm that consists of an array of decision trees, with each tree considering only a certain subset of features ("bootstrapping") and being trained on only a

[7] https://radimrehurek.com/gensim/
[8] https://pyldavis.readthedocs.io/en/latest/readme.html

subset of the data ("bagging"), the prediction being given by majority vote [65].

We obtained the best results for the gender identification sub-problem with a forest of 50 trees 50 nodes deep each, while for stance analysis we obtained a better performance with trees that are 60 nodes deep.

We used the RandomForestClassifier class included in the scikit-learn package in our Python notebooks.

D. SUPPORT VECTOR MACHINE

The support vector machine (SVM) classifier is an error-based classical machine learning algorithm that aims to find the optimal decision boundary, or hyperplane, between classes [65].

We used the SGDClassifier class included in the scikit-learn package in our Python notebooks with the default hyperparameters.

E. FEEDFORWARD NEURAL NETWORK

The feedforward neural network is the simplest neural network architecture found in the literature, and one of the earliest deep learning model families introduced [65]. It is called "feedforward" because it can be represented as an acyclic directed graph through which the inputs are fed "forward", from the input layer, through one or more "hidden" layers, to the output layer [65]. We included this simple network to establish a baseline for future work with deep learning models on sub-word structures.

We used a simple architecture consisting of one single hidden layer containing 512 neurons and used the ReLU activation function (Equation 8) [65]. Feedforward neural networks of such low complexity are not suitable for document classification, especially given the low quantity of the training data. As such, a feedforward neural network was not trained or evaluated for the stance analysis sub-task.

We implemented the neural network using the PyTorch framework.

$$ReLU(z) = \max(0, z) \tag{8}$$

F. LONG SHORT-TERM MEMORY NEURAL NETWORK

Introduced by Hochreiter and Schmidhuber [66] in the late 1990's, the long short-term memory (LSTM) architecture is a special case of a recurrent neural network (RNN) that is designed to solve the exploding/vanishing gradient problem in RNNs.

Because both sub-word n-grams and document-term matrix are sequential data, this architecture is suitable for both problems. We experimented with different architectures and hyperparameters; the best-performing configuration consisted of a linear input layer, an LSTM layer, and a Softmax output layer. In the case of the stance classification task, the inclusion of a dropout layer which probabilistically removes certain neurons from the network for each forward pass, helped improve performance by reducing overfitting. This was not the case for the gender identification task. In both cases, the LSTM appeared to underperform; further tuning and experimentation could improve performance in the future.

We implemented a custom LSTM neural network using PyTorch in our Python notebooks.

G. BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)

BERT is an advanced deep learning architecture based on the Transformer encoder, first introduced by Devlin et al. [59]. It led to the development of a family of pre-trained models based on this architecture, which have been trained on a large volume of data on generic tasks but can be fine-tuned with relatively modest data requirements for any domain-specific task, including classification [61].

BERT is not meant to be used for applications with character-level word representations. Moreover, personal names are likely to be unsuitable for representation as word embeddings due to the lack of immediate semantic content to their constituent parts, and their different etymological origin from most of the rest of English lexicon. As such, BERT was not used for gender identification, though it would likely be a good candidate for implicit gender identification, as in this case its pretraining could be leveraged via transfer learning to offset the issue with identifying the correct features for this task [24], [52]. At the same time, caution should be exercised, as BERT and other large models are known to exhibit certain biases, especially with regards to gender and ethnic background [62].

In our Python notebooks, we used the base BERT model from the HuggingFace provider.

H. ROBUSTLY OPTIMIZED BERT PRETRAINING APPROACH (ROBERTA)

Introduced by Liu et al. [60] as an iteration on the BERT model, RoBERTa aims to improve performance by changing BERT's tokenization scheme as well as certain steps in its training and validation pipeline, such as using different training texts, training for longer on larger batch sizes, skipping the next sentence prediction validation task, and changing the masking pattern applied during masked language modelling.

Similar to BERT, RoBERTa is unsuitable for sub-word level representations for the same reasons (it was not trained for such a task) and, as such, was only used for the stance classification sub-problem, for which it has obtained very good results in the past [2], [15], [40]. Nevertheless, it could also be a good candidate for implicit gender identification, as it has been used successfully in the past for author profiling tasks, including the identification of the author's gender [67].

In our Python notebooks, we used the base RoBERTa model from the HuggingFace provider.

V. RESULTS

Good results have been achieved for both sub-problems. For the gender identification problem, we have obtained the best performance in the literature known to us, although by an

TABLE 3. Performance of the classifiers for the gender identification sub-problem.

Classifier	Vectorization	Class	Recall	F1 score	Balanced accuracy
MNB	CV[b]	*female*	85.25%	0.8354	82.85%
		male	80.45%		
MNB	TF-IDF[c]	*female*	89.83%	0.8197	79.70%
		male	69.57%		
RF	CV	*female*	91.61%	0.8649	84.91%
		male	78.20%		
RF	TF-IDF	*female*	91.14%	0.8581	84.17%
		male	77.20%		
SVM[a]	CV	*female*	88.99%	0.8654	85.69%
		male	82.40%		
SVM	TF-IDF	*female*	87.58%	0.8463	83.62%
		male	79.66%		
FFNN	CV	*female*	80.40%	0.5971	85.26%
		male	93.27%		
FFNN	TF-IDF	*female*	86.36%	0.5872	85.36%
		male	85.39%		
LSTM	CV	*female*	78.14%	0.5783	82.86%
		male	90.39%		
LSTM	TF-IDF	*female*	82.54%	0.5717	83.35%
		male	85.67%		

[a]the best-performing classifier.
[b]character level n-gram counts.
[c]character level n-gram counts, weighted by IDF score.
Note: all representations were on 1024-dimensional vectors of bigrams and trigrams.

admittedly small margin (described in sub-section A). For the stance detection sub-problem, we have obtained results consistent with the current state of the field (discussed in sub-section B). The best performing classifiers for both sub-problems have been used in section C for gender analysis and in section D for trend analysis. Finally, a topic identification is carried out in section E on the purpose of determining if there is or not a strong systematic correlation between gender and stance.

A. GENDER IDENTIFICATION

The model that performed best was a linear SVM learning 1024-dimensional vectors with bigram and trigram count features, obtaining a balanced accuracy score of 85.69% and an F1 score of 0.8654 across 5 folds. To our knowledge, this is one of the best results in the literature for this task.

It must be noted that while this model outperformed some other models in Table 3 only by a small margin, in terms of efficiency it is much less computationally demanding than the deep learning approaches which came close. It is also simpler and more interpretable, which is a significant advantage for research because it leads to valuable insights, such as identifying the sub-word structures that the model found highly predictive of gender (see Table 4).

TABLE 4. 10 features most highly predictive of each gender.

Gender	Features (n-grams)
Female	a$, ah$, een, i$, ko$, jo$, rst, yo$, deb, y$
Male	zac, ben, sea, no$, iy$, abd, bd, ull, sh$, jee

Note: the "$" token marks the start or end of the name.

TABLE 5. Author gender as identified by classifier.

Author type	Gender	Count	Percentage
PERSON	F	124,082	14.83%
	M	295,751	35.34%
UNK[a]	F	90,737	10.84%
	M	326,293	38.99%
Total by gender	F	214,819	25.67%
	M	622,044	74.33%
Total		**836,863**	**100.00%**

[a]unknown, mostly pseudonymous users

A related observation is that the feedforward neural network and the LSTM were unable to exceed the performance of the classical machine learning models, despite the availability of plentiful data of good quality. It is also worthy to note that even the baseline MNB classifier obtained good results, giving even more support to the theory that the sub-word structures of personal names do encode gender information [25], [30], [32].

As can be seen in Table 4, our best-performing model captures some of these structures; common-sense observations, such as female names ending in *-a*, *-i*, or *-y*, as well as more subtle structures within male names reflecting the diverse descent of the names in the data were features which the model deemed highly predictive of each gender.

We used this classifier to annotate the entire dataset with gender information. Since users can specify their names in different orders in the free text field provided for this purpose by Twitter, we associated a weight to each distinct name that has been detected using NER. For users marked as "person", we expected Western name order (*<first name, middle name, last name>*) to be more prevalent, and, since in English names only the first name and possibly the middle name provide gender information, we weighted these with higher weights (0.6 and 0.3, respectively, with the last name weighted with 0.1). For users marked as "unknown", we weighted all tokens equally. The distribution obtained can be seen in Table 5. This is significantly different from the true estimated distribution [68], but it does capture the fact that Twitter is more popular with men than women.

B. STANCE DETECTION

The best performance by far was obtained for the stance detection sub-task by the transformer-based models, BERT

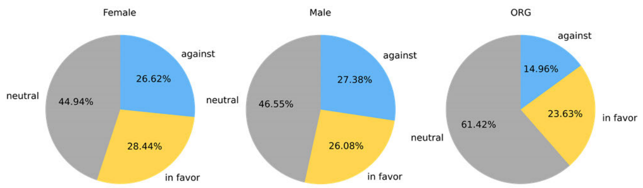


FIGURE 2. The proportion of stances by gender.

and RoBERTa. The best model obtained a very high balanced accuracy score of 93.64% and an F1 score of 0.9364. Given the nature of the task and the small size of the training set, these are some of the best results given similar setups in the field [2], [15], [34], [40].

