How Did Europe's Press Cover Covid-19 Vaccination News? A Five-Country Analysis

David Alonso del Barrio
ddbarrio@idiap.ch
Idiap Research Institute
Switzerland

Daniel Gatica-Perez
gatica@idiap.ch
Idiap Research Institute and EPFL
Switzerland

ABSTRACT

Understanding how high-quality newspapers present and discuss major news plays a role towards tackling disinformation, as it contributes to the characterization of the full ecosystem in which information circulates. In this paper, we present an analysis of how the European press treated the Covid-19 vaccination issue in 2020-2021. We first collected a dataset of over 50,000 online articles published by 19 newspapers from five European countries over 22 months. Then, we performed analyses on headlines and full articles with natural language processing tools, including named entity recognition, topic modeling, and sentiment analysis, to identify main actors, subtopics, and tone, and to compare trends across countries. The results show several consistencies across countries and subtopics (e.g. a prevalence of neutral tone and relatively more negative sentiment for non-neutral articles, with few exceptions like the case of vaccine brands), but also differences (e.g., distinctly high negative-to-positive ratios for the no-vax subtopic.) Overall, our work provides a point of comparison to other news sources on a topic where disinformation and misinformation have resulted in increased risks and negative outcomes for people's health.

CCS CONCEPTS

• **Computing methodologies → Natural language processing**; • **Information systems → Data mining**.

KEYWORDS

Covid-19 vaccination, disinformation, European news, topic modeling, sentiment analysis

ACM Reference Format:
David Alonso del Barrio and Daniel Gatica-Perez. 2022. How Did Europe's Press Cover Covid-19 Vaccination News? A Five-Country Analysis. In *Proceedings of the 1st International Workshop on Multimedia AI against Disinformation (MAD '22), June 27–30, 2022, Newark, NJ, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3512732.3533588

1 INTRODUCTION

Disinformation has rapidly grown as a topic of relevance given its negative effects on society, ranging from politics to health. A huge body of research has been published [29], focusing on information of questionable quality (whether for nefarious purposes or not) and how to deal with it [24] [41]. Specifically in the context of the Covid-19 pandemic, the World Health Organization coined the term infodemic, i.e., "the overexposure of information including false or misleading information in digital and physical environments during a disease outbreak" [40], and significant research on misinformation has been conducted since the beginning of the pandemic, covering information on various issues such as the origin of the coronavirus or the content of vaccines.

While the existing research sheds light on many key aspects of Covid-19 disinformation and how to tackle it, researchers have also emphasized the need of understanding the full information ecosystem [2] [35]. This holistic understanding needs to include the examination of existing high-quality media outlets, which have played a key role during the pandemic period [18] [23] [39]. Comparably less research in computing has been conducted to systematically understand how European media organizations have covered aspects of the pandemic [34], specifically Covid-19 vaccination [10], partly due to the lack of curated multi-country data. For instance, existing research has analyzed the sources used by journalists to report on health news in countries like Spain [7] or Portugal [16]. Other research relates to the analysis of news about Covid-19 vaccination shared on Facebook or Twitter in Italy [31]. These works are also motivated by the relevance of studying how mainstream media is informing the population about vaccination .

Our work aims to address this research gap by posing the following research questions:

RQ1: How can we systematically study the perceptions of Europe's press about Covid-19 vaccination as a multi-faceted subject, given the variety of countries, languages, and newspapers?

RQ2: Can we recover the main perceptions about Covid-19 vaccination by Europe's press through a combination of linguistic analyses? Are these perceptions consistent, or are there any differences across countries?

By addressing the above research questions, our work makes the following contributions:

Contribution 1. We designed and collected a dataset of over 50,000 articles on Covid-19 vaccination from 19 newspapers, 5 European countries, and 4 languages over a period of 22 months. This represents a unique resource to study the subject of interest. We also generated a translated version with all the content in English to facilitate comparisons across countries.

Contribution 2. We conducted our analysis based on a pipeline that included named entity recognition, subtopic modeling, and sentiment analysis using state-of-art tools. We first identified country-specific named entities such as government representatives. We

MAD '22, June 27–30, 2022, Newark, NJ, USA

ACM ISBN 978-1-4503-9242-6/22/06…$15.00
https://doi.org/10.1145/3512732.3533588

then used topic modelling techniques to study subtopics within the Covid-19 vaccination subject, such as its connection to the European Union, the economy, or the no-vax movement. Finally, we performed sentiment analysis at three levels (article, headline, and sentence) to uncover the tone of the European newspapers about Covid-19 vaccination, both in general and regarding key subtopics, finding trends that hold across subtopics and countries but also particularities regarding subtopics and countries. Overall, we believe that linguistic analyses of high-quality press in Europe can contribute to inform the design of tools against disinformation, as a benchmark of what constitutes good information.

The paper is organized as follows. In Section 2, we discuss related work. In Section 3, we present our dataset. In Section 4, we describe the methodology used to analyze European news data. We present and discuss results in Section 5. Finally, we provide conclusions in Section 6 .

2 RELATED WORK

With the advent of Covid-19, many lines of research have been developed in the field of text processing. Focusing on Covid-19 news articles, Ghasiya et al.[11] proposed an analysis of articles from four different countries: UK, India, Japan, South Korea, extracting subthemes and making a sentiment analysis of the headlines. De Melo et al.[9] worked with articles and tweets in Brazil on Covid-19, and performed various analyses such as sub-theme extension with Latent Dirichlet Allocation (LDA)[6], sentiment analysis with VADER [15], previously translated content into English, or Named Entity Recognition (NER) [37] with Spacy. In the work of Aslam et al.[4] the headlines were classified into positive, negative, and neutral sentiments after the calculation of text unbounded polarity at the sentence level score, and incorporating valence shifters. In addition, the National Research Council Canada (NRC) Word-Emotion Lexicon was used to calculate the presence of eight emotions at their emotional weight. The results revealed that the news headlines had high emotional scores with a negative polarity.

There is also abundant literature where the input sources are tweets about Covid-19. As one example, K.Rahul et al.[26] analyzed Covid-19 Vaccine related tweets to study the spread of the virus and the increasing number of daily cases. The method used topic modeling (LDA) to determine popular themes, and VADER and TextBlobfor sentiment analysis [17]. As a second example, Cotfas et al.[8] analyzed the dynamics of opinions regarding Covid-19 vaccination, by considering the one-month period following the first vaccine announcement, until the first vaccination took place in UK, in which civil society manifested a higher interest regarding the vaccination process. As a third example, Ghasiya et al.[21] collected tweet posts by UK and US residents during the pandemic, and designed experiments to answer three questions concerning vaccination. To get the dominant public sentiment, they performed sentiment analysis by VADER.

