# A Computational Framework Using NLP and LLMs for Media Bias Detection in Geopolitical News Reporting

# Aalborg University

# MASTER THESIS

**Submitted By: Mohammad Rakib Hasan**

**4th Semester**

**Business Data Science**

**Supervised By: Primoz Konda**

**Date of Submission: 31 May 2026**

## Abstract

This thesis investigates how Large Language Model (LLM) based analytical systems can be used to detect and evaluate media bias within comparative geopolitical news coverage from Bangladeshi and international media outlets. The research focuses on the Russia-Ukraine war, Iran-Israel conflict and Taiwan Strait tensions in order to examine how political narratives are framed differently across regional media environments and how AI driven systems interpret those differences during automated analysis.

The study develops a computational framework that integrates automated news scraping, Natural Language Processing (NLP), multilingual text processing and OpenAI based LLM analysis within a unified analytical pipeline. News articles were collected from selected Bangladeshi and international outlets including Daily Star, New Age, BBC and The Guardian. The collected data passed through preprocessing stages involving text cleaning, duplicate removal, topic filtering and corpus standardization before entering the analytical stage.

The analytical framework combined lexical analysis, sentiment evaluation and LLM driven comparative narrative interpretation. Instead of relying only on old and traditional classification methods, the system generated contextual summaries and comparative outputs capable of identifying differences in framing, emotional tone, geopolitical emphasis and actor representation across media sources.

The findings demonstrate that the framework successfully identified meaningful variations in reporting structures between Bangladeshi and international media. The analytical framework consistently identified stronger emphasis on humanitarian consequences within Bangladeshi coverage compared to the selected international outlets, while international coverage emphasized diplomacy, geopolitical strategy and broader international implications. Multiple forms of bias including framing bias, selection bias and linguistic bias were identified across both media environments.

Overall, the research demonstrates that NLP and LLM based analytical systems can support scalable and interpretable comparative media analysis. The thesis contributes both a methodological framework for computational bias detection and a practical demonstration of how modern AI driven systems can be applied within media studies and Business Data Science.

# Table of Contents

1. Introduction ..... 5
1.2 Research Question ..... 8
2. Literature Review ..... 9
3.1.1 Ontology ..... 18
3.1.2 Epistemology ..... 19
3.1.3 Quantitative Research Orientation ..... 19
3.2 Validity and Reliability: ..... 20
3.2.1 Internal Validity ..... 20
3.2.2 External Validity ..... 21
3.2.3 Reliability ..... 21
3.3 System Architecture ..... 21
3.3 Data Collection Framework ..... 23
3.4 Data Preprocessing and Corpus Preparation ..... 24
3.5 NLP and LLM Based Analytical Framework ..... 26
3.6 Comparative Bias Detection Process ..... 28
3.7 Streamlit Based Visualization and Interaction System ..... 31
3.8 Validation and Quality Control ..... 32
3.9 Ethical Considerations ..... 33
4. Analysis, Results and Evaluation ..... 34
4.2 Results ..... 38
4.3 Evaluation ..... 42
5. Conclusion ..... 47
6. Limitations and future Work ..... 52
7. References ..... 58
8. Appendix: ..... 60

## 1. Introduction

Freedom of speech is widely recognized as a foundational principle of democratic governance. It enables the exchanges of ideas. It supports political accountability and allows the citizens to participate meaningfully in public debate. Within the democratic theory, freedom of expression is not only an individual right but also a structural condition for the functioning of the press. Media institutions play a central role in shaping political understanding, constructing narratives and influencing how social reality is interpreted. As Barendt (2007) argues, freedom of speech in the media context must be understood as very essential to democratic accountability rather than merely as a personal liberty.

However, the existence of legal protection for free expression does not automatically ensure informational neutrality. Media systems mostly operate within political, economic and technological environments and it can influence editorial priorities and framing practices. Bias in news does not necessarily mean that it was taken as a form of overt censorship. Instead it often appears through selective emphasis, lexical choices, topic salience, tuning and differences in how actors are portrayed. Framing theory suggests that it depends on the way information is presented can shape audience interpretation even when the factual content remains unchanged (Entman, 1993). Similarly, research on media effects indicates that audiences may perceive coverage as ideologically biased depending on contextual framing and narrative structure (Perloff, 2017).

The transformation of the media landscape in the digital era has intensified these dynamics. High numbers of online platforms, algorithmic distribution systems and engagement driven visibility mechanisms have altered traditional gatekeeping structures in a profound way. News visibility is increasingly shaped by technological infrastructures that prioritize certain forms of content moderation and interaction properly. As Napoli (2019) explains, digital media environments continuously introduce new layers of mediation that complicate the assessments of pluralism and public interest. The proliferation of digital outlets has also increased the volume of accessible information. Yet it has also fragmented audiences and amplified the speed at which narratives circulate.

These developments create both conceptual and methodological challenges. If bias is always embedded in patterns of linguistic framing and narrative construction rather than explicit suppression, then identifying it requires systematic analysis of large textual corpora and so on.

Traditional qualitative approaches provide depth but they are often difficult to scale across thousands of articles. Data Science offers tools capable of transforming unstructured text into measurable variables, enabling large scale comparative analysis of discourse patterns.

Recent advances in Natural Language Processing have significantly expanded the possibilities for computational media analysis and beyond. Transformer based architectures such as BERT allow for contextualized word representations that capture semantic nuance more effectively than earlier models (Devlin et al., 2019). These representations can be used to identify sentiment shifts, detect ideological markers and measure linguistic asymmetries across outlets. Furthermore, research has demonstrated that contextualized embeddings can support the identification of biased language through comparative modeling of lexical patterns (Spinde et al., 2021).

Multilingual modeling is extremely important in comparative research that spans different linguistic contexts. A proper Cross lingual transformer architectures explicitly enable the alignment of semantic representations across languages, facilitating structural comparison between media systems operating in different regions (Conneau et al., 2020). These tools eliminate the need for purely manual translation and allow for systematic comparison of discourse features across corpora written in different languages.

At the same time, the emergence of large language models has introduced new analytical opportunities. These systems can properly perform zero shot and few shot classification, summarization and semantic analysis without extensive labeled datasets (Brown et al., 2020). However, their outputs must be interpreted carefully. Because they may reproduce patterns present in their training data. Therefore, the integration of large language models into bias detection frameworks requires methodological transparency and critical evaluation.

Building on these theoretical and technological developments, this thesis investigates how patterns of media representation differ across Bangladeshi and selected European English language news outlets. The comparative designs are motivated by differences in institutional contexts, journalistic norms and regulatory environments that shape media practice. Comparative media system theory suggests that structural and political conditions influence how news is produced, framed and disseminated (Hallin & Mancini, 2004). By examining discourse patterns rather than relying solely on normative claims, this study seeks to provide measurable evidence of variation in narrative construction across regions.

Rather than assuming bias in advance, the project operationalizes bias through quantifiable discourse indicators. These include sentiment polarity dispersion, frequency of evaluative lexical markers, distribution of quoted actors and thematic clustering of frames. That's why bias is conceptualized as structural asymmetry in linguistic representation rather than as intentional distortion. This approach aligns with contemporary research emphasizing computational identification of bias inducing patterns through contextual modeling (Spinde et al., 2021).

From a Business Data Science perspective, the project develops a full analytical pipeline that integrates automated scraping, preprocessing, embedding based representation, classification and evaluation. The objective is not only theoretical insight but also methodological reproducibility. Data driven workflows enable scalable monitoring of narrative trends and sentiment dynamics, demonstrating how computational tools can transform abstract media concepts into structured analytical outputs (Provost & Fawcett, 2013).

A central technical component of the study is the use of multilingual transformer architectures to ensure cross regional comparability. Since the dataset includes news articles originating from different linguistic and cultural environments, semantic alignment is required. Cross lingual embedding models allow textual representations to be mapped into shared vector spaces and it enables systematic comparison without sacrificing contextual nuance (Conneau et al., 2020).

Large language models are further integrated into the framework for classification and summarization tasks. Their capacity for zero shot and few shot inferences allows bias related attributes to be identified without constructing extensive manually labeled datasets. However, LLM outputs are treated as analytical instruments rather than authoritative judgments. Because such systems may inherit latent biases from large scale training data, their predictions are evaluated critically and compared across modeling strategies (Brown et al., 2020).

An important extension of this work is the development of a demonstration web application incorporating an interface. They can choose based on the criteria. The application allows the users to query the dataset and receive structured summaries of sentiment patterns, framing tendencies and comparative discourse indicators across outlets. This applied component illustrates how research insights can be translated into interactive analytical tools for media monitoring, strategic communication and organizational risk assessment. The increasing relevance of algorithmic

mediation in digital information environments highlights the need for such transparent analytical systems (Napoli, 2019).

By situating computational modeling within democratic theory, this thesis treats freedom of speech not solely as a constitutional abstraction but as a measurable feature of discourse diversity and representational balance. While computational metrics cannot capture the full normative complexity of expressive freedom, they provide structured empirical evidence that complements theoretical debate. Understanding how narratives are constructed, emphasized or polarized across media systems contributes to broader discussions about the health of democratic communication (Barendt, 2007).

Ultimately, the study shows a balanced framework which integrates theory, computational methodology and practical application. It demonstrates how multilingual NLP and large language models can be used to detect patterns of framing and sentiment asymmetry while maintaining methodological transparency properly. By bridging democratic theory with applied Data Science, the thesis provides both conceptual insight and a scalable analytical model for evaluating media discourse in contemporary digital ecosystems (Devlin et al., 2019).

## 1.2 Research Question

How can large language model based analytical systems detect and quantify media bias in comparative news coverage between Bangladeshi and International outlets?

## Sub Questions

1. To what extent can multilingual NLP models reliably identify linguistic framing asymmetries, sentiment divergence and ideological lexical markers across cross regional news corpora?
2. Do LLM driven summarization and classification systems reproduce, reduce or amplify existing bias patterns when applied to politically sensitive news content?

## 2. Literature Review

Media bias is a central concern in news analysis because reporting does not only present events but also shapes how those events are interpreted. It also decides what is visible, what is repeated and what is left in the background. In the social science literature, this is not treated as a small stylistic issue. It is treated as a force that can shape how the public understands conflict, power and responsibility. Recent computational reviews also make the same point. They show that media bias has long been studied in the social sciences with richer concepts than many automated systems use and they argue that automated analysis works best when it is linked to those older theories rather than treated as a simple text difference task (Hamborg et al.,2019).

Bias can be defined in a strict and practical way. In the academic sense, bias is not only open propaganda. It is also a systematic pattern of selection, emphasis, wording and omission that makes one interpretation of an event more visible than others. Entman's framing work is central here because it defines framing as the selection of some aspects of reality and their salience in a communication text so that a particular problem definition, causal reading, moral evaluation treatment suggestion is promoted (Entman, 1993). In his later work, he also argues that bias is a core but often undertheorized concept and that it should be measured with explicit theory instead of vague impressions (Entman, 2007).

Framing is the first pillar of this review because it explains how the same event can be told in different ways. Entman's classic definition is especially useful because it links framing to four actions at once. A text can define a problem, diagnose causes, make moral judgments and suggest remedies. That means bias is not limited to false statements. It can also arise when a text repeatedly highlights one side of a conflict and gives less room to other meanings or other actors.This is highly relevant for a comparison between news outlets because the issue is not only whether the facts are correct. The issue is also which facts are made salient and which interpretations are encouraged. Scheufele and Tewksbury later place framing alongside agenda setting and priming and show that these models are related but not identical. Their work helps separate emphasis on meaning from emphasis on issue importance (Scheufele & Tewksbury, 2007).

Agenda setting is the second pillar because the project is also about what is covered and what is not covered. The classic agenda setting argument says that media do not simply tell people what to think, they also tell people what to think about. McCombs and Shaw's original study shows that

editors and newsroom staff shape political reality through the amount of information they give to an issue and the position they give it in the news flow. In later summaries of the theory, agenda setting is described as the influence of media emphasis on the public importance of issues. This matters for this study because a blind spot is not only a missing article, it is also a missing issue from the public agenda. If one outlet barely covers a conflict at all, then that absence can be read as an agenda setting problem as well as a bias problem (McCombs & Shaw, 1972).

The idea of blind spots fits well with the foreign news tradition. Galtung and Ruge's classic study of foreign news argues that not every event becomes news. Their theory of newsworthiness includes factors that make some events more likely to be selected while others remain invisible. They also note that events that satisfy very few of those factors tend not to become news. The point is not that silence always proves bias. The point is that systematic silence can be analytically meaningful when it appears across time and across topics (Galtung & Ruge, 1965).

