Worldwide comparison of the discourse on the FIFA World Cup 2022 in Qatar

Abdulnaser Sabra and Tobias Ehrmeier

University Leipzig - Institute of Informatics
Augustusplatz 10, 04109 Leipzig - Germany

1 Introduction

The FIFA World Cup 2022, held in Qatar, has been the topic of extensive news coverage in recent years. This event has generated significant interest and attention globally, with news reports from various countries discussing different aspects of the tournament, such as infrastructure development, labor practices, security measures, and environmental sustainability. The selection of Qatar as the host country for the World Cup has been controversial, with concerns raised about the country's human rights record, including allegations of exploitation of migrant workers. These concerns have led to increased scrutiny of Qatar's preparations for the tournament, particularly regarding the treatment of migrant workers and the working conditions on World Cup construction sites. Moreover, Qatar's hosting of the World Cup has also raised concerns about the country's ability to handle the influx of visitors and ensure their safety. News reports from different countries have also highlighted the human rights abuses towards people within the LGBTQ community. This report aims to provide a comprehensive analysis of news reports on the Qatar World Cup 2022 from different news outlets from various countries. The paper will compare and contrast different news outlets with data from The Guardian, Al Jazeera, and the BBC, by examining the different issues that have been covered by the media, including labor practices, environmental concerns, human rights and security measures. By analyzing news reports from different countries, this paper seeks to provide a nuanced understanding of the various perspectives and opinions surrounding the World Cup in Qatar and to compare these topics and opinions between different news outlets from different countries.

2 Related Work

In order to conduct a comprehensive analysis of news coverage on a specific topic, such as the World Cup 2022 in Qatar, it is important to examine existing research that has attempted to compare different media outlets. Previous studies have

investigated the differences in perspectives and opinions across various news sources, highlighting the importance of considering multiple viewpoints when analyzing media coverage. One approach to comparing news sources involves examining previous papers that have attempted to analyze and compare different outlets. These papers have employed various methods to compare news sources, including content analysis, discourse analysis, and sentiment analysis. Through these approaches, researchers have sought to identify similarities and differences in the coverage of specific events, as well as patterns of bias and subjectivity in news reporting. Furthermore, studies have also explored how bias can be detected in news coverage, whether it be positive or negative. These studies have shown that bias can occur in various forms, such as selective reporting, framing, and tone. Identifying bias is important as it can provide insights into the agenda-setting of different media outlets and can help to contextualize the coverage of specific events. Finally, papers that have investigated public opinion on the Qatar World Cup have also been examined. These studies have explored how the general public perceives the event, as well as the factors that influence their attitudes towards it. By examining the public opinion, it is possible to gain a deeper understanding of the broader social and cultural context in which the event is taking place.

2.1 Comparing news

In recent times, there have been several global events that have dominated the news cycle, attracting significant media attention and shaping public discourse. In addition to the World Cup 2022 in Qatar, other major events and crises have received extensive news coverage and have been the focus of academic research. One such event is the ongoing conflict between Ukraine and Russia, which has been the subject of intense media scrutiny and has been the focus of research by Durumeric, Hanley, and Kumar (2022)[6]. In their paper, these researchers examined the news coverage of the Ukraine and Russian war across different media outlets. By conducting a comparative analysis of news reports from various countries, they sought to identify differences in the framing, tone, and content of news coverage of the conflict. This comparative analysis of news articles and social media posts related to the Russo-Ukrainian conflict published by Western press, Chinese state media, and Russian propaganda websites. The study collected a total of 11,359 unique news articles published between January 1, 2022, and April 15, 2022, from various sources, including eight popular Western news websites, nine Russian websites identified by the US State Department, and seven Chinese foreign mission news websites using web scraping with algorithms and the Python Selenium package. To find topics in their data set, they use PLDA among other methods[6]. They also conducted sentiment analysis. Their results show that Western, Russian, and Chinese news sources have covered the Russo-Ukrainian war differently. Western media uses Ukrainian spellings and highlights military and humanitarian crises. Chinese media refers to it as a "conflict" and avoids grisly details, resulting in a higher sentiment score. Russian media focuses on reasons for the "special military operation,"

"denazification," and "liberation" of Ukraine's people[6]. Another paper looks at the news coverage of a different recent topic, the Covid-19 pandemic and its corresponding press coverage. De Barrio and Gatica-Perez (2022) conducted a 5 country comparative analysis with 19 newspapers from France, Italy, Switzerland, Spain and the United Kingdom[1]. The data was retrieved by requesting access from the newspapers, this resulted in their selection, ranging from big, international newspapers to smaller regional ones. To work with this data set, they used 3 different methods: Name Entity Recognition, fine grain topic modelling and sentiment analysis (by country and by topic). The analysis is carried out at different levels (article, headline, and sentence) and for different subtopics (European Union, No-Vax, Economy, Education, and Olympics). The authors observe that the proportion of negative sentiment is much higher than positive sentiment, which may be due to humans being predisposed to focus on negative information. They also note that there is a large amount of content classified as neutral, which is expected in news articles that seek objectivity. The analysis shows that the subtopic with the highest percentage of negativity is No-Vax, and that the sentiment is most negative in the United Kingdom when talking about the European Union[1]. In the sentiment towards national government experiment, the researchers analyzed the attitude towards the government of each country and extracted articles that named the president or prime minister and health ministers. The results showed negative sentiment towards the government in general, with France being the most negative and Switzerland being the least negative. They also discovered different sentiments towards vaccination brands[1]. Now that we see how a comparison of news is possible, we look at a new topic that emerged from it.