On a third line of work, social scientists have studied the psychology of news, focusing on why there is a often a feeling that news tend to be more negative than positive, when looking at the news on TV or in the written press [32][33]. Additional work has been conducted on the emotional component of fake news and news in general [27][20].

3 DATA

We constructed from scratch a dataset of European news articles about Covid-19 vaccination. As a first step, we contacted over 30 European newspapers spanning five countries, requesting authorization to extract and analyze articles discussing issues related to Covid-19 vaccination. We obtained the authorization of 19 of them. This includes six newspapers from France (FR) (*Le Monde, Les Echos, Ouest France, La Croix, Liberation, Lyon Capitale*); two from Italy (IT) (*Il Corriere della Sera, Il Sole 24 Ore*); six from Spain (ES) (*ABC, El Español, La Vanguardia, El Mundo, El Diario, 20 Minutos*), three from Switzerland (CH) (*24 Heures, Le Temps, La Liberté*), and two from the United Kingdom (UK) (*The Telegraph, The Irish News*). As the list shows, this set of newspapers includes a combination of major European newspapers (with national and international circulation) and of regional newspapers (with more limited circulation.) Some of the newspapers in our dataset have very wide audiences. For instance, *Le Monde* had a average daily circulation of 393,000 in 2020 and is among the top 50 websites in France in terms of visits [36]. After receiving written authorization from each newspaper, we extracted the articles using Selenium and Beautiful Soup [28] two well know Python scraping libraries. For each article, we extracted the headline, subheadline (if available), main text, authors, date of publication, and link.

The dataset of articles was designed to be bounded with respect to time. For this, we compiled all articles on Covid-19 vaccination for the 22-month period between 01.01.2020 and 31.10.2021. This period resulted in a total of 51320 articles. The total number of articles per country can be seen in Table 1. After normalization, the average number of articles per newspaper varies from 2040 for Swiss newspapers to 3582 for the Italian ones, with an overall average of 2701.0 for the studied 22-month period, this corresponds to an average of 122.7 articles per newspaper per month, or about four articles per newspaper per day. This also corresponds to an average of 27 sentences per article (note, however, that the variance for these estimates is large.)

The dataset is multilingual, since it contains articles in English, French, Spanish, and Italian, according to the origin of the extracted news. We translated both headlines and main text of the whole dataset into English using DeepL [1], to be able to work with both the multilingual and the English datasets. We have considered the translation to a common language as a necessary task, since it allows the comparison of results across countries and opens more analysis options, e.g., in terms of availability of pre-trained models.

Table 1: Characteristics of the dataset per country: number of newspapers, number of articles, number of sentences, and average number of articles per newspaper

Country	#Newspapers	#Articles	#Sentences	#Articles/Newspaper
FR	6	14513	412077	2418.8
IT	2	7164	210388	3582.0
ES	6	18748	417200	3124.6
CH	3	6121	219554	2040.3
UK	2	4774	134060	2387.0
TOTAL	19	51320	1393279	2701.0

[1]https://www.deepl.com/es/pro-api?cta=header-pro-api/

4 METHODOLOGY

With our curated dataset of European news about Covid-19 vaccination, we apply a pipeline of analytical methods to discover key content trends. The methodology includes named entity recognition, fine-grained topic analysis, and sentiment analysis about key subtopics. Each step is described in the rest of the section.

4.1 Named Entity Recognition

Named Entity Recognition (NER), also known as entity extraction, is an information extraction task that seeks to locate and classify named entities into predefined categories, such as people, organisations, places, time expressions, and quantities. The aim of this analysis was to identify individuals who were the most frequently named, such as presidents and health ministers, as well as the most often named organisations including the different vaccine brands, or the cities and countries that appeared most frequently in the articles of each country. This provides a first insight into the concept of local news and the similarities and differences between countries when referring to entities. For this task, we used the Spacy library, [13] which is a free, open-source library for advanced Natural Language Processing (NLP) in Python. This was applied on the original multilingual dataset.

4.2 Fine-grained topic modeling

Next, we aimed to know if any subtopics within the articles could be identified. While the main article was vaccination, we wanted to extract possible subtopics covered in conjunction with the main topic. Looking at the related work, alternatives included using a classical LDA approach, or using embedding vectors and clustering techniques with algorithms like Top2Vec [3] and BERTopic [12]. The advantage of models based on document embeddings is that, unlike LDA, we do not have to define a priori the number of subtopics, nor do any preprocessing.

BERTopic for subtopic exploration. In a first approximation, we used BERTopic and Top2Vec applied to the English version of the dataset. In particular, BERTopic is a powerful tool to visualize graphs of the results and facilitate their interpretation. BERTopic creates an embedding for each article, then reduces the embedding size and does clustering of similar embeddings. Finally, it uses TF-IDF to extract the most important words from each cluster. Empirically, we considered that BERTopic had better performance than Top2Vec.

Despite seeing better results in BERTopic, there were subtopics that were defined by very common words such as articles or prepositions, so we decided to eliminate stopwords and words such as 'vaccine' or 'Covid-19', as we wanted to extract the subtopics within the Covid-19 vaccination theme, and to keep only nouns and proper nouns that appeared in the articles, so that the input text for BERTopic was a list of words for each article, rather than the full article. An illustration of the results generated by BERTopic appears in Table 2. This exploration is used to define specific subtopic 'queries' for further analysis.

Query extraction for further analysis. Having a list of words that define a subtopic, we can establish subsets of news articles from a filter using queries. More specifically, from the original corpus with all articles for a given country, we can define a query using 3-4 words that define a subtopic, and select all the articles that contain, either in the headline or in the main text, any of the words in that query, thus defining a subset of articles that belong to a subtopic. For example, we can define a subtopic to be the European Union, and this subtopic is defined by the words. 'EU', 'European Union', 'Eurozone', 'European Commission'; in this way we have a set of articles that deal with the European Union, and similarly for other subtopics.