This is also where comparative news analysis becomes powerful. Hamborg, Donnay and Gipp explain that studies of omission and source selection usually need at least two corpora because bias often appears only when one article is compared with another article or with a non media baseline. Their review also shows that social science research has long studied bias by commission and omission through content analysis and frame analysis. This supports a direct comparison between news reporting on the same war or diplomatic crisis. The point is not to ask whether one outlet is simply "good" and the other is simply "bad." The point is to ask how each outlet constructs the event world. One may stress national relevance and emotional conflict, while the other may stress diplomacy, strategic balance  institutional response. That is exactly the kind of difference the comparative narrative summaries are designed to capture (Hamborg et al., 2019).

A useful detail from the media bias literature is that bias by word choice and labeling can be subtle but still important. Hamborg, Zhukova and Gipp show that different phrases for the same concept can create different emotional effects and can lead readers toward different interpretations. This is highly relevant for conflict reporting because terms like invasion, strike, blockade, retaliation, aggression  confrontation are not neutral labels. They guide interpretation. In this thesis, this means that lexical choice should not be treated as a cosmetic issue. It should be treated as one of the core analytical dimensions of bias. When one coverage repeatedly uses more loaded labels than other

coverage or vice versa, the difference can be discussed as a framing difference with possible evaluative bias (Hamborg et al., 2019).

The computational literature also supports a hybrid method. Hamborg and colleagues argue that computer science has useful automated tools, but that many systems still describe media bias too vaguely. They recommend combining social science theory with automated NLP methods. That point fits this workflow very well. The current pipeline already uses structured collection, topic filtering and TF-IDF based text representation. TF-IDF remains a standard and transparent baseline in text classification research because it turns words into interpretable weighted features. At the same time, the later LLM layer can be framed as a second stage that reads the filtered corpora and produces comparative narrative analysis. This is a strong thesis design because the baseline keeps the process auditable, while the LLM adds higher level interpretation (Lai et al., 2022).

For the argument itself, the cleanest synthesis is this, bias is not only a matter of falsehood. It is also a matter of emphasis, exclusion and recurring narrative structure. Framing explains how meaning is built. Agenda setting explains how importance is assigned. News values explain why some issues are visible and others are absent. Comparative corpus analysis then makes those differences measurable across Bangladeshi and international media. This provides a solid academic bridge from theory to method. It also gives a strong justification for later labels such as right, left and neutral because those labels can be introduced only after the broader narrative and lexical differences have been established first. The integration of large language models into computational text analysis represents a major shift in how complex social phenomena such as media bias are studied. Earlier approaches in natural language processing relied heavily on feature extraction techniques and supervised classification. These methods required predefined labels and large annotated datasets. While they were effective for structured tasks, they struggled to capture deeper narrative meaning and contextual relationships. Recent work shows that language models that are based on transformer architectures are able to encode context, semantics and discourse level information in ways that were not possible with earlier models. This shift is particularly relevant for media bias analysis, where meaning is often embedded in how events are described rather than in isolated words (Devlin et al., 2019).

In computational media analysis, the design of the data pipeline plays a critical role in determining the quality of the final output. The process begins with large scale data acquisition from different

multiple news sources. Research in different computational social science emphasizes that automated data collection must be carefully structured to ensure consistency and comparability across many sources. Articles are collected from different outlets and stored with structured attributes such as publication date, topic, source and region. This structured representation allows the system to align articles properly across similar geopolitical events. Such alignment is very essential because bias can only be meaningfully analyzed when comparable texts are examined side by side (Lazer et al., 2009).

Web based data extraction methods form the foundation of many modern text analysis systems. These methods can enable the automated collection of large volumes of textual data from online sources very well. In academic research web scraping is often discussed as part of data engineering rather than taking it as a core analytical method. However its role is mostly critical because it defines the scope and quality of the dataset. Studies often highlight that differences in scraping strategies can often lead to differences in data coverage. Which in turn affect analytical results. In this context the use of multiple scraping routines for different news sources ensures that the dataset captures diverse perspectives. This supports the comparative nature of the analysis and reduces the risk of source specific bias in the dataset itself (Mitchell, 2018).

Data preprocessing and filtering are very essential steps in preparing textual data for analysis. These processes include removing irrelevant contents, aligning topics and properly ensuring consistency across documents. Topic based filtering is particularly important in comparative studies. It ensures that differences in analysis reflect variations in framing rather than underlying topic differences. Without such alignment, observed differences may arise from content variation rather than bias. Research in text mining consistently shows that preprocessing steps such as cleaning, normalization and filtering have a significant impact on downstream performance because they reduce noise and improve the reliability of both statistical and model based analyses as discussed by Jurafsky & Martin (2020).

The use of TF-IDF as an initial representation method provides a transparent and interpretable view of the textual data. TF-IDF assigns weights to words based on their frequency within a document and their rarity across the corpus. This allows the identification of terms that are particularly characteristic of specific articles or sources. In the context of bias detection such terms can reveal differences in emphasis and framing. For example, one source may always use more emotionally

charged language while another one could use more neutral terminology. Although TF-IDF does not capture full semantic meaning, it provides a strong baseline for understanding lexical variation and supports further analysis at higher levels (Ramos, 2003).

The use of large language models for comparative narrative analysis represents an emerging direction in natural language processing. Unlike traditional NLP approaches that mainly focus on classification tasks, generative models enable the production of structured interpretations of text, including summarization, identification of key actors and comparison of narrative framing across sources. This shift reflects a broader trend in NLP research as models are increasingly applied not only for prediction but also for reasoning and explanation. Prior studies indicate that such models could capture discourse level patterns and create coherent summaries that reflect underlying narrative structures, as demonstrated by Brown et al. (2020).

Prompt based interaction has become a central paradigm in working with large language models as tasks are specified through properly and carefully designed instructions rather than additional training. Prior research demonstrates that prompt design always plays a critical role in shaping model outputs, especially in the tasks such as summarization, comparison and analytical reasoning. For example Liu et al. (2023) highlight that well structured prompts could improve consistency, coherence and task alignment in created responses. Furthermore, structured prompting approaches which guide models through multiple steps such as summarization and comparison, have been shown to enhance interpretability and analytical depth. These findings suggest that prompt engineering is not only a technical tool but also a methodological framework for controlling model behavior in complex analytical tasks.

The shift from classification to comparative reasoning is a key contribution of this approach. Traditional bias detection systems often reduce complex texts to simple labels. While such labels are very useful, they do not capture the richness of narrative differences. In contrast, the use of a language model always allows the system to explain detailed explanations of how and why two sets of articles may differ. This aligns with recent research that highly emphasizes the importance of explainability in AI systems. In domains such as media analysis, users require not only results but also justifications. Narrative outputs provide a form of explanation that is closer to human reasoning and interpretation (Rudin, 2019).

The ability of language models to properly integrate information from multiple documents into a single analytical output has become an important area of research in natural language processing. Rather than focusing on individual texts, multi document approaches enable the identification of patterns that emerge from different sources  such as recurring themes, consistent tone and repeated framing strategies. Research on multi document summarization demonstrates that language models can effectively synthesize information from diverse inputs and produce coherent, unified summaries. For instance, Liu & Lapata (2019) show that neural models can capture cross document relationships and generating structured summaries that reflect shared narrative elements. This capability is particularly important for tasks such as comparative bias analysis, where the objective is to understand broader narrative trends rather than isolated articles

The system reflects a layered analytical design in which different techniques are used at different stages. Data collection and filtering ensure that the input is relevant and comparable. TF-IDF provides an initial view of lexical patterns. The language model then performs higher level interpretation and comparison. This layered approach is consistent with the best practices in data science. Where complex problems are addressed through a combination of methods rather than a single technique. It allows the system to balance efficiency, interpretability and analytical depth (Provost & Fawcett, 2013).

Finally, the integration of large language models into bias analysis raises some very important considerations about reliability and evaluation. While these models are extremely powerful, their outputs are often influenced by specific prompt design and other training data. This means that results has to be interpreted carefully and whenever possible, validated against known patterns or theoretical expectations. Recent studies emphasize that human oversight and critical evaluation remain essential when using generative models in research. The structured design of prompts and the use of comparative analysis help guide the model toward consistent outputs. This supports the use of language models as analytical tools while maintaining awareness of their limitations (Bommasani et al., 2021).

The increasing use of large language models in text analysis has introduced a new direction in computational research where interpretation becomes as important as prediction. In earlier stages of natural language processing the systems were designed to classify or extract specific features from text. These approaches required clear task definitions and structured outputs. However, recent

research shows that many real world problems such as media bias cannot be fully addressed through rigid classification schemes. Bias often exists in subtle forms that emerge across multiple documents and through patterns of narrative construction. Large language models address this limitation by enabling systems to write explanations and comparative insights rather than fixed labels. This shift is especially relevant , where the goal is to compare news narratives rather than assign simple categories (Bommasani et al., 2021).

A key capability of large language models is their ability to perform zero shot and few shot reasoning. These models could analyze text and produce meaningful outputs without requiring extensive labeled datasets. This particularly is important in media bias research because often labeled data is difficult to obtain and might reflect subjective judgments. Instead of relying on predefined annotations, the model uses its general language understanding to interpret the content of articles. This allows the system to process collections of news articles and generated structured comparative analyses without prior training on a specific bias dataset. Studies have shown that such capabilities enable flexible and scalable analysis across different domains and topics (Brown et al., 2020).

The concept of narrative analysis plays a very crucial role in understanding how bias operates within the texts. Narratives often are not simply a summary of events, it is a structured representation that connects actors, actions and outcomes. Different sources may construct different narratives even when describing the same event. These differences can reflect variations in emphasis, interpretation and perspective. Research in computational linguistics has begun to explore methods for extracting and comparing such narrative structures. Large language models are particularly effective in this context because they can process entire documents and identify relationships that extend beyond sentence level patterns. The model can write separate summaries for articles and then produces a comparative interpretation. This aligns with recent work on multi document understanding and narrative synthesis (Liu et al., 2023).

Another important dimension of this approach is the treatment of bias as a relative and context dependent phenomenon. Traditional methods often attempt to define bias in absolute terms. However, recent literature suggests that bias is more effectively identified through comparison. A text may appear neutral in isolation but reveal bias when contrasted with alternative accounts of the same event, Where two corpora are analyzed in parallel. The model identifies differences in

tone, framing and emphasis between sources. This comparative framework allows a more nuanced understanding of bias that reflects real world media dynamics (Hamborg, Donnay, & Gipp, 2019).

The notion of omission or blind spot bias is also highly relevant in this context. Unlike framing or lexical bias, omission bias is characterized by the absence of coverage. This makes it difficult to detect using traditional text analysis methods. However, different analytical and comparative approaches provide ways to identify these patterns. If a topic appears well and frequently in one corpus but is largely absent in another, this difference can indicate selective coverage. In this project, the absence of certain geopolitical topics in specific Bangladeshi sources can be interpreted as a potential blind spot. This aligns with theoretical perspectives from agenda setting research, which emphasize that media influence is exercised not only through what is presented but also through what is excluded (McCombs, 2005).

The use of prompt based systems introduces both flexibility and complexity into the analytical process. Prompts act as a form of task specification that guides the behavior of the language model. Research shows that small changes in prompt design can lead to significant differences in output. This highlights the importance of careful prompt construction in ensuring consistent and meaningful results. In this project, there are prompts that are carefully designed to follow a structured format. This includes narrative summary, framing comparison and bias identification. This structure reduces ambiguity and it also helps the model focus on relevant aspects of the text. It also aligns the output with theoretical concepts from media studies. Thereby bridging the gap between computational methods and social science theory (Liu et al., 2023).

Another significant aspect of this approach is the emphasis on explainability. In many machine learning systems outputs are difficult to interpret because they are based on complex internal representations. This creates challenges in domains where transparency is essential. Media bias analysis is one such domain because conclusions must be supported by clear reasoning. Large language models offer a partial solution by generating textual explanations that describe their interpretations. While these explanations are not perfect, they provide a level of transparency that is not available in traditional classification models. This makes them particularly suitable for exploratory and comparative analysis tasks (Rudin, 2019).

The integration of structured data with unstructured text analysis further strengthens the analytical framework. Metadata such as publication date, source and region is used alongside textual content.

This allows for more controlled comparisons and helps properly support the identification of patterns across time and also different sources. Research in data science emphasizes the importance of combining different types of data for the purpose of improving analytical outcomes. Structured data provides context and constraints. While unstructured text provides rich information about content and also meaning. The combination of these elements enables a more comprehensive analysis of media bias (Provost & Fawcett, 2013).

Despite the substantial body of research on media bias, several limitations remain in existing studies as usual. A large part of the literature mostly focuses either on qualitative analyses of framing and agenda setting or sometimes on computational approaches applied to single datasets or specific geopolitical contexts (Horne et al., 2018). While these studies provide valuable insights, they often lack comparative perspectives across multiple regions and conflicts. Thus, limiting their ability to capture broader patterns of bias. In addition, computational approaches frequently emphasize detectable features such as sentiment or word frequency but give less attention to omission bias and disparities in coverage, which are central to agenda setting theory (McCombs & Shaw, 1972).

