Does the World Notice?

Analyzing Media Coverage of Global Humanitarian Crises Through Data Science

Ashritha Narne & Nikhitha Nagabhyru

April 2026

ABSTRACT

This project investigates whether international media coverage of humanitarian crises aligns with their measurable severity. Using a structured dataset of 78,667 news articles spanning ten crises from 2009 to 2025, we apply exploratory analysis, feature engineering, and regression modeling to identify the predictors of monthly media attention. We find that coverage is primarily driven by the timing of a crisis rather than its humanitarian scale: the onset period of a crisis is the strongest temporal predictor, while longer-running crises receive systematically less attention over time. Crises with larger internationally recognized funding needs attract more coverage, while the number of people affected shows little predictive power. For the two crises with available narrative data — Gaza and Ukraine — framing and sentiment analysis reveals stark asymmetries in how actors are portrayed. These findings contribute to a growing body of evidence that media agenda-setting in humanitarian contexts follows geopolitical and editorial logic rather than humanitarian need.

1. Introduction

In any given year, dozens of humanitarian crises unfold simultaneously across the world. Millions of people face displacement, famine, armed conflict, and disease in contexts that rarely reach the front pages of major newspapers. Meanwhile, other crises dominate international headlines for months on end, generating public outcry, donor mobilization, and diplomatic pressure. The question this project asks is simple but consequential: is media attention to humanitarian crises proportional to their severity?

The relationship between media coverage and humanitarian response is well established in the journalism and policy literature. When crises receive sustained media attention, they are more likely to attract funding, political engagement, and international intervention. When they do not, they risk being overlooked by the very institutions responsible for responding to them. Understanding what drives media salience is therefore not merely an academic exercise; it has real implications for how resources are allocated and whose suffering is acknowledged.

This project builds on the *Humanitarian Crisis Coverage Report* published by the Media and Journalism Research Center (Dragomir, 2025), which provided the underlying dataset for this analysis. That report documented patterns of

unequal coverage across ten crises using descriptive statistics and qualitative framing analysis. Our contribution is to extend that work using the tools of data science: systematic feature engineering, correlation analysis, and predictive modeling applied to a monthly-level dataset of 734 observations. We ask not only whether coverage is unequal, but which measurable characteristics of crises explain that inequality, and how well those characteristics can predict monthly article volumes across a decade of reporting.

RESEARCH QUESTIONS

- Does media coverage of humanitarian crises correlate with their humanitarian severity?
- What crisis characteristics best predict monthly media attention?
- How do narrative framing and entity portrayal differ between high-coverage crises?

2. Data

2.1 Source Dataset

The data for this project comes from the Media and Journalism Research Center's *Humanitarian Crisis Coverage Report* (Dragomir, 2025), which collected and processed 78,667 news articles from English-language outlets across the United States, United Kingdom, Canada, Australia, Ireland, and three internationally prominent non-Western outlets (Al Jazeera, RT, and IRNA). Articles were retrieved using the Google News API via SerpAPI, filtered for relevance using a combination of rule-based and GPT-4o-assisted classification (achieving an F1-score of 93%), and enriched with humanitarian metadata from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

2.2 Crises Covered

The dataset covers ten humanitarian crises selected on the basis of overall humanitarian impact as defined by the UN OCHA 2025 Global Humanitarian Overview. The ten crises, along with their key severity indicators, are summarized in Table 1.

Crisis	Start Date	People Affected (M)	Funding Required ($B)	Total Articles
Gaza and the Occupied Palestinian Territories	Oct 2023	3.3	4.00	29,020
Ukraine	Feb 2022	12.7	3.32	21,440
Syria	Mar 2011	16.7	8.58	10,542
Afghanistan	Aug 2021	22.9	3.04	8,302
Yemen	Sep 2014	19.5	2.50	1,973
Myanmar	Aug 2017	19.9	1.10	1,856
Sudan	Apr 2023	30.4	10.28	1,749
Democratic Republic of the Congo	Mar 2022	21.2	3.23	1,580
Ethiopia	Nov 2020	10.0	2.00	1,511
Chad	Jul 2009	7.8	1.50	349

Table 1. Humanitarian crisis overview with severity indicators and total article counts.

2.3 Database Structure

Raw CSV files were loaded into a relational SQLite database with six linked tables: crises, monthly_coverage, coverage_by_outlet, framing, sentiment, and victim_causor. This structure allowed analytical flexibility while maintaining referential integrity across the dataset. Framing, sentiment, and victim/causor data were only available for Gaza and Ukraine, as full-text extraction for the qualitative analysis was limited to these two crises by the original study.