Another approach we considered when defining the subsets of articles was to apply the query only to the headlines, that is, to select only those articles that contained some of the keywords in the headlines. In this way, we could ensure that the articles that formed a subset were focused on that subtopic, since by applying the query to the headline and the text, we select articles that may only have a single sentence or two with the keywords in our subset. One disadvantage of this approach is that it represents a very restrictive filter, and thus the number of samples per subtopic is considerably reduced, leaving out articles that deal with a subtopic although the headline might not directly express it. Imagine the anti-vaccine movement subtopic, which we define with words like 'no-vax', anti-vaccine', 'anti-vaccination', and an article that deals with anti-vaccination, but with headline: "How to convince your friend to get vaccinated?" Following the approach of applying the query only to the headline, we would be discarding this article, while based on its content, the article would indeed be about the subtopic and therefore would have to be included for analysis.

Finally, we decided on a less restrictive approach, in which we apply the queries to both the headline and the text. One final step before further analysis was to remove articles containing more than 3000 words. The rationale is the following. After manually examining our news corpus, we found that newspapers tend to write a type of article that summarizes the most important news of the week, in the form of a concatenation of summarized news. These articles tend to be very long, typically over 3000 words. The concatenation of weekly news turn these articles into a mixture of different subtopics, which is unsuitable for our further analysis. More specifically, concatenated articles may contain the keywords for a given subtopic, but any analysis at the document level, e.g. sentiment analysis, would not reflect the sentiment towards the subtopic of interest, but rather towards several subtopics. Given this, we excluded these articles for further analysis. Finally, when defining a topic, there have been cases of metonymy, for example the word 'Brussels' implies the European Union as an alternative way of designating it, with sentences such as: "Brussels drafts a "green digital passport", "Brussels approves Pfizer's vaccine".

Table 2: Sample of detected topics

Topic	Words per Topic
European Union	commission, eu, brussels, european union, europe, von der leyen, coordination, eurozone
Economy	gdp, bank,growth, economy,economist, spending, business, investment
Education	education, pupil, school, classroom, minister,learning, child, teacher, weir, executive
Olympics	olympics, tokyo, suga,yoshihide,japan,olympic, prefecture, emergency, osaka,games

4.3 Sentiment analysis

After defining subtopics, we proceeded to their sentiment analysis. We used three ways of representing the sentiment of newspaper articles:

1. Sentiment of Headline: Headlines have a key importance in a digital article, as they motivate readers to click on a news item. We extracted the percentage of positive, negative, and neutral headlines for each country.

2. Sentiment of Article: We analyzed the sentence-by-sentence sentiment for each article, and extracted the percentage of positive, negative, and neutral sentences for each of them.

3. Sentiment at Sentence-Level. We extracted the number of positive, negative, and neutral sentences containing any of the words that define a subtopic.

The difference between representations 2 (article) and 3 (sentence) lies in the perspective with which the sentiment of a piece of text is analyzed. The sentiment of article takes into account the article as a whole, i.e., all its sentences, even if not all of them have words that define a subtopic. In contrast, the sentiment at sentence-level only takes into account all those sentences that contain one of the keywords that define a subtopic, and allows to examine in greater depth the sentiment for those sentences that contain a keyword.

These representations are used for temporal and country-based analysis. For the first one, we wanted to understand how sentiment on a subtopic evolves over time, i.e., to examine if there are fluctuations or if the tone remains constant. For the second one, we conducted descriptive analysis of sentiment by country, i.e., we considered all articles belonging to each country and extracted the average percentage of positive, negative, and neutral sentences, as well as the percentage of positive, negative, and neutral headlines for each country. As part of this analysis, we also compute the *negative-to-positive ratio* of these variables, to indicate how many times the negative sentiment of a piece of news is larger (or smaller) than the positive sentiment, for each of the three representations (headline, article, sentence.)

To perform sentiment analysis, we used the Hugging Face transformers package [38] which is a popular Python library that provides pre-trained models that are useful for a variety of NLP tasks. Specifically, we used a model called BERTsent [25]: a fine-tuned, BERT-based sentiment classifier for English language tweets. BERTsent was trained with the SemEval 2017 corpus, and is based on bertweet-base, which was trained on 850M English Tweets and additional 23M Covid-19 English Tweets. We chose this model because of the content (Tweets about Covid-19), despite being trained only with English content. We would have liked to use a model trained on press articles, but did not find a suitable one. We consider that the approach of dividing the text into sentences and applying the model based on Tweets is a reasonable approach. Single sentences can be seen as being similar to Tweets, and in addition we avoid BERT's 512-token limitation, as we do not expect to have single sentences of such length. In contrast, if we were to use the whole article as input text, there would be many articles with more than 512 tokens, and the text would have to be truncated.

At first, we performed the sentiment analysis using a multilingual model [22], but the results did not seem appropriate since the multilingual model had not been trained with Covid-19-related content, and we observed that the inferred sentiment of sentences was often incorrect. Based on this, we decided to applied the BERTsent (English) model to the English translation of the full corpus, which allows to make a more realistic comparison across countries. Other popular options for sentiment analysis like VADER or TextBlob were also tested on UK articles, but the results were not considered to be sufficiently accurate in sentences with content about Covid-19, and were finally ruled out.

5 RESULTS AND DISCUSSION

In this section, we present the results obtained from each analysis and their corresponding discussion.

5.1 Named Entity Recognition

We extracted the most frequently named people, organizations, and places in the newspapers of each country, and we observe a repeated pattern in all countries. Among the most often named people are the top health officer of each country, the president of the government, and the president of the United States; in terms of organizations, the top entities are the European Union, the WHO, and the pharmaceutical companies that developed the main vaccines; in terms of places, national press names most frequently its own country, and several local cities. In Figure 1, we show the most frequent entities by country. Examining the results, it is interesting

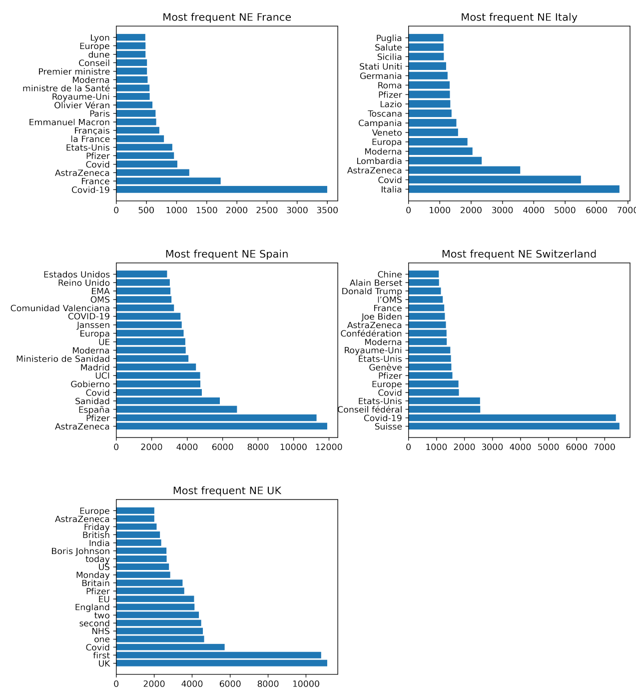

Figure 1: Top-20 most frequent entities by country. In each plot, x-axis: frequency; y-axis: ordered from the bottom up.

