# Religion in the News

### Betsy Fridman

### December 10th, 2025

## 1    Introduction

The representation of religion in contemporary news media plays an important role in shaping public opinion. News coverage often promotes certain narratives, which may contribute to stereotypes or biased portrayals of different religious groups. Understanding these patterns is essential for analyzing how major religions such as Islam, Christianity, and Judaism are discussed in the news, and for identifying potential framing biases.

This study presents an analysis of English-language news articles that contain keywords relating to religion. The goal is to investigate the frequency of mentions of major religions, and the meaning and sentiment of the context in which these terms are used. This analysis seeks to examine how each of the three religions is framed and how these framing patterns vary across religious groups. The research addresses several key questions. Which religion is mentioned most frequently in this corpus of news articles? In what contexts are these terms mentioned? Does the context around each religion carry more positive or negative sentiment? What insights do these patterns provide about the narratives that the news articles are promoting?

The analysis combines lexical frequency and statistical association measures (PMI), and concludes with a sentiment analysis of the immediate context of the religions. By analyzing the statistically significant collocates, the study provides a deep understanding of which ways religions are covered most, and what affective tone the media uses when covering them. The result of the analysis was that all three religions experienced a net negative affective framing, but Islam was associated with the most intense negative language and the greatest volume of negative contextual vocabulary.

# 2    Data

The corpus utilized in this study is from the Guardian News Articles dataset on Kaggle, which contains 149,839 articles published between 2016 and 2022. The data was filtered to retain the articles that contained at least one keyword related to Islam, Christianity, or Judaism. The resulting corpus consisted of 14,387 articles with a total of 14,294,669 raw tokens. The initial preparation of the analysis involved standardizing the data. Errors in the text encoding were corrected using character-replacement functions to ensure that the data was consistent. All text was converted to lowercase, tokenized, and stripped of punctuation and symbols.

Before the analysis, the initial keyword counts were as follows: Islam had 4,407 mentions, Christianity had 6,673 mentions, and Judaism had 3,689 mentions. After inspecting the context surrounding "christian", it was clear that the word frequently appeared as a proper name. Hence, "christian" was removed from the keyword set, leaving only "Christians" and "Christianity" in this group. This reduced the total count for the Christianity group to 2,782 mentions, to keep the focus on the religious context.

# 3    Methods

The objective of this analysis was to compare the context that the three religions are used in, and whether their immediate context has more of a positive or negative connotation. To ensure a more accurate representation, more than one variant of saying each religion was used. All lexical variants were normalized to a single representative token for each religion: Islam, Muslim, Muslims, and Islamic were all replaced with Islam; Christianity and Christians were replaced with Christianity; Jewish, Jew, Jews, and Judaism were replaced with Judaism.

The first part of the analysis examined contextual patterns. To find words strongly associated with each religion, also known as the collocations of a word, a statistical measure called Pointwise Mutual Information (PMI) was used. PMI identifies words that co-occur with a target term more frequently than expected under an assumption of independence, providing a reliable measure of how strongly those words are linked. A window of five words to the left and five words to the right of the religious term was used for our collocation analysis. The results were strictly filtered to keep only the collocates with a minimum of 20 occurrences and a PMI $\geq 5$. From this, we now had three lists of statistically significant collocates, which were used to understand the qualitative context around each religion. Visualizations were created to display the strongest collocated words and their association strengths.

The analysis then focused on affective tone. The emotional quality of the sta-

tistically significant collocates was measured using the AFINN Sentiment Lexicon. This lexicon is an expert-developed resource that assigns a score to common English words, ranging from -5 (highly negative connotation) to 5 (highly positive connotation). A score of 0 indicates a neutral term in the lexicon, as well as collocates that are not present in the lexicon. The AFINN score was then merged with the list of collocates. The primary metric used to compare the religions, the mean sentiment score, computed as the average AFINN value of all associated words. This measure provides a direct way to look at the overall affective tone of the language surrounding the three religions.