As for the other models, they had a poor performance, as can be seen in Table 6. Especially low performance was seen in the case of the LSTM neural network; overfitting was the main issue in this case, and the addition of a dropout layer did improve performance, although not sufficiently to compete with the transfer learning models. The lack of sufficient training data was probably another factor that has affected the performance of the LSTM classifier.

This difference in performance highlights the utility of transfer learning for highly context-dependent tasks such as stance analysis, even for training datasets of small size. Contrast this to the gender identification sub-task, where context is not available, and the potential for ensemble approaches, combining multiple classifiers, to author profiling tasks becomes clear.

After classifying all tweets with the best model, we have obtained the following class distribution: 50.61% of tweets have been predicted as *neutral* to vaccination, 25.79% *in favor* of vaccination, and 23.60% as *against* vaccination.

In particular, in the case of female authors, it has been observed that the *in favor* class included slightly more tweets (28.44%) than the *against* class (26.62%), generally following the overall class distribution with an approximately up to 3% difference (Figure 2).

In the case of male authors, the distribution of the tweets presents a shift in the proportion between the *in favor* and *against* tweets, namely the percentage of *against* tweets (27.38%) exceeds the percentage of *in favor* tweets (26.62%). It can thus be noted that there is a difference between the female and male authors from the perspective of the hierarchy of the three classes: while the percentage of *neutral* tweets holds the first place for both genders, a change in order is recorded for the percentage of *in favor* and *against* tweets.

As for the organization authors, a higher percentage of tweets have been included in the *neutral* class (61.42%), with an especially lower percentage of *against* tweets (14.96%). The percentage of *in favor* tweets in the case of organization users (23.63%) is lower than in the case of both female and male authors.

TABLE 6. Performance of the classifiers for the stance identification sub-problem.

Classifier	Vectorization	Class	Recall	F1 score	Balanced accuracy
MNB	*CV*	*neutral*	70.72%	0.6634	66.40%
		in favor	58.93%		
		against	69.52%		
MNB	*TF-IDF*	*neutral*	67.17%	0.6703	67.00%
		in favor	62.45%		
		against	71.39%		
MNB	*spaCy*	*neutral*	69.00%	0.5498	55.49%
		in favor	41.42%		
		against	56.06%		
RF	*CV*	*neutral*	65.75%	0.6083	60.80%
		in favor	58.39%		
		against	58.25%		
RF	*TF-IDF*	*neutral*	60.49%	0.5999	60.48%
		in favor	52.55%		
		against	53.08%		
RF	*spaCy*	*neutral*	64.73%	0.5834	58.31%
		in favor	53.50%		
		against	56.70%		
SVM	*CV*	*neutral*	65.42%	0.6443	64.40%
		in favor	63.37%		
		against	64.44%		
SVM	*TF-IDF*	*neutral*	66.86%	0.6738	67.38%
		in favor	64.40%		
		against	70.86%		
SVM	*spaCy*	*neutral*	65.16%	0.5856	59.16%
		in favor	55.33%		
		against	57.00%		
LSTM	*CV*	*neutral*	68.00%	0.6412	64.27%
		in favor	59.70%		
		against	65.46%		
LSTM	*TF-IDF*	*neutral*	74.25%	0.6006	62.34%
		in favor	51.45%		
		against	65.72%		
LSTM	*spaCy*	*neutral*	62.44%	0.6019	58.52%
		in favor	56.92%		
		against	56.19%		
BERT	-	*neutral*	92.99%	0.9115	91.16%
		in favor	90.77%		
		against	89.84%		
RoBERTa[a]	-	*neutral*	93.89%	**0.9364**	**93.64%**
		in favor	93.13%		
		against	94.07%		

[a]the best performing classifier, trained for 20 epochs with a batch size of 60, sentences truncated to 256 tokens, and a learning rate of 0.00003.

TABLE 7. ***In Favor*** **Tweets (selection).**

Gender	Tweet
female	After experiencing mild symptoms, I tested positive for #COVID19, and am isolating at home. I notified close contacts and will continue to follow public health guidance. This is a reminder to please do your part, get vaccinated, and get boosted.
female	We have to meet people where they are in order to get Oregonians vaccinated. Thank you @uoregon for helping our communities defeat COVID-19. #MyVaccineReason https://t.co/Bh8SZc2Loj
female	Just had another COVID-19 booster! This time it was the Moderna bivalent vaccine. I'm so thankful for vaccinations.("You're just going to keep getting jabbed? When does this end?" This ends when COVID-19 is extinct.)
male	Please consider getting vaccinated. I lost my eldest sister Saturday, January 1st, 2022, and my mother Wednesday, January 5th, 2022 Both to COVID-19 w/ pneumonia. Neither wanted to be vaccinated due to the crap pushed out on social media instead of listening to their physicians. https://t.co/4SkLEo1Eek
male	When I get frustrated that people are doing terrible risk assessments and expressing vaccine hesitancy I like to remind myself that basic math isn't intuitive for everyone. Vaccines are *thousands of times safer* than covid. Vaccine hesitancy puts all of us at risk. https://t.co/CEBTnIgi9
male	A patient's father just told me a conversation with me is what convinced his now fully vaccinated wife to get the COVID vaccine, and she passed on the conversation to her friends, and so at least three people got vaccinated as a result, and I am over the mooooooooon.
ORG	With @MidwivesRCM we are urging all pregnant women to have their flu jab and COVID-19 vaccine this winter. It is possible to be infected with flu and COVID-19 at the same time and this could make pregnant women seriously ill. Find out more: https://t.co/d6cnZKWhXg https://t.co/v2CBNCHey7
ORG	Did you know that people are more likely to get vaccinated & boosted if they know a loved one has gotten their vaccines? Join us next week for a training on how to talk to your friends and family about getting vaccinated against COVID-19.https://t.co/YkcemZvKM6
ORG	If you haven't received your first booster, get it as soon as possible!Your protection against getting a COVID-19 infection has likely decreased if it has been more than 6 months since you received your initial vaccine series and you haven't had COVID-19 in the past 3 months. https://t.co/aqL53oXepr

TABLE 8. ***Against*** **Tweets (selection).**

Gender	Tweet
female	There seem to be an awful lot of middle aged men in the news this weekend who are *amazed* to discover that they "survived" Covid. Yes, the vaccine will have cut their risk of severe illness or death, but they still had a 99%+ chance of "surviving" the virus. 🤷‍♀️
female	She got vaccinated last month and died this week of covid, my question is what are they injecting people now?🤷‍♀️
female	@ThaiPBSWorld The first year of covid, before there was a vaccine, only 88 people died of covid. Total. In a year. What has happened since then that would explain the huge increase in deaths?
male	@DutchHockeyMom @MargaretChant The uk pay £120,000 if a person is 60% disabled. Good luck proving it. People with death certificates stating the cause of death to be a covid vaccine - STILL HAVNT BEEN PAID OUT. I don't need to say but £120,000 is not enough for severe disablement or death. My dads certificate. https://t.co/JpdKw7Zg3r
male	Biden's COVID Strategy:1. Force people—through threats—to take vaccines against their informed consent2. Censor, silence, de-platform, and even cancel doctors, scientists, and thought leaders who question "the science"3. Hide vaccine documents for 75 years4. Force-masking
male	When you're a baby you get "one" vaccine shot per immunization, and during flu season you only get "one" flu shot; but with the Covid jab you're required to get 3 - 4 - or 5 shots and even then you can still get the virus!! Glad I didn't fall for this big pharma cash grab!! https://t.co/McYyjslNTj
ORG	They are coming after our kids!😱"The ACIP panel voted 15-0 for the CDC to recommend that children get the COVID-19 "vaccines and boosters."The actual vote to add this experimental COVID-19 mRNA injection to the childhood vaccination schedule is tomorrow."~Dr. Robert Malone
ORG	If the vaccine manufacturers were subject to the liability of their COVID products..They would all be bankrupt and in prison by now.
ORG	Conspiracy theory: We get a 99% successful vaccine for COVID-19 but it leaves 70% of vaccinated population sterile. How's that for population control! :)

C. GENDER ANALYSIS

In this section we have considered the three types of tweet authors, namely females, males and organizations, in order to better understand if there is a difference in the focus of the discourse, the arguments brought in favor or against vaccination, the length of the messages and the most frequently used terms.

First, a selection of *in favor* tweets has been made for the three genders, as presented in Table 7. From this selection, it can be observed that the *in favor* tweets posted by female and male authors are mainly focusing on presenting personal experiences, while the tweets published by organization authors are more general, urging the public to vaccinate or informing it about the available vaccination points.

Next, a selection of *against* tweets is presented in Table 8. Based on the selection, it can be noticed that, in the case of all three types of tweet authors, the messages focus on the hesitancy reasons rather than on the personal experiences and include calls against vaccination.

Regarding the length of the tweets posted by each of the three classes of authors (Figure 3), it can be observed that when considering the number of tokens, the length of the *in favor* tweets is greater than that of the *against* tweets, with the average ranging between 36.80 and 37.60 tokens in the case of *in favor* tweets versus an average ranging between 33.68 and 34.97 tokens in the case of *against* tweets (Figure 3).

