## Introduction:

The report presents a detailed analysis of Al Jazeer’s coverage of the Gaza conflict from two angles: article length trends and the use of conflict versus peace-related language. Using two different datasets, we looked at how media focuses and narrative have changed over time. The first part examines article lengths and how often they were published to understand the volume and intensity of coverage. The second part analyzes the frequency of specific conflict peace words in text using 1-gram analysis.

By combining these two methods, the report shows not only how much attention Gaza received but also which themes were most prominent. This double approach gives important insight into journalistic choices and how conflict is framed by a major news outlet.

This report also includes documentation for each part of the analysis, explaining how the work was carried out. In addition, it provides an overview of the folder structure of the project repository to help readers navigate the files and understand how everything is organized.

## Documentation:

Name: Shabir Karim

Shabir organized the workflow for the project. He began by forking the project repository and setting up the team’s Trello board to manage tasks. He ensured an organized and collaborative workflow. During the data selection phase, Shabir reviewed all available data frames to identify which would best support a meaningful analysis. After careful consideration, he selected the length dataframe for further exploration and analysis.

Firstly, he worked with the “length-year-month.csv” file. In his initial script, he explore the data, and created three visualizations:

1. A line graph of total article length over time.
2. A line graph of average article length.
3. A bar chart showing the top 5 months by total length.

However, after reviewing the data and visuals, he concluded that the visuals lack a sufficient argument, they are not showing something meaningful.

In his second exploration script, he conducted the following things.

1. Combined year, month, and day columns into a datetime object.
2. Created a bar chart showing the number of articles published per day.
3. Calculated and plotted average article length by month, both as a line chart and bar chart.
4. Identified that data from 2017 to early 2021 was sparse and therefore decided to exclude it in the final version.

For the final presentation script, he filtered the dataset to include the records from May 2021 onwards, as it gives clear visuals. In final presentation script, he produced the following three visualizations:

1. **Bar Chart** of daily articles counts with a color gradient and annotated start point.
2. **Line Chart** showing the trend of average article length over time, with an annotation marking the peak.
3. **Bar Chart** of average article length per month, annotated to highlight the end-of-year peak.

All these plots were saved to an appropriate folder. Moreover, the code was written using IDLE with structured and clear comment, AI Documentation is also included.

Shabir's contribution gives the project a ground dimension. It shows how article volume and length fluctuate over time, which aids in understanding the rhythm of media attention across the timeline of the Gaza conflict.

Name: Atiqa Rani

Atiqa analyzed the presence and evolution of conflict and peace related terms in the Gaza article corpus using 1-gram. In the exploration part, she loaded the dataset and created a new date column from the separate year, month, and day fields to enable filtering. She defined two word categories, “conflict” and “peace.” It was based on manually selected keywords such as “war,” “attacks,” “military,” and “united,” “nations,” respectively. Each 1-gram was labeled accordingly using a custom function. To understand how these word categories appeared over time, she grouped the data by year and plotted bar charts using Plotly Express to visualize total frequencies.

In the final presentation script, she refined her analysis by discussing both total and relative frequency charts, by yearly and monthly. The relative frequencies were calculated by dividing the category-specific counts by the total 1-gram counts for each time unit. She used material from course slides and used AI to adapt code and do plotting. The work provided a focused lens into how the discussion of conflict and peace changed over time within Gaza-related reporting.

## Folder Structure:

**Important note: While working we saved our project, not in a folder. After completing, in order to organize we placed our script in a script folder, and from there it is no reading dataframes, even it is not giving option to go to data frame folder and files. If we wish to run it again, we have to move it out from the folder and then we can run the script. Please consider this problem, Thank You.**

|  |
| --- |
| ## Folder Structure  FASDH25-portfolio3/  ├── Scripts/ # Python scripts for n-gram and length analysis  │ ├── 1-gram\_Script/  │ │ ├── Exploration\_script\_1gram.py  │ │ └── Final\_presentation\_script\_1gram.py  │ └── length\_Script/  │ ├── Final\_Length\_Presentation\_Script.py  │ ├── Final\_Length\_Exploration\_Script.py  │ └── Initial\_Exploration\_length-year-month.py  │  ├── AI Documentation/ # AI documentation written by team members  │ ├── AI Documentation\_Atiqa.docx  │ └── AI Documentation\_Shabir.docx  │  ├── Visuals/ # All visualizations and figures  │ ├── Final\_Visuals/  │ │ ├── 1-gram\_Visuals/  │ │ │ ├── bar\_relative\_monthly.html  │ │ │ ├── bar\_relative\_yearly.html  │ │ │ ├── bar\_total\_monthly.html  │ │ │ └── bar\_total\_yearly.html  │ │ └── Length\_final\_visuals/  │ │ ├── avg\_article\_length\_by\_year\_month.html  │ │ ├── articles\_per\_day.html  │ │ └── trend\_article\_length\_over\_time.html  │ └── Test\_Visuals/  │ ├── fig1\_total\_coverage.PNG  │ ├── fig1\_total\_coverage.html  │ ├── fig3\_Top5\_months.PNG  │ ├── fig3\_Top5\_months.html  │ ├── fig2\_average\_length.PNG  │ └── fig2\_average\_length.html  │  ├── Report/ # Final written report  │ └── Report\_Mini\_Project3.docx  │  └── data/ # Datasets and supporting files  ├── dataframes/  ├── utilities/  └── articles/ |