2.2 News bias

News bias refers to the tendency of media outlets to present news stories in a way that reflects their own political or ideological beliefs or biases. While it is impossible for any news source to be completely objective, scientific research has shown that certain media outlets tend to have a more pronounced ideological bias than others. Studies have also shown that news bias can have a significant impact on the public's perception of political issues, and can even affect their voting behavior. Understanding the nature and extent of news bias is therefore an important area of research for both academics and journalists, as it can have important implications for the functioning of democratic societies. Regarding this paper, we are looking at the first World Cup that is held by a Muslim country. Muslims still face racial and religious bias, especially in western countries. Which is why we examine papers focusing on news bias on Muslim topics. First, a study by Connor Nickerson (2019)[12] on a comparison of media portrayal of Muslims and terrorist attacks that happened in a western country, France, and a Muslim country, Turkey, in the lens of US media. The study finds that the US media is not neutral in its coverage of terrorist attacks in different countries and the parties responsible. Rather, the framing of news reports is driven by security threat perceptions and reinforces an Orientalist perspective, perpetuating

prejudice against Muslims and a belief in the increased threat of terrorism to US national security[12]. The media's biased framing of events leads to a misalignment between current policies, perceptions, and statistical evidence. The article argues that rectifying this problem requires addressing the unconscious biases prevalent in media coverage and expanding the conceptual limits of the community to include those deemed as "global others" excluded from having rights in individual national justice systems[12]. A different paper by Jörg Matthes et al. (2020) [11] looks at who differentiates between Muslims and Islamic terrorists. The study conducted a quantitative content analysis of terrorism news coverage in three German-speaking countries (Austria, Germany, and Switzerland) between January 2015 and November 2017. The study focused on this time frame as it marked the beginning of a new wave of Islamist terrorist attacks in Europe, starting with the Charlie Hebdo shooting in Paris, France. The study selected news articles that contained the terms "terror" and "Muslim" from 12 news outlets, including both conservative and liberal newspapers, and employed a coding procedure to determine the main topic of the articles. The unit of analysis was the textual content of a news article, and pictures or picture captions were not included. The study found that there was a high degree of undifferentiated news coverage in the selected news articles, with undifferentiated statements being more prevalent than differentiated statements. The study also found that undifferentiated news coverage was more prevalent in tabloid newspapers than in quality newspapers, and that there were differences in undifferentiated news coverage across the three countries studied[11].

2.3 World Cup and Qatar

A world cup puts the hosting nation into a worldwide spotlight. This can for example be measured by looking at Google searches. A study by Kobierecki and Strożek [9] uses Google Trends data to analyze the number of searches for countries before and after hosting events such as the Olympic Games and the FIFA World Cup. The study finds that hosting such events has a significant impact on the number of Google searches for a country, and the impact varies depending on the country and the type of event [9]. So it did 2018 with Russia, however not always in a positive way, with news surrounding topics like LGBTQ rights. A study by Nadezhda Ozornina Alexander Mannin [13] analyzes how Russia was portrayed in British media during the 2018 FIFA World Cup, using J. Druckman's competing frames theory. The researchers examined 751 articles from British tabloids and broadsheets, identifying five frames: 'Russia as the enemy of Great Britain,' 'Russia as the potential partner of Great Britain,' 'Russia as the unsafe state,' 'Russia as the worthy host of the World Cup,' and 'Russia as a country with a rich culture.'[13] The results show that negative perceptions of Russia dominated in the media, particularly in broadsheets. Although there was a slight positive shift during the World Cup, by the end of the tournament, the negative framing had returned to pre-championship levels. Furthermore, we look at how the Qatar World Cup was perceived by the world. A study by Susan Dun et al. [5] collected and geotagged over 4 million tweets containing "Qatar"

with the result that the sentiment of Qatar's hosting is negative, especially in northern countries. A similar result was found by Arianto and Dewi [4] by using sentiment analysis on tweets on Qatar. The study finds that before the awarding of the World Cup, the positive sentiment on Qatar was at 88.46 percent, declining to 79.38 percent after the awarding and slightly rising to 83.72 percent during the World Cup [4].

3 Research question

The literature review conducted shed light on the ways in which a comparative analysis of news related to the Qatar World Cup can be carried out effectively. It also brought attention to the dominant biases that exist in media coverage and public opinions towards host nations, specifically Qatar. Through the analysis of various sources, the review provided valuable insights into the nature of media bias and its impact on shaping public perception. The study highlighted the importance of critically evaluating news coverage and considering the different perspectives that may be at play in the reporting of such events. This leads to the research questions of this paper.

RQ: Is there a difference in reporting on the FIFA World Cup 2022 by different media outlets from different countries and backgrounds?

To find an answer to this question, we form hypotheses. First, an investigation of topics is conducted.

H1: The three media outlets cover similar topics with comparable frequencies.

Here the dependent variable is the topics, the independent variable is the news outlets. Furthermore, we need to examine the topics covered by the three observed media outlets and the frequency with which they address them. Initially, we assume that the three media outlets cover comparable topics with comparable frequencies and then we can reject this hypothesis, if we notice variations in either the topics covered or their frequencies. The hypothesis considers the frequencies of the addressed topics, because it is possible for some media outlets to publish only a small number of articles, such as two or three, on a specific subtopic.

Secondly, we want to look at how articles are written and if opinions can be deducted.

H2: The three media outlets display similar positive and negative sentiments in their coverage of the World Cup 2022.

We aim to find differences in the sentiment score by looking at the average positive and negative percentage of the articles and also their headlines. Initially, we assume that the three media outlets show the same positive and negative tone and then this hypothesis can be rejected if there is significant difference detectable.

Third, we aim to find out if there are different sentiment within the topics produced in H1.

H3: The average positive and negative sentiments do not vary by topics
produced in H1 within the news outlets.

We will look at the common topics produced earlier to see if the general sentiment towards them varies.

Fourth, we are looking for entities in the data set.

H4: The named entities in the different data sets are the same by news outlets.

Additionally, we lastly look at the entities produced by Named Entity Recognition methods. These will be investigated by looking at their background and they will be categorised to find patterns within the news outlets.