# 4    Results

Taken together, the collocates in Tables 1a, 1b, and 1c revealed distinct thematic patterns for each religious group. As seen in Table 1a, the top collocates for the Islam node were heavily concentrated in terms related to internal group identity and conflict, including specific sects and political associations (e.g., *ahmadiyya*, *khorasan*, *brotherhood*, and *militancy*). This suggests a dominant media focus on political activity and sectarianism within Islamic communities. In contrast, the collocates for Christianity, seen in Table 1b, highlighted themes of persecution and conversion dynamics (e.g., *persecuted*, *persecution*, *converted*), indicating that the media narrative often focuses on the group's vulnerability and also their expansion. The strong co-occurrence with *hindus* suggests a particular emphasis on interfaith conflict or activity in South Asia. Judaism's top associations, as seen in Table 1c, involved terms related to history, culture, and social structure (e.g., *orthodox*, *synagogues*, *ghetto*, *rabbi*), alongside legal terms like *deputies*. This framing suggests a media narrative centered on institutional structure and history.

Table 1: Top 9 collocates for the three religious groups, ranked by PMI

(a) Islam

| Collocate | Collocation Freq | Total Freq | PMI |
|---|---|---|---|
| ahmadiyya | 23 | 25 | 9.889935 |
| khorasan | 35 | 39 | 9.854110 |
| brotherhood | 130 | 182 | 9.524802 |
| registry | 34 | 60 | 9.190801 |
| fundamentalism | 33 | 67 | 8.988534 |
| slimani | 51 | 110 | 8.901295 |
| militancy | 31 | 78 | 8.679023 |
| hindus | 63 | 164 | 8.629957 |
| sunni | 86 | 225 | 8.622713 |

(b) Christianity

| Collocate | Collocation Freq | Total Freq | PMI |
|---|---|---|---|
| evangelical | 105 | 535 | 10.722856 |
| hindus | 22 | 164 | 10.173885 |
| persecuted | 24 | 210 | 9.942722 |
| persecution | 37 | 520 | 9.259091 |
| converted | 20 | 361 | 8.898079 |
| religion | 37 | 1835 | 7.439894 |
| christianity | 30 | 1660 | 7.281929 |
| islam | 146 | 13861 | 6.503086 |
| religious | 39 | 3754 | 6.483195 |
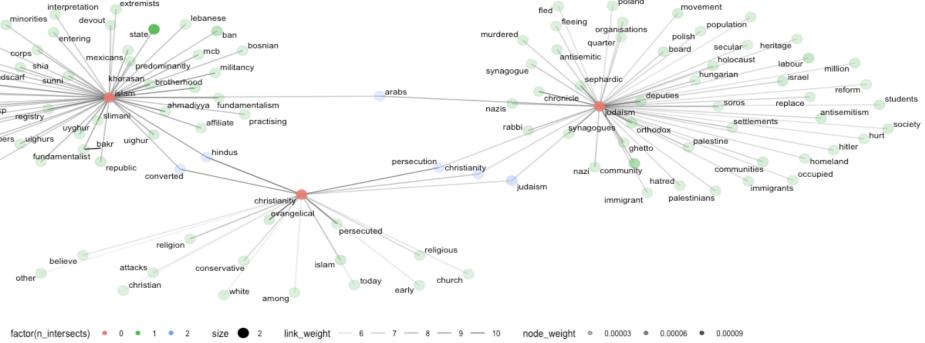
(c) Judaism

| Collocate | Collocation Freq | Total Freq | PMI |
|---|---|---|---|
| deputies | 91 | 142 | 10.144740 |
| sephardic | 26 | 44 | 10.027700 |
| chronicle | 39 | 84 | 9.679777 |
| orthodox | 130 | 368 | 9.285498 |
| synagogues | 22 | 126 | 8.268844 |
| ghetto | 20 | 133 | 8.053338 |
| arabs | 24 | 172 | 7.945390 |
| soros | 28 | 228 | 7.761157 |
| nazis | 75 | 647 | 7.677889 |