Furthermore, although recent advances in natural language processing particularly transformer based models have improved the ability to analyze textual data at scale, their application in cross source, multi conflict media comparison remains limited (Devlin et al., 2019; Bender et al., 2021). Existing studies also tend to rely on either fully automated methods or purely qualitative interpretation, with relatively few integrating structured data collection, statistical filtering and large language model based reasoning within a unified analytical framework.

This thesis addresses existing gaps by developing a comparative analytical pipeline that combines automated news scraping, keyword based filtering and large language model analysis to examine how different conflicts are represented across media sources. By analyzing multiple corpora simultaneously the study aims to identify the differences in tone, framing, coverage intensity and thematic focus. By doing this, it extends prior research by incorporating both representation bias and omission bias within a proper unified analytical framework. In this way it offers a more comprehensive understanding of how media narratives are constructed across different geopolitical contexts. It enables a way where this integrated approach allows the system to handle large volumes of data while maintaining a clear connection between theoretical concepts and computational

methods. It also demonstrates how modern data science techniques can be applied to complex social questions in a structured and meaningful way (Lazer et al., 2009).

## 3. Methodology

### 3.1 Philosophy of Science

This research was designed as a comparative computational media analysis framework for detecting and evaluating bias in geopolitical news coverage through the use of Natural Language Processing and Large Language Model based analytical systems. The study focused on the way political conflicts are framed by different regional media outlets and how AI driven systems interpret these narratives during automated analysis tasks.

#### 3.1.1 Ontology

The ontological position of this research assumes that news articles and media narratives represent observable textual realities that can be analyzed computationally. The study views media bias not as a single absolute entity but as a multidimensional phenomenon which may appear through linguistic framing, narrative emphasis, emotional tone, lexical selection and contextual prioritization. It states that This position aligns with critical realist perspectives which acknowledge that social phenomena may contain underlying structures and patterns that can be observed indirectly through empirical analysis (Bhaskar, 1978).

The research therefore shows that geopolitical news coverage contains underlying discursive patterns that can be identified through computational comparison between different media sources. These patterns may not always represent objective truth in an absolute sense but they often can still be systematically observed, modeled and analyzed through textual analysis and computational interpretation. The study also recognizes that media narratives are socially constructed representations of reality rather than neutral reflections of events. This understanding is consistent with social constructionist perspectives in media and communication research, where meaning is shaped through language, framing and institutional context (Berger & Luckmann, 1966).

Because of this, the ontology of the study combines observable computational structures with interpretive awareness of how media discourse constructs political meaning across different regional environments.

### 3.1.2 Epistemology

The epistemological position of the study is primarily empirical and analytical. Knowledge was developed through automated data collection, NLP based textual processing, comparative analysis and LLM driven interpretation of geopolitical news articles. The study relied on structured computational procedures in order to investigate reporting differences across selected media outlets. Comparative knowledge was therefore generated through observation of measurable linguistic features, contextual interpretation and narrative comparison rather than through personal opinion or purely theoretical discussion.

This position is closely related to positivist and post positivist traditions in computational and data driven research because the framework attempts to identify observable patterns through systematic analytical procedures (Creswell & Creswell, 2018). At the same time, the research also acknowledges that interpretation of media discourse cannot be treated as fully objective because language and framing often contain contextual and ideological dimensions that require interpretive understanding. This reflects elements of interpretivist thinking within the broader analytical framework (Bryman, 2016).

The research further acknowledges that interpretation generated by AI systems may itself contain limitations, uncertainty or embedded bias inherited from training data and model architecture. Because of this, LLM outputs were treated as analytical support mechanisms rather than objective sources of truth. Human interpretation and comparative evaluation therefore remained important during the analytical process. This aligns with contemporary discussions in AI research emphasizing the importance of critical evaluation and transparency when applying large language models in social and computational research environments (Bommasani et al., 2021).

### 3.1.3 Quantitative Research Orientation

This study primarily followed a quantitative computational research orientation because the analytical framework relied heavily on measurable textual patterns and comparative processing across large collections of news articles. Several quantitative elements were integrated throughout the framework including sentiment related evaluation, lexical frequency analysis, topic filtering, comparative keyword structures, publication trend observation and source based comparative

grouping. The use of automated preprocessing and structured analytical pipelines also contributed to the quantitative nature of the research design.

The quantitative orientation of the research is consistent with data driven approaches in computational social science, where large textual corpora are transformed into measurable analytical variables for systematic comparison and interpretation (Lazer et al., 2009). The study also followed principles commonly associated with quantitative content analysis because recurring linguistic and narrative patterns were examined comparatively across different media environments (Neuendorf, 2017).

At the same time, the framework also incorporated interpretive analytical components through the use of Large Language Models for contextual summarization, comparative narrative interpretation and framing analysis. These components introduced limited qualitative interpretive dimensions into the analytical workflow because the system attempted not only to measure textual variation but also to explain contextual differences in media narratives. The study therefore combined quantitative computational processing with contextual interpretive support in order to improve analytical depth during comparative media evaluation. This mixed analytical orientation reflects the increasing integration of computational and interpretive methodologies within modern digital media research (Bryman, 2016).

## 3.2 Validity and Reliability:

### 3.2.1 Internal Validity

Several procedures were implemented in order to improve internal validity throughout the research framework. All collected articles passed through standardized preprocessing pipelines including duplicate removal text normalization date standardization and topic filtering before analytical processing began. Manual validation of randomly selected articles was also performed in order to verify extraction quality topical relevance and structural consistency across sources.

The study also maintained consistent analytical procedures across all selected news outlets in order to reduce methodological inconsistency during comparative evaluation.

### 3.2.2 External Validity

The external validity of the study remains limited to the selected geopolitical topics time period and media outlets included within the research corpus. The findings therefore should not be generalized to all global media systems without further investigation.

However the computational framework itself was designed in a modular and scalable structure which allows future integration of additional media sources geopolitical topics and analytical models. This improves the broader applicability of the framework for future comparative media analysis research.

### 3.2.3 Reliability

Reliability was improved through automation and standardization of the overall research pipeline. The use of repeatable scraping procedures structured preprocessing workflows consistent filtering mechanisms and unified storage schemas improved operational consistency throughout the study.

The GitHub Actions based scheduling system also supported reproducibility by allowing the collection pipeline to operate continuously under the same procedural structure during repeated executions.

## 3.3 System Architecture

The overall system architecture was designed as a modular analytical pipeline where each stage performs a specific processing task before transferring the output into the next stage of the workflow. The architecture was developed to support continuous article collection automated preprocessing LLM based analysis and interactive visualization within the same environment. The system mainly consisted of five major layers. These included the data acquisition layer, preprocessing layer, analytical processing layer, storage layer and visualization layer. Each layer was separated logically in order to improve maintainability, scalability and processing efficiency.

The data acquisition layer was responsible for collecting articles from the selected news outlets. Since each website used different page structures rendering methods and metadata formats separate scraping strategies were implemented for each source. Browser automation based extraction API based retrieval and HTML parsing methods were combined depending on the characteristics of the target website.

After collection the extracted articles moved into the preprocessing layer. This stage handled text cleaning duplicate removal formatting normalization date standardization and topic filtering. The preprocessing pipeline transformed raw news articles into a structured corpus suitable for NLP and LLM based analysis.

The analytical layer formed the core component of the research framework. This stage processed the cleaned corpus through various NLP and LLM based operations. OpenAI based models were integrated into the pipeline for tasks such as summarization classification narrative comparison and bias related interpretation. The analytical layer was designed to compare reporting patterns across sources rather than analyzing articles independently.

The architecture also included a structured storage layer. Processed articles were stored in both CSV and PKL formats. CSV files improved readability portability and manual inspection while PKL files improved loading speed during repeated analytical execution. Existing datasets were reloaded before each scraping session so that newly collected articles could be appended without reprocessing the entire corpus repeatedly. An important aspect of the architecture involved incremental updating. Since the dataset grows daily the system was designed to identify previously collected URLs and avoid duplicate storage. Publication dates and article identifiers were checked continuously before adding new entries into the corpus. This improved long term operational efficiency and reduced redundant processing.

The system architecture also included automated scheduling functionality. A GitHub Actions workflow was configured to execute the scraping pipeline automatically every day. This allowed continuous dataset expansion without requiring manual execution of the collection scripts. The updated corpus could then be committed automatically into the repository after successful execution.

The visualization layer was implemented through Streamlit. This layer allowed the processed corpus and analytical outputs to be explored interactively through a web based interface. Users could filter articles by topic source keyword and date range before running comparative analysis tasks. The Streamlit interface therefore functioned as the interaction layer between the analytical pipeline and the final user environment.

Another important architectural consideration involved flexibility. The pipeline was intentionally designed so that additional news sources topics analytical methods or models could later be integrated without major structural changes. This modular design improved the extensibility of the framework and supported future research expansion. The architecture also attempted to balance automation with interpretability. Rather than creating a fully black box system the framework preserved intermediate outputs throughout the workflow. Cleaned articles topic assignments summaries and processed outputs could therefore be inspected separately during validation and analysis stages. The final system architecture therefore functioned as an integrated computational environment for large scale comparative geopolitical news analysis through NLP and LLM based methodologies.

## 3.3 Data Collection Framework

The data collection framework was designed as an automated multi source acquisition system for gathering geopolitical news articles related to the selected research topics. Since the study focused on comparative media analysis it was important to ensure that articles from different regional outlets could be collected consistently and transformed into a unified analytical corpus.

The collection process began from 1 June 2025 and was designed as a continuously updating pipeline rather than a one time extraction process. New articles were collected daily from the selected sources which allowed the dataset to evolve continuously during the research period. Different scraping strategies were implemented depending on the structure of each news outlet. A single extraction approach was not considered reliable because all sources differed in content delivery structure metadata formatting and rendering behavior.

For BBC browser automation based scraping was implemented through Playwright. BBC pages often contain dynamically rendered elements which makes static HTML extraction less reliable in several situations. A headless Chromium browser was therefore used to render pages before extracting article links and metadata. Candidate links were collected from topic related pages and filtered before detailed extraction began. For The Guardian an API based collection strategy was used. Structured queries were sent to the Guardian content API with topic related search parameters and publication date filters. This approach improved extraction consistency and reduced parsing errors compared to direct HTML scraping.

Daily Star and New Age were collected primarily through HTML parsing and structured page traversal. Listing pages related to world news coverage were visited sequentially and candidate article links were extracted directly from the page structure. Each candidate article was then processed individually for metadata and full text extraction. Several filtering mechanisms were integrated directly into the collection framework. Opinion articles editorials live blogs and unrelated media pages were excluded whenever possible because the study focused mainly on comparative reporting patterns within standard news articles.

Topic relevance filtering also formed an important part of the collection pipeline. Articles were matched against predefined geopolitical keyword structures associated with Russia Ukraine Iran Israel and Taiwan Strait related reporting. Weighted matching procedures were also applied in some pipelines to improve contextual relevance and reduce false topic assignment. The extraction process preserved full article text instead of only collecting titles or summaries. This was considered important because ideological framing narrative emphasis and lexical asymmetries often emerge gradually throughout the article body. Full text preservation therefore improved the reliability of later NLP and LLM based analytical stages.

Duplicate management procedures were also integrated during collection. Existing URLs were checked continuously before new articles were stored. This reduced repeated entries and improved corpus consistency during long term daily updates. The collected articles were finally transformed into a standardized tabular structure containing publication date source region topic article title URL and full text content. This unified schema improved compatibility across downstream analytical stages and simplified comparative processing inside the Streamlit environment.

## 3.4 Data Preprocessing and Corpus Preparation

After the article collection stage the dataset passed through several preprocessing and cleaning procedures before analytical processing began. This stage was considered essential because raw web collected news data often contains structural inconsistencies duplicated content formatting artifacts broken encoding patterns and irrelevant textual elements that may negatively affect NLP and LLM based analysis.

The preprocessing pipeline was designed to improve textual consistency while preserving the semantic structure and contextual meaning of the articles. Instead of aggressively reducing or

simplifying the collected text the workflow attempted to maintain as much narrative information as possible because political framing and ideological emphasis often appear gradually across the full article body.

The first preprocessing stage focused on text extraction and normalization. During scraping many articles contained residual HTML fragments navigation elements advertisement related content embedded formatting tags and unnecessary whitespace patterns. BeautifulSoup based parsing methods were therefore used to isolate the primary article body from surrounding webpage components. Additional regular expression cleaning procedures were also applied to remove remaining formatting artifacts and inconsistent spacing structures.

Unicode normalization and text repair procedures were also integrated into the preprocessing pipeline. Web collected articles occasionally contained corrupted symbols encoding inconsistencies invisible Unicode characters or malformed text patterns created during browser rendering and content extraction. Such issues may reduce embedding quality interfere with tokenization and affect downstream analytical interpretation. Text normalization therefore improved overall corpus consistency before NLP processing began.