4 Data

4.1 Choice of news outlets

First, we need to choose news outlets that we want to compare, and that have different aspects and backgrounds. Also, there has to be enough articles in the English language to give us sufficient data to work with. Our first choice was The Guardian. The Guardian is a British daily newspaper known for its progressive and liberal stance on political and social issues[15]. It was founded in 1821 as The Manchester Guardian and is now published in London[10]. The Guardian covers a wide range of topics including politics, business, science, technology, the environment, culture, and sports. Its reporting is known for its in-depth analysis and investigative journalism[17]. The Guardian has a strong online presence, with its website being one of the most visited news sites in the world. It also publishes a range of digital content including podcasts, videos, and interactive graphics. Our second choice is the BBC. The BBC (British Broadcasting Corporation) is a publicly-funded broadcaster based in the United Kingdom. It is the world's oldest national broadcaster, founded in 1922, and operates a wide range of television, radio, and online services both domestically and internationally[2][14]. The BBC's online presence is extensive, with a range of websites and apps covering news, sports, entertainment, and education. Its news website, BBC News, is one of the most widely read news sources in the world, providing up-to-the-minute coverage of national and international news events, as well as in-depth analysis and investigative reporting. One of the unique aspects of the BBC's online presence is its commitment to impartiality and accuracy. As a public service broadcaster, the BBC is committed to providing objective, trustworthy news and information to its audience. It also has strict editorial guidelines designed to ensure that its content is fair and balanced[3]. Opposed to our first source, the private owned The Guardian, is publicly founded[2]. Our third source is Al Jazeera. Al Jazeera is a news network headquartered in Doha, Qatar, that operates a global network of television channels and digital platforms. It was launched in 1996 and quickly established itself as a major player in the international news media[7]. Al Jazeera's online presence is extensive, with a range of websites and apps covering news, sports, documentaries, and other content.

Its main news website, Al Jazeera English, provides coverage of national and international news events from a global perspective, with a focus on issues that are often overlooked by other news outlets. One of the unique aspects of Al Jazeera's online presence is its commitment to providing a platform for diverse voices and perspectives. It regularly features opinion pieces and commentary from a range of contributors, including activists, academics, and experts in various fields. Al Jazeera is also known for its coverage of the Middle East and North Africa, with a range of specialized content covering politics, culture, and society in the region. Its Arabic-language news channel, Al Jazeera Arabic, is widely watched throughout the Arab world and has played a key role in shaping public opinion in the region[18]. Al Jazeera is owned and by the Qatari state[7]. We chose the English language website of Al Jazeera. Overall, we have to British news outlets and one from Qatar. Only The Guardian is privately owned.

4.2 Data acquisition

To gather data for the World Cup analysis, we obtained data sets from the category pages of the websites dedicated to covering the event. These pages are repositories for all articles related to the World Cup and contain information dating back to 2010. We employed a web scraping technique to extract the data from these pages, allowing us to retrieve a large volume of articles that could be analyzed for insights and trends. Once the data was collected, we undertook a cleaning process to ensure that the data set was accurate and relevant. This involved the removal of articles that contained only videos or photo galleries, as these did not contain sufficient textual content to be analyzed. By eliminating these extraneous articles, we were able to focus on those that provided substantive content and insights into the World Cup and its impact on global audiences. The cleaning process was an essential step in ensuring the accuracy and usefulness of the data set. By carefully curating the data to include only relevant articles, we were able to generate insights that were valuable for understanding the World Cup and its impact on society.

4.3 Data set

This resulted in a data set of a total of 2966 articles. The most came from The Guardian, 1916. 824 articles were in the data set of Al Jazeera and 226 of the BBC.

5 Methodology

Our approach to resolving the research questions involved a range of methods that drew from multiple sources. Some of these methods were based on the lessons we learned in the course and were an obvious choice given our research objectives. We also looked to the findings of related work to guide our comparative analysis, incorporating insights from these sources into our approach. In addition, we remained open to new and innovative methods as they arose during

the project. As we encountered new options, we evaluated their potential for contributing to our research goals and pursued those that showed promise. This allowed us to continually refine our approach and adapt to emerging opportunities throughout the research process.

5.1 Topic Modelling

Topic modelling is a computational technique used to identify themes or topics in a large collection of texts. It involves a statistical analysis of the frequency and distribution of words in the text to identify patterns and themes. By using topic modelling, we can identify and group together similar content, allowing us to gain insights into the themes, trends, and patterns that emerge from the data. We can compare these findings within the different media outlets. This method was a topic in our course, and it was also used frequently in the research presented in the related work. Our approach involved utilizing Bertopic, a pretrained model, to extract subtopics from news articles. Bertopic generates an embedding vector for each article, which is then used in clustering algorithms to group articles with similar embedding vectors into clusters. Finally, TF-IDF techniques are employed to extract the most significant words from each cluster. These words provide valuable insight into the content and potential subtopic addressed within each cluster.

5.2 Sentiment analysis

Sentiment analysis is used to identify the emotional tone of a news article, among others. It involves using natural language processing (NLP) algorithms to analyze the text and determine whether it expresses positive, negative, or neutral sentiment. Sentiment analysis is useful in our projects because it allows us to understand the attitudes and opinions of these articles. This was also part of our initial approach, as it occurred in our course and also in the related work we investigated.

5.3 Sentiment analysis on topics

With the topics gathered in our first method, we want to apply sentiment on those specific topics. This gives on insights on how the actual topics are discussed in the news papers. In addition, we can compare the sentiments on topics between the different news outlets. This was a result of our first feedback discourse with the teacher.

5.4 Named entity recognition

Named entity recognition (NER) is used to identify and classify named entities in a text, such as people, organizations, locations, and products. It involves using NLP algorithms to analyze the text and identify words or phrases that correspond to specific categories of named entities. We are using NER to identify and track specific entities allowing us to gain insights into patterns and trends

related to those entities. NER was first discovered during the investigation of related work, and became even more interesting after first results, so we chose to apply it to our data set.

6 Results

In this section , we showcase the results derived from our analysis of the articles published by the three observed media outlets.

6.1 Overall Topic modelling by media outlet

As explained in the section 5.1, Bertopic was employed to extract significant subtopics from the three observed media outlets. These subtopics were characterized by specific sets of words, as it was also elaborated in the same section. After analyzing Al Jazeera data, the subtopics identified were "worker rights" and "corruption allegations".

Al Jazeera: sample of detected topics

Topic	Words per topic	Frequency(%)
Worker rights	Workers_labour_rights_migrant	0.04
Corruption allegations	Investigation_ibhais_committee_officials	0.01

Fig. 1: *Al Jazeera* sample of detected topics

It should be noted that those are not the only subtopics were extracted using the Bertopic model , but there also other topics , but alot of them were focused very much on the matches , players , winners during the event , therefore we did not take them into consideration.