2.4 Limitations

The dataset has several important limitations that should be noted throughout. First, it is weighted toward English-language Anglophone media from the Global North, which means the coverage patterns observed reflect the editorial priorities of a specific media ecosystem rather than global journalism as a whole. Second, narrative-level features (framing and sentiment) are available for only two of the ten crises, limiting the scope of the qualitative analysis. Third, the underlying data was collected and processed by the MJRC; we did not control the original keyword selection, relevance filtering, or outlet inclusion criteria.

3. Methodology

3.1 Data Wrangling and Feature Engineering

The original dataset was structured at the crisis level, with one aggregated row per crisis. This was insufficient for regression modeling, as it produced only ten observations. We restructured the master dataset at the *monthly* level using the monthly coverage table, which records article counts per crisis per month and contains 734 rows spanning the full observation period from 2009 to 2025. This restructuring enabled proper train/test evaluation and allowed us to model temporal dynamics within each crisis.

We engineered the following features from the restructured dataset:

Feature	Description	Type
months_since_start	Months elapsed since each crisis began	Continuous
is_onset	Binary flag: 1 if within first 3 months of crisis	Binary
log_coverage	Log-transformed monthly article count (modeling target)	Continuous
top3_outlet_ratio	Share of crisis coverage from the top 3 outlets	Continuous
post_onset_ratio	Ratio of average post-onset to onset coverage (decay measure)	Continuous
framing_ratio_humanitarian	Proportion of humanitarian-framed articles (Gaza/Ukraine only)	Continuous
framing_ratio_geopolitical	Proportion of geopolitical-framed articles (Gaza/Ukraine only)	Continuous
framing_ratio_economic	Proportion of economic-framed articles (Gaza/Ukraine only)	Continuous

Table 2. Engineered features used in the modeling dataset.

Raw monthly coverage counts were log-transformed ($\log(x + 1)$) to reduce the strong right skew caused by the extreme article volumes of Gaza and Ukraine relative to other crises. Framing ratios were zero-filled for the eight crises without framing data, and this limitation is acknowledged in the feature interpretation.

3.2 Modeling Approach

We compared four regression models on the 734-row monthly dataset: Linear Regression, Ridge Regression (L2 regularization, $\alpha = 1.0$), Decision Tree (max depth = 4), and Random Forest (200 estimators, max depth = 6). All features were standardized using z-score normalization prior to fitting. An 80/20 temporal train/test split was used with shuffle disabled, so that the test set represents the most recent months of coverage rather than a random sample. This more closely resembles real-world prediction tasks for time-structured data.

Model performance was evaluated using three metrics: R^2 (proportion of variance explained), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), all computed on the held-out test set. Five-fold cross-validation was also reported on the training set to assess consistency.

3.3 Exploratory and Narrative Analysis

Exploratory analysis examined coverage distributions, temporal trends, outlet-level patterns, and the relationship between coverage and severity indicators across all ten crises. For Gaza and Ukraine, we conducted a narrative analysis using framing ratios and entity-level sentiment data, examining how coverage composition varies by outlet and how key actors are portrayed.

4. Findings

4.1 Coverage Is Deeply Unequal Across Crises

The most immediate finding from the data is the scale of disparity in coverage between crises. Gaza received 29,020 total articles over its observation period and Ukraine received 21,440, together accounting for over 65% of all articles in the dataset. At the other end, Chad received just 349 articles across a crisis that has been ongoing since 2009.