The preprocessing stage also included duplicate removal and structural validation procedures. Duplicate URLs were identified and removed before storage. Articles with missing publication dates invalid timestamps extremely short content or empty extraction results were also excluded from the final corpus. These filtering procedures improved dataset reliability and reduced analytical noise during later processing stages.

Another important preprocessing task involved topic alignment and relevance filtering. Since geopolitical reporting often overlaps across multiple international events some articles contained indirect references to more than one conflict. Topic assignment therefore required contextual filtering rather than simple keyword detection alone.

Different topic filtering strategies were implemented depending on the source characteristics. Some pipelines relied on direct keyword validation while others used weighted scoring approaches. Strong geopolitical indicators contributed more heavily toward topic classification while weaker contextual references were treated as secondary indicators. This structure improved contextual flexibility while still preserving topic specificity.

The preprocessing pipeline also attempted to exclude opinion oriented material whenever possible. Editorial sections opinion columns commentary pages and analytical blogs were filtered during scraping and preprocessing stages because the study focused mainly on comparative framing inside standard news reporting. Removing explicitly opinion driven material reduced the probability of analytical distortion during comparative evaluation.

Publication dates extracted from different sources were standardized into unified datetime structures through pandas based conversion procedures. Since the selected news outlets used different date formatting styles a standardized temporal representation was necessary for chronological filtering trend analysis and date based comparative visualization. The cleaned dataset was then organized into a structured tabular format containing publication date topic source region article title URL and full article text. Topic ordering and chronological sorting were also applied to maintain consistency during later analytical stages and visualization tasks.

An important objective of the preprocessing stage involved preservation of contextual richness. Many NLP pipelines aggressively reduce text through stemming lemmatization or heavy stopword removal. However this study focused strongly on narrative framing ideological wording and contextual interpretation. Excessive reduction of the article body could therefore remove meaningful linguistic patterns associated with political framing. Because of this the preprocessing stage remained relatively conservative regarding semantic reduction.

The preprocessing layer therefore functioned not only as a cleaning mechanism but also as a corpus preparation stage designed specifically for downstream comparative media analysis and LLM based interpretation.

## 3.5 NLP and LLM Based Analytical Framework

After preprocessing the cleaned corpus entered the analytical stage of the research framework. This stage combined Natural Language Processing techniques with OpenAI based Large Language Model analysis in order to evaluate narrative framing sentiment divergence lexical asymmetry and comparative reporting patterns across the selected news outlets.

The analytical framework was designed as a layered processing structure where traditional NLP operations and LLM driven interpretation complemented each other instead of functioning as isolated components. This combined approach was selected because media bias often contains both

measurable linguistic signals and deeper contextual framing patterns that cannot always be captured through conventional statistical methods alone.

The first analytical stage focused on corpus level textual preparation for model processing. Articles selected through Streamlit filtering were grouped according to topic source and date range. This allowed the analytical pipeline to compare articles discussing similar geopolitical events across different regional outlets.

OpenAI based language models were then integrated into the workflow through API driven processing. The LLM layer was used for comparative summarization classification narrative interpretation and bias related analytical evaluation. Instead of treating the language model only as a generative system the methodology positioned the model as an analytical instrument for identifying differences in framing emphasis and reporting structure across news sources.

The summarization process attempted to condense article content while preserving the main political narrative and contextual structure. This allowed comparative observation of how similar geopolitical events were described differently by different outlets. Since summarization systems may themselves introduce framing shifts this stage also supported investigation of whether LLM generated outputs reproduce amplify or reduce existing media bias patterns.

The analytical framework also included comparative narrative interpretation procedures. Articles discussing the same geopolitical topic from different outlets were analyzed comparatively in order to identify variations in wording emotional emphasis actor representation and contextual framing. Differences in narrative structure were treated as potential indicators of media bias or regional perspective divergence.

Keyword based and lexical analysis procedures were also integrated into the workflow. Topic specific language patterns military terminology diplomatic wording and ideological references were examined comparatively across the corpus. This allowed identification of linguistic asymmetries between Bangladeshi and international reporting structures.

The framework also supported sentiment oriented comparative analysis. Political conflicts are often reported with emotionally charged language and also some varying levels of intensity depending on the source perspective. Sentiment related processing most of the times helped evaluate whether

certain actors, conflicts or geopolitical positions were framed more positively, negatively or neutrally across different outlets.

An important aspect of the analytical framework involved contextual preservation. Instead of analyzing isolated sentences alone the pipeline preserved larger sections of article content during processing. This was considered important because political framing often emerges through narrative progression contextual ordering and selective emphasis rather than through individual words alone.

The analytical stage was intentionally designed to support interpretability rather than only prediction. Instead of producing simple classification labels the framework attempted to expose differences in narrative construction framing emphasis and contextual representation across sources. This approach aligned more closely with the research objective of investigating media bias through computational analysis.

Another important consideration involved scalability and flexibility. Since the corpus continues to grow daily the analytical framework was developed in a modular structure where additional NLP methods models or comparative procedures could later be integrated without major architectural modification.

The final analytical framework therefore combined automated NLP preprocessing OpenAI based language model interpretation comparative narrative analysis and sentiment oriented evaluation within a single computational media analysis environment.

## 3.6 Comparative Bias Detection Process

The comparative bias detection process formed the central analytical component of the research framework. This stage focused on identifying differences in reporting patterns between Bangladeshi and international news outlets through computational comparison of articles discussing similar geopolitical events. The process was designed around the assumption that media bias may appear through multiple linguistic and structural dimensions rather than through direct factual inaccuracies alone. Differences in emotional tone narrative emphasis actor representation lexical selection contextual prioritization and thematic framing were therefore treated as important analytical indicators.

The bias detection workflow began after articles were filtered according to topic source and temporal range inside the Streamlit environment. Articles discussing similar geopolitical conflicts were grouped together for comparative analysis. This grouping structure improved contextual consistency during LLM based interpretation and comparative evaluation. Comparative analysis was then performed across multiple dimensions. One dimension focused on lexical framing differences. Variations in diplomatic terminology military wording conflict description and political labeling were examined comparatively between sources. Certain outlets emphasized humanitarian framing while others focused more heavily on strategic military narratives diplomatic positioning or geopolitical stability.

Another dimension focused on narrative emphasis. Different sources often prioritized different aspects of the same event. Some articles emphasized civilian impact while others focused primarily on political leadership military escalation or international diplomatic response. These variations were treated as potential indicators of framing divergence across regional media environments. Sentiment divergence also formed an important part of the comparative process. Differences in emotional intensity positive or negative framing and descriptive tone were evaluated comparatively across the selected outlets. This stage helped identify whether specific actors governments or geopolitical positions were described differently depending on the source perspective.

The LLM based analytical layer also contributed to comparative bias interpretation through contextual analysis of article summaries and narrative structures. Since large language models are capable of interpreting broader semantic relationships they provided an additional analytical perspective beyond conventional keyword frequency analysis alone. Another important objective

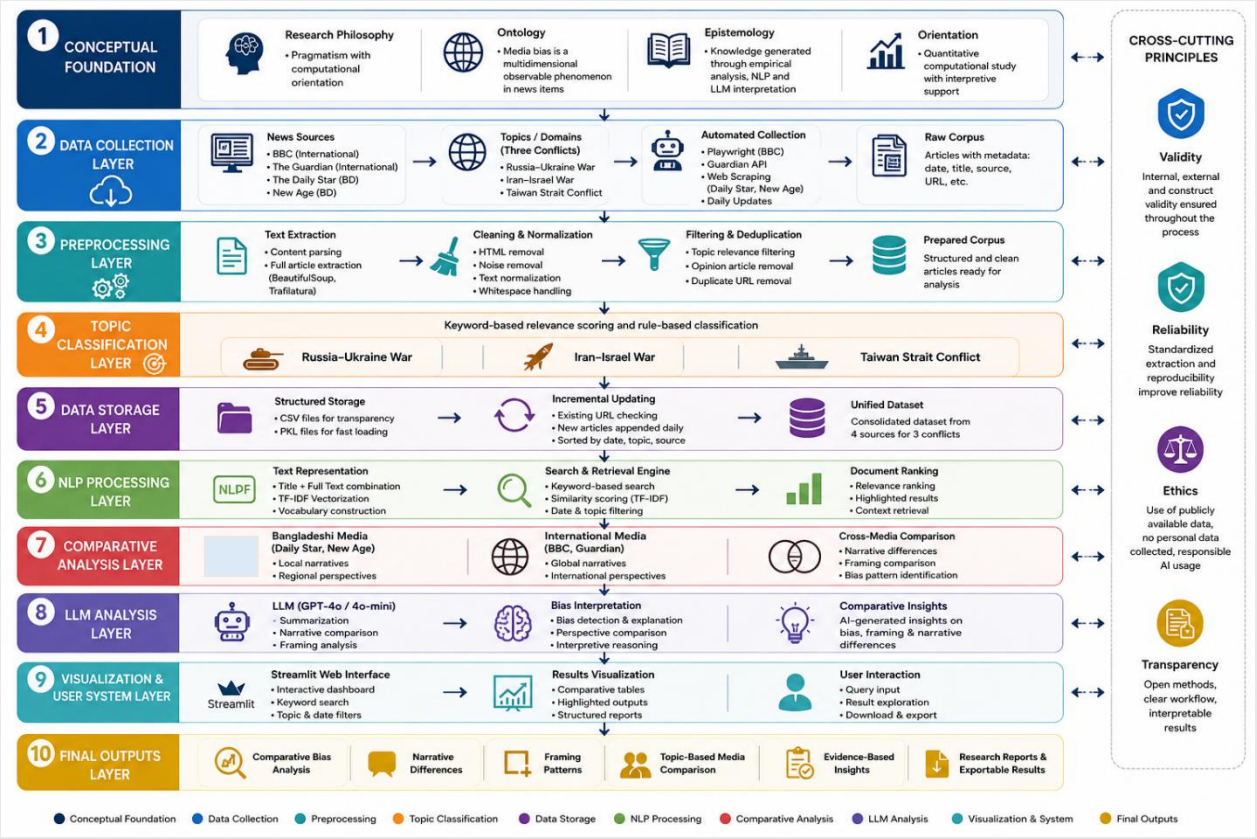


**Figure 1:** Overall architecture of the proposed computational media bias analysis framework

of the bias detection process involved examination of amplification effects created by AI systems themselves. The methodology therefore did not assume that LLM generated outputs are neutral representations of source material. Instead the framework examined whether summarization and interpretation stages preserve existing framing patterns or unintentionally strengthen certain narratives during automated processing.

The comparative workflow was intentionally designed to support exploratory interpretation rather than deterministic labeling. Media bias is highly contextual and politically sensitive. Because of this the framework attempted to expose narrative asymmetries and linguistic divergence rather than assigning absolute binary bias classifications. The final comparative bias detection framework therefore functioned as a computational analytical environment for examining how geopolitical events are framed differently across regional media systems and how AI driven analytical models interpret these differences during automated language processing tasks.

## 3.7 Streamlit Based Visualization and Interaction System

The final stage of the framework focused on visualization interaction and presentation of the analytical outputs through a Streamlit based web application. The Streamlit environment was developed as the primary user interaction layer of the system and functioned as an interface between the processed news corpus and the NLP and LLM based analytical pipeline. The interface was designed to support comparative exploration of geopolitical news coverage across different media outlets. Instead of presenting static analytical results the system allowed dynamic filtering and interactive analysis based on user selected conditions. This structure improved interpretability and made the analytical workflow more flexible during experimentation and evaluation.

The Streamlit application was connected directly with the processed dataset and analytical modules. Users could select geopolitical topics choose news sources define date ranges and search through article content using keyword based filtering. These filtering operations allowed the corpus to be narrowed into context specific subsets before analysis was performed. The interface also displayed processed article information including source topic publication date article title and cleaned text content. This made it possible to inspect the original reporting material before examining comparative analytical outputs. Preserving access to the original article context was considered important because interpretability and transparency formed major objectives of the research framework.

After article selection the Streamlit system passed the filtered corpus into the analytical layer where OpenAI based processing and comparative evaluation tasks were performed. The created outputs were then displayed directly inside the interface environment. This created a connected workflow where collection preprocessing analysis and visualization operated together inside the same system architecture. The visualization layer was also designed to support comparative observation between Bangladeshi and international reporting structures. Articles discussing similar geopolitical events could be explored side by side which improved interpretability during framing analysis and narrative comparison tasks.

Another important aspect of the Streamlit environment involved usability and accessibility. Since the project combined multiple computational stages including scraping preprocessing NLP analysis and LLM driven interpretation a simplified interaction layer was necessary for practical exploration

of the results. Streamlit provided a lightweight framework for building this environment while maintaining compatibility with Python based processing pipelines.