After the BBC data was analyzed, the subtopics that were deemed significant are "LGBTQ discrimination", "worker rights", and "alcohol ban" as shown in the table in the fogure 2 .Among these subtopics, "LGBT discrimination" received the most attention, while "alcohol ban" received the least attention. The latter can be explained by the short-term nature of the ban, since it was only announced days before the event[8].

Compared to the other two media outlets, The Guardian has covered a broader range of topics. Specifically, it has published articles on five subtopics, including "Worker rights", "corruption", "LGBTQ discrimination", "Armband", and "alcohol ban" as shown in the table in the figure 3 . The most addressed subtopic by The Guardian was "Worker rights", while the least addressed subtopic was "alcohol ban".

The analysis of the three tables indicates that the topic of "worker rights" was the common addressed subtopic across all three media outlets. Additionally,

BBC: sample of detected topics

Topic	Words per topic	Frequency(%)
LGBTQ discrimination	lgbt_rights_gay_human	0.16
Worker rights	Workers_labour_rights_migrant_amnesty_ibhais	0.07
Alcohol ban	Alcohol_minute_sale_ban	0.01

Fig. 2: *BBC* sample of detected topics

it can be noticed that the Guardian showed a greater interest in "worker rights" (13 percent) and less interest in "LGBTQ discrimination" (2 percent) compared to the BBC, whereas the BBC demonstrated more interest in "LGBTQ discrimination" subtopic (16 percent) and less interest in "worker rights" subtopic (7 percent) than the Guardian.

In summary, it can be noted that Al Jazeera placed a significant emphasis on

The Guardian: sample of detected topics

Topic	Words per topic	Frequency(%)
Worker rights	Workers_migrant_labour_rights al_qatari_doha_workers rights_human_workers_kit	0.13
Corruption	Committee_gracia_bid_executive swiss_investigation_gracia_evidence	0.05
LGBTQ discrimination	lgbtq_gay_rights_rainbow	0.02
Armband	armband_wear_armbands_sancations	0.01
Alcohol ban	alcohol_budweiser_beer_fan	0..005

Fig. 3: *The Guardian* sample of detected topics

covering the matches during the 2022 World Cup and paid comparatively less attention to other topics when compared to the BBC and the Guardian.

Looking at our first hypothesis, it is apparent that the number of topics addressed by each media outlet differs. Al Jazeera covered only two topics, BBC addressed three, and The Guardian addressed five, making it difficult to claim that they addressed the same topics. Moreover, we observed significant differences in the frequencies of the common topics, particularly between British newspapers and Al Jazeera. It was noted that Al Jazeera had the lowest frequency of articles published regarding the subtopic "Worker rights." We also observed variations in the emphasis on topics between BBC and The Guardian. The frequency of the "LGBTQ discrimination" topic was higher for BBC at 16 percent compared to only 0.02 percent in The Guardian data set. However, The Guardian published nearly twice as many articles on the subtopic "Worker rights" than BBC, which only had a frequency of 0.07. Therefore, we must reject the first hypothesis based on these findings. This means the first hypothesis can be rejected.

6.2 Overall sentiment analysis by media outlet

Figure 4 presents the results of a comprehensive sentiment analysis of the various media outlets. The analysis reveals that, in relation to the World Cup 2022, all three media outlets displayed a similar positive tone. It is noteworthy that Al Jazeera's articles regarding the World Cup 2022 had a lower negative tone compared to those of the other two media outlets (BBC and The Guardian). This may not be surprising, given that Al Jazeera is funded by the Qatari government, which could explain the shortage of negative sentiments in their articles. Finally, it is worth noting that the negative tone at the article level is lower than that at the title level, as media outlets often include more negative sentiment in their headlines to capture the attention of their viewers.

Generic sentiment analysis by media outlet

Analysis	Al Jazeera	BBC	The Guardian
AVG.POSITIVE PART(%)	0.16	0.14	0.15
AVG.NEGATIVE PART(%)	0.07	0.10	0.10
AVG.POSITIVE PART TITLE (%)	0.14	0.13	0.13
AVG.NEGATIVE PART TITLE (%)	0.07	0.17	0.12

Fig. 4: *Generic* sentiment analysis by media outlet

Considering the subject matter we examined, it is not surprising that the two British media outlets, BBC and The Guardian, exhibited a higher proportion of negative sentiment, as expected. This is because there were numerous concerns raised about the treatment of construction workers and migrant workers, as well as LGBTQ rights issues. Qatar banned the display of rainbow flags in stadiums and, just days before the tournament started, they also prohibited alcohol consumption during the matches of the World Cup. In contrast, such issues are not present in Europe, where there are laws that protect the rights of construction workers and LGBTQ individuals and allow them to practice their sexual orientation without fear of punishment by the state. Therefore, due to these reasons, the Western media would strongly criticize and condemn the actions taken by the Qatari government, whether they relate to worker deaths, low wages, or the ban on displaying LGBTQ flags in stadiums. This could explain the higher negative sentiment displayed by BBC and The Guardian. In conclusion of these results we can reject H2 in at least some regards. Again there is a greater difference between British and Qatari news outlets but also the BBC produces way more negative headlines than The Guardian.

6.3 Sentiment analysis by the subtopic (worker rights) by the three media outlets

The topic of "Worker Rights" appears to be the most commonly addressed topic among the three media outlets. An analysis of the sentiments expressed in articles related to "worker rights" was carried out across three different media outlets. It is apparent that the treatment of the "Worker Rights" topic varies across the three observed media outlets, with The Guardian having the highest frequency, the BBC having half the frequency of The Guardian, and Al Jazeera publishing the least about this topic, which is not surprising.

Additionally, it is worth noting that Al Jazeera exhibits a slightly higher positive tone regarding the topic at both the article and title levels. However, it is surprising that Al Jazeera displays a higher negative tone at the title level compared to the BBC. Meanwhile, The Guardian exhibits the highest negative tone at the title level.