to observe the local component of each country. In Spain, cities such as Madrid, Barcelona, and Valencia are mainly mentioned, as well as presidents of autonomous communities and those responsible for public health in the country. In Switzerland, the Federal Council and cities like Geneva and Bern are named. In Italy, regions such as Veneto, Lombardy, Rome, government representatives, and ministers are mentioned. On the other hand, we also observe common entities across countries such as the European Union, the WHO, the

United States, or Russia. In qualitative terms, we are satisfied with the results obtained in terms of the accuracy of the NER tool, as we see the main organisations, people, and locations most frequently mentioned in each country, although some mistakes are made, such as considering 'AstraZeneca/Pfizer' or 'Covid-19' as people.

5.2 Overall sentiment analysis by country

The results obtained from a overall analysis by country is shown in Table 3. We observe that all countries have similar percentages of positive and negative content at article level (16-21% for negative, and 4-6% for positive), as well as headlines (18-25% for negative, and 2-4% for positive). The proportion of negative sentiment is 2.8 to 4.2 times higher than positive sentiment for articles, and 4.7 to 8.2 times higher for headlines.

Table 3: Generic sentiment analysis by country

ANALYSIS	FR	CH	IT	ES	UK
AVG. NEGATIVE ARTICLES (%)	18.4	21.0	17.7	16.5	20.1
AVG. POSITIVE ARTICLES (%)	4.4	6.4	6.2	4.8	6.9
NEGATIVE HEADLINES (%)	18	21.1	19.8	20.1	25.8
POSITIVE HEADLINES (%)	2.2	4.5	2.8	3.2	4.6

From the viewpoint of the central issue we investigate, the higher proportion of negative sentiment may seem surprising. All articles deal directly or indirectly with the issue of Covid-19 vaccination, an issue that, seen in perspective, could be regarded as positive: vaccines have saved countless lives and are the key to a "return to normality" [1]. What can explain this markedly negative tone (compared to the positive tone) in our corpus? Recently published works such as [30] state that news sources tend to give audiences what they want, as several previous studies affirm that negative news capture in a greater way the attention of people. This leads to another question: why do negative news impact us more? Plausible answers can be found in studies like [33] and [32], which claim that humans may be neurologically or psychologically predisposed to focus on negative information. The argument is rooted in an evolutionary explanation of how humans decide what to pay attention to. It is advantageous to prioritize negative information, the argument goes, because the potential costs of negative information outweigh the benefits of positive information. Consequently, people might be predisposed to focus on negative information.

Also noteworthy is the large amount of content classified as neutral (roughly 70-80%). According to [5], this is to be expected, as newspapers tend to seek a sense of objectivity in their articles (unless they are opinion pieces), and journalists often avoid using vocabulary that can be classified as positive or negative. This is different than having product reviews or tweets as input text. In those cases, the goal might be the expression of a personal opinion, rather than searching for objectivity, and thus subjective opinions might result in more positive/negative sentiment being present.

5.3 Sentiment analysis by subtopic

Table 4 shows the average percentage of negative tone at article, headline, and sentence level, and the ratio between the average percentage of negative and positive tone. Cases where the ratio is greater than 5 are marked in red, and cases where the positive tone is greater than the negative tone (ratio below 1) are marked in green. The results of Table 4 shows that there is a higher percentage of negative sentences than positive ones, as all ratios are greater than one; the same happens with the headlines in all the subtopics. In no case positive sentiment is larger than negative one.

As for the frequency with which the subtopics are treated, we can see that the values are very similar across countries for a given topic, with a few exceptions. On the other hand, the frequency of each subtopic can differ greatly (e.g., No-Vax vs. Economy.). More specifically:

- European Union: With frequency between 14-18%, we see that in the United Kingdom the sentiment is the most negative, especially if we look at headlines and specific sentences that talk about the EU, although the ratio is bigger in Italy and France at headline level.
- No-Vax: With a comparatively low frequency (1-7%), this is the subtopic with the highest percentage of negativity in all the analyses performed, shown by the high ratios at all levels.
- Economy: With frequency between 15-35%, we observe similar ratios at article level and headline level, suggesting that the headline similarly represents the sentiment of the main text.
- Education: With frequency between 13-29%, we observe very high ratios at headline level, while the ratios at sentence level and article level are similar.
- Olympics: This is the issue with the lowest frequency. We see that France at headline level and the UK at sentence level have a rather negative tone.