The interface also improved flexibility during experimentation because analytical outputs could be regenerated dynamically after changing filters or selecting different topic ranges. This structure allowed repeated comparative evaluation without modifying the underlying codebase manually during each analytical iteration. The Streamlit layer therefore functioned not only as a visualization environment but also as an operational analytical interface connecting the corpus preprocessing framework and LLM based comparative analysis system into a unified research platform.

## 3.8 Validation and Quality Control

Several validation and quality control procedures were integrated throughout the research pipeline in order to improve corpus reliability extraction accuracy and analytical consistency. Since the framework relied heavily on automated scraping preprocessing and LLM based interpretation continuous validation was considered important during all stages of the workflow.

The first validation stage focused on article extraction quality. Randomly selected articles from each news source were inspected manually after scraping in order to verify title accuracy publication date consistency and correctness of the extracted article body. This step was necessary because automated web extraction may occasionally capture incomplete content duplicated sections or unrelated webpage elements.

Topic relevance validation was also performed during corpus construction. Although keyword based and weighted filtering procedures were implemented automatically manually reviewing selected article samples helped verify whether the collected articles actually belonged to the intended geopolitical categories. This was particularly important because some political articles contain indirect references to multiple conflicts simultaneously.

Duplicate management procedures formed another important part of the quality control framework. Existing URLs were checked continuously before new articles were added into the corpus. This reduced repeated entries and improved consistency during longitudinal dataset expansion. Additional preprocessing validation procedures were also performed after text cleaning stages. Articles with extremely short content broken formatting invalid timestamps or missing metadata

were removed automatically from the dataset. These procedures reduced analytical noise and improved downstream model processing reliability.

The research also attempted to maintain consistency across different scraping strategies. Since BBC The Guardian Daily Star and New Age all used different webpage structures separate collection methods were implemented for each source. Despite these differences all extracted articles were transformed into a unified schema before analytical processing began. This standardization improved comparability across the final corpus.

The LLM based analytical stage also required careful validation because language models may occasionally generate inconsistent or contextually weak interpretations. Because of this analytical outputs were examined comparatively alongside the original article content rather than treated as fully autonomous conclusions. This approach improved interpretability and reduced overreliance on purely automated interpretation. Another important quality consideration involved reproducibility. The pipeline structure storage format and preprocessing workflow were designed so that the dataset could be updated repeatedly while preserving consistency across later executions. Incremental update procedures and automated scheduling also improved operational stability during long term data collection.

The validation framework therefore combined automated filtering structural consistency checks manual inspection procedures and comparative review mechanisms in order to improve overall reliability of the research environment.

## 3.9 Ethical Considerations

Several ethical considerations were taken into account during development of the research framework. The study relied exclusively on publicly accessible news articles collected from openly available online news sources. No private user data personal identifiers or restricted information were collected during the research process.

The project focused on computational analysis of media narratives rather than evaluation of individual journalists organizations or political actors. The objective was to investigate comparative reporting patterns and AI driven interpretation processes rather than assign absolute political judgments to specific news outlets.

Another important ethical consideration involved responsible use of Large Language Models. Although OpenAI based systems were integrated into the analytical framework LLM generated outputs were not treated as objective truth. Language models may contain embedded biases inherited from training data model architecture or contextual prompting behavior. Because of this the study examined LLM outputs critically rather than assuming complete neutrality.

The research also acknowledged that media bias itself is a highly contextual and subjective concept. Different audiences political systems and regional environments may interpret the same article differently depending on ideological perspective cultural background or geopolitical position. Because of this the framework attempted to identify comparative framing asymmetries and narrative divergence rather than produce rigid binary classifications of bias.

## 3.10 Use of AI Assisted Tools

During preparation of the thesis, limited AI assisted tools and Large Language Model based systems were used to support grammatical correction, sentence refinement, writing clarity improvement and limited assistance during parts of the coding process. AI assistance was also used for some visual illustrations and technical explanations during documentation of the system architecture. All generated outputs, code suggestions and written content were manually reviewed, modified and verified before inclusion in the final thesis in order to maintain academic consistency and research accuracy.

# 4. Analysis, Results and Evaluation

### 4.1 Analysis

The analysis stage focused on examining the structure, distribution and comparative characteristics of the continuously collected geopolitical news corpus. Unlike many traditional datasets used in computational media research, the developed framework operated as a continuously updating analytical environment where new articles were collected automatically every day through GitHub Actions based scheduling. Because of this, the dataset evolved gradually throughout the research period and reflected changing geopolitical reporting patterns over time instead of remaining fixed to a single static collection period.

On 29 May 2026 the system contained a total of 6,165 processed geopolitical news articles collected from four selected media outlets including BBC, The Guardian, The Daily Star and New Age. The analysis showed that The Daily Star contributed the largest portion of the dataset with 3,318 articles, followed by New Age with 1,234 articles, The Guardian has 908 articles and BBC has 705 articles. And this number keeps increasing each day. This distribution suggested that the Bangladeshi news sources generated a substantially larger quantity of collected geopolitical reporting compared to the selected international outlets during the collection timeline.

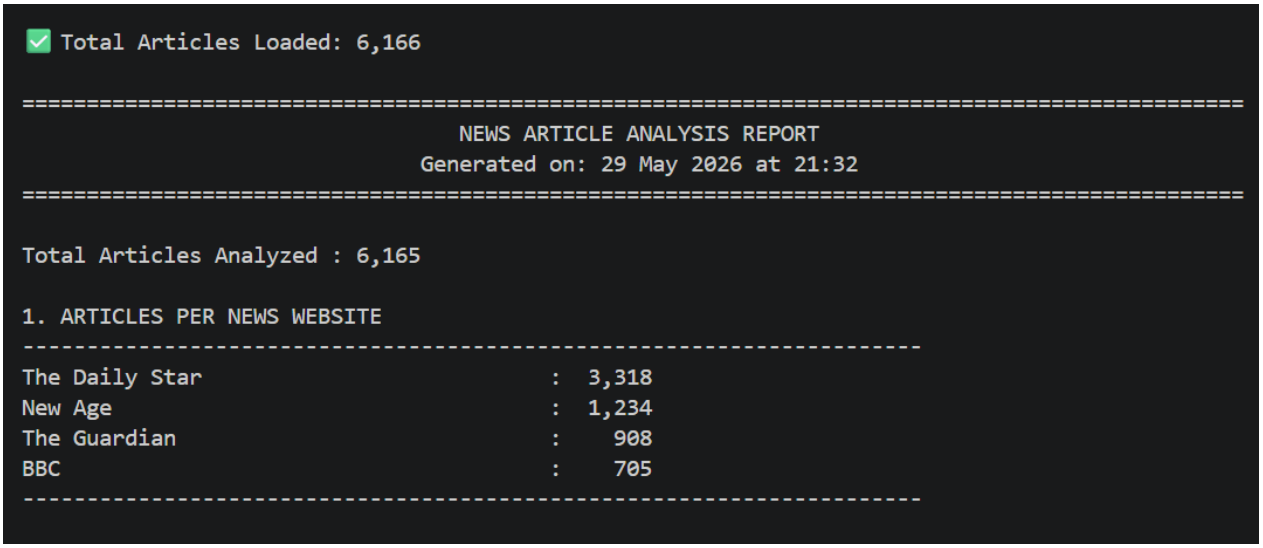


Figure 2: Analysis of Articles

The topic distribution analysis revealed that the Iran Israel conflict dominated the overall corpus very strongly. It also suggests that current trend gets the most news coverage. Out of the total 6,165 articles, 4,136 articles were about the Iran Israel war, representing approximately 67.1 percent of the entire dataset. In comparison, 1,976 articles were associated with the Russia Ukraine war, accounting for approximately 32.1 percent of the corpus. Only 53 collected articles were related to Taiwan Strait tensions, representing less than 1 percent of the total dataset.

This topic imbalance reflected important changes in global media attention during the later stages of the collection period. Since the framework continuously collected articles through automated daily execution, the dataset naturally adapted to real world geopolitical developments and changing reporting intensity across international media systems. As tensions surrounding the Iran Israel conflict escalated further during the research timeline, the frequency of related reporting increased

```
1. ARTICLES PER NEWS WEBSITE
------------------------------------------------------------------------
The Daily Star                          :  3,318
New Age                                 :  1,234
The Guardian                            :    908
BBC                                     :    705
------------------------------------------------------------------------

2. BREAKDOWN BY TOPIC (Overall)
------------------------------------------------------------------------
Iran Israel war                         :  4,136 ( 67.1%)
Russia Ukraine war                      :  1,976 ( 32.1%)
Taiwan Strait conflict                  :     53 (  0.9%)
------------------------------------------------------------------------
```

Figure 3: Breakdown of Articles

significantly across all selected outlets. The system therefore captured evolving media attention patterns dynamically rather than relying on manually balanced sampling procedures.

The comparative topic breakdown per website also revealed several noticeable differences in reporting priorities between media organizations. The Daily Star showed the strongest concentration toward Iran Israel related reporting, where approximately 72 percent of collected articles focused on that conflict. New Age demonstrated a very similar reporting pattern with nearly 73 percent of its articles connected to the same topic. These findings suggested that Bangladeshi media coverage during the observed period became highly concentrated around developments in the Middle East.

BBC and The Guardian displayed comparatively more balanced topic distributions between the Iran Israel and Russia Ukraine conflicts. The Guardian showed the most balanced coverage structure among all selected sources, with 49.3 percent of articles related to the Iran Israel conflict and 48.7 percent connected to the Russia Ukraine war. This indicated that the selected international outlets maintained broader geopolitical attention across multiple ongoing conflicts rather than concentrating predominantly on a single geopolitical event.

```
3. TOPIC BREAKDOWN PER WEBSITE
================================================================================================

BBC (705 articles):
   • Iran Israel war                    :   401 ( 56.9%)
   • Russia Ukraine war                 :   287 ( 40.7%)
   • Taiwan Strait conflict             :    17 (  2.4%)

NEW AGE (1234 articles):
   • Iran Israel war                    :   899 ( 72.9%)
   • Russia Ukraine war                 :   326 ( 26.4%)
   • Taiwan Strait conflict             :     9 (  0.7%)

THE DAILY STAR (3318 articles):
   • Iran Israel war                    : 2,388 ( 72.0%)
   • Russia Ukraine war                 :   921 ( 27.8%)
   • Taiwan Strait conflict             :     9 (  0.3%)

THE GUARDIAN (908 articles):
   • Iran Israel war                    :   448 ( 49.3%)
   • Russia Ukraine war                 :   442 ( 48.7%)
   • Taiwan Strait conflict             :    18 (  2.0%)

================================================================================================
                              Analysis Completed Successfully!
```

Figure 4: Topic breakdown per website

Another important analytical observation involved the relative absence of Taiwan Strait related reporting across all selected media sources. Despite the geopolitical importance of Taiwan related tensions within global political discourse, only a very small number of related articles appeared inside the collected dataset. This pattern may reflect a form of agenda imbalance where certain geopolitical issues receive substantially lower media visibility compared to active military conflicts receiving continuous international attention. From a comparative media perspective, this observation aligns closely with agenda setting theory because media influence operates not only through what is reported frequently but also through what receives limited visibility or reduced narrative emphasis.

The continuously updating nature of the framework also introduced an important methodological characteristic into the overall analytical environment. Since the dataset expanded automatically every day through GitHub Actions scheduling, the corpus remained temporally dynamic rather than historically static. This improved the realism of the research because the system continuously

adapted to emerging geopolitical developments, changing narrative structures and evolving media priorities over time.

The preprocessing and duplicate management procedures additionally contributed significantly to analytical consistency and dataset reliability. Existing URLs were checked continuously before insertion into the corpus, which reduced repeated storage of identical articles and preserved structural consistency during long term automated collection. Topic filtering procedures further improved dataset relevance by excluding unrelated political material and maintaining analytical focus on the selected geopolitical conflicts.

Overall, the analysis demonstrated that the developed computational framework was capable of maintaining a large scale continuously evolving geopolitical news corpus suitable for comparative NLP and LLM based media bias analysis. The resulting dataset reflected both differences in reporting priorities across regional media systems and broader shifts in international geopolitical attention throughout the research timeline.

## 4.2 Results

The developed analysis produced noticeable differences between Bangladeshi and international media coverage of the Russia Ukraine war. Different reporting patterns, framing styles and narrative focuses were identified from the selected articles. Although the same conflict was discussed by both media groups, the presentation style and narrative emphasis varied significantly.

The Bangladeshi media coverage focused mainly on immediate conflict events and humanitarian consequences. Greater attention was given to missile attacks, infrastructure destruction, civilian casualties and direct military actions. They put emphasis mostly on the emotional sides. Most of the retrieved Bangladeshi articles discussed the physical impact of the conflict rather than broader geopolitical implications. Due to this reporting pattern, the Bangladeshi narrative appeared more conflict centered and event focused.

