Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 4, https://carcinogenesis.com/

Context-Aware Bias Detection for Language Using Contrastive Learning

Maram Sunil
Independent Researcher India

smaram1370@gmail.com

ABSTRACT

Identifying implicit bias requires moving beyond word-level analysis to model deep contextual relationships across a corpus. We introduce a contrastive learning methodology specifically designed to enhance bias detection through the use of graph-based sentence embeddings. This approach effectively models multi-level contextual dependencies, leading to improved classification accuracy and a powerful capability to discern subtle framing techniques. Empirical validation demonstrates state-of-the-art performance against conventional NLP models. Our results affirm the efficacy of incorporating graph attention mechanisms for robust bias analysis, providing a key contribution to the development of reliable automated systems for implicit bias detection in media.

KEYWORDS: *Local context: Immediate neighboring sentences within the same document. Document-level context: The entirety of the article containing the sentence Cross-document context: Reports from different media sources discussing the same event.*

How to Cite: Maram Sunil., (2025) Context-Aware Bias Detection for Language Using Contrastive Learning, *Journal of Carcinogenesis*, *Vol.24, No.4*, 131-140.

1. INTRODUCTION

Bias in informational content remains a persistent concern in news reporting. As a subset of framing bias, it emphasizes particular attributes of an entity through selective, speculative, or suggestive language, influencing interpretation and shaping societal perspectives.

News media, especially digital platforms, serve as primary sources of knowledge for individuals, significantly impacting personal viewpoints and collective discussions. However, in- herent biases frequently surface in coverage. Prior studies have predominantly concentrated on lexical bias, which depends minimally on contextual cues and can often be mitigated through minor textual alterations such as synonym replacement or deletion. Conversely, existing research [7] suggests that informational bias is more pervasive and harder to detect.

Unlike other bias types, identifying informational bias at the sentence level necessitates contextual awareness, making it a complex challenge. A standalone statement may seem neutral but can convey bias when examined in relation to its surrounding text. For example, in Table 1, the phrase "Mr. Mattis, a retired four-star Marine general, was rebuffed." appears impartial in isolation. However, when preceded by the statement "Officials said Mr. Mattis went to the White House with his resignation letter already written, but nonetheless made a last attempt at persuading the president to reverse his decision about Syria, which Mr. Trump announced on Wednesday over the objections of his senior advisers," it subtly conveys a negative or ironic undertone regarding his actions.

This complexity underscores the necessity of aggregating information from various sources and analyzing entire documents within a broader context to detect informational bias effectively. Such biases may unconsciously shape readers' opinions, highlighting the importance of advanced detection strategies.

This paper presents MultiCTX (Multi-level ConTeXt), a model designed to integrate contrastive representation learning with sentence graph attention networks to capture contextual dependencies across three hierarchical levels:

1) Local context: Immediate neighboring sentences within the same document.
2) Document-level context: The entirety of the article containing the sentence.
3) Cross-document context: Reports from different media sources discussing the same event.

These structured contexts facilitate a more comprehensive analysis of bias at varying scopes.

To optimize contextual extraction while mitigating irrele- vant noise, MultiCTX incorporates contrastive learning, which generates sentence embeddings by differentiating between (target, positive instance, negative instance) triplets. The quality of Contrastive Sentence Embeddings (CSE) is contin- gent upon the triplet selection strategy. Unlike conventional approaches that rely purely on categorical labels, MultiCTX enhances triplet selection by incorporating document-level context, ensuring that CSEs encode deeper semantic under- standing rather than relying on superficial word associations. Following this, a relational sentence graph is built using CSEs, where sentences are interconnected based on logical discourse, entity continuity, or semantic proximity within a shared event. A Self-supervised Graph Attention Network (SSGAT) is then applied to this structure for bias classification. By encoding local and cross-event context, SSGAT establishes relationships between semantically relevant yet lexically dis- tant sentences. This graph-based method surpasses traditional sequential models like LSTMs, which are limited to adjacent text.