Table 4: Sentiment analysis by subtopic

SUBTOPIC		FR	CH	IT	ES	UK
EUROPEAN UNION	NEG(%);NEG/POS ARTICLE RATIO	16.3 ; 3.7	20.1 ; 3.65	15.7 ; 2.53	14.5 ; 2.74	20.3 ; 3.33
	NEG(%);NEG/POS HEADLINE RATIO	13.9 ; 6.3	17.9 ; 5.1	19.8 ; 7.0	14.5 ; 4.2	25.2 ; 5.6
	NEG(%);NEG/POS SENTENCE RATIO	9.8 ;2.65	13.0 ; 3.02	7.6 ; 1.38	10.4 ; 2.12	21.0 ; 8.08
	FREQUENCY (%)	17.70	18.70	14.02	16.17	18.41
NO VAX	NEG(%);NEG/POS ARTICLE RATIO	28.4 ; 8.3	29.9 ; 6.3	23.5 ; 5.6	29.9 ; 7.1	30.4 ; 5.2
	NEG(%);NEG/POS HEADLINE RATIO	32.6 ; 54.3	30.5 ; 14.5	37.1 ; 28.5	40.0 ; 12.5	49.1 ; 27.2
	NEG(%);NEG/POS SENTENCE RATIO	52.6 ; 87.6	66.6 ; 222.0	37.4 ; 28.7	59.7 ; 85.2	50.2 ; 167.3
	FREQUENCY (%)	3.7	3.8	7.6	1.5	3.7
ECONOMY	NEG(%);NEG/POS ARTICLE RATIO	18.6 ; 3.38	20.1 ; 2.79	14.3 ; 1.61	17.1 ; 2.44	21.3 ; 2.84
	NEG(%);NEG/POS HEADLINE RATIO	15.9 ; 4.4	17.4 ; 3.1	15.6; 3,06	18.3 ; 3.5	27.1 ; 3.8
	NEG(%);NEG/POS SENTENCE RATIO	14.3 ; 1.9	16.2 ; 1.8	11.4 ; 1.05	15.1 ; 1.5	17.8 ; 2.3
	FREQUENCY (%)	22.6	35.2	18.0	15.1	21.9
EDUCATION	NEG(%);NEG/POS ARTICLE RATIO	19.9 ; 4.5	22.1 ; 3.4	16.1 ; 2.78	16.9 ; 3.6	21.6 ; 3.22
	NEG(%);NEG/POS HEADLINE RATIO	18.6 ; 8.8	20.2 ; 5.1	19.0 ; 7.9	20.4 ; 6.8	24.0 ; 4.3
	NEG(%);NEG/POS SENTENCE RATIO	16.4 ;3.7	17.3 ; 2.7	13.4 ; 2.0	12.3 ; 2.0	17.4 ; 2.8
	FREQUENCY (%)	23.4	22.2	29.6	13.1	20.0
OLYMPICS	NEG(%);NEG/POS ARTICLE RATIO	20.1 ; 4.1	20.1 ; 3.0	17.7 ; 2.1	17.0 ; 2.3	22.8 ; 3.8
	NEG(%);NEG/POS HEADLINE RATIO	13.9 ; 9.2	20.7 ; 4.4	7.1 ; 2.0	16.9 ; 3.3	22.4 ; 2.7
	NEG(%);NEG/POS SENTENCE RATIO	12.5 ; 1.8	12.5 ; 1.7	16.0 ; 1.8	16.6 ; 3.5	14.9 ; 8.7
	FREQUENCY (%)	2.3	3.8	1.2	0.7	1.1

As for the analysis of positive and negative tone over time, we have generally observed that both tones maintain a signal level that tends to be constant after the first wave of vaccinations, although for some subtopics, the early period of uncertainty about vaccines generated more fluctuations in negative/positive tone. In Figure 2, we illustrate the case of the UK and the European Union subtopic. The vertical line marks the date of the first vaccination in the UK. The figure shows how the fluctuations that existed at the beginning of the pandemic tend to decrease once the vaccination of the population begins; a similar trend can be seen in all subtopics and countries (not shown for space reasons.)

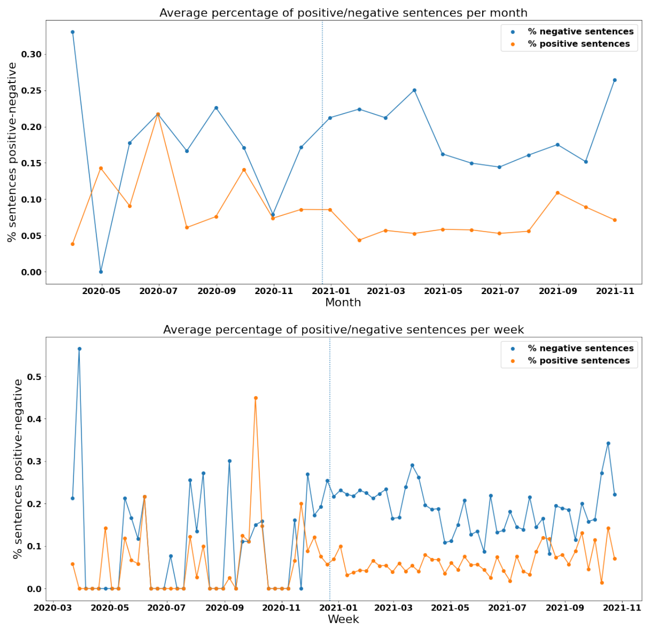

Figure 2: Positive and negative tone in UK about European Union per month and per week.

Regarding the interpretation of the results, in the case of the European Union subtopic, we see that the non-EU countries under study (Switzerland and United Kingdom) present higher values of negativity; if we look at the results at the sentence level, we see that the United Kingdom has an even more negative tone. One could hypothesize this to be linked to the intention of UK press to mark differences with Europe, due to its recent exit from the EU, but such hypothesis would have to be investigated in the future. In the case of the anti-vaccine movement subtopic, we see that the headlines have very high negative values; this could be explained by the fact that there has been strong opposition to anti-vaccine groups, and the press has sought to promote vaccination as a solution to the pandemic situation. As some researchers have discussed [14], social media has significantly enhanced our understanding of anti-vaccination movements and the potential impact on public health attitudes and behaviors regarding vaccination, and how innovative methods of analysis of social media have an important role in developing health promotion to counter anti-vaccination movements. Regarding the economy, the fact that there are no large ratios may indicate that this is a more neutral issue. In the case of the education subtopic, the high ratios at the headline level could be explained by the observation that headlines often open with news about the number of teacher absences or of outbreaks of Covid-19 within a school, discuss whether children should be vaccinated, or discuss whether teachers are at risk and should be vaccinated with first priority. Finally, as for the Olympic Games subtopic, the Japanese public opposition to the event, or the fact that the public was not allow to attend them, might explain the negative tone.

5.4 Sentiment analysis by country

We carried out two experiments concerning sentiment and countries. On one hand, we wanted to extract the sentiment towards the individual governments of the countries under study; on the other hand, we wanted to understand the sentiment towards other countries that appeared as a recurring theme in our previous analysis of subtopics.