Al Jazeera vs BBC vs The Guardian: sentiment analysis by the subtopic (worker rights)

SUBTOPIC		Al Jazeera	BBC	The Guardian
Worker rights	Avg pos part (%)	0.14	0.12	0.12
	Avg neg part (%)	0.10	0.12	0.12
	Avg pos part title(%)	0.13	0.08	0.10
	Avg neg part title(%)	0.17	0.14	0.21
	Frequency(%)	0.04	0.07	0.13

To delve deeper into articles related to "worker rights", we performed co-occurrence analysis on the texts to identify the most common words appearing
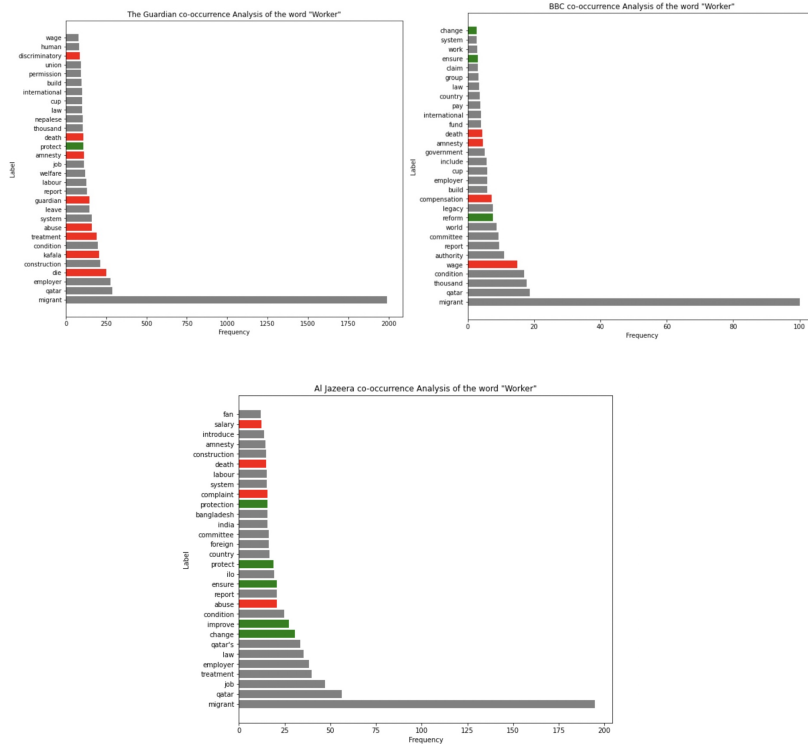
Fig. 5: *Co-occurrence* Analysis of the word "Worker" in the three media outlets

alongside with the word "worker". Our findings revealed that in articles published by The Guardian, the most frequent co-occurring words with "worker" were negative words such as "die", "kafala", "abuse", and "death". In contrast, in articles published by Al Jazeera, the most common words co-occurring with "worker" were words related to legal reforms, such as "change", "improve", "ensure", "protect", and "protection". Although the articles from Al Jazeera also contained negative words such as "death" and "abuse", they were not as prevalent as in The Guardian articles.

As for the BBC, we also observed some negative words like "death" appearing alongside the word "worker", but to a lesser extent than in Al Jazeera articles. Moreover, we noticed some positive words like "reform", "ensure", and "change" in the articles, but not as frequently as in Al Jazeera articles. These findings could potentially provide an explanation for the reason why Al Jazeera tends to have a higher degree of negativity in its headlines compared to the BBC, but a lower level of negativity compared to The Guardian. Additionally, these results suggest that Al Jazeera has a more positive tone in its headlines compared to both of these media outlets.

6.4 Sentiment analysis by the subtopics ("LGBTQ rights" and "Alcohol ban") by The Guardian and BBC

SUBTOPIC		BBC	The Guardian
	BBC vs The Guardian: sentiment analysis by the subtopics (LGBTQ discrimination and alcohol ban)		
LGBTQ discrimination	Avg pos part (%)	0.15	0.16
	Avg neg part (%)	0.12	0.11
	Avg pos part title(%)	0.11	0.13
	Avg neg part title(%)	0.15	0.16
	Frequency(%)	0.16	0.02
Alcohol ban	Avg pos part (%)	0.23	0.17
	Avg neg part (%)	0.13	0.08
	Avg pos part title(%)	0.28	0.15
	Avg neg part title(%)	0.38	0.21
	Frequency(%)	0.01	0.005

Fig. 6: *Sentiment Analysis* by The Guardian and BBC regarding the topics "LGBTQ rights" and "alcohol ban"

The topics of "LGBTQ rights" and the "alcohol ban" during the World Cup 2022 are covered by both The Guardian and BBC. The Qatari government implemented alcohol ban just days before the tournament began, which explains why there are very few articles addressing this topic in both media outlets[8]. In contrast, the topic of LGBTQ rights has been addressed for a long time and is more prominent in the BBC's coverage than in The Guardian's. The Guardian mainly focused on the topic of "worker rights," which had the highest frequency among all topics covered by the outlet. The frequency of articles published by the BBC on LGBTQ rights was much higher than that of The Guardian.

The table in the figure 6 shows that both media outlets have a similar tone, both positive and negative, in their coverage of "LGBTQ discrimination."
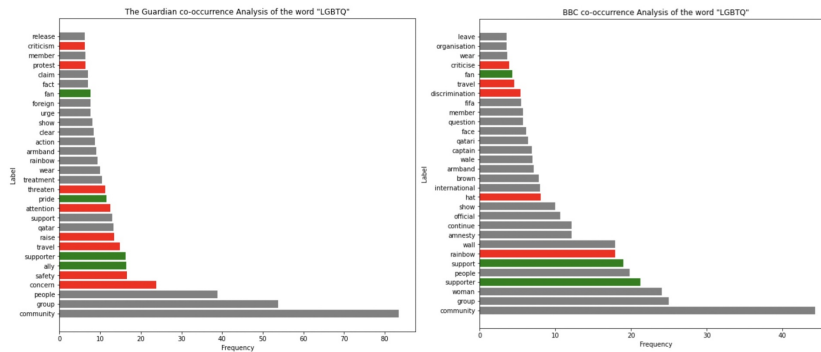