## Analysis:

1. **1-gram Frequency Analysis of Conflict vs. Peace Discourse - Atiqa Rani**

This analysis explores the language used in Al Jazeera Gaza corpus. It focuses on how terms related to conflict compare to those associated with peace and diplomacy. The foundation of this analysis is the 1-gram dataset (1-gram.csv). This data contains the frequency counts of individual words (1-grams) across all articles. The data was then cleaned by converting the year, month and day columns into single datetime format. To visualise clearly, only articles from May 2021 onwards are selected, because the data before 2021 was insignificant there was less data and didn’t conclude a trend there.

To study my theme, one important step was grouping certain individual words, or 1-grams. These words were sorted into two main categories: those related to conflict and those linked to peace. The conflict group includes terms like ‘war', 'attacks', 'attack', 'children', 'ground', 'palestinians', 'military', 'forces', 'besieged', 'hamas', and 'civilians'. These words were chosen for their direct association with hostilities, military actions, and the human impact of conflict. In comparison to this, the peace group includes terms like 'united', 'nations', 'international', 'well', and 'end'.

The method of categorizing and analyzing 1-gram frequencies, especially through relative frequency. This method is valuable for distant reading. Its main strength is that it allows researchers to track broad shifts in journalistic focus and discourse over time using quantitative data.

By looking at how often conflict-related terms appear compared to peace-related ones, we can infer the narrative in the corpus. This comparison helps in highlighting which themes were emphasized more in the coverage. Here, the relative frequency was essential because it accounts for changes in total number of articles. This way, any trends we see are based on actual shifts in word usage.

However there are limitations to this approach as well. Relying only on 1-grams can oversimplify language. Words often get their meaning from context, which this method misses. For example, the word “end” could mean “end of conflict” (peace) or “end of lives” (conflict). Our method can't tell the difference.

The word “Palestinians” is neutral, but we included it in the conflict category. This is because it often appears in conflict-related articles, not because the word itself is conflict-related. Furthermore, the 'peace' word list is smaller and less clear than the 'conflict' list. Some peace words are vague or relate more to diplomacy. This may reflect the source’s focus on conflict but might also miss some peace language, exaggerating the imbalance.