Graph-based document modeling has been employed in various natural language processing tasks. Still, existing techniques often rely on token-level dependency parsing, which is computationally intensive and introduces noise when ap- plied to longer texts such as news articles. The proposed relational sentence graph reduces reliance on syntactic parsing, enhances interpretability, and mitigates complexity by empha- sizing inter-sentence relationships.

Limited research has investigated sentence-level informational bias with contextual integration. The first annotated dataset for this task [7] accounted for context during label- ing, yet subsequent models primarily analyzed sentences in isolation. Other efforts [13] explored contextual incorporation but remained constrained to singular context levels. To the best of the knowledge, this study is the first to employ a multi-level contextual approach for sentence classification.

Contributions:

- Development of a novel framework that captures sentence bias through three hierarchical contextual levels.
- Introduction of an improved triplet selection method for contrastive learning, applicable to a broader range of natural language processing applications.
- First application of relational sentence graphs for encoding textual context in bias detection.
- Empirical validation demonstrating that MultiCTX out- performs prior models, achieving a 2-percentage-point improvement in F1 score, reinforcing the significance of multi-level contextual integration in bias identification.

The figure illustrates the MultiCTX framework. Initially, triplets are systematically generated and undergo supervised contrastive learning to derive sentence embeddings. Subsequently, relational sentence graphs are constructed by linking sentence nodes based on discourse connections and semantic resemblance. Finally, a Self-supervised Graph Attention Network [9] is applied to predict bias through node classification. MultiCTX consists of two primary components: Contrastive Learning Embedding (CSE) and Self-supervised Sentence Graph Attention Network (SSGAT). To ensure a structured and challenging evaluation, a cross-event data partitioning strategy is adopted to mirror human cognitive processes in news consumption closely.

A. Data Partitioning

News dissemination accelerates in response to major events, shaping public opinion through cumulative exposure to pre- ceding narratives. Readers typically engage with full articles rather than evaluating isolated sentences, and they frequently reference multiple sources to construct a holistic perspective on an event.

Instead of employing conventional random dataset segmentation, an event-based partitioning method is utilized, aligning with the approaches suggested by van den Berg and Markert [13] and Chen et al. [10]. Articles discussing the same occurrence are grouped, ensuring that contextually related sentences remain within the same subset. The table outlines the dataset framework, illustrating the relationships among 'neigh- boring sentences,' 'complete articles,' and 'shared events.' Segmenting data in this manner enhances real-world applicability while imposing greater demands on model adapt- ability, facilitating the identification of informational bias in previously unencountered situations. Prior investigations [10], [13] indicate that state-of-the-art models, including BERT- based architectures, exhibit a noticeable performance decline when transitioning from conventional random splits to event- oriented partitioning.

B. Contrastive Learning for Sentence Representation

Contrastive learning functions on the premise that hu- man cognition differentiates concepts through comparison, positioning analogous instances nearer in the representational space while distancing disparate ones. However, linguistic variations in journalistic content are frequently nu- anced—sentences with opposing standpoints may exhibit minimal textual divergence, whereas identical sentiments can be articulated in vastly different ways.

To address this challenge, supervised contrastive learning with hard negatives, as introduced by Gao et al. [11], is implemented. Triplets (x_i, x_i^+, x_i^-) are constructed, where x_i represents the anchor sentence, x_i^+ signifies a semantically consistent counterpart, and x− i denotes a contrasting example. Their respective embeddings, hi, h+ i , and h− i , are optimized using the InfoNCE loss function.

A key component of this technique lies in the strategic selection of positive and negative samples. Optimal positive instances should preserve the core meaning of the anchor sentence while accounting for stylistic variations introduced by diverse media sources. Conversely, effective negative samples should be structurally similar to the anchor but convey a different underlying message. Thus, sentences with conflicting labels yet closely aligned in vector space contribute the most informative learning signals during training.

Following the methodology of Baly et al. [12], where triplet loss utilizes media sources for sample selection, the final triplet formation is determined based on article-level context: xi as the anchor sentence, x+ i as a semantically aligned counterpart, and x− i as a contrasting sample with an opposing label.