Fig. 7: *Co-occurrence* Analysis of the word "LGBTQ" in BBC and The Guardian

In order to conduct a more detailed analysis, we performed an co-occurrence analysis on the word "LGBTQ" to identify the most frequently co-occurring

words. We found that the word "travel" was the most common co-occurring word in both The Guardian and BBC. However, The Guardian's articles also had negative co-occurring words such as "concern", "safety", "attention", "threaten", and "criticism" associated with "LGBTQ." In contrast, for BBC, some of the top co-occurring words were "rainbow", "hat", "discrimination", "criticize", and "travel". It was interesting to note that "travel" was a common co-occurring word in both media outlets when addressing "LGBTQ rights," which made sense given the safety concerns raised in Western countries about LGBTQ people traveling to Qatar.
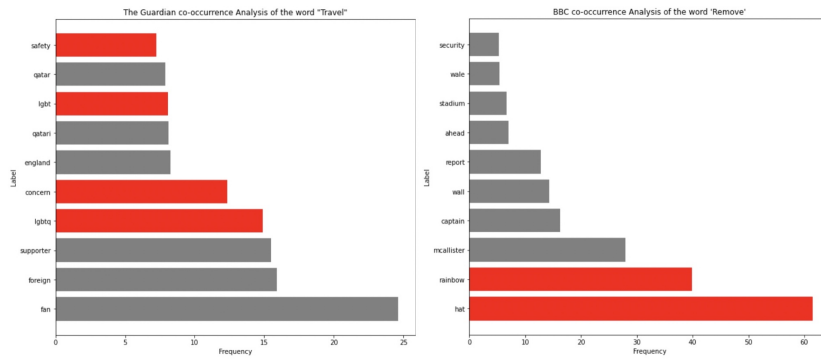

Fig. 8: *Co-occurrence* Analysis of the words "Travel" and "Remove" in BBC and The Guardian

Additionally, we observed that the words "LGBTQ","concern" and "safety" were highly co-occurring with the word "travel" in the articles published by the Guardian regarding the subtopic "LGBTQ rights", as shown in the above plot on the left side. This is because concerns were raised about the safety of LGBTQ people traveling to Qatar. We also noted that the two words "rainbow" and "hat" were the most common co-occurring words with the word "remove" in the BBC's articles about "LGBTQ rights". This is interesting because fans were frequently asked to remove rainbow flags and hats before entering the stadiums.

On a positive note, we found that many positive words such as "ally," "supporter," "pride," and "fan" were the most common co-occurring words with the word "LGBTQ" in The Guardian's articles. Similarly, for the BBC, words like "supporter," "support," and "fan" were top co-occurring words with the word "LGBTQ" as shown in the figure 7. This may explain the positive tone of both media outlets when it comes to reporting on "LGBTQ rights."

The table in figure 6 reveals that the BBC has expressed both positive and negative attitudes towards the "alcohol ban" subtopic more frequently compared to The Guardian.
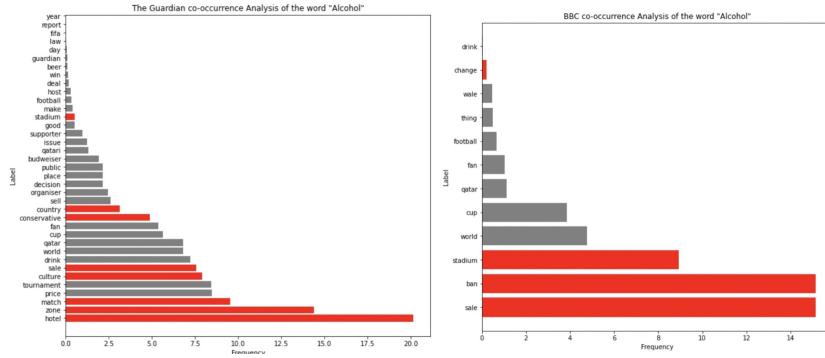

Fig. 9: *Co-occurrence* Analysis of the word "Alcohol" in BBC and The Guardian

In articles published by The Guardian that address the "alcohol ban" topic, the top co-occurring words with "alcohol" are "hotel," "zone," and "match." These words suggest that hotels and specific zones may offer alcohol for sale. Other interesting co-occurring words that were observed in the Guardian articles related to the "alcohol ban" topic include "culture" and "conservative," which refer to the Qatari culture and its conservative nature due to the majority of its citizens being Muslims, and Islam forbids alcohol consumption.

On the other hand, in articles published by the BBC addressing the "alcohol ban" topic, the top co-occurring words with "alcohol" are "sale," "ban," and "stadium." These words indicate that the articles are discussing the ban on selling alcohol in stadiums. It can be inferred that the word "ban" is a top negative co-occurring word with "alcohol," which may explain the higher level of negativity in BBC articles compared to those in The Guardian.

6.5 Sentiment analysis by the subtopic ("corruption allegations") by The Guardian and Al Jazeera

To summarize the corruption allegations shortly, Qatar was incriminated of buying their World Cup bid in 2010 by bribing FIFA officials and several members of football associations worldwide to "sportwash" their image as a nation [16]. The Guardian and Al Jazeera both covered the topic of "corruption allegations" related to Qatar's winning bid to host the 2022 World Cup. The Guardian, however, reported on the topic significantly more frequently than Al Jazeera did as shown in the table in the table in the figure 10. Interestingly, both news outlets used a similar mix of positive and negative tones when covering the allegations.

In order to gain a deeper understanding of the findings, we performed co-occurrence analysis on the articles published by both media outlets which discuss the subject of "corruption allegations."