**Visuals and their description:**

**Relative Frequency of Conflict and Peace Words (Yearly) (fig 1)**

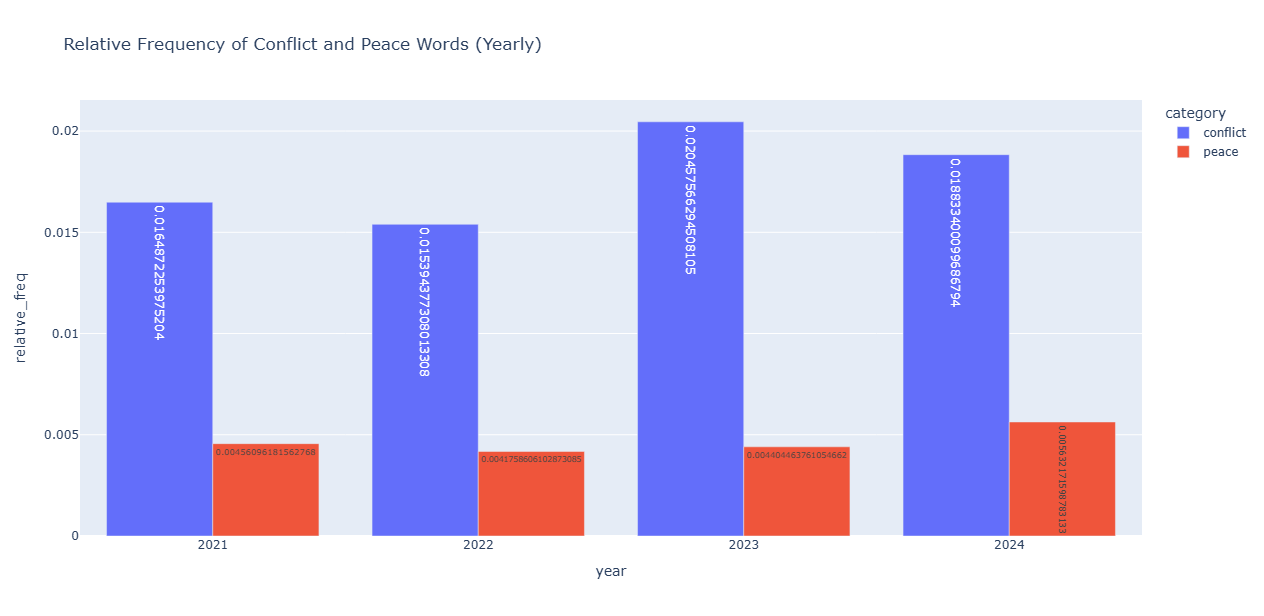
****

Fig 1 presents a bar chart showing the relative frequency of conflict and peace words each year, adjusted for the total number of words. This normalization helps provide a clearer picture of the themes over time. The conflict category consistently represents a much larger portion of the language used compared to the peace category. While the total number of words varies from year to year, the proportion of conflict terms often stays between 5% and 7% or higher. In contrast, peace terms remain much lower, usually under 1% to 2%. This pattern shows that even in years with fewer articles, Al Jazeera’s coverage maintained a strong focus on conflict, highlighting its thematic importance in the corpus.

**Relative Frequency of Conflict and Peace Words (Monthly) (fig 2)**

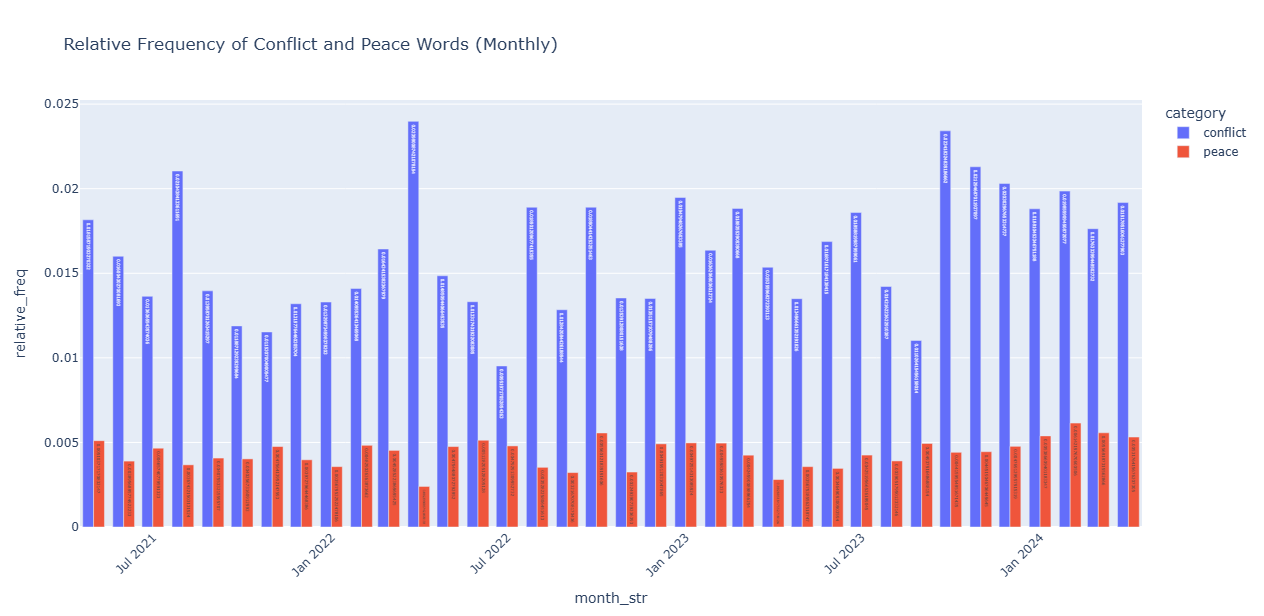
****

Fig 4 focused in on the relative frequency of conflict and peace words by month, giving a more detailed view of language use over time. The chart clearly shows that during major events, like May 2021, conflict-related words spike dramatically, making up a much larger share of the coverage than usual.

Even in quieter months, conflict terms still appear far more often than peace terms.

This highlights how Al Jazeera’s reporting on Gaza consistently focuses heavily on conflict. However, peace and diplomacy language remain much less prominent, even when accounting for overall article volume.

Al Jazeera’s coverage of Gaza, especially from May 2021 onward, is strongly focused on conflict. Words related to peace or diplomacy get much less attention, both in how often they appear and in their overall share of language. The high frequency of conflict terms shows a clear journalistic focus on reporting hostilities, military actions, and their human impact. During periods of intense activity, the use of conflict-related language spikes dramatically. This highlights how coverage tends to concentrate on confrontation.