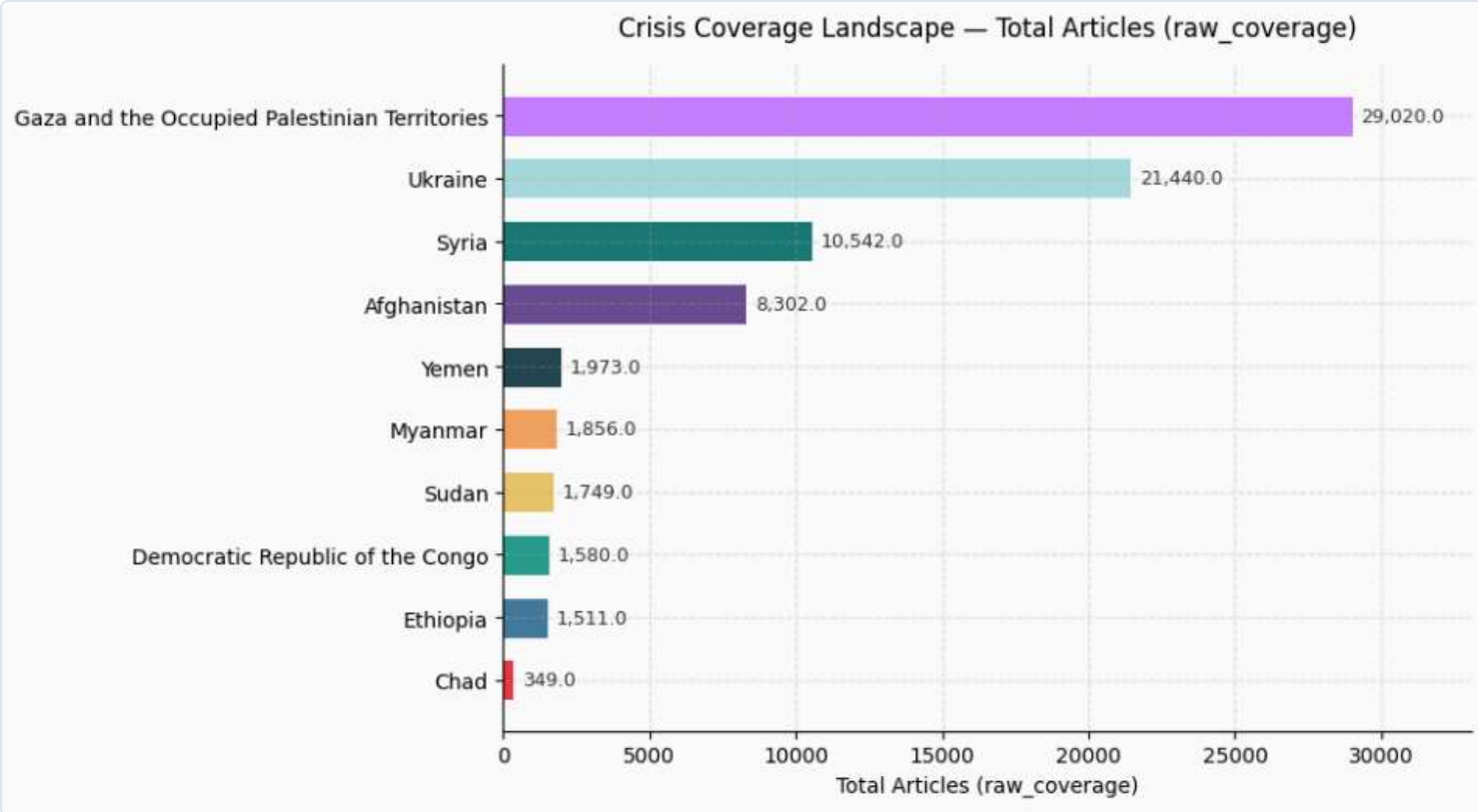

Figure 1. *Total article count per crisis. Gaza and Ukraine account for the majority of coverage, while Chad, Ethiopia, and the Democratic Republic of the Congo receive a small fraction despite large affected populations.*

Crucially, this disparity does not reflect humanitarian severity. Gaza, which affects approximately 3.3 million people, received more than 18 times the coverage of Sudan, which has 30.4 million people in need. The Democratic Republic of the Congo, with 21.2 million people requiring aid and one of the largest funding requirements in the dataset, received

fewer than 1,600 articles in total. When coverage is normalized by people affected, Gaza receives 8,794 articles per million people, compared to just 57.5 for Sudan and 74.5 for the DRC.

> *"Gaza received 83 times more articles per person affected than Sudan, despite Sudan having nine times more people in need of humanitarian assistance."*

4.2 Severity Metrics Are Poor Predictors of Coverage

Scatter plot analysis of raw coverage against the three available severity indicators confirms that humanitarian scale is a weak predictor of media attention at the crisis level. People affected ($R^2 = 0.271$) shows the strongest but still modest relationship, and in the wrong direction: crises affecting more people tend to receive less coverage, not more, because the most affected crises (Sudan, DRC, Afghanistan) are precisely those that receive the least media attention. Funds required ($R^2 = 0.010$) shows virtually no relationship with coverage, and crisis duration ($R^2 = 0.136$) is negatively associated, meaning longer-running crises receive less coverage per month over time.