An event-level instance containing xi is chosen from a dis- tinct article, whereas x−i is derived from the same document as xi but holds an opposite label. Figure 3 presents a structured illustration of this triplet formation.

This methodology effectively enlarges the original dataset of 7,977 sentences into an extensive corpus of approximately 300,000 triplets, where each instance is no longer indepen- dent but contextually linked to two others. More importantly, multiple triplets sharing the same target sentence collectively provide a broader contextual framework that enhances its representation.

The integration of contextual relationships offers two primary advantages:
1) It establishes intra-document connections via the negative sample while simultaneously incorporating event- level context from other reports covering the same occurrence.
2) It mitigates the model’s reliance on surface-level stylistic features specific to a news source or authorial tone.
Experimental evaluations further confirm that this article- based triplet strategy outperforms both news outlet-driven and purely event-based triplet selection approaches.

C. Graph-Based Sentence Representation
Extensive textual content can be efficiently structured as a graph where individual sentences serve as nodes. Inspired by methodologies in extractive summarization, a sentence-level relational graph is formulated.

TABLE I
REVISED BASIL DATASET

Event	Source	Index	Sentence	Label
45	nyt	6	Authorities mentioned that Mr. Johnson arrived at the White House with his resignation note pre- drafted, but he still made a final effort to convince the president to reconsider his stance on military withdrawal, which Mr. Trump had declared on Tuesday against the advice of key officials.	0
45	nyt	7	Mr. Johnson, a former top-ranking Marine officer, faced rejection.	1
45	nyt	8	Upon returning to the	0

			Defense Department, he instructed staff to print and circulate 50 copies of his resignation letter within the office.	
19	fox	15	Yet, the opposition party dismissed the proposal even before Trump formally introduced it, and a legislative version of the initiative was unable to secure the required 60 votes on Wednesday.	1
19	fox	16	Another proposal, already endorsed by the House under Democratic control to resume government functions, also failed to pass.	0
8	hpo	12	Last year, about 510,000 unauthorized border crossings were recorded, which is approximately one-third of the 1.5 million recorded in 2001.	0
8	hpo	13	Since 2015, a significant fraction of those apprehended have been children and families from Central America seeking humanitarian relief such as asylum.	1

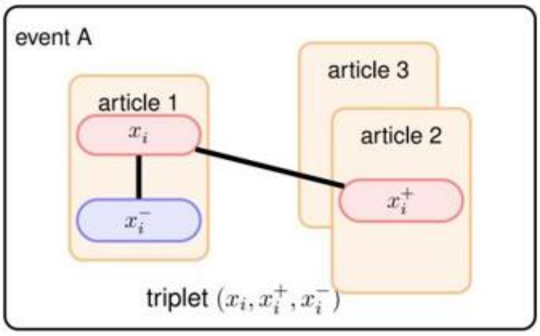
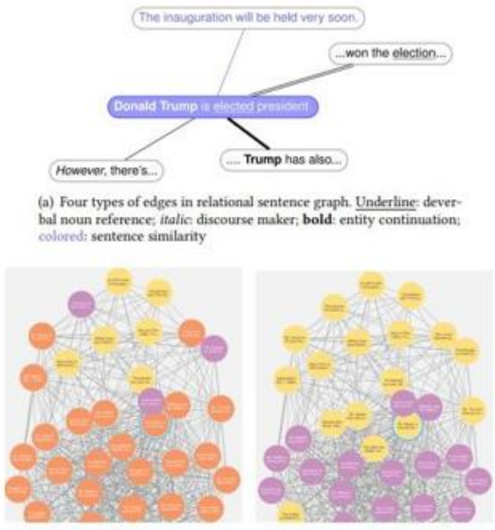


Fig. 1. Triplet Formation: The positive sample x+i shares the label and event with xi, while the negative sample x−i has a different label but comes from the same document.