5.4.1 Sentiment towards national government. To analyze the attitude towards the government of each country, we extracted those articles that name the president (or prime minister) and the health ministers of each country. The results are shown in Table 5, including the average percentage of negative and positive tone at article, headline, and sentence level, the ratio between the average percentage of negative tone and positive tone, and the frequency of occurrence. As before, red color indicates cases where the negative-to-positive ratio is greater than 5. A few trends are noticeable:

Table 5: Sentiment towards each national government

SUBTOPIC		FR	CH	IT	ES	UK
NATIONAL GOVERNMENT	NEG;POS ARTICLE (%)	18.3 ; 3.6	18.8 ; 6.7	14.8 ; 6.9	14.4 ; 4.4	18.7 ; 7.4
	NEG;POS HEADLINE (%)	16.7 ; 1.4	13.2 ; 4.2	14.3 ; 2.3	17.2 ; 2.4	22.7 ; 4.5
	NEG;POS SENTENCES (%)	10.8 ; 3.7	10.1 ; 7.6	6.2 ; 7.0	9.2 ; 4.1	9.1 ; 5.6
	NEG/POS ARTICLE RATIO	5.0	2.8	2.1	3.2	2.5
	NEG/POS HEADLINE RATIO	11.9	3.1	6.2	7.1	5.0
	NEG;POS SENTENCES RATIO	2.9	1.3	0.8	2.2	1.6
	FREQUENCY (%)	24.8	11.2	17.8	13.5	25.9

- France: We observe the overall most negative tone towards to its government, especially at the headline level, and it is one of the two countries with the highest relative frequency (24.8%).
- Switzerland: This is the only country without any particularly high or low ratios, and has the lowest relative frequency of the five countries (11%).
- Italy: It presents high ratio at headline level and low ratio at sentence level.
- Spain: It presents a high ratio at headline level, only behind France.
- United Kingdom: This is the case with the highest relative frequency (25.9%), with sentiment trends similar to France, Italy, and Spain (high headline ratio.)

In general, we can see that there is negative sentiment, which could be hypothesized as a criticism of the government' management of vaccination in particular and of Covid-19 in general. Headlines tend to be more negative. We can also assume that the results might be affected by the political leanings of the newspapers involved, and whether these are aligned with the governing political parties. Testing these hypotheses would require additional future work.

5.4.2 Sentiment towards other countries. Table 6 shows the corresponding sentiment measures for three countries that emerged from the subtopic analysis: the US, Russia, and India. Once again, the negative sentiment is predominant:

- United States: This case has ratios above five in three of the five countries at headline level, and at sentence level it has levels close to five. It reaches the highest relative frequency (38%) for Switzerland.

- Russia: This case has ratios above five at headline and sentence level for all five countries, with highest relative frequency (9.8%) also for Switzerland.
- India: For this case, we observe similar trends as for Russia (ratios above five at sentence level for all five countries, and at headline level for four of them), with highest relative frequency (11.2%) for the UK.

Table 6: Sentiment towards other countries

SUBTOPIC		FR	CH	IT	ES	UK
USA	NEG(%);NEG/POS ARTICLE RATIO	19.0 ; 3.88	22.0 ; 3.6	16.0 ; 2.2	19.1 ; 3.3	19.1 ; 3.8
	NEG(%);NEG/POS HEADLINE RATIO	17.9 ; 5.4	22.6 ; 5.2	15.4 ; 3.4	22.8 ; 5.1	26.3 ;4.3
	NEG(%);NEG/POS SENTENCE RATIO	19.0 ; 4.1	23.0 ; 4.8	14.9 ; 2.6	18.8 ; 4.70	14.9 ; 3.6
	FREQUENCY (%)	26.2	38.1	15.6	18.2	23.3
RUSSIA	NEG(%);NEG/POS ARTICLE RATIO	22.2 ; 5.5	24.5 ; 4.71	17.3 ; 3.20	19.2 ; 4.09	21.5 ; 4.2
	NEG(%);NEG/POS HEADLINE RATIO	19.1 ; 9.10	22.1 ; 5.02	14.9 ; 6.21	22.3 ; 5.07	30.8 ; 5.40
	NEG(%);NEG/POS SENTENCE RATIO	25.0 ; 9.2	26.1 ; 6.8	22.8 ; 9.5	17.0 ; 5.0	23.2 ; 12.8
	FREQUENCY (%)	8.2	9.8	4.6	4.5	4.7
INDIA	NEG(%);NEG/POS ARTICLE RATIO	22.1 ; 5.3	24.0 ; 4.6	17.7 ; 3.22	22.4 ; 4.9	20.3 ; 3.8
	NEG(%);NEG/POS HEADLINE RATIO	21.7 ; 13.5	25.9 ; 6.8	14.5 ; 4.1	25.0 ; 7.3	24.2 ; 5.7
	NEG(%);NEG/POS SENTENCE RATIO	27.3 ; 8.8	27.4 ;10.1	23.0 ; 10.9	26.3 ; 10.9	20.4 ; 10.2
	FREQUENCY (%)	9.8	9.8	4.4	4.3	11.2

We observe that the sentiment is predominantly negative for the three countries. In the case of India, the negative tone is likely due to the impact of the Delta variant, while the United States and Russia have both developed their own vaccines, but have been hit hard by the pandemic.

5.5 Sentiment analysis by vaccine brand

To analyze the sentiment towards each of the main vaccine brands, we decided to do a stricter extraction of subtopics. For each of the vaccine brands, we only considered articles whose headline contained the brand but did not mention any of the other brands, so that we are left with articles that mainly talk about one brand, thus allowing to be more rigorous in our analysis. This, however, caused the number of samples belonging to each brand to be lower when applying a stricter filter. Table 7 shows the results; as a reminder, red color indicates negative-to-positive ratio above 5, while green color indicates a ratio below 1.

Table 7: Sentiment analysis of vaccines

VACCINES		FR	CH	IT	ES	UK
ASTRAZENECA	NEG(%);NEG/POS ARTICLE RATIO	16.7 ; 4.7	20.9 ; 4.2	16.1 ; 3.8	14.0 ; 3.8	16.0 ; 2.9
	NEG(%);NEG/POS HEADLINE RATIO	9.7 ; 2.4	11.4 ; 1.6	12.6 ; 4.8	10.7 ; 3.9	17.3 ; 5.7
	NEG(%);NEG/POS SENTENCE RATIO	12.5 ; 3.57	17.1 ; 3.23	13.0 ; 5.9	9.9 ; 3.19	13.0 ; 2.60
	FREQUENCY (%)	4.0	1.1	4.4	5.2	5.1
PFIZER	NEG(%);NEG/POS ARTICLE RATIO	11.9 ; 1.9	12.7 ; 1.81	12.5 ; 2.0	9.8 ; 1.6	7.8 ; 0.9
	NEG(%);NEG/POS HEADLINE RATIO	6.8 ; 0.9	5.4 ; 0.7	9.7 ; 1.6	7.2 ; 1.1	4.5 ; 0.4
	NEG(%);NEG/POS SENTENCE RATIO	6.4 ; 0.7	6.0 ; 0.8	5.1 ; 0.7	4.5 ; 1.2	4.8 ; 0.5
	FREQUENCY (%)	1.8	0.9	1.9	2.3	2.4
MODERNA	NEG(%);NEG/POS ARTICLE RATIO	11.5 ; 1.9	12.4 ; 2.1	5.9 ;0.5	10.4 ; 1.4	5.5 ; 0.4
	NEG(%);NEG/POS HEADLINE RATIO	8.5 ; 1.6	8.2 ; 0.8	3.3 ;0.4	11.4 ; 2.3	4.0 ; 1
	NEG(%);NEG/POS SENTENCE RATIO	6.5 ; 0.8	4.9 ; 0.6	3.5 ; 0.4	8.4 ; 0.7	2.4 ; 0.2
	FREQUENCY (%)	1.0	1.0	0.4	0.9	0.5