These findings align with established theory in journalism studies, particularly the concept of "news values" identified by Galtung and Ruge (1965), which emphasizes that newsworthiness is determined by factors such as proximity to powerful nations, dramatic imagery, and elite involvement rather than by the scale of human suffering.

4.3 Coverage Is Episodic and Decays Over Time

The monthly timeline reveals a consistent pattern across all crises: coverage spikes sharply at crisis onset and then decays rapidly. Afghanistan's highest monthly coverage came in August 2021, immediately following the Taliban's seizure of Kabul, and fell by more than 80% within three months. Ukraine's coverage peaked in early 2022 following the full-scale Russian invasion and has declined gradually since. Gaza shows the most dramatic onset spike in the dataset, exceeding 7,000 articles in a single month following the October 2023 escalation.

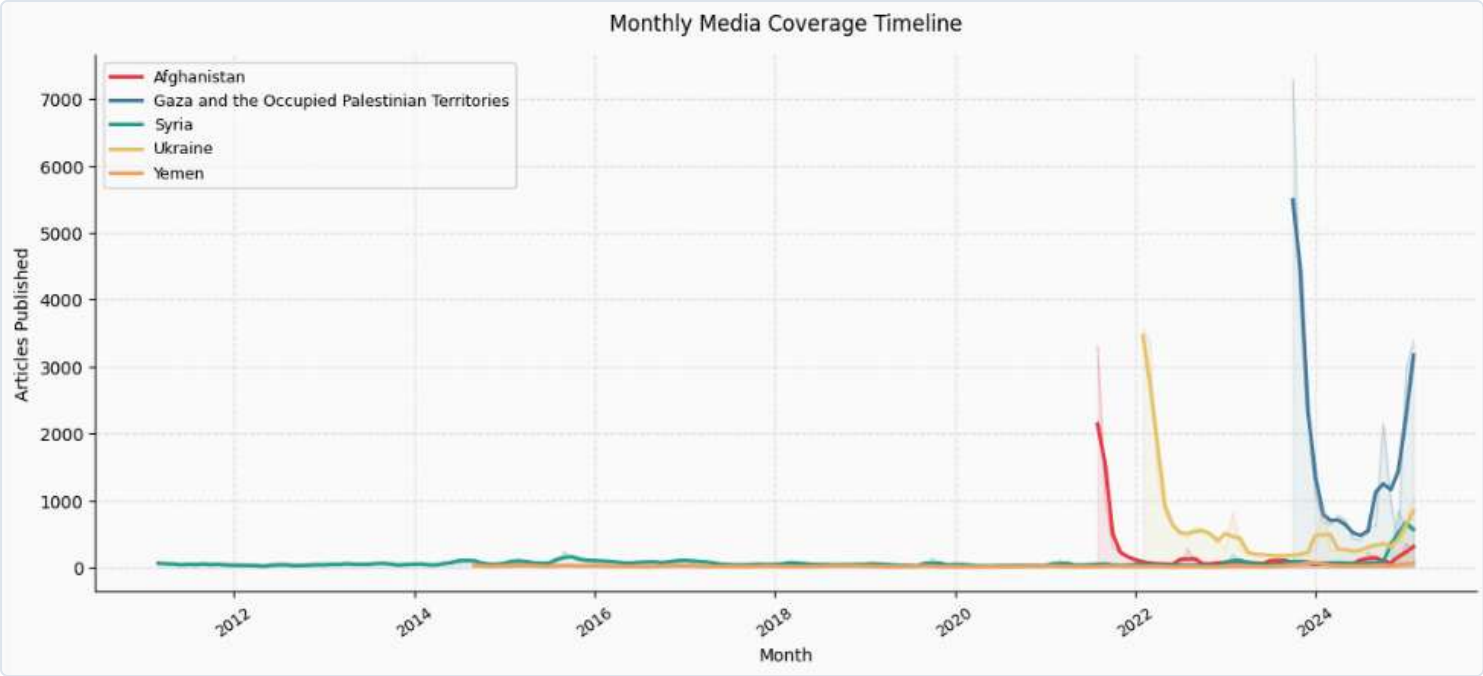

Figure 2. *Monthly media coverage timeline for the five most-covered crises (3-month rolling average). Coverage is highly episodic, with sharp onset spikes followed by rapid decay. Syria is a notable exception with more sustained long-term attention.*

This episodic pattern has practical consequences. Humanitarian organizations rely on sustained media attention to maintain donor engagement and political pressure. When crises disappear from headlines after the initial news cycle, they risk losing both public and institutional support precisely when long-term needs are greatest.