Additionally, based on Figure 3, it can be seen that the variability of the length in the case of *against* tweets is greater than in the case of *in favor* tweets. When taking into account the gender of the authors, it can be noticed that the tweets with female authors are slightly longer than the ones published by male authors, which in turn slightly exceed the length of the organization tweets.

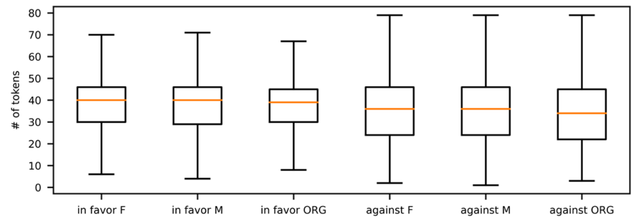


FIGURE 3. **The average number of tokens for in favor and against tweets for each gender.**

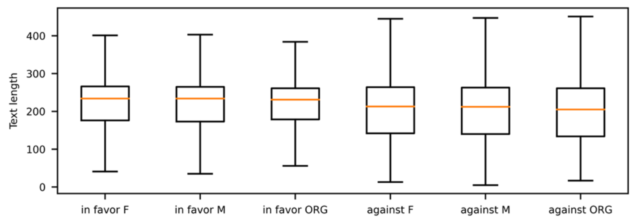


FIGURE 4. **The average number of characters for in favor and against tweets for each gender.**

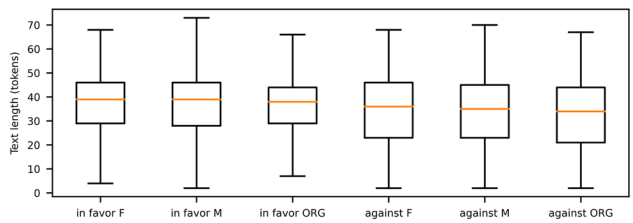


FIGURE 5. **The average number of tokens for in favor and against tweets for each gender after removing uninformative entities.**

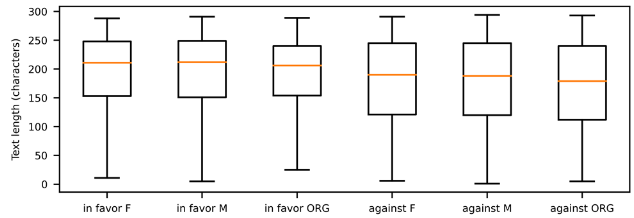


FIGURE 6. **The average number of characters for in favor and against tweets for each gender after removing uninformative entities.**

If instead of the length in tokens of the tweets, we consider the length in characters, the same observations hold true, namely the average length of *in favor* tweets (216.96 to 218.22 characters) is higher than the one of the *against* tweets (198.54 to 203.62 characters). Furthermore, as expected, the length in characters of the female authored tweets is greater than the one of male and organization authored tweets (Figure 4).

Based on the observation that tweets frequently contain mentions, URLs, hashtags, emoji, punctuation and duplicated spaces (Table 9), we have decided to remove them, since they can have a considerable impact on the length of the tweets, without representing actual text created by the author of the tweet for supporting the expressed view towards COVID-19 vaccination. As a result of this process, the average length in tokens of the tweets ranges between 35.91 and 36.87 tokens for *in favor* tweets and between 32.96 and 34.40 tokens for *against* tweets (Figure 5). Even in this case, the female authored tweets slightly exceed in length the male and organization authored tweets.

Regarding the number of characters for *in favor* and *against* tweets for each gender after removing these uninformative entities, the same observations hold true, namely the *in favor* tweets contain more characters than the *against* tweets, while the female authored tweets are longer than both male and organization authored tweets (Figure 6).

Even though similar results have been obtained both when the uninformative entities have been kept and when they have been removed, a difference in length associated with the presence of uninformative entities can be noticed, especially in the case of the number of characters, where the length has been reduced with up to 12.48%. The reduction in terms of characters has been higher on average in the case of *against* tweets (11.70%) than in the case of *in favor* tweets (10.35%), showing that the *against* tweets are more prone to contain uninformative entities. Regarding the gender of the tweets authors, it has been observed that a higher reduction of the tweets length has been recorded for the organization authors tweets (2.42% when the tokens have been counted and 12.48% when the characters have been counted).

Further, an n-gram analysis has been performed for each gender of the authors, by considering the bigrams and tri-grams present in the tweets. When establishing the top-10 n-grams we have not taken into account the n-grams that are specific to vaccination and covid-19 (e.g., "covid 19", "19 vaccine", "covid 19 vaccine", etc.).

The top-10 n-grams for *in favor* tweets are listed in Table 10, along with their number of appearances in the collected dataset between brackets. While some n-grams are common to all the three types of authors (e.g., "wear mask", "safe effective", "side effects" and "public health"), other are specific to female and male authors (e.g., "please get", "people get", "tested positive" and "died covid"). As it can be noticed, more n-grams are common between female and male authors, than between female / male and organization authors. Still, a difference can be observed between the n-grams corresponding to female and male authors, as "please get vaccinated" has been included in the top-10 n-grams for female authors, while "herd immunity" and "fight covid" have been included for male authors. Furthermore, while female and male authored tweets include n-grams focusing on asking others to get the vaccine (e.g., "please get", "please get vaccinated", "fight covid") and on the consequences of contacting COVID-19 (e.g., "died covid"), the tweets authored by organizations are more informative, focusing on protecting others (e.g., "help protect") and taking the vaccine at a vaccination clinic (e.g., "vaccine clinic", "vaccination clinic", "best way" and "make sure").

TABLE 9. Tweets containing uninformative entities (selection).

Tweet	Length in tokens / characters	Length in tokens / characters after removing uninformative entities
@FemmeVoilet @AngryFleas @EXPOSECORRUPT11 @account_walt @David4677484957 @Forsolace @TheWreckingBal5 @WCSmith50952576 @8BrianVogel4 @canyon_free @teamphil @HoundTheLibs @nikkillavender @sundog723 @Billenjoylife @Julie_NoNumbers @ever_texas4 @LadyB01427871 @IanB_67 @FamousSheymus1 @PogueMoran @LadyBB85125431 @PhilipPower13 @Hirdman @garygreno @Paul4dogs @marvin_hill123 @Marc62300483 @EviemaRee777 @FD03955088 @Birdseye1 @Fueldog3 @DuckworthVSTers @Dave92442188 @Tom_Neverwinter @sandraclunie @goodtroubless @spiralwrap @NumbersMean @ChrisMc47547598 @Chamele01374771 @NtRdytoMkNice @CanYouHearMeN15 @JMDelga82664720 @TyroneWJackson @durandundee7105 @_RaiseTheBlack_ @planetdoge @YouTube @AE911Truth The Pfizer shot suppressed white blood cells for 7 days after vaccination allowing for untold amounts of people to contract Covid. Those same infections were counted as "unvaccinated" <14 days. It was never a pandemic of the unvaccinated and you were never in my league. 😂	96 / 975	46 / 267
A colleague of mine who happened to be vaccinated, 3 shots so far used to make fun of me for not taking the JAB, he just tested positive for COVID-19 today.😂😂😂😂😂breathe 😂😂😂😂😂😂😂😂 🤣🤣🤣🤣🤣	33 / 183	33 / 161

TABLE 10. Top 10 N-grams by Gender for *In-favor* Tweets.

Female	Male	ORG
wear mask (1044)*	wear mask (2511)*	safe effective (1364)*
safe effective (768)*	safe effective (2093)*	wear mask (1220)*
side effects (753)*	side effects (2001)*	vaccine clinic (1165)
please get (746)+	people get (1789)+	loved ones (1135)
public health (702)*	please get (1762)+	vaccination clinic (1134)
people get (635)+	public health (1718)*	best way (1123)
tested positive (635)+	tested positive (1612)+	make sure (1032)
loved ones (621)	died covid (1389)+	side effects (1016)*
died covid (508)+	herd immunity (1378)	public health (868)*
please get vaccinated (486)	fight covid (1367)	help protect (847)

*common for all types of authors
+common for female and male authors

TABLE 11. Top 10 N-grams by Gender for *Against* Tweets.

Female	Male	ORG
side effects (1099)*	side effects (3096)*	side effects (1011)*
died covid (771) / die covid (421) / dying covid (419)*	vaccine mandates (2326)*	vaccine mandates (671)*
vaccine mandates (727)*	died covid (2174) / die covid (1302) / dying covid (1156)*	bill gates (587)*
bill gates (612)*	bill gates (1958)*	died covid (551) / die covid (385) / dying covid (344)*
spread covid (606)*	long term (1742)*	spread covid (499)*
long term (584)*	spread covid (1716)*	long term (458)*
herd immunity (545)*	herd immunity (1677)*	herd immunity (451)*
immune system (471)*	natural immunity (1514)+	immune system (444)*
natural immunity (454)+	immune system (1464)*	wear mask (328)
vaccine passports (410)+	vaccine passports (1226)+	big pharma (318)

*common for all types of authors
+common for female and male authors

For the *against* tweets the top-10 n-grams by gender are included in Table 11. It can be observed that the n-grams highlight various hesitancy reasons related to COVID-19 vaccination, such as the existence of side effects (e.g., ''side effects'', ''long term''), the risk of dying (e.g., ''died covid'', ''die covid'', ''dying covid''), the fact that the pandemic is a scam (e.g., ''big pharma'', ''bill gates''), the fact that vaccinated people can still spread COVID-19 (e.g., ''spread covid''), the existence of alternatives (e.g., ''herd immunity'', ''natural immunity'', ''immune system'', ''wear mask''). Additionally, several n-grams express the opposition against some of the actions put in place by governments (e.g., ''vaccine mandates'', ''vaccine passports'').