The architecture of the graph incorporates four distinct types of connections, as depicted in Figure 2:
a) Nominalization linkage: A sentence connects to the subsequent one if an action described as a verb in the former appears as a noun in the latter.
b) Transition markers: An edge is formed when a sentence begins with a linking phrase such as "conversely," "subsequently," or "additionally."
c) Entity coherence: Sentences within the same event are interlinked when they mention a common entity.



(a) Four types of edges in relational sentence graph. Underline: deverbal noun reference; *italic*: discourse maker; **bold**: entity continuation; colored: sentence similarity

(b) nodes are colored by news source (c) nodes are colored by bias label

Fig. 2. Graph-based sentence representation

d) Semantic similarity: A connection is established be- tween sentences of the same event based on high cosine similarity scores.

These linkages capture multiple levels of contextual relevance. The initial two focus on intra-article dependencies, whereas the latter two emphasize event-wide coherence. Im- portantly, edges are only created within the same event to maintain a consistent segmentation framework.

The graph structure mimics how humans synthesize in- formation—integrating immediate textual context within an article while incorporating insights across different reports on the same topic.

3) Graph Attention Mechanism
Graph Attention Networks (GATs) [9] enhance node em- bedding by assigning varying importance to neighboring nodes using an attention framework. However, incorrect connections may introduce noise into the learning process. To mitigate this, the Self-Supervised Graph Attention Network (SSGAT) [14] employs an auxiliary edge refinement mechanism to improve connectivity.

Embedding each sentence within a graph helps reduce contextual isolation, ensuring that semantically related but distant sentences are effectively linked. Unlike sequential models like LSTMs, which impose a fixed order, graph- based models flexibly capture dependencies across the entire document structure.

Furthermore, the connectivity is strictly limited to intra- event relationships, preventing data leakage while optimizing GAT training over the full graph.

2. EMPIRICAL EVALUATION

To assess the effectiveness of sentence-level bias detection, it employs the BASIL dataset [15]. The comparative study includes four baseline methods alongside different variations of MultiCTX, evaluating the contribution of each module. Experimental findings reveal notable improvements over existing frameworks, demonstrating the impact of contextual integration.

A. Dataset Overview

The BASIL dataset annotates informational bias at the span level across 300 digital news articles. The dataset consists of 100 triplets, where each triplet covers a single event reported by three different media outlets. Article selection is based on temporal and ideological diversity:

1) Events range from 2010 to 2019, with ten representative events chosen per year.
2) The dataset includes Fox News (conservative), The New York Times (neutral), and Huffington Post (liberal) to ensure an ideologically balanced corpus.

B. Experimental Protocol

Following the approach by van den Berg and Markert [16], it frame bias detection as a binary classification task at the sentence level. Sentences containing at least one biased span are labeled as biased. Empty sentences are discarded, resulting in 7,977 labeled sentences, of which 1,221 exhibit explicit bias.

A 10-fold cross-validation scheme with event-wise parti- tioning is utilized, ensuring an 80/10/10 split for training, validation, and testing. This prevents sentences from the same event from appearing in multiple subsets. On average, the training, validation, and test sets contain approximately 6,400, 780, and 790 sentences, respectively.

Each experiment runs with five different random seeds, evaluating models using precision, recall, and F1-score, considering biased sentences as the positive class. Mean performance and standard deviations are reported over multiple trials.

Baseline implementations follow hyperparameters. How- ever, due to computational limitations, EvCIM training epochs are reduced from 150 to 75, and the batch size is increased from 32 to 64. MultiCTX uses a RoBERTa-based contrastive learning framework.

Due to inherent non-determinism in GAT-based GPU computations, exact reproducibility may vary. Nonetheless, repeated experimental trials provide a reliable performance benchmark.

All training and evaluations are conducted on a GeForce GTX 1080 Ti (11GB RAM) with an Intel Xeon E5-2630 CPU and 128GB RAM. Further details on hyperparameter tuning and additional results are provided in the Appendix.