4.4 Outlet Coverage Is Concentrated but Systematically Biased

Across all crises, Al Jazeera published the most total articles (5,688), followed by Reuters (4,632) and the New York Times (4,472), with a relatively steady decline across the remaining outlets. The coverage distribution across outlets is not heavily concentrated at the global level, but the per-crisis heatmap reveals a more striking pattern: every outlet in the dataset directed its highest relative attention toward Gaza. The Gaza column is the most intensely colored across all rows, meaning that regardless of editorial identity or geographic base, Gaza dominated each outlet's coverage within the period of the dataset.

By contrast, Chad, the DRC, and Ethiopia are uniformly pale across all outlets, indicating that their neglect is not the editorial preference of any single newsroom but a systemic feature of the dataset's media ecosystem. This supports the interpretation that coverage hierarchies are structural rather than idiosyncratic.

4.5 Regression Modeling: What Actually Predicts Monthly Coverage?

The regression analysis on the 734-row monthly dataset produced the following results across four models:

Model	Test R^2	RMSE	MAE
Linear Regression	0.653	0.760	0.615
Ridge Regression	0.656	0.757	0.612
Decision Tree (depth 4)	0.674	0.737	0.602
Random Forest (200 trees)	0.472	0.937	0.769

Table 3. Model performance on the held-out test set (log-transformed monthly coverage).

Decision Tree achieved the highest test R^2 (0.674), with Linear and Ridge Regression close behind (0.653 and 0.656). Random Forest underperformed at 0.472 despite being the most complex model. This counterintuitive result is consistent with the data structure: the relationships between features and log-coverage are largely linear, and the framing features are sparse zero values for eight of the ten crises, which limits the signal available to ensemble tree methods.