In several retrieved Bangladeshi articles, strong emphasis was placed on destruction of energy facilities, emergency conditions and civilian suffering. The reporting style remained closely

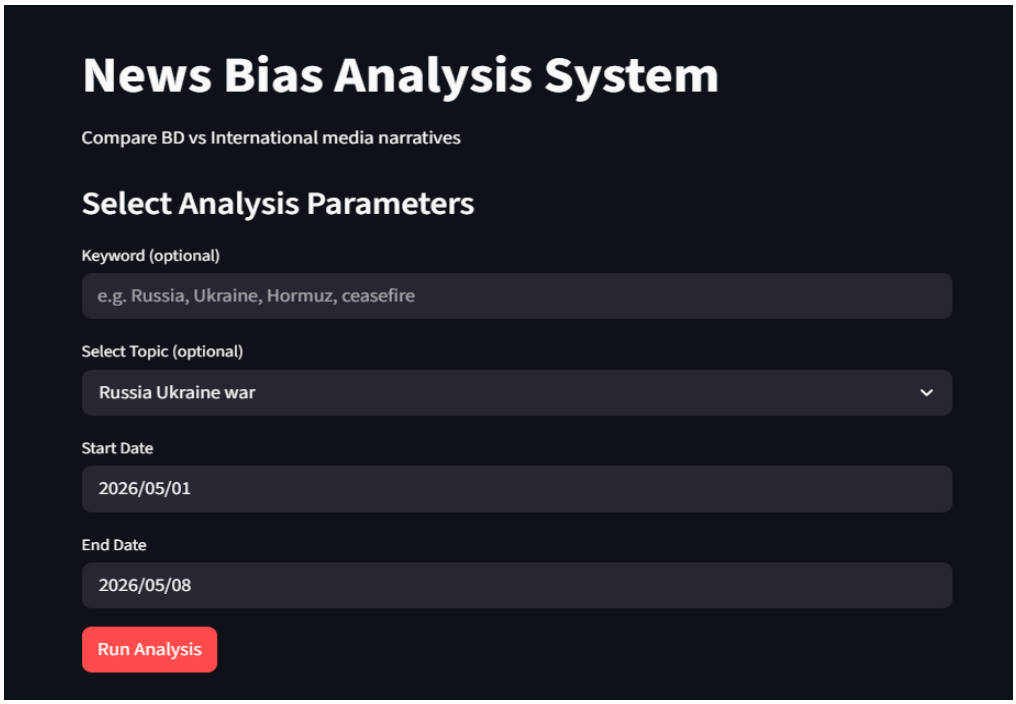

Figure 5: Streamlit-based interface used for selecting geopolitical topics, date ranges and analysis parameters for comparative media evaluation

connected to visible consequences of war. Political interpretation and strategic analysis were discussed less frequently compared to direct battlefield developments. As a result, the generated summaries reflected a more immediate and humanitarian oriented narrative structure.

In contrast, the international media coverage presented a broader interpretive narrative. Along with battlefield developments, significant attention was also given to diplomatic negotiations, geopolitical consequences, military strategy and symbolic political events. The international reporting frequently connected war events with larger international implications and future political outcomes.

The first major result was produced through the comparison of Bangladeshi and international reporting on the Russia Ukraine war. In the Bangladeshi coverage, the conflict was mainly described through the direct effects of war. Military strikes, casualties, infrastructure damage and the human cost of the conflict were placed at the center of the narrative. The tone was factual and conflict centered. The selected Bangladeshi articles gave strong attention to the immediate

Figure 6: Generated results within the selected articles

consequences of Russian attacks on Ukrainian energy systems, while the issue of ceasefire was presented with visible skepticism. Zelensky's criticism of the ceasefire offer was also emphasized, which made the narrative appear more suspicious of Russian intentions than neutral diplomatic reporting would usually be. The result was a coverage pattern that focused on harm, destruction and distrust. The international coverage produced a broader frame. The same war was not only presented as a battlefield event, but also as a geopolitical and symbolic struggle. Different topics about coverage of ceasefire negotiations, accusations of violations, the meaning of Victory Day in Russia and wider international effects was included in the narrative in many ways. The international reporting also placed more attention on strategic questions, such as military performance, diplomatic pressure and the possible role of other states in the conflict. Because of this wider framing, the international narrative appeared more analytic and more contextual than the Bangladeshi one. It did not remain fixed on casualties alone, but moved toward interpretation of political and military consequences.The main framing difference was therefore found in scope. The Bangladeshi media was observed to frame

the conflict in a narrower and more immediate way. Attention was placed on violence, damage and the suffering caused by the war. The international media was observed to frame the same conflict in a broader way. Attention was placed not only on violence, but also on diplomacy, symbolism, strategic performance and global implications. This meant that the Bangladeshi coverage was more

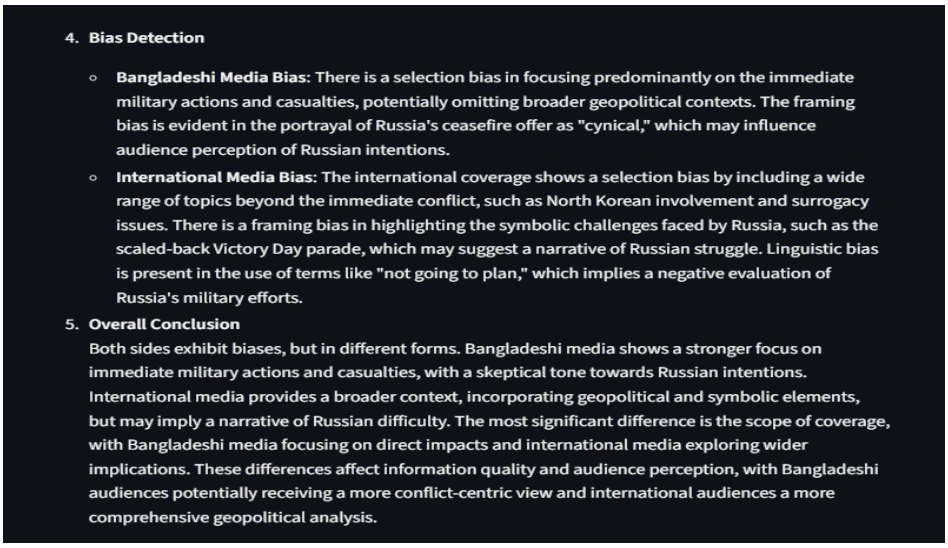

Figure 7: Generated Results within the selected articles

event centered, while the international coverage was more context centered. The two frames did not only differ in detail. They differed in what was treated as the main story itself. Several bias

patterns were then identified in the analysis. In the Bangladeshi coverage, selection bias was noted because immediate military action was emphasized more strongly than wider geopolitical context. Framing bias was also noted because the ceasefire proposal was presented in a skeptical way, which could shape audience perception of Russian intentions. In the international coverage, selection bias was also present because a wider range of themes was selected, including issues that moved beyond the direct battlefield.

Framing bias was observed in the presentation of Russian difficulty and symbolic weakness, especially where the broader meaning of events was highlighted. Linguistic bias was also present in the international framing because evaluative language gave a negative interpretation to Russian performance. The framework identified potential indicators of framing, selection and linguistic bias, but the type of bias was different in each case.

The overall result of the comparison was that the system was able to produce a structured comparison between two media ecosystems and then explain their differences in a clear form. The Bangladeshi narrative was found to be more concentrated on the direct human and infrastructural effects of the war. The international narrative was found to be more interpretive and more geopolitically broad. The system therefore did not simply classify articles by topic. It also reconstructed the dominant story shape in each media environment. For a thesis in business data

science, this was an important outcome because it showed that a text based retrieval and language model pipeline could be used to turn raw news articles into a structured narrative comparison.

The second important result was the academic evaluation layer. The system was able to assign criterion based scores to the analysis. In the sample shown in the application, the overall quality was placed in the moderate to good range. The scores suggested that the analysis was faithful to the source texts, reasonably balanced and clearly structured, although the evidence depth and the balance between the two media spheres could still be improved. This meant that the output was not treated as final truth. Instead, it was treated as a machine generated interpretation that still required academic checking. That design choice made the system more suitable for research use than for casual summarization.

A further result was found in the way the data pipeline was organized for repeated use. The scrapers were built to skip already collected URLs, to stop when old dates were reached and to save results in both CSV and pickle formats. The files were also sorted so that the newest date appeared first and the topic order remained stable. This reduced noise in later analysis and made repeated runs more reliable. It also meant that the language model was fed with cleaner and more consistent article blocks. The output quality was therefore supported not only by the prompt design, but also by the upstream data engineering work.

## 4.3 Evaluation

The outputs were further evaluated in order to examine the quality, consistency and reliability of the developed bias analysis system. The evaluation process focused on how accurately the analysis reflected the original news articles and whether the produced comparisons remained balanced, understandable and academically useful. Several evaluation criteria were used during this process, including fidelity to texts, balance and fairness, evidence quality, framing analysis depth, bias detection accuracy, structure and clarity and overall objectivity.

The evaluation results showed that the analysis maintained strong consistency with the retrieved articles. Most important themes, events and reporting patterns from the original news texts were reflected correctly in the shown summaries. The major conflict related developments discussed in the selected articles were preserved during the analysis generation process. Due to this, the system achieved a relatively strong score in fidelity to texts. The narrative summaries remained closely connected to the original reporting context and avoided major factual deviation.

Figure 8: Evaluation module assessing fidelity, fairness, framing depth and analytical consistency of generated outputs.

The evaluation also showed that the system was capable of producing reasonably balanced comparisons between Bangladeshi and international media. Both media groups were analyzed separately before comparative conclusions were generated. The system identified bias patterns from both sides instead of presenting only one side as biased. This improved the fairness of the outputs. However, small differences in narrative depth were still observed between the two sides. International coverage sometimes appeared more detailed because broader geopolitical discussions were included in the retrieved articles. As a result, slight imbalance could occasionally be observed in the final comparative interpretation.

Another important part of the evaluation was evidence quality. The analysis used examples and reporting patterns taken from the retrieved articles in order to support the final conclusions. Direct references to missile attacks, ceasefire proposals, infrastructure damage, geopolitical interpretation

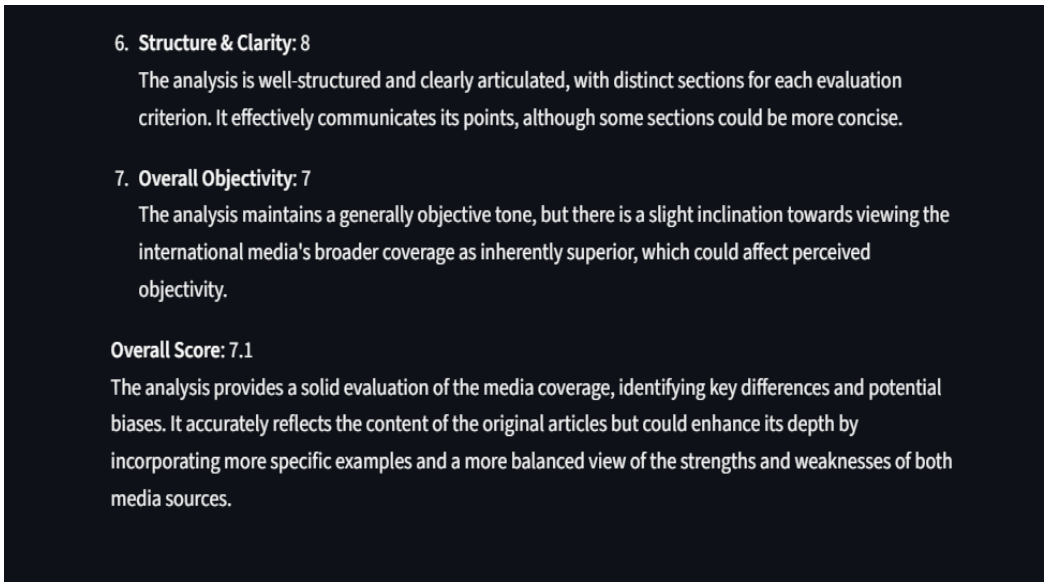


Figure 9: Evaluation module and overall score

and strategic discussions were included within the summaries. These examples strengthened the credibility of the produced analysis. However, the evaluation also showed that some conclusions could have been supported with more direct article based references and stronger textual examples. Because of this, the evidence quality score remained moderate instead of extremely high.

The framing analysis depth was also evaluated carefully. The outputs successfully identified differences in how the same conflict was presented across different media environments. Bangladeshi reporting was mainly associated with direct humanitarian and military consequences, while international reporting was associated with broader geopolitical interpretation. The framing comparison section clearly explained these differences. The system therefore demonstrated the ability to move beyond simple summarization and produce interpretive comparative analysis. At the same time, some deeper political and historical framing dimensions were not explored extensively, which slightly reduced the overall framing depth score.