Al Jazeera vs The Guardian: sentiment analysis by the subtopic (Corruption allegations)

SUBTOPIC		Al Jazeera	The Guardian
Corruption	Avg pos part (%)	0.09	0.10
	Avg neg part (%)	0.12	0.09
	Avg pos part title(%)	0.07	0.08
	Avg neg part title(%)	0.17	0.15
	Frequency(%)	0.01	0.05

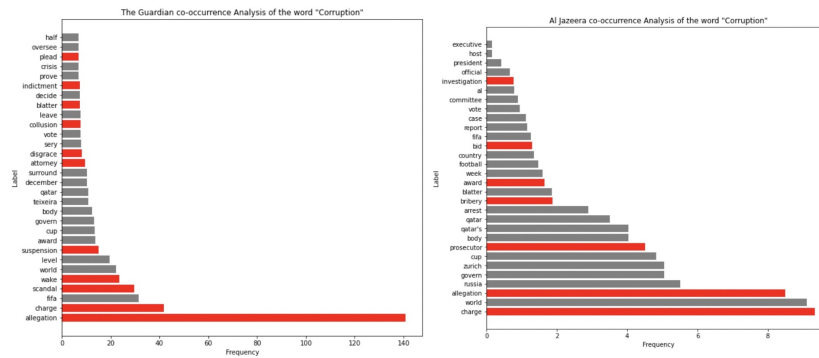Fig. 10: *Sentiment Analysis* by The Guardian and Al Jazeera regarding the subtopic "corruption allegations"

Fig. 11: *Co-occurrence* Analysis of the word "Corruption" in Al Jazeera and The Guardian

When examining Al Jazeera articles that discuss the topic of "corruption allegations," it is evident that the most frequently occurring words that are associated with "corruption" are "charge," "allegation," "prosecutor," "bribery," "award," "bid," and "investigation". Similarly, in The Guardian's coverage of the topic, the words "charge" and "allegation" are also frequently co-occurring with "corruption". Besides the fact that The Guardian covered the topic of "corruption allegations" in significantly more articles than Al Jazeera did, it also employed stronger language when discussing the matter, utilizing terms such as "scandal", "wake", "suspension", "disgrace", "collusion", "Blatter", "indictment", and "plead" as evidenced in the right-hand side of the above plot.

Concluding the section of sentiment analysis on specific topics, the mixed results make it difficult to confirm or reject our H3 regarding this part. On one hand, very different sentiments can be seen regarding the topic of workers' rights, on the other hand topics like corruption allegations have similar scores within the news outlets.

6.6 Named Entity Recognition

NER methods were used to extract the most mentioned **people**, **organisations** and **media outlets** in the articles published by each media outlet. we observed a recurrent trend across all three media outlets. The persons most frequently mentioned include **FIFA officials**, **politicians** of the host country and other countries, **former footballers**; In terms of organisations, the top entities are the **supreme committee** and the **FBI**; in terms of media outlets, **Al Jazeera** and **BBC** were frequently mentioned.
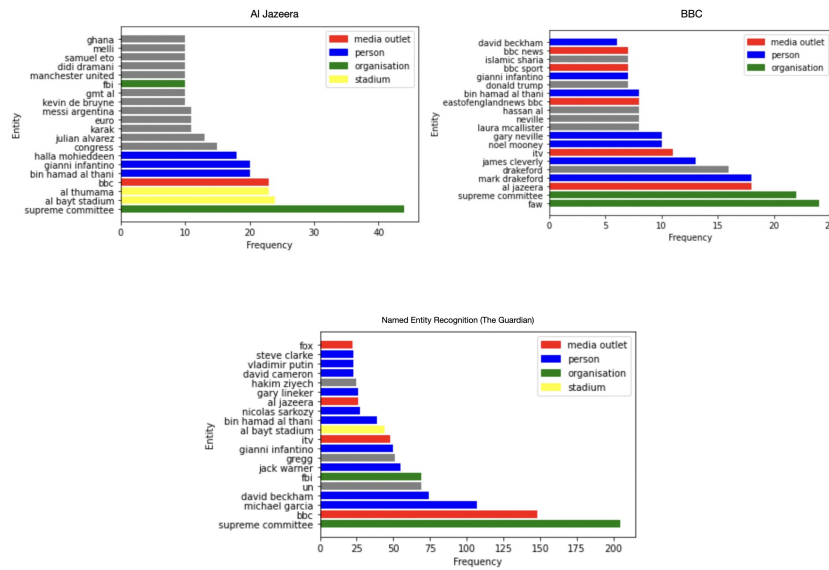

Fig. 12: *Top-20* most frequent entities by media outlet. In each plot, x-axis: frequency; y-axis: ordered from the bottom up.

Figure 12 illustrates the entities that were mentioned most frequently by different media outlets. Examining the results, it is noteworthy to observe the local inclination of each media outlet. For instance, Al Jazeera emphasized organizations like the **Supreme Committee** and **FBI**, particularly the **Supreme Committee**; it should be mentioned that the **Supreme Committee** is the organization responsible for delivering the necessary infrastructure and planning for Qatar's hosting operations. Regarding individuals, FIFA officials such as **Infantino**, politicians like the **Emir of Qatar**, and journalists such as **Halla Mohideen** were also named by Al Jazeera. Additionally, two stadiums, namely **Al Thumama** and **Al Bayt**, used for hosting the World Cup were also mentioned many times. The media outlet **BBC** was also referred to multiple times by Al Jazeera.

Regarding the BBC, organisations such that **FAW** (The football association of Wales) and the **Supreme Committee** are named, as well as several individ-

uals such as politicians like the **Emir of Qatar** and **Mark Drakeford** (a Welsh politician),**James Cleverly** (British politician and Army Reserve officer), **Noel Mooney** (the chief executive of the Football Association of Wales (FAW)). Former footballers like **Gary Neville** and **David Beckham** were also mentioned many times. Additionally, the BBC referenced multiple media outlets such as **Al Jazeera**, **ITV**, **East of England News BBC**, **BBC Sport**, and **BBC News**

The Guardian named numerous entities and individuals, including governmental bodies like the **Supreme Committee** and the **FBI**, as well as political figures such as the **Emir of Qatar**, **Nicolas Sarkozy**, **David Cameron**, and **Vladimir Putin**. The articles also referenced FIFA officials, specifically **Michael Gracia** and **Gianni Infantino** multiple times. In addition, former footballers, like **David Beckham**, **Gary Lineker**, and **Steve Clarke**, were named repeatedly. The media outlets, **BBC**, **ITV**, and **Fox**, were also mentioned several times. Furthermore, the articles made multiple references to **Al Bayt Stadium**, which was inaugurated in time for the matches during the 2022 World Cup.