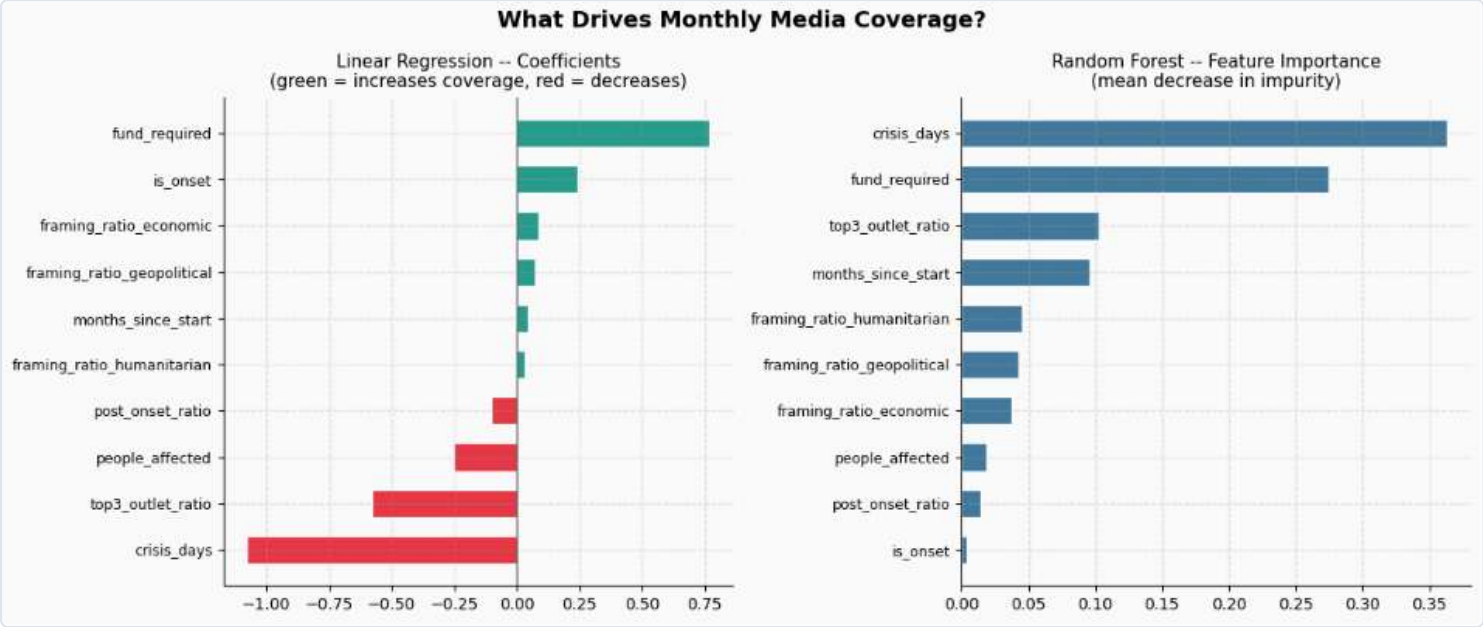

Figure 3. *Feature importance from Linear Regression (standardized coefficients, left) and Random Forest (mean decrease in impurity, right). Both models identify crisis_days and fund_required as the dominant predictors, while people_affected shows little importance.*

Feature importance analysis reveals consistent results across both the linear and tree-based approaches. **crisis_days** is the single most important feature in the Random Forest and carries the largest negative coefficient in Linear Regression: longer-running crises receive less monthly coverage, reflecting the decay pattern visible in the timeline data. **fund_required** is the strongest positive predictor in both models, suggesting that crises with larger internationally recognized funding requirements attract more coverage, likely because they have greater institutional visibility and are more frequently referenced by major humanitarian organizations.

Importantly, **people_affected** shows a negative coefficient in Linear Regression and low importance in Random Forest. Crises affecting more people do not receive more coverage. This is the clearest quantitative evidence in our analysis that humanitarian scale does not drive media attention.

The **is_onset** flag and **top3_outlet_ratio** also contribute meaningfully. Onset months receive more coverage regardless of the crisis, consistent with the event-driven spike pattern identified in Section 4.3. The negative effect of top3_outlet_ratio confirms that crises dominated by a small number of outlets tend to have lower total monthly article volumes than those that attract diverse coverage.