The bias detection performance of the system produced positive evaluation results. Different forms of bias such as selection bias, framing bias and linguistic bias were identified successfully from both media groups. The analysis explained how certain topics were emphasized more strongly in one reporting environment compared to another. Differences in wording, topic selection and narrative structure were also detected effectively. This showed that the developed pipeline was capable of identifying patterns connected to media bias rather than only producing general summaries of news events.

The structure and clarity of the outputs received strong evaluation results. The analysis was divided into organized sections including narrative summaries, framing comparison, bias detection and overall conclusion. This structure improved readability and made the outputs easier to follow. Each section focused on a specific analytical purpose, which reduced confusion and improved the overall presentation quality. The evaluation showed that the reports remained understandable even when large amounts of information from multiple articles were processed together.

The overall objectivity of the analysis was also examined during the evaluation stage. The outputs generally maintained a neutral academic tone while discussing both media groups. Strong emotional language was mostly avoided in the explanations. Biases were identified through reporting patterns and framing differences rather than through personal opinion. This improved the objectivity of the final analysis. However, slight preference toward broader international contextual interpretation could still be noticed in some generated outputs because international reporting contained larger thematic diversity and more geopolitical detail.

The evaluation process additionally demonstrated that the system could generate stable outputs across repeated runs with similar query conditions. When similar topics and time periods were selected, comparable narrative patterns and framing differences continued to appear in the analysis. This consistency suggested that the retrieval and analysis pipeline remained relatively stable during operation. The repeated appearance of similar comparative patterns strengthened confidence in the reliability of the produced outputs.

Another important observation from the evaluation was related to topic relevance. Most retrieved articles remained strongly connected to the selected conflict topics, which improved the quality of the summaries and comparisons. Topic filtering and keyword matching reduced unrelated article inclusion during the retrieval stage. Because of this, the analysis remained focused on the selected geopolitical conflict instead of moving toward unrelated discussions. This contributed positively to both clarity and reliability.

The evaluation also revealed several limitations within the system. The quality of the analysis remained dependent on the quality and diversity of the retrieved articles. If limited articles were available within a selected date range, the final narrative comparison also became narrower. In addition, large language model generated analysis could occasionally produce interpretations that

were broader than the original article content. Although major factual inconsistency was not observed in the tested outputs, careful human review still remained necessary for academic use.

Another limitation identified during evaluation was related to source imbalance. International media sources generally contained longer articles with more geopolitical explanation, while Bangladeshi sources often contained shorter and more event focused reporting. Because of this structural difference, the comparative analysis occasionally contained more detailed interpretation for the international side. This limitation was connected to differences in source material rather than direct system failure.

The evaluation process further showed that the system functioned effectively as an exploratory analytical tool rather than as a fully autonomous truth verification system. The outputs were useful for identifying narrative differences, framing patterns and possible bias structures across media environments. However, the results still required academic interpretation and human verification before final conclusions could be accepted without further review. This was considered important because media bias analysis involves interpretive dimensions that cannot always be measured perfectly through automated systems alone.

Despite these limitations, the evaluation results demonstrated that the developed system achieved its primary research objective successfully. Comparative media narratives were generated clearly, major framing differences were identified effectively and multiple forms of bias were detected from both Bangladeshi and international reporting environments. The final outputs remained structured, readable and analytically useful for comparative media studies.

Overall, the evaluation findings suggested that the integration of automated article retrieval, topic filtering and large language model based interpretation can support meaningful comparative bias analysis within modern news environments. The developed system showed that large collections of news articles could be transformed into organized analytical outputs capable of identifying framing patterns, thematic emphasis and reporting differences across media sources.

The evaluation scores were generated using the same LLM-based framework employed during analysis. Therefore, these scores should be interpreted as model based assessments rather than independent validation measures. Future research should incorporate human evaluators or multiple independent models to strengthen evaluation reliability.

## 5. Conclusion

This thesis investigated how Large Language Model based analytical systems can be used to detect, quantify and critically examine media bias within comparative geopolitical news coverage from Bangladeshi and international media outlets. The study combined automated news collection, Natural Language Processing, multilingual textual analysis and LLM driven interpretation within a unified computational framework designed specifically for large scale comparative media analysis. By focusing on reporting related to the Russia-Ukraine war, Iran-Israel conflict and Taiwan Strait tensions, the research explored how political narratives are constructed differently across regional media environments and how AI systems interpret those differences during automated analytical processing.

The overall findings of the research demonstrate that computational analytical systems can successfully identify meaningful variations in framing, narrative emphasis, lexical structure and sentiment orientation across news corpora originating from different regional contexts. The developed framework was able to transform large collections of continuously updated geopolitical news articles into a properly structured analytical outputs that revealed important differences in reporting behavior between Bangladeshi and international media sources. Rather than treating media bias as a simple binary concept, the thesis approached bias as a multidimensional phenomenon that emerges through language selection, contextual prioritization, emotional intensity, actor representation, thematic focus and narrative structure. This broader interpretation allowed the system to move beyond simplistic classification approaches and support deeper comparative analysis.

One of the major contributions of the study was the successful development of a fully integrated analytical pipeline capable of handling the entire workflow from data collection to visualization and interpretation. The system demonstrated that automated scraping, preprocessing, topic filtering, comparative NLP analysis and LLM driven interpretation can operate together within a stable and scalable research environment. Articles from BBC, The Guardian, Daily Star and New Age were collected continuously through source specific scraping strategies and transformed into a unified corpus structure suitable for comparative analysis. Despite major differences in webpage architecture, metadata formatting and article structures, the framework maintained consistency across the dataset through preprocessing standardization and structured validation procedures.

The findings showed that multilingual and NLP based analytical procedures were effective in identifying linguistic framing asymmetries, sentiment divergence and ideological lexical patterns across the selected geopolitical conflicts. Comparative analysis consistently revealed that international outlets often provided broader geopolitical interpretation, strategic framing, diplomatic analysis and historical context within their reporting. In contrast, Bangladeshi outlets more frequently emphasized immediate humanitarian consequences, direct conflict developments, casualty information and event centered reporting structures. These differences were visible not only through narrative interpretation but also through recurring linguistic patterns, topic emphasis and contextual organization throughout the articles.

The first sub research question asked to what extent multilingual NLP models could reliably identify framing asymmetries, sentiment divergence and ideological lexical markers across cross regional news corpora. The findings indicate that the combined NLP and LLM framework was largely successful in this objective. The system repeatedly detected variations in emotional tone, military terminology, diplomatic wording and political framing between sources discussing similar events. Comparative analysis demonstrated that certain outlets framed geopolitical actors more strategically while others framed them more emotionally or humanitarianly. Differences in actor representation, conflict emphasis and thematic prioritization were also observed consistently throughout the analytical outputs.

At the same time, the findings also showed that purely statistical or keyword based methods alone were not sufficient for capturing deeper narrative structures. Contextual interpretation provided by Large Language Models significantly improved the ability to detect subtle framing differences that traditional frequency analysis might overlook. Preserving full article text instead of reducing articles into heavily simplified representations also improved analytical depth because many framing patterns emerged gradually through contextual progression rather than through isolated keywords alone. Therefore, the study concludes that multilingual NLP systems are effective for identifying measurable discourse asymmetries, but their analytical performance becomes significantly stronger when combined with contextual LLM driven interpretation.

The second sub research question examined whether LLM driven summarization and classification systems reproduce, reduce or amplify existing bias patterns when analyzing politically sensitive news content. The findings suggest that LLM systems generally preserved many of the original

narrative tendencies present in the source articles and occasionally amplified certain dominant framing structures during summarization and interpretation stages. The models often reflected the broader contextual richness available within the retrieved articles. Since international articles frequently contained longer geopolitical explanations and more extensive contextual information, generated outputs occasionally produced deeper interpretive discussion for international reporting compared to Bangladeshi reporting.

However, the evaluation process did not reveal major factual distortion or intentional ideological manipulation generated by the models themselves. Instead, amplification effects mainly appeared through selective emphasis, contextual expansion or interpretive weighting of already dominant narratives within the source material. This finding is highly important because it demonstrates that Large Language Models should not be treated as neutral or fully objective analytical instruments. Their outputs remain dependent on source structure, training data tendencies, prompt design and contextual distribution within the analyzed corpus. Consequently, the thesis argues that LLM outputs should function as interpretive analytical support mechanisms rather than autonomous truth verification systems.

Another important finding of the research involves the relationship between automation and interpretability. Many machine learning systems used in text analysis operate as highly opaque classification structures where users receive outputs without understanding how conclusions were reached. In contrast, the developed framework emphasized interpretability throughout the workflow. Comparative summaries, framing explanations, narrative observations and sentiment related outputs remained visible and inspectable during analysis. This improved transparency and made the analytical process more academically useful because outputs could be evaluated directly alongside original article content.

The Streamlit based interaction system also contributed significantly to the practical value of the research. The interface transformed the analytical framework from a purely experimental environment into a functional exploratory media analysis platform. Users could filter articles according to topic, date range, keyword and source before generating comparative outputs dynamically. This interactive structure improved usability and demonstrated how computational media analysis can be integrated into applied analytical environments for journalism research, strategic communication, policy observation and media monitoring.

The study also demonstrated the importance of continuous and scalable data collection within geopolitical media analysis. Political narratives evolve rapidly, especially during active international conflicts. Static datasets often fail to capture changing narrative structures over time. By designing the framework as a continuously updating corpus beginning from June 2025, the research preserved temporal flexibility and supported ongoing longitudinal analysis. Automated scheduling through GitHub Actions further improved operational consistency and reduced dependence on manual data collection.

Despite these contributions, the thesis also identified several important limitations. The first limitation involved source imbalance. International outlets generally produced longer articles with richer geopolitical context and more extensive diplomatic analysis, while Bangladeshi outlets often focused more narrowly on immediate developments. This imbalance occasionally influenced comparative interpretation depth. Another limitation involved the interpretive nature of media bias itself. Bias is not an objectively measurable entity in all situations. Because audience interpretation, political ideology, cultural perspective and regional context influence how narratives are perceived. That's why, The framework was intentionally designed for exploratory comparative analysis rather than rigid binary classification.

The research also acknowledged limitations connected to LLM based interpretation. Although the generated outputs remained largely coherent and analytically useful, occasional overgeneralization and contextual broadening still occurred during summarization stages. Human review therefore remained necessary throughout evaluation and interpretation. The system should not be considered a fully autonomous analytical authority. Instead, it should be understood as a computational support environment that assists researchers in identifying potential framing patterns, narrative asymmetries and comparative discourse structures.

Another limitation involved corpus scope. The study focused only on four news outlets and three geopolitical conflicts within a specific temporal range. Consequently, the findings cannot automatically be generalized to all global media systems or all political reporting environments. However, the modular structure of the framework allows future expansion into additional languages, regions, media systems and analytical models. Within a theoretical perspective, the thesis contributes to ongoing discussions surrounding media bias, framing theory, computational discourse analysis and democratic communication. The findings support the argument that media

bias is often embedded within subtle linguistic and contextual structures rather than direct factual distortion alone. The research also demonstrates that Data Science methodologies can operationalize abstract theoretical concepts such as framing, narrative asymmetry and representational imbalance into measurable analytical indicators. By integrating computational methods with media theory, the thesis bridges an important gap between social science interpretation and scalable analytical automation.

From a technical perspective, the study demonstrates how modern NLP systems and Large Language Models can support large scale comparative media analysis without requiring fully labeled datasets or manually coded corpora. That is the reality of it. The combination of automated scraping, structured preprocessing, contextual interpretation and interactive visualization created a flexible and extensible framework capable of supporting future research development. The modular architecture also allows additional analytical components such as multilingual translation systems, advanced sentiment classifiers, topic modeling techniques or real time monitoring modules to be integrated later without major redesign.

The overall conclusion of this thesis is that LLM driven analytical systems can successfully support comparative media bias analysis when combined with structured preprocessing, contextual preservation, validation procedures and critical human interpretation. The developed framework demonstrated strong capability in identifying framing differences, sentiment divergence, narrative emphasis and comparative reporting asymmetries across Bangladeshi and international news environments. At the same time, the study also showed that AI systems themselves are not completely neutral analytical observers. It depends a lot in their training. Large Language Models may preserve or amplify dominant contextual patterns already embedded within source material, making critical evaluation and transparency essential during analytical use.

Ultimately, this thesis demonstrates that computational media analysis can move beyond simple keyword statistics and become a meaningful interdisciplinary tool for understanding geopolitical discourse in modern digital environments. The tools that were used here and by integrating NLP, LLM based interpretation, automated corpus construction and interactive visualization within a single framework, the research provides both methodological and conceptual contributions to the growing field of AI assisted media analysis. The developed system therefore represents not only a technical achievement but also a broader analytical model for investigating how narratives are

constructed, distributed and interpreted across increasingly complex global information ecosystems.