In the case of vaccines, although the non-neutral sentiment at the article level is still mostly negative, we observe a few different trends for headline and sentence levels:

- AstraZeneca: It has higher negative sentiment in all analyses, as well as the highest relative frequency (up to 5%).
- Pfizer: It has low ratios at headline and sentence level, indicating a more positive tone.
- Moderna: It also has low ratios. We highlight Italy, where all ratios are below one.

These results are not entirely surprising, because although the AstraZeneca vaccine was an attractive option due to its low cost and ease of storage, it was controversial in some countries, which limited its use (e.g. to patients of a certain age.) This could explain why this vaccine was more talked about and has more negative sentiment. In the case of Pfizer and Moderna, we can see that in some countries such as the UK, France and Switzerland, there is a higher percentage of positive headlines than negative ones, with the same results obtained in the analysis at sentence level. This could be because these vaccines were not as controversial as Astrazeneca, and overall had an better public opinion. These results can be contrasted by those obtained in the work of [19]. While that research was conducted with tweets in English, the results are similar in that Pfizer and Moderna have a positive sentiment, while AstraZeneca is negative, and the opinion expressed on Twitter can be assumed to be influenced by the written press and radio/television news, among many other sources.

5.6 Comparison with related work

As a final step, we position our results in the context of previous work. We have not found any work dedicated to the analysis of full news articles focused on Covid-19 vaccination, but we found several analyses of news articles on Covid-19 in general. We consider this as a good starting point to contrast our results.

First, in the work by Ghasiya et al.[11], we observe that when analyzing the sentiment of headlines of different sub-themes like economy or education, the results in each country are different. More specifically, they note that some countries have more positive headlines than negative ones, which is something we do not observe in our analysis. We hypothesize that our dataset is perhaps more homogeneous, as all countries are European, while the data used in [11] comes from countries from two different continents, covering the Covid-19 issue in general. Second, the work by De Melo et al.[9], which was focused on Brazil's media, shows that social media tended to have a more negative polarity for all themes, while formal media seemed to present almost neutral polarity on average. These results present a somewhat similar trend to what we observed in our study, even though they use VADER for sentiment analysis. Finally, the work by Aslam et al.[4] found that the preponderant class are negative headlines. This paper uses data from newspaper headlines in the early part of the pandemic, when fear, uncertainty, anger or sadness were very present, and this may be the reason for such trend. These results are consistent with what one would expect from the Covid-19 pandemic in its early stages, when the number of deaths and the uncertainty and fear of how to deal with the situation affected the population. They are also consistent with the view that the press tends to be negative to keep readers alert to an adverse situation, according to psychological studies [32] and [33] .

6 CONCLUSIONS

This paper presented a first analysis of a new dataset of European articles on Covid-19 vaccination. The pipeline of entity, subtopic identification, and sentiment analysis used in our work, allowed to uncover some of the characteristics of high-quality press in Europe, and also to compare the results across the different countries involved. We found that a neutral tone predominates in the studied press articles, which can be seen as evidence of the search for objectivity by the studied newspapers, although the tone is more negative than positive for non-neutral articles. These findings may complement the ongoing efforts to better understand the nature of disinformation in other news sources, whose aim is often the elicitation of strong emotions, generally negative, as messages are conveyed to audiences. We also found that despite having a news dataset from five European countries with great cultural differences marked by language, the results across countries followed homogeneous trends in terms of sentiment. Moreover, substantial differences have been appreciated in few cases, for instance in the analysis of identities where the local component associated with each country is distinct. To conclude, we advocate for the computational analysis of high-quality news as enablers of benchmarks to contrast and validate other information sources, as part of a systematic understanding of disinformation and misinformation around health topics.

Acknowledgments. This work was supported by the AI4Media project, funded by the European Commission (Grant 951911) under the H2020 Programme ICT-48-2020. We also thank the newspapers that agreed to share their online articles.