Visualizing these associations provides further insight into the structural framing. Figure 1, which compares the networks of the three religions, highlights vocabulary overlap and makes clear some possible themes shared between them. The visual analysis indicates that there is limited vocabulary overlap, which indicates that the narratives told of each religion are fairly different. Christianity shares links with Islam through *hindus* and *islam*, while it shares the term *judaism* with the Judaism node. Judaism and Islam share only the term *arabs*. This shared vocabulary, although minimal, suggests that when themes are shared it is often linked through specific geopolitical contexts or regional interfaith conflict.

Figure 1: Collocation network showing the unique and similar collocates of
Islam, Christianity, and Judaism



To assess the emotional quality of these contexts, the AFINN Sentiment Lexicon
was applied to the collocate lists. The aggregated sentiment results are summarized
in Table 2.

Table 2: Islam has a lower mean sentiment score and has a stronger total
weighted PMI

| Religion | Mean Sentiment Score | Total Weighted PMI | Negative Words | Positive Words |
|---|---|---|---|---|
| Islam | -0.181 | -182.65 | 118 | 3 |
| Christianity | -0.150 | -25.53 | 2 | 0 |
| Judaism | -0.071 | -38.59 | 5 | 3 |

The mean sentiment score revealed context with negative sentiment across all
three religions, although the degree of negativity varied. Islam recorded the most
negative score (Mean Sentiment = -0.181), followed closely by Christianity (Mean
Sentiment = -0.150), and then Judaism (Mean Sentiment = -0.071), which can
be seen in Figure 2. This negative framing is reinforced by a total weighted PMI
of -182.65, which confirms that the most statistically strong associations are pre-
dominantly negative. Furthermore, Islam's negative affective context is driven by
a larger number of negative words, 118, compared to positive words, 3. Tables
3a and 3b present the most extreme sentiment-scored collocates, illustrating the
specific high-PMI words responsible for this affective tone.

Figure 2: Islam has the lowest average sentiment of strongest collocates
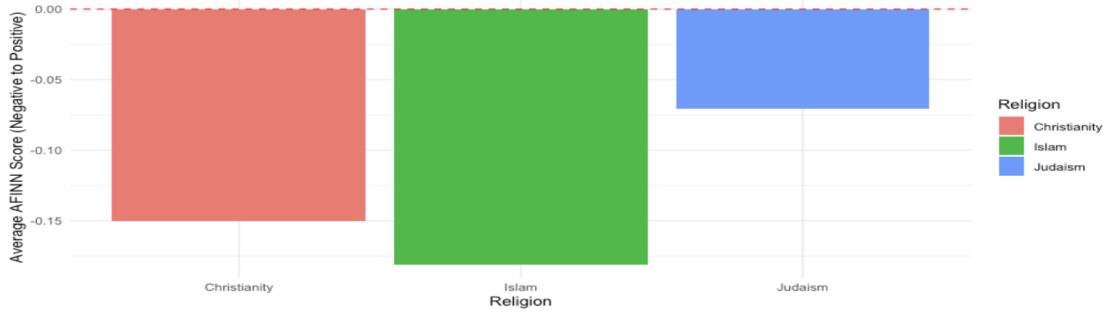


Table 3: Islam and Judaism have the most extreme values of sentiment in their collocates

(a) AFINN Score: -3

| Collocate | Religion | Score | Weighted PMI | Collocation Freq |
|---|---|---|---|---|
| violent | Islam | -3 | -17.122569 | 68 |
| conspiracy | Judaism | -3 | -16.437792 | 25 |
| terror | Islam | -3 | -16.347340 | 61 |
| evil | Islam | -3 | -15.831830 | 27 |
| killing | Judaism | -3 | -15.750113 | 31 |
| hate | Islam | -3 | -15.561751 | 82 |
| kill | Islam | -3 | -15.206344 | 33 |