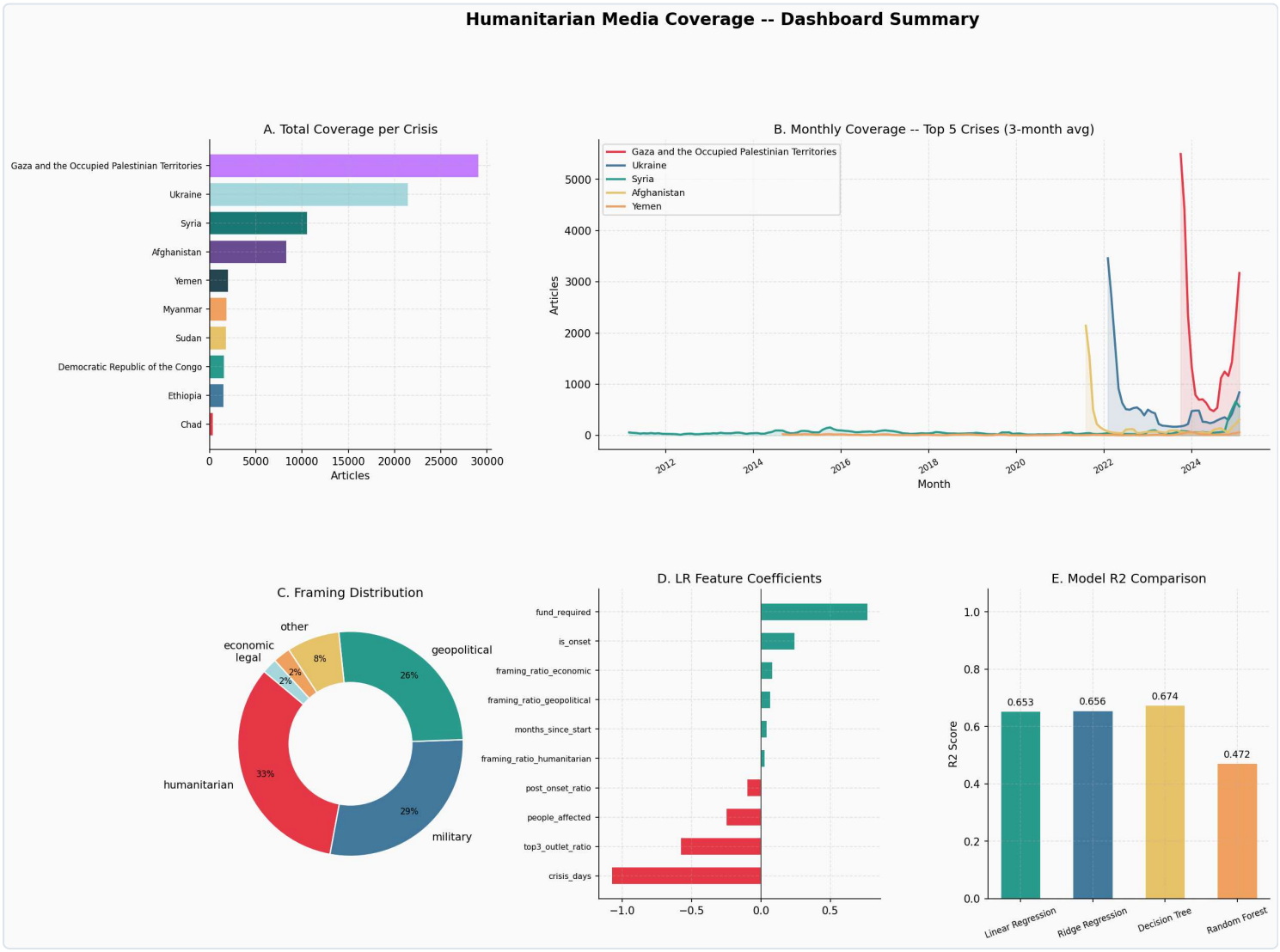

Figure 4. *Dashboard summary panel combining total coverage per crisis, monthly timeline, framing distribution, linear regression coefficients, and model R^2 comparison.*

4.6 Narrative Framing: Gaza and Ukraine

For the two crises with available framing data, the global distribution shows that humanitarian framing dominates across all articles (33%), followed by military framing (29%) and geopolitical framing (26%). Legal and economic frames together account for less than 5% of all articles, suggesting that accountability and structural economic perspectives are largely absent from the coverage of these crises.

The per-crisis breakdown reveals meaningful differences between Gaza and Ukraine. Gaza coverage is 40% humanitarian in framing, reflecting the intense focus on civilian suffering, displacement, and basic needs. Ukraine coverage is more heavily militarily framed (33%), consistent with its portrayal as an active battlefield with rapidly evolving frontlines and strategic significance for NATO allies. Geopolitical framing is more prominent in Ukraine (32%) than in Gaza (21%), reflecting the extent to which the Ukraine conflict has been narratively positioned as a confrontation between democracy and authoritarianism.

4.7 Sentiment and Entity Portrayal

Entity-level sentiment analysis across the four key actors in the dataset reveals stark asymmetries in how individuals are portrayed. Zelensky is the only net-positively portrayed actor, with substantially more positive mentions than negative ones. Hamas receives the most overwhelmingly negative portrayal in the dataset, with a net sentiment of approximately -25,000 mentions. Putin and Netanyahu are both net-negative, though to different degrees.

When filtered to the Gaza crisis specifically, both Hamas and Netanyahu are net-negative, but Hamas is portrayed far more negatively. In the Ukraine crisis, Putin is heavily negative while Zelensky is strongly positive. This pattern holds regardless of which outlet is examined, suggesting that these portrayals reflect a consistent editorial stance across the Anglophone media ecosystem rather than the preference of any individual newsroom.

> *"Entity portrayal is not crisis-specific — it is systemic. Actors on Western-aligned sides are consistently portrayed more favorably across all outlets in the dataset."*

5. Discussion

Taken together, the findings of this project paint a consistent picture: international media coverage of humanitarian crises follows editorial and geopolitical logic rather than humanitarian need. The crises that receive the most attention are those involving Western geopolitical interests, dramatic escalations, and high-profile actors, not those affecting the most people or requiring the most resources.

Our regression modeling adds a quantitative dimension to this observation. By moving from a 10-row crisis-level dataset to a 734-row monthly dataset, we were able to model the temporal dynamics of coverage in a way that earlier analyses could not. The dominance of crisis_days as a negative predictor, and is_onset as a positive one, confirms that media attention follows a predictable lifecycle: intense interest at the onset of a crisis, followed by rapid decay as the crisis continues. This cycle disadvantages long-running crises, many of which are in the Global South, relative to acute escalations that receive enormous initial attention.

The positive effect of fund_required as a predictor is more nuanced. It may reflect the fact that larger funding requirements are associated with greater institutional visibility in the humanitarian system, which in turn attracts more journalistic attention. Alternatively, it may be that crises requiring more funding are those involving state-level actors or geopolitical dimensions that independently drive coverage. Disentangling these mechanisms would require more granular data than is currently available.