On the other hand, we have noticed shared elements among the three media platforms. These include the **Supreme committee**, the **FBI**, the **emir of Qatar**, the names of **FIFA officials**, the names of the **stadiums** used during the World Cup, and media organizations like **Al Jazeera and BBC.** Rejecting the fourth hypothesis presents a significant challenge due to the numerous shared entities found across the three media outlets. For instance, "the supreme committee" is ranked as a top named entity by all three media outlets, alongside many other entities that are similarly named. Additionally, our visualization is limited to the top 20 named entities of each media outlet, meaning that certain entities that appear within the top 20 of one media outlet may have lower frequencies within other media outlets.

7   Discussion

After carrying out our research, we conducted a second analysis of our results to determine if and how we had answered our research question. Our findings indicate that we were able to confirm our research questions, as we identified various topics, sentiments, and entities within the data sets of the three news outlets that we analyzed. However, it is essential to take into consideration the backgrounds of these news outlets and how they may impact their content.

For instance, we observed that Al Jazeera, which is owned by the World Cup host Qatar, is unlikely to publish negative content about their own nation and owners. This is in stark contrast to the two British news outlets that we examined, particularly The Guardian, a left-leaning outlet known for its investigative reports rather than sports news, which may produce a diverse range of topics. However, it is critical to note that we were able to confirm our initial suspicions while also shedding light on new findings, such as differences between the British news outlets and similarities in reporting across all three outlets.

Despite our successes, we faced certain limitations during our research. A larger data set with more media outlets from various countries and backgrounds would have greatly enhanced our project, given additional time and resources. Although we initially investigated other news outlets, we discovered that they either focused exclusively on sports or had insufficient data available for analysis.

References

[1] David Alonso del Barrio and Daniel Gatica-Pérez. 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, page 35â43, New York, NY, USA, 2022. Association for Computing Machinery.

[2] The BBC. Learn more about what we do - about the bbc, 2022. Last accessed 12-March-2023.

[3] The BBC. Policies and guidelines - production and delivery. `https://www.bbc.com/delivery/policies-guidelines`, 2023. Last accessed 12-March-2023.

[4] Syarafina Dewi and Dede Brahma Arianto. Twitter sentiment analysis towards qatar as host of the 2022 world cup using textblob. *Journal of Social Research*, 2(2):443–455, 2023.

[5] Susan Dun, Hatim Rachdi, Shahan Ali Memon, Rohith Krishnan Pillai, Yelena Mejova, and Ingmar Weber. Perceptions of fifa menâs world cup 2022 host nation qatar in the twittersphere. *International Journal of Sport Communication*, 1(aop):1–10, 2022.

[6] Hans W. A. Hanley, Deepak Kumar, and Zakir Durumeric. "a special operation": A quantitative approach to dissecting and comparing different media ecosystems' coverage of the russo-ukrainian war, 2022.

[7] Al Jazeera. About us. `https://www.aljazeera.com/about-us`, 2023. Last accessed 12-March-2023.

[8] Shaimaa Khalil. World cup 2022: Alcohol sales banned at world cup stadiums in qatar. `https://www.bbc.com/sport/football/63674631`, Nov 2022. Last accessed 14-March-2023.

[9] Michał Marcin Kobierecki and Piotr Strożek. Sports mega-events and shaping the international image of states: How hosting the olympic games and fifa world cups affects interest in host nations. *International Politics*, 58(1):49–70, 2021.

[10] The University of Manchester Library. Guardian (formerly manchester guardian) archive. `https://web.archive.org/web/20210428161019/https://www.library.manchester.ac.uk/rylands/special-collections/exploring/a-to-z/collection/?match=Guardian%2B%28formerly%2BManchester%2BGuardian%29%2BArchive`, Apr 2021. Last accessed 12-March-2023.

[11] Jörg Matthes, Ruta Kaskeleviciute, Desirée Schmuck, Christian von Sikorski, Claudia Klobasa, Helena Knupfer, and Melanie Saumer. Who differentiates between muslims and islamist terrorists in terrorism news coverage? an actor-based approach. *Journalism Studies*, 21(15):2135–2153, 2020.

[12] Connor Nickerson. Media portrayal of terrorism and muslims: a content analysis of turkey and france. *Crime, Law and Social Change*, 72:547–567, 2019.

[13] Nadezhda Ozornina and Alexander Mannin. Framing the image of russia in the british media during the world cup 2018. *Russian Journal of Communication*, 12(2):121–136, 2020.

[14] House of Parliament. Chapter 2: The importance of the bbc - publications.parliament.uk. `https://publications.parliament.uk/pa/ld200506/ldselect/ldbbc/50/5005.htm`, 2005. Last accessed 12-March-2023.

[15] Matthew Smith. How left or right-wing are the uk's newspapers? `https://yougov.co.uk/topics/politics/articles-reports/2017/03/07/how-left-or-right-wing-are-uks-newspapers`, Mar 2017. Last accessed 12-March-2023.

[16] Freya Thomson. Exposing the qatar world cup corruption. `https://www.openaccessgovernment.org/exposing-qatar-world-cup-corruption/147871/`, Nov 2022. Last accessed 14-March-2023.

[17] Jim Waterson. Guardian most trusted newspaper in britain, says industry report. `https://www.theguardian.com/media/2018/dec/17/guardian-most-trusted-newspaper-in-britain-says-industry-report`, Dec 2018. Last accessed 12-March-2023.

[18] Robert F. Worth and David Kirkpatrick. Seizing a moment, al jazeera galvanizes arab frustration. `https://www.nytimes.com/2011/01/28/world/middleeast/28jazeera.html`, Jan 2011. Last accessed 12-March-2023.

Allocation of work

Abdulnaser Sabra: Web Scraping, Topic modelling, sentiment analysis, Named Entity Recognition, Report: Research Question, Results

Tobias Ehrmeier: Web Scraping, Report: Introduction, Related Work, Research Question, Methodology, Data, Results (small additions), Discussion