Looking at the language over time suggests the main story told by Al Jazeera about Gaza focuses on ongoing conflict. At the same time, themes of resolution or diplomacy are noticeably less present. This pattern likely reflects not only real events but also the editorial choices and focus of this influential news outlet.

1. **Article Length Analysis – Shabir Karim**

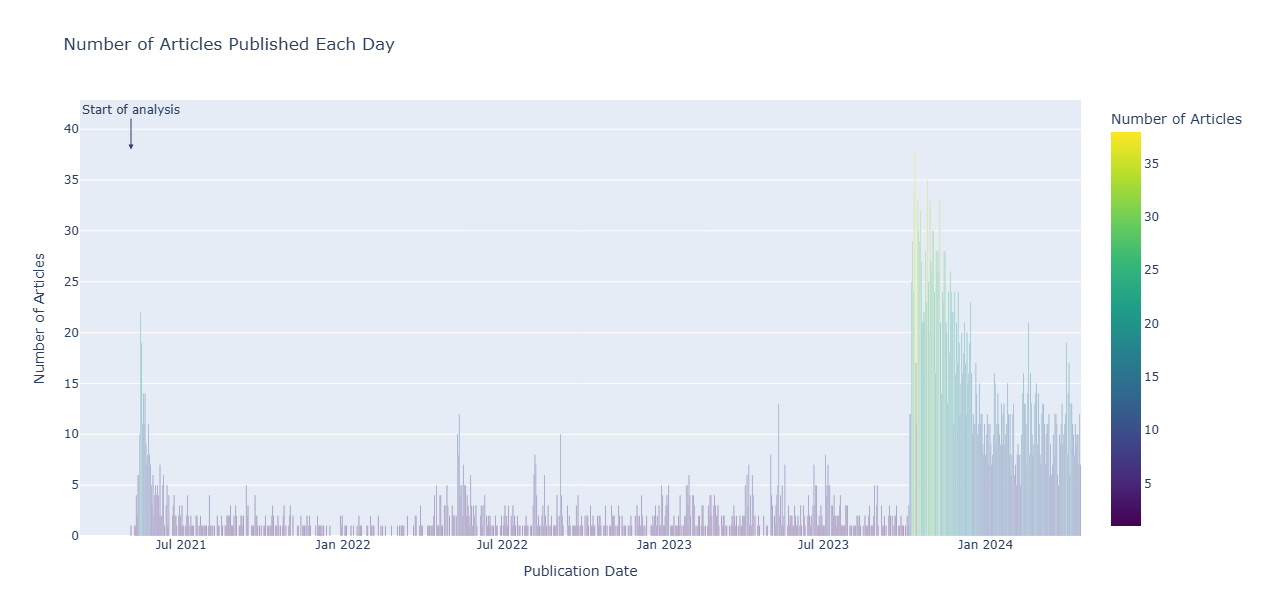
This is a dataset of articles, collected from Al Jazeera. The data is saved in a file called length.csv. It includes the publication year, month, day, and the length of each article.

The information shows two things: how many articles were published each day and how the average article length changes over time. It uses the Pandas library for data manipulation and Plotly for creating interactive visualizations.

This method helps visualize trends in how many articles are published and their lengths over time. By grouping articles by date and calculating averages, it uncovers patterns that aren’t obvious in the raw data. Using Plotly makes the charts interactive, which makes it easier to explore and understand the information. There are many limitations of this method, like this analysis only used the data in the particular file. Without using information like article topics, it is hard to link trends to real world events. Merging this dataset with topic model data frame could be a good option for analysis.

**Visual and their description:**

**Number of Articles Published Each Day (fig 3)**

****

The bar chart clearly shows the number of articles Al Jazeera published about Gaza each day. Between July 2021 and July 2023, the publication rates were relatively low. This period likely reflects regular reporting on humanitarian issues, political developments, or occasional tensions. However, it doesn’t show the intense coverage that comes with full-scale conflict. The sharp increase in articles published in November 2023 matches the major escalation of the conflict in Gaza after October 7th. As a leading regional news source, Al Jazeera ramped up its coverage. It reported on the Israeli military actions, the growing humanitarian crisis, and the fast-moving geopolitical reactions.

## Conclusion:

The findings of this project reveal a clear and consistent narrative in Al Jazeera’s coverage of Gaza. The article length analysis shows shifts in media attention, with spikes during major escalations in the conflict. At the same time, the 1-gram frequency analysis points to a strong focus on conflict-related language, while peace-related terms appear much less often. This pattern highlights how the coverage prioritizes hostilities, military actions, and their human impact.

While there are some limitations, such as the simplification in using 1-grams and the lack of topic-level analysis in the article length data. The combined findings offer a meaningful view of how the conflict is framed. Overall, the project shows how editorial choices shape public understanding of complex geopolitical events like the Gaza conflict.