# 6. Limitations and future Work

## 6.1 Limitations

Despite the successful development and implementation of the proposed framework, several limitations were identified throughout the research process. Which is usual for these kinds of models. These limitations are important to acknowledge properly because they influence the interpretability, scalability and generalizability of the findings produced by the system.

One of the primary limitations of the study was the restricted corpus size and source diversity. The research focused on only four news outlets, specifically BBC, The Guardian, Daily Star and New Age, while analyzing three selected geopolitical conflicts. Although these sources provided sufficient variation for comparative analysis. They do not fully represent the broader global media ecosystem. It was a limitation for many reasons. Media organizations operate within highly different political, cultural, linguistic and institutional environments, meaning that reporting structures and framing behaviors may vary significantly across additional regions and publication systems. Consequently, the findings of this study cannot be generalized universally to all forms of international media reporting.

Another important limitation involved the imbalance in reporting depth between regional and international news sources. International outlets frequently published longer articles containing extensive geopolitical explanations, diplomatic analysis and historical context, while Bangladeshi outlets often focused more on immediate developments and humanitarian consequences. This structural imbalance occasionally affected comparative interpretation because larger contextual datasets naturally provided richer material for NLP processing and LLM based summarization. As a result, some analytical outputs contained deeper narrative interpretation for international articles compared to shorter regional reports.

The study also faced limitations related to automated web scraping and dynamic website structures. News websites frequently modify HTML layouts, metadata organization and content delivery mechanisms. Which can interrupt scraping pipelines or create inconsistencies in data extraction. Although preprocessing and validation mechanisms were implemented to minimize these issues,

occasional missing metadata, duplicate content or incomplete article retrieval still occurred during the data collection process. Maintaining long term scraping stability therefore remains a continuous technical challenge for automated media analysis systems.

Another limitation emerged from the interpretive nature of media bias itself. Bias is not always directly measurable through objective computational indicators because interpretations of neutrality, framing and ideological positioning can differ according to audience perspective, political orientation, cultural context and prior beliefs. While the framework successfully identified measurable linguistic and contextual asymmetries, the system cannot definitively determine absolute ideological intent behind every article. Therefore, the generated outputs should be interpreted as comparative analytical observations rather than final judgments regarding journalistic objectivity.

Limitations were also observed in relation to Large Language Model based interpretation and summarization. Although the models produced coherent and contextually meaningful outputs, occasional overgeneralization, interpretive expansion and narrative emphasis were still present within generated summaries. LLMs may unintentionally reinforce dominant contextual patterns contained in the source material, particularly when certain narratives appear repeatedly across articles. This means that the models themselves cannot be treated as fully neutral analytical observers. Human evaluation and critical interpretation therefore remained necessary throughout the research process to validate the generated insights.

The multilingual dimension of the research also introduced certain constraints. While the system successfully handled comparative English language analysis across regional and international sources, it did not fully incorporate native multilingual processing for Bangla language articles at scale. Some Bangladeshi reporting structures and contextual nuances may therefore not have been captured completely through English centered analytical procedures. A more extensive multilingual architecture could potentially improve cultural and linguistic sensitivity within future implementations.

Another limitation involved computational resource constraints. Running continuous scraping pipelines, preprocessing large text corpora and generating LLM based interpretations required significant processing time and memory usage. Due to hardware and deployment limitations, certain optimization decisions had to be made regarding article storage, model selection and

frequency of automated processing. More advanced models or real time large scale monitoring systems may require substantially greater computational infrastructure than what was available during this research.

Finally, the study was conducted within a specific temporal period beginning from June 2025 and focused on active geopolitical conflicts during that timeframe. It is obvious that a different timeline would give different kinds of importance on different news. Political narratives evolves way too rapidly over time, especially during international crises. Consequently, the framing patterns and reporting behaviors observed within this thesis may shift in future geopolitical conditions or during different stages of conflict development. So these limitations remains within the study. The findings should therefore be understood within the temporal context in which the dataset was collected and analyzed.

## 6.2 Future Work

Although the developed framework successfully demonstrated the effectiveness of NLP and LLM based comparative media analysis, several opportunities remain for future improvement, expansion and deeper research exploration. The modular design of the system was intentionally developed to support extensibility, allowing future works to be integrated with additional analytical capabilities, larger datasets and more advanced AI models into the framework.

One important direction for future work involves expanding the diversity of news sources and geopolitical topics included within the corpus. This study focused on four news outlets and three major geopolitical conflicts to maintain analytical consistency and manageable system complexity. There can be a lot of future research that could incorporate a significantly larger collection of international, regional and independent media organizations from different political systems and cultural environments. Including sources from Europe, the Middle East, South Asia, East Asia, Africa and Latin America would allow broader comparative analysis and provide a more globally representative understanding of media framing behavior.

Another major area for future development is multilingual media analysis. While this research primarily analyzed English language content, future implementations could integrate direct Bangla language processing alongside other multilingual NLP pipelines. Incorporating transformer based multilingual models capable of understanding native linguistic structures, idioms and contextual

nuances would improve analytical accuracy and cultural sensitivity. This extension would also allow the framework to compare how the same geopolitical event is framed differently across multiple languages rather than only across English translated reporting structures.

Future research may also focus on improving bias detection methodologies through more advanced hybrid modeling approaches. The current framework relied heavily on contextual LLM interpretation combined with traditional NLP techniques such as sentiment analysis, keyword extraction and topic comparison. Future systems could integrate stance detection, narrative progression analysis, discourse relation modeling, entity relationship extraction and emotion classification models to produce deeper multi dimensional interpretations of political framing. Combining symbolic reasoning techniques with neural language models may further improve interpretability and analytical precision.

Another promising direction involves integrating real time monitoring and streaming analysis capabilities into the system. At present, the framework collects and processes articles periodically through scheduled automation. Future versions could implement continuous live data ingestion pipelines capable of monitoring geopolitical narratives in near real time. Such systems could provide immediate comparative analysis during breaking international events, elections, military escalations or diplomatic crises. This would significantly increase the practical value of the framework for journalism researchers, policymakers, communication analysts and media monitoring organizations.

Future work could also explore more sophisticated visualization and user interaction systems. While the Streamlit interface developed in this thesis successfully supported interactive exploration and comparative analysis, more advanced dashboards could improve usability and analytical depth further. Interactive network graphs, narrative evolution timelines, geopolitical heatmaps, comparative framing matrices and real time sentiment tracking systems could provide richer exploratory experiences for users analyzing complex media ecosystems.

Another valuable extension would involve incorporating credibility assessment and misinformation analysis modules. Media bias and misinformation are closely related but analytically distinct phenomena. Future systems could combine factual consistency verification, source reliability scoring, propaganda pattern identification and misinformation detection mechanisms alongside comparative framing analysis. This would create a more comprehensive

computational media intelligence framework capable of examining not only narrative differences but also factual reliability and information integrity.

Human evaluation mechanisms also represent an important area for future improvement. Overall even though people have different opinions , human judgement of a major number of people often considered as the best ones. While this research included interpretive validation through comparative review, future studies could incorporate structured expert evaluation, journalist feedback, intercoder reliability testing and audience perception analysis to assess the accuracy and usefulness of AI generated interpretations more systematically. Combining computational outputs with human centered validation procedures would improve analytical reliability and strengthen interdisciplinary applicability.

From a technical perspective, future research may investigate the use of more efficient and specialized Large Language Models optimized specifically for media discourse analysis. As open source LLM development continues to evolve rapidly, future systems may achieve higher contextual understanding, lower hallucination rates, faster processing speeds and improved bias sensitivity while requiring fewer computational resources. As we can use better models the better the results will become in the future. Fine tuning domain specific models on curated geopolitical news datasets may further improve analytical consistency and interpretive depth.

Future implementations may also benefit from integrating temporal narrative tracking systems capable of analyzing how media framing evolves over extended periods of time properly. A lot of things are possible such as instead of focusing only on comparative snapshots of reporting behavior, longitudinal analysis could reveal how narratives shift during different stages of conflict escalation, diplomatic negotiation, humanitarian crises or political transition. The options are unlimited. Such analysis would provide deeper understanding of dynamic media influence and evolving geopolitical discourse patterns.

Finally, future work could explore the broader ethical implications of AI assisted media analysis systems. That is what everyone hopes. As with time Large Language Models become increasingly integrated into journalism, political communication and public information ecosystems, questions surrounding algorithmic transparency, model bias and responsible AI deployment will become increasingly important. Future research should therefore continue investigating how AI systems

can support media analysis responsibly while minimizing risks related to misinformation amplification, interpretive distortion or automated ideological reinforcement in a better way.

Overall, the future potential of this research extends far beyond the scope of the current implementation. The framework developed in this thesis establishes a very strong foundation for scalable, multilingual and AI assisted comparative media analysis. With further expansion, optimization and interdisciplinary integration, such systems may become valuable tools for researchers, journalists, policymakers and communication analysts seeking deeper understanding of how global narratives are constructed and distributed within increasingly complex digital information environments.

## 7. References


1. Barendt, E. (2007). Freedom of speech (2nd ed.). Oxford University Press.
2. Bender, E. M., Gebru, T., McMillan Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability and Transparency. https://doi.org/10.1145/3442188.3445922
3. Berger, P. L., & Luckmann, T. (1966). The social construction of reality: A treatise in the sociology of knowledge. Anchor Books.
4. Bhaskar, R. (1978). A realist theory of science (2nd ed.). Harvester Press.
5. Bommasani, R., et al. (2021). On the Opportunities and Risks of Foundation Models. https://doi.org/10.48550/arXiv.2108.07258
6. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few shot learners. Advances in Neural Information Processing Systems.
7. Bryman, A. (2016). Social research methods (5th ed.). Oxford University Press.
8. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., et al. (2020). Unsupervised cross lingual representation learning at scale. Proceedings of ACL 2020. https://aclanthology.org/2020.acl-main.747/
9. Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.
10. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre training of deep bidirectional transformers for language understanding. https://doi.org/10.18653/v1/N19-1423
11. Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x
12. Entman, R. M. (2007). Framing bias: Media in the distribution of power. https://doi.org/10.1111/j.1460-2466.2006.00336.x
13. Galtung, J., & Ruge, M. (1965). The structure of foreign news. https://doi.org/10.1177/002234336500200104
14. Hallin, D. C., & Mancini, P. (2004). Comparing media systems: Three models of media and politics. Cambridge University Press. https://www.cambridge.org/core/books/comparing-media-systems
15. Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated identification of media bias in news articles: An interdisciplinary literature review.

16. Hamborg, F., Zhukova, A., & Gipp, B. (2019). News bias and labeling. https://doi.org/10.13140/RG.2.2.10120.06402
17. Horne, B. D., Khedr, S., & Adalı, S. (2018). Sampling the news producers. Proceedings of ICWSM 2018.
18. Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing.
19. Lai, V., et al. (2022). Human AI collaboration in explainable NLP.
20. Lazer, D., et al. (2009). Computational social science. https://doi.org/10.1126/science.1167742
21. Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. https://doi.org/10.18653/v1/D19-1387
22. Liu, Y., et al. (2023). A Survey of Large Language Models https://doi.org/10.48550/arXiv.2303.18223
23. McCombs, M. (2005). A look at agenda setting: Past, present and future. https://doi.org/10.1080/00909880500250438
24. McCombs, M. E., & Shaw, D. L. (1972). The agenda setting function of mass media.
25. Mitchell, R. (2018). Web Scraping with Python.
26. Napoli, P. M. (2019). Social media and the public interest: Media regulation in the disinformation age. Columbia University Press.
27. Neuendorf, K. A. (2017). The content analysis guidebook (2nd ed.). SAGE Publications
28. Perloff, R. M. (2017). The hostile media effect. Mass Communication and Society.
29. Provost, F., & Fawcett, T. (2013). Data science for business. O'Reilly Media.
30. Ramos, J. (2003). Using TF-IDF to determine word relevance
31. Rudin, C. (2019). Stop explaining black box models. https://doi.org/10.1038/s42256-019-0048-x
32. Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting and priming.
33. Spinde, T., Rudnitckaia, L., Hamborg, F., & Gipp, B. (2021). Detection of biased language in news articles using contextualized word embeddings. Information Processing & Management. https://www.sciencedirect.com/science/article/pii/S0306457321000157

# 8. Appendix:

### Source Code Repository

The complete source code, implementation files, and project documentation are available in the GitHub repository:

GitHub Repository: https://github.com/RakibHasan221b/Bias-Detector

### Web Application

The deployed Streamlit application developed for this research can be accessed through the following link:

Streamlit Application: https://bias-detector-j8flh2gexcqqdneh3dotz5.streamlit.app/