(b) AFINN Score: 1 and 2

| Collocate | Religion | Score | Weighted PMI | Collocation Freq |
|---|---|---|---|---|
| prominent | Judaism | 2 | 11.213567 | 24 |
| prominent | Islam | 2 | 10.830565 | 36 |
| solidarity | Islam | 2 | 10.327198 | 22 |
| responsible | Judaism | 2 | 10.323653 | 26 |
| faith | Islam | 1 | 5.859020 | 125 |
| faith | Judaism | 1 | 5.095963 | 43 |

The intensity is evident when examining the most extreme negative words (AFINN Score -3) in Table 3a, which are heavily linked to conflict: *violent*, *terror*, *hate*, and *kill*. For comparison, the few positive words found across all religions are generally mild, such as *prominent* and *faith* (Table 3b). The number of negative words also shows a major bias: Islam is associated with 118 different negative words, while Christianity has only 2 (Table 2). This means the media uses a much larger and more varied negative vocabulary when discussing Islam. Furthermore, Christianity has zero positive words, suggesting that positive stories about that religion are either not covered or are described using strictly neutral language.

# 5 Discussion

The analysis shows that media coverage of major religions differs significantly in both the story it tells and the affective tone. These different framing patterns likely impact public perception. Islam was the most frequently mentioned religious topic in the final analysis corpus with 4,407 mentions, and the study finds that Islam experiences the most intense negative framing.

The qualitative analysis of the top collocates confirms that the media narratives speak of different aspects of each religion. Coverage of Islam appears tightly linked to political activity and sectarian issues, with top collocates such as *brotherhood*, *sunni*, and *militancy* suggesting a focus on internal conflict and geopolitical concerns. The narrative surrounding Christianity frequently emphasizes institutional vulnerability, with high-PMI terms like *persecuted*, *persecution*, and *hindus*. The mention of Judaism primarily emphasizes cultural identity and structure, reflected by collocates including *orthodox* and *synagogues*. These differences suggest that the news media segment religious discourse, positioning Islam as political and conflict-driven, Christianity as an issue of institutional defense and social vulnerability, and Judaism as a matter of history and community structure.

The quantitative findings on affective tone reveal a difference in distribution of bias. The mean sentiment score confirmed a net negative context across all three religions, but Islam recorded the most negative score. These patterns may have real-world implications, given prior evidence that media framing can influence stereotypes and public attitudes.

The intensity of the bias is evident, where the most extreme negative language is strongly associated with Islam. The volume of negative sentiment is far greater for Islam, as it has dramatically more negative collocates than the other religions. This difference in negative vocabulary volume indicates that while negative events may be covered for all groups, the news employs a much larger emphasis on reinforcing negative associations when discussing Islam. Furthermore, the absence of positive words for Christianity may show a media tendency to report on that religion using neutral or non-affective language.

Several limitations should be noted when interpreting these findings. First, the corpus may not be fully representative of global news coverage. Since all articles were sourced from The Guardian (2016–2022), the results are likely influenced by British editorial perspectives and cultural interests. The merging of religious words into one main keyword was necessary to count religion mentions, but may have obscured nuanced distinctions in how specific terms are used, or may not have fully captured the context of each religion as there are more related words that were unused. Finally, the use of the generalized AFINN lexicon may not fully capture the sentiment of the collocations, which could lead to mislabeling or underestimating certain biases.

7

This study establishes a baseline for affective framing in news media, although there are many possibilities for future research. One valuable next step would be to conduct a temporal analysis and to analyze this same corpus over time year-by-year, to determine if major events in the world correlate with observable changes in positive or negative sentiment, or just in general changes in thematic focus for specific religious groups. Additionally, an essential step would be to compare different news sources to see how they speak of each of these religions differently. Future research could also examine institutional or structural factors that might shape these framing patterns.

# References

[1] A. Kharosekar, *Guardian News Articles Dataset*, 2025. `https://www.kaggle.com/datasets/adityakharosekar2/guardian-news-articles/data`. Accessed: 2025-12-10.