In terms of similarity between genders, it can be noticed that most of the top-10 n-grams are the same across all the considered author types, albeit a difference in the order of the number of appearances. Thus, while some specific n-grams are present only in the case of the tweets published by organizations (e.g., ''wear mask'', ''big pharma''), in the case of male and female authors we have encountered the same n-grams, but in a different order.

D. TREND ANALYSIS

We plotted the results of the stance analysis as a time series showing the daily frequency for each stance-gender combination, divided by the total number of tweets with that stance posted during that day (Figure 7). This was done to counter the gender imbalance we have noted. We observed some interesting long-term trends – during the first few months of the pandemic, the majority of the tweets have been *neutral*, with a spike in polarized opinion (that is, a rise in both *in favor* and *against* classes) around February 2020 (E1), which declined rapidly. For E1 it has been observed that the percentage of *in favor* tweets for the considered types of authors (female, male and organization) range between 12.19% and 30.44%, while the percentage of *against* tweets range between 14.84% and 37.74%. The next 8 months were characterized by relatively constant proportions in all classes (approximately 20% *in favor,* 24% *against* and 56% *neutral*). Starting from October 2020 (E2), a pattern of increase in the *in favor* class and decline in the *against* class prevailed for 7 months, until May 2021 (E3), when the *against* class begun an increasing trend. Between February and May 2022, the *against* class continued to increase – for female (noted with F in Figure 7) and male users (noted with M in Figure 7) it passes the *in favor* class around February (E4.1), while for organizational users (noted with ORG in Figure 7) around May (E4.2). In the case of the latter, the difference is not so radical. This pattern of increase in opposition appears to continue towards the present day. As for immediately observable gender differences, it can be noted that the majority of tweets have been posted by users identified as male (see Table 5), but when the ratio between *neutral*, *in favor* and *against* tweets presented in Figure 7 is considered, there is no indication that there is a significant difference between genders.

This observation is supported by the results of the correlation analysis between stance-gender pairs presented in 8. Considering the gender of the users it can be noticed that high correlation values are recorded between genders for each stance, rather than within genders.

Similarly, as can be seen in Figure 8, there is much more correlation within stances than there is between stances. Within stances, it can be observed that the highest positive correlations are recorded for the number of *against* tweets when considering the three categories of accounts – female, male and organizations – ranging between 0.91 and 0.94. High values are recorded also for the number of *neutral* tweets (ranging between 0.79 and 0.85) and for the number of *in favor* tweets (ranging between 0.73 and 0.84) – Figure 8.

Between stances, it can be noticed that the number of *neutral* tweets is negatively correlated with the number of *against* tweets. The highest negative correlation has been recorded between the number of *neutral* tweets published by organizations and the number of *against* tweets published by male users, with a value of −0.64 (Figure 8). Relatively high negative correlations have been noticed between the number of *neutral* tweets published by organizations and the number of *against* tweets posted by female accounts (−0.61) and by organization accounts (−0.59). Lower negative correlations are observed between the number of *neutral* tweets published by either female or male users and the number of *against* tweets posted irrespective of the type of account.

Moreover, we plotted the squares of the difference between stance-gender time series in Figure 9. We can see that the residuals in the second and third plots, where the stance is varied and gender is held constant, are an order of magnitude larger than the ones where gender is varied. We conclude therefore that gender does not appear to be a significant factor with regards to stances concerning vaccination.

Finally, we plotted the normalized trend of the tweets against normalized case and death counts, as shown in Figure 10. We can see that despite the death count steadily declining after the introduction of vaccines (December 2020, denoted by E_v) [2], the trend of the *against* stance is increasing. The case and death spike around late 2021-early 2022 might be relevant to E4.

E. TOPIC IDENTIFICATION

Topics were generated for each gender-stance pair using LDA. Four topics were generated for each pair for each event in order to facilitate identification of the relevant news stories. Using domain knowledge and the information from the generated topics, we used Google News Search to identify the events. A selection of topics can be seen in Table 12, while the events identified can be seen in Table 13.

To check for topic content correlations between stance-gender pairs, we built a vocabulary consisting of all terms contained within the topics, then built the document-term matrix for each topic, comparing them using Cramer's V-measure [69], which yields a score between 0 and 1 indicating correlation for categorical data (0 means no correlation, 1 means perfect correlation). The results for E3, which is similar to other events as well, can be seen in Figure 11.

With these results we have seen that for all four events, from a gender perspective, there was no systematic correlation between gender and stance. A strong systematic correlation did exist between stances, especially for organizational users. Thus, we observe that the contents of the discussion depended more on the stance the user has with regards to vaccination rather than their gender.

VI. DISCUSSION

As observed in other findings in the literature [2], [15], [34], [40], we found that most tweets are *neutral* towards vaccination. We have not found any support for the existence of gender differences between the quantity of tweets that cannot be explained by the abovementioned gender imbalance on Twitter, nor have we found significant differences in content between genders based on the topic analysis, although slight differences in wording and perspective were apparent between genders.

Likewise, the evolution of stances appears to respond closely to the occurrence of relevant events, denoted E1-4 in Figure 7. An initial spike in interest (E1) was followed by a period of reduced interaction (except for the US, where the vaccine issue was taken up by both sides during the 2020 US Presidential Election – E2), until the introduction of several measures, including vaccination, mask mandates, and several iterations on the idea of COVID-19 certificates (E3). Finally, a spike in deaths and cases (E4) provoked increased interaction once again. The current trend appears to show an increase in the number of tweets against vaccination.

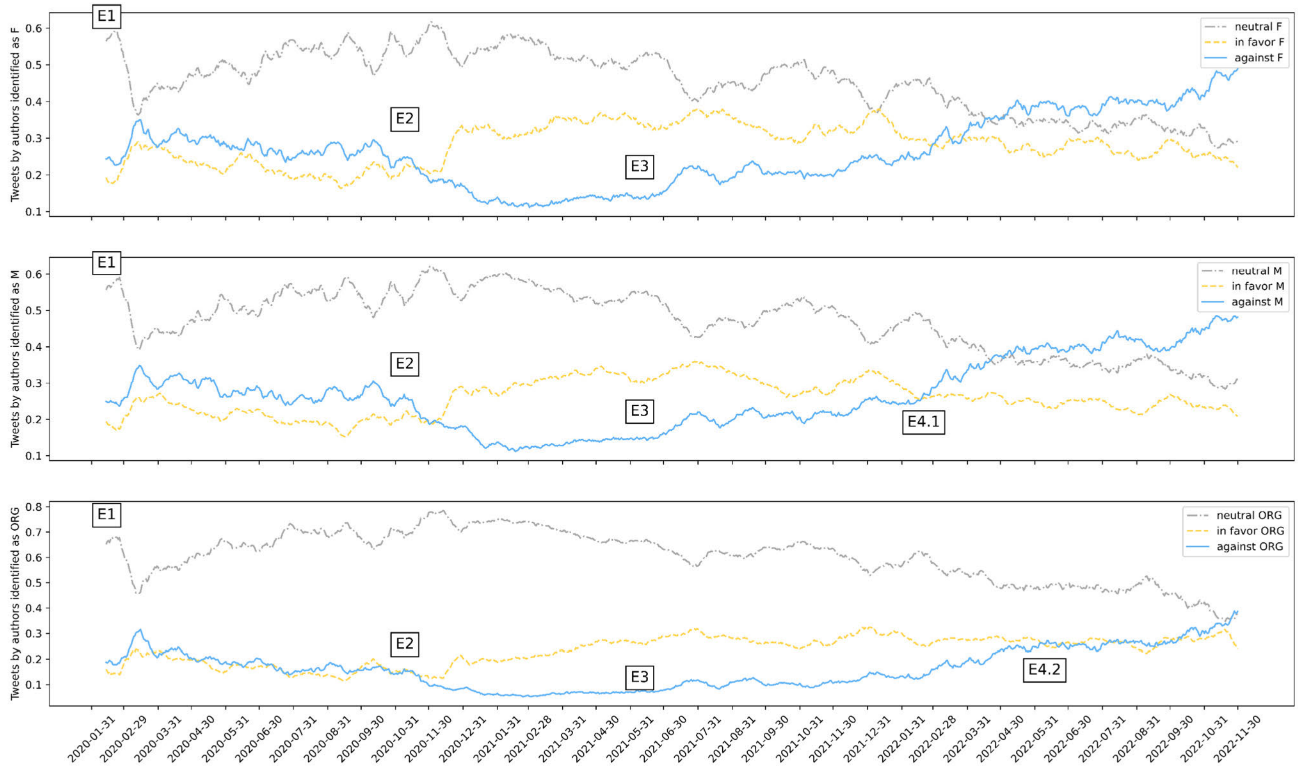


FIGURE 7. The evolution of each stance-gender combination as ratio to total number of tweets from users identified as each gender (15-day rolling average).