REFERENCES

[1] 2021. WHO/ECDC: Nearly half a million lives saved by COVID-19 vaccination in less than a year. (2021). https://www.ecdc.europa.eu/en/news-events/who-ecdc-nearly-half-million-lives-saved-covid-19-vaccination
[2] Jennifer Allen, Baird Howland, Markus Mobius, David Rothschild, and Duncan Watts. 2020. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* 6 (04 2020), eaay3539. https://doi.org/10.1126/sciadv.aay3539
[3] Dimo Angelov. 2020. Top2Vec: Distributed Representations of Topics. arXiv:2008.09470 [cs.CL]
[4] M-Faheem Aslam, Tahir Awan, Jabir Hussain Syed, Aisha Kashif, and Mahwish Parveen. 2020. Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications* 7 (07 2020). https://doi.org/10.1057/s41599-020-0523-3
[5] Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2013. Sentiment Analysis in the News. arXiv:1309.6202 [cs.CL]
[6] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3, null (mar 2003), 993–1022.
[7] Daniel Catalan-Matamoros and Carlos Elías. 2020. Vaccine Hesitancy in the Age of Coronavirus and Fake News: Analysis of Journalistic Sources in the Spanish Quality Press. *International Journal of Environmental Research and Public Health* 17, 21 (nov 2020), 8136. https://doi.org/10.3390/ijerph17218136
[8] Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin, Corina Ioanăş, Dana Simona Gherai, and Federico Tajariol. 2021. The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics From Tweets in the Month Following the First Vaccine Announcement. *IEEE Access* 9 (2021), 33203–33223. https://doi.org/10.1109/ACCESS.2021.3059821
[9] Tiago de Melo and Carlos M S Figueiredo. 2021. Comparing News Articles and Tweets About COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach. *JMIR Public Health Surveill* 7, 2 (10 Feb 2021), e24585. https://doi.org/10.2196/24585
[10] Federico Germani and Nikola Biller-Andorno. 2021. The anti-vaccination infodemic on social media: A behavioral analysis. *PloS one* 16 (03 2021), e0247642. https://doi.org/10.1371/journal.pone.0247642
[11] Piyush Ghasiya and Koji Okamura. 2021. Investigating COVID-19 News Across Four Nations: A Topic Modeling and Sentiment Analysis Approach. *IEEE Access* 9 (2021), 36645–36656. https://doi.org/10.1109/ACCESS.2021.3062875
[12] Maarten Grootendorst. 2020. BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics. https://doi.org/10.5281/zenodo.4381785
[13] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). To appear.
[14] Luis Saboga-Nunes-EUPHA (HP) and Patty Kostkova-UK. 2019. Round table: Online anti-vaccination movements: the role of social media in public health communications. *European Journal of Public Health* 29, Supplement_4 (2019), ckz185–681.
[15] C. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media* 8 (May 2014), 216–225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550
[16] Andrea Langbecker and Daniel Catalan-Matamoros. 2021. Who Guides Vaccination in the Portuguese Press? An Analysis of Information Sources. *International Journal of Environmental Research and Public Health* 18, 4 (2021). https://doi.org/10.3390/ijerph18042189
[17] Steven Loria. 2018. textblob Documentation. *Release 0.15* 2 (2018).
[18] Katharine Mach, Raul Salas Reyes, Brian Pentz, Jennifer Taylor, Clarissa Costa, Sandip Cruz, Kerronia Thomas, James Arnott, Rosalind Donald, Kripa Jagannathan, Christine Kirchhoff, Laura Rosella, and Nicole Klenk. 2021. News media coverage of COVID-19 public health and policy information. *Humanities and Social Sciences Communications* 8 (09 2021), 220. https://doi.org/10.1057/s41599-021-00900-z
[19] Robert Marčec and Robert Likic. 2021. Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal* (08 2021), postgradmedj–2021. https://doi.org/10.1136/postgradmedj-2021-140685
[20] Kembrew McLeod Melissa Zimdars (Ed.). 2020. *Fake News: Understanding Media and Misinformation in the Digital Age* (1st. ed.). The MIT Press, Chapter 5. https://doi.org/10.7551/11807.001.0001
[21] Tao Na, Wei Cheng, Dongming Li, Wanyu Lu, and Hongjian Li. 2021. Insight from NLP Analysis: COVID-19 Vaccines Sentiments on Social Media. *ArXiv* abs/2106.04081 (2021).
[22] NLP Town. 2020. bert-base-multilingual-uncased-sentiment. https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment.
[23] Gulsan Ara Parvin, Reazul Ahsan, Md. Habibur Rahman, and Md. Anwarul Abedin. 2020. Novel Coronavirus (COVID-19) Pandemic: The Role of Printing Media in Asian Countries. *Frontiers in Communication* 5 (2020). https://doi.org/10.3389/fcomm.2020.557593
[24] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104* (2017).
[25] Rabindra Lamsal. 2021. Sentiment Analysis of English Tweets with BERTsent. https://huggingface.co/rabindralamsal/finetuned-bertweet-sentiment-analysis.
[26] Kumar Rahul, Bhanu Raj Jindal, Kulvinder Singh, and Priyanka Meel. 2021. Analysing Public Sentiments Regarding COVID-19 Vaccine on Twitter. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Vol. 1. 488–493. https://doi.org/10.1109/ICACCS51430.2021.9441693
[27] Eryn J. Newman Rainer Greifeneder, Mariela E. Jaffé and Norbert Schwarz (Eds.). 2020. *The Psychology of Fake News* (1st. ed.). Routledge, London, Chapter 9. https://doi.org/10.4324/9780429295379
[28] Leonard Richardson. 2007. Beautiful soup documentation. *April* (2007).
[29] Nicola Righetti. 2021. Four Years of Fake News. A Quantitative Analysis of the Scientific Literature. (03 2021). https://doi.org/10.31235/osf.io/buemr
[30] Bruce Sacerdote, Ranjan Sehgal, and Molly Cook. 2020. *Why Is All COVID-19 News Bad News?* Working Paper 28110. National Bureau of Economic Research. https://doi.org/10.3386/w28110
[31] Alfonso Semeraro, Salvatore Vilella, Giancarlo Ruffo, and Massimo Stella. 2022. Writing about COVID-19 vaccines: Emotional profiling unravels how mainstream and alternative press framed AstraZeneca, Pfizer and vaccination campaigns. *arXiv preprint arXiv:2201.07538* (2022).
[32] Stuart Soroka, Patrick Fournier, and Lilach Nir. 2019. Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences* 116, 38 (2019), 18888–18892. https://doi.org/10.1073/pnas.1908369116
[33] Stuart Soroka and Stephen McAdams. 2015. News, Politics, and Negativity. *Political Communication* 32, 1 (2015), 1–22. https://doi.org/10.1080/10584609.2014.881942
[34] Santiago Tejedor, Laura Cervi, Fernanda Tusa, Marta Portales, and Margarita Zabotina. 2020. Information on the COVID-19 Pandemic in Daily Newspapers' Front Pages: Case Study of Spain and Italy. *International Journal of Environmental Research and Public Health* 17, 17 (aug 2020), 6330. https://doi.org/10.3390/ijerph17176330
[35] Duncan J. Watts, David M. Rothschild, and Markus Mobius. 2021. Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences* 118, 15 (2021). https://doi.org/10.1073/pnas.1912443118
[36] Wikipedia. 2022. Le Monde. https://en.wikipedia.org/wiki/Le_Monde [Online; accessed 14-Febrary-2022].
[37] Wikipedia. 2022. Named-entity recognition — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=Named-entity%20recognition&oldid=1073163208. [Online; accessed 12-April-2022].
[38] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie

Brew. 2019. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *ArXiv* abs/1910.03771 (2019).

[39] Niaz Mahmud Zafri, Sadia Afroj, Imtiaz Mahmud Nafi, and Md. Musleh Uddin Hasan. 2021. A content analysis of newspaper coverage of COVID-19 pandemic for developing a pandemic management framework. *Heliyon* 7, 3 (2021), e06544. https://doi.org/10.1016/j.heliyon.2021.e06544

[40] John Zarocostas. 2020. How to fight an infodemic. *The lancet* 395, 10225 (2020), 676.

[41] Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–40.