The framing and sentiment findings for Gaza and Ukraine reinforce the broader pattern. Both crises are framed in ways that align closely with Western foreign policy positions: Ukraine as a heroic resistance, Gaza as a humanitarian catastrophe. Entity sentiment follows this logic consistently across outlets, suggesting that media portrayals of individual actors are shaped more by geopolitical alignment than by journalistic assessment of individual behavior.

5.1 Limitations of This Study

Several limitations should be noted. The modeling dataset, while substantially larger than the original crisis-level table, remains relatively small by machine learning standards. Framing features are zero-filled for eight crises, limiting their interpretive value in the regression models. The dataset is also heavily weighted toward Anglophone media from the Global North, which means the conclusions about media behavior apply to that specific ecosystem and may not generalize to Arabic-language, French-language, or other media environments. Finally, our models explain up to 67% of the variance in monthly log-coverage; the remaining 33% reflects factors not captured in our feature set, including specific events, political developments, and editorial decisions that are difficult to operationalize quantitatively.

6. Conclusion

This project set out to ask whether international media coverage of humanitarian crises aligns with their severity. The answer, supported by both exploratory and predictive analysis, is clearly that it does not. Coverage is driven primarily by the timing of events, the geopolitical salience of a crisis, and outlet behavior, not by the number of people affected or the scale of humanitarian need.

Crises like Chad, the Democratic Republic of the Congo, and Ethiopia have been ongoing for years or decades, affect tens of millions of people, and require billions of dollars in humanitarian funding. They receive a fraction of the media attention directed at Gaza and Ukraine. Our modeling confirms that longer-running crises receive systematically less monthly coverage, creating a structural disadvantage for the world's most protracted emergencies.

These findings have practical implications. If media coverage shapes donor behavior and political advocacy, as a substantial body of research suggests, then the crises receiving the least attention are also the most likely to be deprioritized in funding and policy decisions. Addressing this requires not just greater awareness within newsrooms but structural changes to how humanitarian journalism is resourced, incentivized, and distributed.

From a data science perspective, this project demonstrates the value of restructuring aggregated data to unlock temporal dynamics, using log-transformation to manage skewed distributions in media datasets, and treating feature importance across multiple models as a more robust signal than any single model's coefficients. Future work could extend this analysis to non-English media ecosystems, incorporate social media amplification as an additional signal, and explore causal identification strategies that distinguish between media agenda-setting and humanitarian funding outcomes.

References

Dragomir, M. (2025). *Humanitarian Crisis Coverage Report*. Media and Journalism Research Center: Tallinn/Santiago de Compostela/London. https://doi.org/10.13140/RG.2.2.13831.87203

Galtung, J., & Ruge, M. H. (1965). The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research, 2*(1), 64–90.

Harcup, T., & O'Neill, D. (2001). What is news? Galtung and Ruge revisited. *Journalism Studies, 2*(2), 261–280.

Moeller, S. D. (1999). *Compassion fatigue: How the media sell disease, famine, war and death*. Routledge.

United Nations Office for the Coordination of Humanitarian Affairs. (2025). *Global Humanitarian Overview 2025*. United Nations. https://www.unocha.org/publications/report/world/global-humanitarian-overview-2025

Chouliaraki, L. (2006). *The spectatorship of suffering*. Sage.

Entman, R. M. (1993). Framing: Towards clarification of a fractured paradigm. *McQuail's Reader in Mass Communication Theory*, 390–397.

Appendix: Technical Summary

Repository Structure

File / Folder	Description
notebooks/milestone_2/data_wrangling.ipynb	Feature engineering and monthly master table construction
notebooks/milestone_2/data_modeling.ipynb	Regression model training, evaluation, and feature importance
notebooks/milestone_2/data_visualization_static.ipynb	Interactive ipywidgets dashboard (8 sections)
notebooks/milestone1_analysis.ipynb	Data loading, database creation, and exploratory analysis
data/raw/	Original CSV datasets from MJRC
data/processed/monthly_model_data.csv	734-row monthly modeling dataset
humanitarian.db	SQLite relational database
diary/	Weekly reflection entries documenting project decisions

Table 4. Repository structure summary.

Software and Dependencies

Python 3.10 | pandas 2.x | numpy | scikit-learn | matplotlib | seaborn | ipywidgets | sqlite3