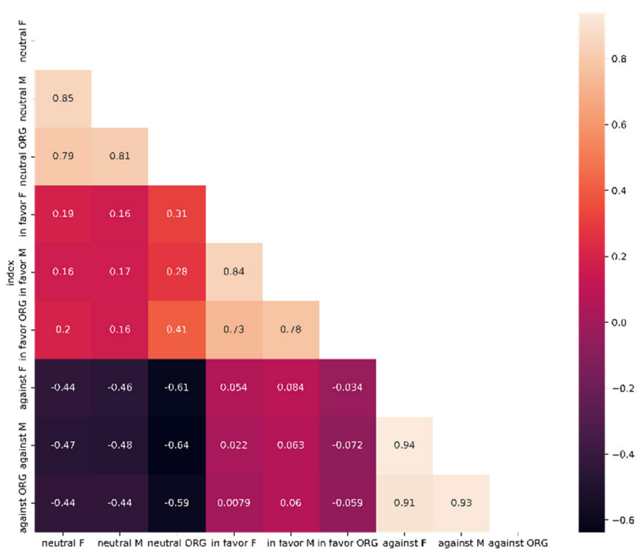


FIGURE 8. Correlogram showing Cramer's V-measure for correlation between n-grams used in each topic for E3.

The topics for the *neutral* and *in favor* stances appear to be consistent with previous findings when applied to COVID-19 [2], [15], [40]. They also appear quite consistent across the four events, stressing the fact that the vaccines have been extensively tested ("safe", "effective"), the need for the vaccine in order to resume normal life ("work", "family", "community"), and the risk from the disease itself ("risk", "severe", "death").

The topics generated for the against stance also suggest a preoccupation with themes already familiar in the literature, such as conspiracy theories (especially visible in E1 topics, which contain n-grams such as "china", "created", "lab" etc.), possibly referring to the then-prevalent idea that the novel virus could be a biological weapon created by China, but also much more reasonable concerns, such as side-effects ("risk"), the new technology used to develop some of the vaccines ("mrna", "experimental"). The n-gram analysis also confirms these preoccupations. These themes also appear similar to previous findings related to vaccine hesitancy with regards to COVID-19, but are quite distinct from "traditional" vaccine hesitancy topics [34]. Additionally, from the topics generated, as well as the n-gram analysis, it can be seen that opposition to vaccines was associated with opposition to other measures against the pandemic such as lockdowns and mask mandates ("mask", "lockdown", "mandate").

Additionally, a very interesting dynamic observed in both *in favor* and *against* tweets during E2 was the prevalence of both anti-Trump rhetoric, including blaming the pandemic on the Trump administration by directly addressing tweets with anti-vaccine contents ("world", "plan", "government") to Trump's official Twitter account (topic "realdonaldtrump"), as well as crediting and praising Trump for the development of the vaccines ("president", "working"). During

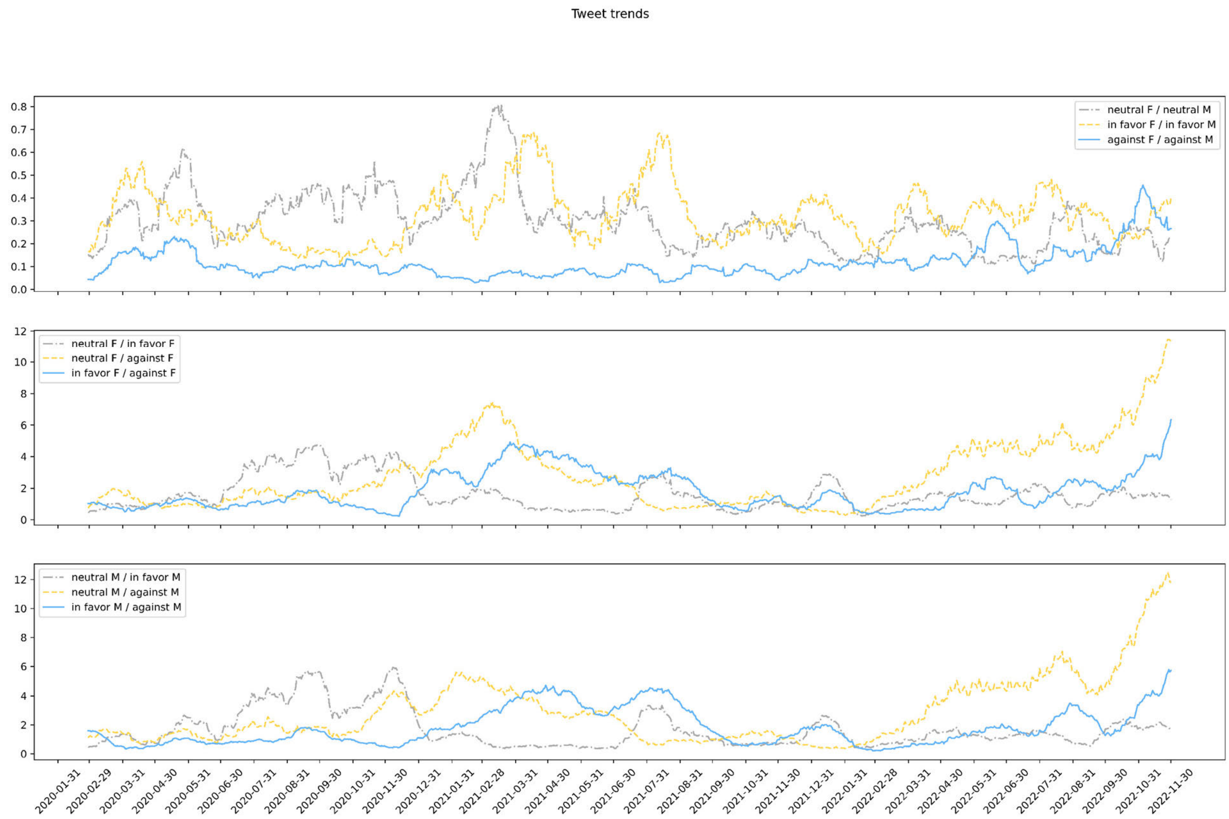


FIGURE 9. Squared residuals of stance-gender pairs. The first plot shows the situation when gender is varied and stance is held constant, while the following two show the gender being held constant while the stance is varied.

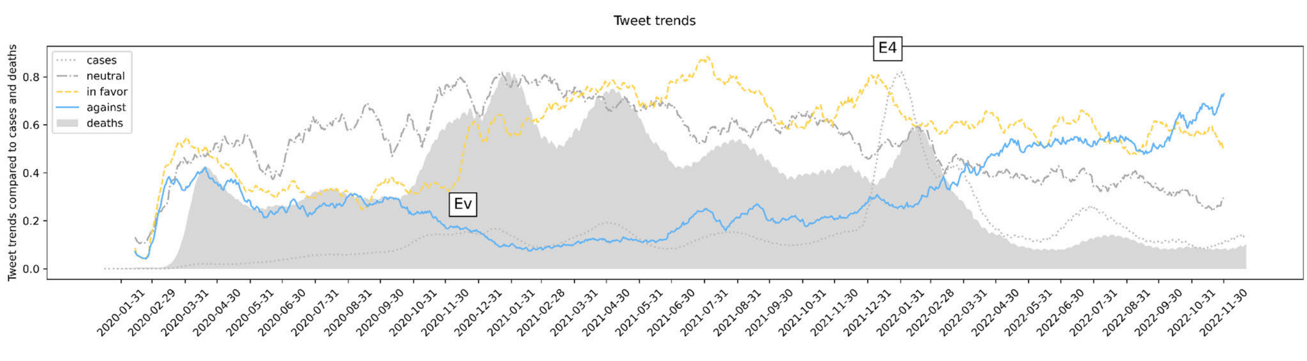


FIGURE 10. Normalized daily frequencies for each stance plotted against normalized daily cases per million and daily deaths per million. COVID-19 data source: Mathieu et al. [69].

E3, the vaccination of children also caused controversy (“child”).

Another interesting area that was not pursued extensively is the existence of organizational actors (users identified as “ORG”) that have systematically presented an *against* stance. The topic analysis suggests the topics discussed by these users are very similar to the general *against* tweets. It would be of interest to learn more about these organizations and their relationship to the larger ecosystem of discourse critical of vaccination against COVID-19.

A prevalent theme in *against* tweets we feel must be stressed is the opposition to coercive or authoritarian measures that have not been democratically or openly discussed, such as mask mandates, lockdowns, and vaccine passports. These concerns cannot be simply associated with vaccine hesitancy and dismissed out of hand; rather, the limitations of technocratic policymaking must be carefully and critically addressed when such measures are considered, and researchers and authorities must explicitly address these concerns in their communication with the public.

TABLE 12. Topics identified (selection).

Event	Stance	Gender	n-grams
E1	*against*	ORG	lab, created, china, hour, health, disease, make, pharma, virus, month
E2	*neutral*	M	oxford, astrazeneca, response, immune, trial, elderly, month, time, produce, university
E2	*neutral*	F	year, oxford, pfizer, work, response, elderly, time, immune, moderna, produce
E2	*in favor*	M	safe, effective, world, day, trump, trial, research, scientist, biden, election
E2	*in favor*	F	trump, life, made, available, president, effective, work, working, safe, health
E2	*against*	ORG	election, mask, health, world, plan, virus, realdonaldtrump, dangerous, testing, question
E3	*against*	F	virus, high, death, died, case, spread, government, centre, rate, health
E3	*neutral*	ORG	state, million, health, worker, government, staff, shot, employee, country, news
E3	*against*	ORG	year, experimental, government, risk, make, taking, life, child, death, work
E3	*against*	ORG	child, mask, week, work, health, mrna, time, year, fully, lockdown
E4.1	*against*	M	year, unvaccinated, death, mandate, fully, doe, immunity, fact, data, immune
E4.1	*against*	F	child, time, unvaccinated, test, case, life, infection, boosted, data, symptom
E4.1	*in favor*	M	day, fully, case, boosted, symptom, positive, test, month, unvaccinated, severe
E4.1	*in favor*	M	risk, death, long, infection, rate, health, world, country, immunity, kid
E4.1	*in favor*	F	protect, year, health, today, family, long, work, life, community, sick
E4.2	*neutral*	ORG	case, india, administered, update, rate, country, read, currently, data, information
E4.2	*neutral*	ORG	dose, child, booster, aged, year, clinic, pfizer, state, age, available
E4.2	*in favor*	ORG	risk, health, protect, safe, death, community, protection, infection, severe, illness
E4.2	*against*	ORG	effect, pfizer, unvaccinated, data, side, shot, booster, month, study, risk

Failing to do so can result in the proliferation of vaccine-hesitant and science-skeptical discourse that might harm future efforts to combat large-scale issues such as pandemics and climate change. As can be seen in Figure 7, the vaccine-hesitant discourse is on the rise, even though the absolute number of tweets regarding the subject of vaccination is decreasing. This suggests that rhetoric exhibiting the *against* stance has become an institutionalized discourse

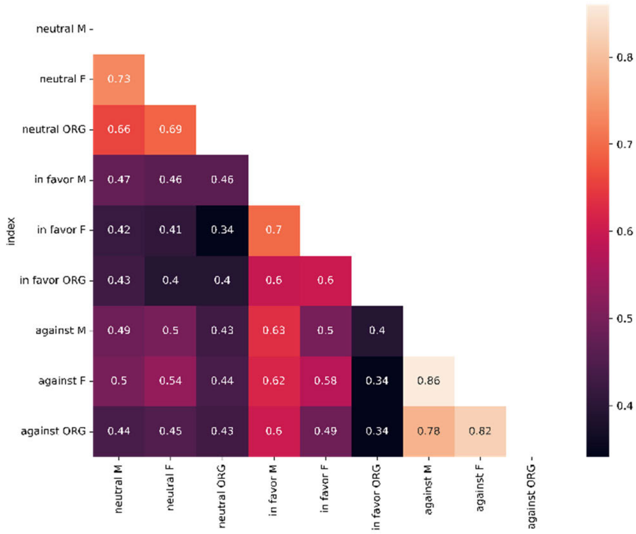


FIGURE 11. Correlogram showing Cramer's V-measure for correlation between n-grams used in each topic for E3.

TABLE 13. Event definition and identification.

Event number	Period	Events
E1	10 Feb 2020 – 20 Feb 2020	Start of the COVID crisis
E2	26 Oct 2020 – 5 Nov 2020	Development of certain vaccines (Pfizer, Moderna) in advanced stage; 2020 Presidential Election in the US;
E3	26 May 2021– 5 June 2021	Implementation of the COVID-19 digital pass in EU [70]; Several countries hit by the Delta variant;
E4 (E4.1, E4.2)	23 Feb 2022 – 5 June 2022	A worldwide spike in cases and COVID-19-related deaths (Figure 10); Moderna, Pfizer and BioNTech request authorization of their respective vaccines for used in children;
E_v	8 Dec 2020	The start of the first vaccination campaign in the UK.

Sources: Google News Search

within certain social media spaces focusing on the issue of vaccination. It is possible in this way that the legacy of how the COVID-19 pandemic was handled is in part responsible for a revitalization of the vaccine-hesitant community, which was provided with powerful new arguments and themes quite distinct from the pre-COVID-19 ones focusing on side-effects and autism [34]. It is crucial to critically reassess both the measures themselves and how they were implemented and presented to the public; otherwise, the public's already shaken trust in institutions will continue to decline. Continuing to track and study this emergent social media discourse might be a good first step to identify what concerns are most relevant.

VII. LIMITATIONS OF THE STUDY

The gender distribution we have observed appears to be different from the actual estimated distribution. We believe this is in part an artifact of the name parsing method we have used. We have seen evidence in the data, especially in the distribution of the genders when compared to the results from polling, that the NER parser performed a sub-optimal identification, causing the final results to be skewed towards the male gender. Improvements can be made to the parsing of personal names. The weighting of the names was also done using an intuitive heuristic; more principled approaches would surely improve the quality of the results. Machine learning techniques adapted to the domain data can be applied in both these areas as well to improve performance.

Perhaps the greatest limitation of our study is the fact that the task of gender identification obviously and necessarily involves a reductionist approach to gender, and as such we could not hope to include the whole breadth of what is otherwise a very complex aspect of our social reality. Notably, its treatment of gender identification as a binary classification problem reduces gender to the binary opposition rooted in language and biological sex, moreover in our case based on its tenuous relationship to human onomastics. Thus, it should be evident that any of our classifiers will utterly fail to capture important nuances such as gender expression and identity. It is clear then that any reasoning based on the conclusions of our study must take great care to avoid the ecological fallacy, that is, to extrapolate the results we observe in aggregate to any individual author of any individual tweet.

Finally, it must also be noted that the identification of gender deception is beyond the scope of this paper, as is any treatment of gender expression or identity. Cases in which the reported gender differs from onomastic gender, such as in the case of un-gendered names, are also by necessity overlooked. The question of the actual underlying demographics [68] is also a relevant one, as has been highlighted previously in similar computational approaches [13], [14], [40]. Data quality and availability is also a challenge – as discussed within the paper, only about one third of the users could be assigned a gender based on their name with some certainty, as the rest consisted of organizational or pseudonymous users. These are to be considered open research questions, to be included in future treatments of this complex and intriguing issue.

VIII. CONCLUSION

We have succeeded in using social media data to develop computational tools that are able to automatically identify social media users as female or male based on their public information. Likewise, we have developed improved tools to identify social media texts' stance towards vaccination. Combining these two approaches, we have not distinguished any differences between users identified as female and those identified as male with regards to their stance towards vaccination, but we did find small differences between the contents and textual characteristics of their tweets. We have thus found support for the theory that stance towards vaccination as expressed in social media discourse on Twitter is independent of gender.

REFERENCES

[1] *Coronavirus Disease (COVID-19)*. Accessed: Dec. 15, 2022. [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19

[2] L.-A. Cotfas, C. Delcea, I. Roxin, C. Ioanăş, D. S. Gherai, and F. Tajariol, "The longest month: Analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement," *IEEE Access*, vol. 9, pp. 33203–33223, 2021, doi: 10.1109/ACCESS.2021.3059821.

[3] Twitter. (Feb. 10, 2022). *Twitter Announces Fourth Quarter and Fiscal Year 2021 Results*. Accessed: Dec. 13, 2022. [Online]. Available: https://s22.q4cdn.com/826641620/files/doc_financials/2021/q4/Final-Q4'21-earnings-release.pdf

[4] M. P. Cristescu, R. A. Nerisanu, D. A. Mara, and S.-V. Oprea, "Using market news sentiment analysis for stock market prediction," *Mathematics*, vol. 10, no. 22, p. 4255, Nov. 2022, doi: 10.3390/math10224255.

[5] Statista. *Most Popular Social Media Apps in U.S.* Accessed: Feb. 26, 2022. [Online]. Available: https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/

[6] K. Drake. (Oct. 6, 2021). *Pandemic Mental Health: The Role of Social Media*. Accessed: Dec. 27, 2022. [Online]. Available: https://www.medicalnewstoday.com/articles/how-has-social-media-affected-mental-health-during-the-pandemic

[7] C. A. Melton, O. A. Olusanya, N. Ammar, and A. Shaban-Nejad, "Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the reddit social media platform: A call to action for strengthening vaccine confidence," *J. Infection Public Health*, vol. 14, no. 10, pp. 1505–1512, Oct. 2021, doi: 10.1016/j.jiph.2021.08.010.

[8] F. Cascini, A. Pantovic, Y. A. Al-Ajlouni, G. Failla, V. Puleo, A. Melnyk, A. Lontano, and W. Ricciardi, "Social media and attitudes towards a COVID-19 vaccination: A systematic review of the literature," *eClinicalMedicine*, vol. 48, Jun. 2022, Art. no. 101454, doi: 10.1016/j.eclinm.2022.101454.

[9] Reuters. (Jan. 1, 2021). *Fact check: Nurse Who Fainted After COVID-19 Vaccine Did Not Die*. Accessed: Sep. 29, 2022. [Online]. Available: https://www.reuters.com/article/uk-factcheck-nurse-covid-vaccine-dead-idUSKBN29629G

[10] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, and L. Derczynski, "SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours," in *Proc. 13th Int. Workshop Semantic Eval.*, Minneapolis, MN, USA, 2019, pp. 845–854, doi: 10.18653/v1/S19-2147.

[11] A. Abilov, Y. Hua, H. Matatov, O. Amir, and M. Naaman, "VoterFraud2020: A multi-modal dataset of election fraud claims on Twitter," Apr. 2021, *arXiv:2101.08210*. Accessed: Jun. 3, 2022.

[12] M. Cinelli, G. Etta, M. Avalle, A. Quattrociocchi, N. Di Marco, C. Valensise, A. Galeazzi, and W. Quattrociocchi, "Conspiracy theories and social media platforms," *Current Opinion Psychol.*, vol. 47, Oct. 2022, Art. no. 101407, doi: 10.1016/j.copsyc.2022.101407.

[13] I. Price, J. Gifford-Moore, J. Fleming, S. Musker, M. Roichman, G. Sylvain, N. Thain, L. Dixon, and J. Sorensen, "Six attributes of unhealthy conversation," Oct. 2020, *arXiv:2010.07410*. Accessed: Mar. 14, 2022.

[14] E.-R. Kovacs, L.-A. Cotfas, and C. Delcea, "From unhealthy online conversation to political violence: The case of the January 6^{th} events at the capitol," in *Advances in Computational Collective Intelligence*. Cham, Switzerland, 2022, pp. 3–15, doi: 10.1007/978-3-031-16210-7_1.

[15] L.-A. Cotfas, C. Delcea, and R. Gherai, "COVID-19 vaccine hesitancy in the month following the start of the vaccination process," *Int. J. Environ. Res. Public Health*, vol. 18, no. 19, p. 10438, Oct. 2021, doi: 10.3390/ijerph181910438.

[16] G. Dobriţa, A. Bara, S.-V. Oprea, C. Baroiu, and D.-C. Barbu, "Mobility, COVID-19 cases and virus reproduction rate data analysis for Romania using machine learning algorithms," in *Proc. 26th Int. Conf. Syst. Theory, Control Comput. (ICSTCC)*, Oct. 2022, pp. 244–251, doi: 10.1109/ICSTCC55426.2022.9931806.

[17] *Why a Gender Lens is Needed for the COVID-19 Response*. Accessed: Dec. 27, 2022. [Online]. Available: https://www.gavi.org/vaccineswork/why-gender-lens-needed-covid19-response

[18] L. S. Flor et al., "Quantifying the effects of the COVID-19 pandemic on gender equality on health, social, and economic indicators: A comprehensive review of data from March, 2020, to September, 2021," *Lancet*, vol. 399, no. 10344, pp. 2381–2397, Jun. 2022, doi: 10.1016/S0140-6736(22)00008-3.

[19] A. Vloo, R. J. M. Alessie, J. O. Mierau, M. H. Boezen, J. O. Mierau, L. Franke, J. Dekens, P. Deelen, P. Lanting, J. M. Vonk, I. Nolte, A. P. S. Ori, A. Claringbould, F. Boulogne, M. X. L. Dijkema, H. H. Wiersma, R. Warmerdam, and S. A. Jankipersadsing, "Gender differences in the mental health impact of the COVID-19 lockdown: Longitudinal evidence from The Netherlands," *Population Health*, vol. 15, Sep. 2021, Art. no. 100878, doi: 10.1016/j.ssmph.2021.100878.

[20] *COVID-19 Public Health Emergency of International Concern (PHEIC) Global Research and innovation Forum*. Accessed: Dec. 17, 2022. [Online]. Available: https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum

[21] E. Sezerer, O. Polatbilek, and S. Tekir, "A Turkish dataset for gender identification of Twitter users," in *Proc. 13th Linguistic Annotation Workshop*, Florence, Italy, Aug. 2019, pp. 203–207, doi: 10.18653/v1/W19-4023.

[22] J. Soler and L. Wanner, "A semi-supervised approach for gender identification," in *Proc. 10th Int. Conf. Lang. Resour. Eval. (LREC)*, Portorož, Slovenia, May 2016, pp. 1282–1287. Accessed: Dec. 12, 2022. [Online]. Available: https://aclanthology.org/L16-1204

[23] M. Bartl and S. Leavy, "Inferring gender: A scalable methodology for gender detection with online lexical databases," in *Proc. 2nd Workshop Lang. Technol. for Equality, Diversity Inclusion*, Dublin, Ireland, May 2022, pp. 47–58, doi: 10.18653/v1/2022.ltedi-1.7.

[24] N. Cheng, R. Chandramouli, and K. P. Subbalakshmi, "Author gender identification from text," *Digit. Invest.*, vol. 8, no. 1, pp. 78–88, Jul. 2011, doi: 10.1016/j.diin.2011.04.002.

[25] A. S. Slater and S. Feinman, "Gender and the phonology of North American first names," *Sex Roles*, vol. 13, nos. 7–8, pp. 429–440, Oct. 1985, doi: 10.1007/BF00287953.

[26] "Law restricts parents from naming their children as they please—Portugal," Portuguese Amer. J., Apr. 2011. Accessed: Dec. 13, 2022. [Online]. Available: https://portuguese-american-journal.com/law-restricts-parents-from-naming-their-children-as-they-please-portugal/

[27] A. Hungary. (May 8, 2019). *Hungarians Can Now Name Their Child After Daenerys From GoT*. Accessed: Dec. 13, 2022. [Online]. Available: https://abouthungary.hu//news-in-brief/hungarians-can-now-name-their-child-after-daenerys-from-got

[28] J. van de Weijer, G. Ren, J. van de Weijer, W. Wei, and Y. Wang, "Gender identification in Chinese names," *Lingua*, vol. 234, Jan. 2020, Art. no. 102759, doi: 10.1016/j.lingua.2019.102759.

[29] D. Offord, *Using Russian: A Guide to Contemporary Usage*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[30] V. Nastase and M. Popescu, "What's in a name? In some languages, grammatical gender," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Singapore, Aug. 2009, pp. 1368–1377. Accessed: Dec. 13, 2022. [Online]. Available: https://aclanthology.org/D09-1142

[31] S. Moorthy, R. Pogacar, S. Khan, and Y. Xu, "Is Nike female? Exploring the role of sound symbolism in predicting brand name gender," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Brussels, Belgium, Oct. 2018, pp. 1128–1132, doi: 10.18653/v1/D18-1142.

[32] S. Malmasi, "A data-driven approach to studying given names and their gender and ethnicity associations," in *Proc. Australas. Lang. Technol. Assoc. Workshop*, Melbourne, VIC, Australia, Nov. 2014, pp. 145–149. Accessed: Dec. 12, 2022. [Online]. Available: https://aclanthology.org/U14-1021

[33] A. ALDayel and W. Magdy, "Stance detection on social media: State of the art and trends," *Inf. Process. Manage.*, vol. 58, no. 4, Jul. 2021, Art. no. 102597, doi: 10.1016/j.ipm.2021.102597.

[34] E. D'Andrea, P. Ducange, A. Bechini, A. Renda, and F. Marcelloni, "Monitoring the public opinion about the vaccination topic from tweets analysis," *Expert Syst. Appl.*, vol. 116, pp. 209–226, Feb. 2019, doi: 10.1016/j.eswa.2018.09.009.

[35] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, May 2014, vol. 8, no. 1, pp. 216–225, doi: 10.1609/icwsm.v8i1.14550.

[36] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," Aug. 2013, *arXiv:1308.6297*. Accessed: Feb. 26, 2022.

[37] A. Borg and M. Boldt, "Using VADER sentiment and SVM for predicting customer response sentiment," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113746, doi: 10.1016/j.eswa.2020.113746.

[38] R. Wang, D. Zhou, M. Jiang, J. Si, and Y. Yang, "A survey on opinion mining: From stance to product aspect," *IEEE Access*, vol. 7, pp. 41101–41124, 2019, doi: 10.1109/ACCESS.2019.2906754.

[39] Z. Lyu and H. Takikawa, "Media framing and expression of anti-China sentiment in COVID-19-related news discourse: An analysis using deep learning methods," *Heliyon*, vol. 8, no. 8, Aug. 2022, Art. no. e10419, doi: 10.1016/j.heliyon.2022.e10419.

[40] E.-R. Kovacs, L.-A. Cotfas, and C. Delcea, "COVID-19 vaccination opinions in education-related tweets," in *Eurasian Business and Economics Perspectives*. Cham, Switzerland, 2022, pp. 21–41, doi: 10.1007/978-3-031-15531-4_2.

[41] Z. Zhang, G. Feng, J. Xu, Y. Zhang, J. Li, J. Huang, B. Akinwunmi, C. J. P. Zhang, and W.-K. Ming, "The impact of public health events on COVID-19 vaccine hesitancy on Chinese social media: National infoveillance study," *JMIR Public Health Surveill.*, vol. 7, no. 11, Nov. 2021, Art. no. e32936, doi: 10.2196/32936.

[42] Q. Liu, W.-S. Lee, M. Huang, and Q. Wu, "Synergy between stock prices and investor sentiment in social media," *Borsa Istanbul Rev.*, vol. 23, no. 1, pp. 76–92, Jan. 2023, doi: 10.1016/j.bir.2022.09.006.

[43] S. Shen, L. Xia, Y. Shuai, and D. Gao, "Measuring news media sentiment using big data for Chinese stock markets," *Pacific-Basin Finance J.*, vol. 74, Sep. 2022, Art. no. 101810, doi: 10.1016/j.pacfin.2022.101810.

[44] C. Qian, N. Mathur, N. H. Zakaria, R. Arora, V. Gupta, and M. Ali, "Understanding public opinions on social media for financial sentiment analysis using AI-based techniques," *Inf. Process. Manage.*, vol. 59, no. 6, Nov. 2022, Art. no. 103098, doi: 10.1016/j.ipm.2022.103098.

[45] N. Kyriazis, S. Papadamou, P. Tzeremes, and S. Corbet, "The differential influence of social media sentiment on cryptocurrency returns and volatility during COVID-19," *Quart. Rev. Econ. Finance*, Mar. 2023, doi: 10.1016/j.qref.2022.09.004.

[46] G. Andrade, "Predictive demographic factors of COVID-19 vaccine hesitancy in Venezuela: A cross-sectional study," *Vacunas*, vol. 23, pp. S22–S25, Aug. 2022, doi: 10.1016/j.vacune.2022.08.002.

[47] E. Lasher, G. Fulkerson, E. Seale, A. Thomas, and A. Gadomski, "COVID-19 vaccine hesitancy and political ideation among college students in central New York: The influence of differential media choice," *Preventive Med. Rep.*, vol. 27, Jun. 2022, Art. no. 101810, doi: 10.1016/j.pmedr.2022.101810.

[48] K. A. Schmidtke, L. Kudrna, A. Noufaily, N. Stallard, M. Skrybant, S. Russell, and A. Clarke, "Evaluating the relationship between moral values and vaccine hesitancy in great Britain during the COVID-19 pandemic: A cross-sectional survey," *Social Sci. Med.*, vol. 308, Sep. 2022, Art. no. 115218, doi: 10.1016/j.socscimed.2022.115218.

[49] D. Casale and D. Posel, "Gender inequality and the COVID-19 crisis: Evidence from a large national survey during South Africa's lockdown," *Res. Social Stratification Mobility*, vol. 71, Feb. 2021, Art. no. 100569, doi: 10.1016/j.rssm.2020.100569.

[50] M. T. J. Ansari and N. A. Khan, "Worldwide COVID-19 vaccines sentiment analysis through Twitter content," *Electron. J. Gen. Med.*, vol. 18, no. 6, Nov. 2021, Art. no. em329, doi: 10.29333/ejgm/11316.

[51] P. Suhem. (Feb. 9, 2022). Introducing the sort_order parameter for search endpoints in the Twitter API V2—Announcements. Twitter Developers. Accessed: Feb. 26, 2022. [Online]. Available: https://twittercommunity.com/t/introducing-the-sort-order-parameter-for-search-endpoints-in-the-twitter-api-v2/166377/2

[52] A. Sboev, I. Moloshnikov, D. Gudovskikh, A. Selivanov, R. Rybka, and T. Litvinova, "Automatic gender identification of author of Russian text by machine learning and neural net algorithms in case of gender deception," *Proc. Comput. Sci.*, vol. 123, pp. 417–423, Jan. 2018, doi: 10.1016/j.procs.2018.01.064.

[53] Social Security Administration. *Baby Names From Social Security Card Applications—National Data*. Accessed: Dec. 16, 2022. [Online]. Available: https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-national-data

[54] A. Rao, "UCI machine learning repository," School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, Tech. Rep., 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[55] S. Vajjala, B. Majumder, A. Gupta, and H. Surana, *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. Sebastopol, CA, USA: O'Reilly Media, 2020.

[56] F. de Saussure and W. Baskin, *Course in General Linguistics*. New York, NY, USA: Columbia Univ. Press, 2011.

[57] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543, doi: 10.3115/v1/D14-1162.

[58] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," Presented at the Int. Conf. Learn. Represent., Jan. 2013. Accessed: Dec. 15, 2022. [Online]. Available: https://www.semanticscholar.org/paper/Efficient-Estimation-of-Word-Representations-in-Mikolov-Chen/330da625c15427c6e42ccfa3b747fb29e5835bf0

[59] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[60] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," Jul. 2019, *arXiv:1907.11692*. Accessed: Feb. 26, 2022.

[61] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," May 2018, *arXiv:1810.04805*. Accessed: Feb. 26, 2022.

[62] R. Wolfe and A. Caliskan, "Low frequency names exhibit bias and overfitting in contextualizing language models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Punta Cana, Dominican Republic, Nov. 2021, pp. 518–532, doi: 10.18653/v1/2021.emnlp-main.41.

[63] C. P. George and H. Doss, "Principled selection of hyperparameters in the latent Dirichlet allocation model," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 5937–5974, 2018. Accessed: Apr. 10, 2022. [Online]. Available: http://jmlr.org/papers/v18/15-595.html

[64] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "SemEval-2016 task 6: Detecting stance in tweets," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, San Diego, CA, USA, Jun. 2016, pp. 31–41, doi: 10.18653/v1/S16-1003.

[65] J. D. Kelleher, B. M. Namee, and A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, 2nd ed. Cambridge, MA, USA: MIT Press, 2020.

[66] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[67] *Gender Screening on Question-Answering Communities | Elsevier Enhanced Reader*. Accessed: Dec. 14, 2022. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S0957417422024241?token=B5DD4B80E93BA9747A7269939F6E17B33E0E04EEC4DF1E1641AB4628BE393D424E6826683727740858ED99CE13C8FFDB&originRegion=eu-west-1&originCreation=20221214211941

[68] Statista. *Global Twitter User Distribution by Gender 2022*. Accessed: Dec. 16, 2022. [Online]. Available: https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/

[69] A. M. Liebetrau, *Measures of Association*. Newbury Park, CA, USA: Sage, 1983.

[70] CNN. *EU Digital COVID Certificate: Everything You Need to Know*. Accessed: Dec. 27, 2022. [Online]. Available: https://www.cnn.com/travel/article/eu-covid-certificate-travel-explainer/index.html

ERIK-ROBERT KOVACS received the M.Sc. degree in economic informatics from the Bucharest University of Economic Studies, Bucharest, Romania, where he is currently pursuing the Ph.D. degree in economic informatics, with a research project focusing on natural language processing using machine learning and deep learning techniques with applications in the social science.

He is also a Software Developer with Endava România. He is also an Adjunct Associate Professor with the Economic Cybernetics and Informatics Department, Bucharest University of Economic Studies.

LIVIU-ADRIAN COTFAS received the Ph.D. degree in economic informatics from the Bucharest University of Economic Studies, Bucharest, Romania.

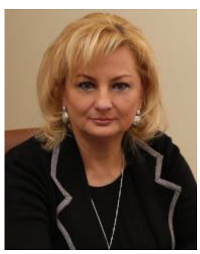
He was an Invited Professor with the University of Technology Belfort-Montbéliard, France, and the University of Franche-Comté, France. He is currently a Professor with the Department of Economic Cybernetics and Informatics, Bucharest University of Economic Studies. His research interests include machine learning and deep learning techniques, natural language processing, social media analysis, agent-based modeling, recommender systems, semantic web, and grey systems theory.

Dr. Cotfas has received several research awards, including the Georgescu Roegen Award for excellent scientific research. His research results were mentioned in popular media, including the *WIRED* magazine and *The New York Times* newspaper.

CAMELIA DELCEA received the Ph.D. degree in economic cybernetics and statistics from the Bucharest University of Economic Studies, Bucharest, Romania.

She is currently with the Economic Cybernetics and Informatics Department, Bucharest University of Economic Studies. Her research interests include agent-based modeling, operations research (optimizing the airplane boarding methods and improving the evacuation process), grey systems theory, artificial intelligence systems, companies' financial and non-financial analysis, risk management, nonlinear and dynamic systems, consumer's behavior, online social networks, and sentiment analysis.

Dr. Delcea is an Active Member of the Grey Uncertainty Analysis Association. She received the 19 international and national awards, including the Best Paper Award, the Georgescu Roegen for Excellent Scientific Research Award, the Excellent Paper Award, and the Top Reviewer Award. She was invited to deliver a keynote speech on grey systems themes at the IEEE GSIS Conference, in 2013, 2016, and 2017; and the GSUA Conference, in 2018. She is a member of the Editorial Advisory Board of the *Grey Systems* journal (Emerald).



MARGARETA-STELA FLORESCU is a Full Professor with the Department of Administration and Public Management, Bucharest University of Economic Studies, where she is also the Director of the Institute for Advanced Research, with over 25 years of experience in management of research projects and strategic management, collaborating with various international institutions in the field of education and research. She has published over 14 books, 127 scientific papers held at various prestigious symposia and international conferences, 133 scientific articles published in specialized international academic journals, 75 scientific research projects funded by international and national research programs, seven patents, and 55 articles in ISI listed journals. She received seven international awards.