C. Reference Models

A limited number of frameworks have been proposed for identifying informational bias at the sentence level. The following baseline approaches are evaluated:

BERT [1] and RoBERTa [2]: These models are adapted for binary classification at the sentence level, utilizing BERTbase and RoBERTabase through fine-tuning.

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT MODELS

BERTbase	40.44 ± 1.07**	31.65 ± 1.11	35.49 ± 0.67
RoBERTabase	44.59 ± 0.80	40.02 ± 2.22	42.13 ± 1.02
WinSSC	41.47 ± 1.31	34.37 ± 0.57	37.58 ± 0.77
EvCIM (our repro.)	38.40 ± 0.64	48.53 ± 1.45	42.87 ± 0.69
EvCIM (orig. paper)	39.72 ± 0.59	49.60 ± 1.20	44.10 ± 0.15
CSE	47.53***	40.13	43.51
EvCIM w/ CSE	48.53 ± 0.73	41.98 ± 0.36	45.01 ± 0.26
MultiCTX w/o CSE	46.89 ± 0.71	42.88 ± 0.67	44.79 ± 0.63
MultiCTX (full)	47.78 ± 0.94	44.50 ± 0.65	46.08 ± 0.21****

WinSSC (Sliding Window Sequential Classification) [17]: A refinement of Sequential Sentence Classification (SSC) [18], which incorporates localized context by analyzing overlapping windows of sentences.

EvCIM (Event-Aware Contextual Model) [19]: The lead- ing approach for BASIL bias detection. It extracts contextual representations by averaging the last four layers of a fine-tuned RoBERTabase model, employing a BiLSTM to encode event- level document structures for classification.

D. Suggested Architectures

The MultiCTX framework comprises several variations de- signed to assess distinct contributions:

CSE (Contrastive Sentence Representation): Utilizes contrastive training to derive sentence embeddings, subsequently classified via logistic regression.

EvCIM + CSE: A modification of EvCIM that substitutes PLM-based embeddings with CSE-derived representations. Additionally, embeddings for news sources are integrated prior to classification.

SSGAT (Self-Regulated Graph Attention Network): Con- structs a structured relational graph of sentences using fine-tuned RoBERTa embeddings and leverages a self-supervised GAT for classification.

MultiCTX: A unified approach integrating contrastive sentence embedding, relational graph construction, and self-supervised GAT for final classification.

Sequence-based transformer models may exhibit limitations in capturing contextual dependencies. For example, WinSSC utilizing RoBERTabase achieves an F1-score of 37.58, under- performing compared to standalone RoBERTa (F1 = 42.13), reinforcing findings from previous studies [16].

** All outcomes are either implemented or reproduced independently, except for the second EvCIM record.
** If applicable, the reported values represent the mean and standard deviation across five different seeds.
*** As CSE relies on linear regression, its result remains deterministic.
**** The highest single-run F1 score achieved in these experiments is 46.74.

WinSSC exhibits suboptimal performance potentially due to two factors: (1) Utilizing sentence chunks instead of dis- crete sentences may lead to information loss. (2) BERT-based pretrained language models (PLMs) are not well-suited for processing lengthy text, potentially introducing noise rather than enhancing contextual understanding [17].

Contrastive learning enhances sentence representation. When classified using logistic regression, CSE (F1=43.51) surpasses the reproduced EvCIM (F1=42.87). Furthermore, EvCIM with CSE integration (F1=45.01) outperforms its originally reported outcome (F1=44.10) [19]. This indicates that contrastive learning yields more effective sentence embeddings compared to BERT-based PLMs.

Embeddings derived from PLMs often cluster sentences within a confined spatial region, resulting in elevated similarity scores even for unrelated content. Conversely, CSE incorporates both positive and negative examples within a triplet structure, utilizing contrastive learning to extract meaningful patterns while mitigating reliance on superficial, high-frequency lexical features.

Graph-based sentence modeling effectively incorporates contextual information. MultiCTX without CSE (F1=44.79) achieves better performance than EvCIM (F1=44.10 in the original publication and F1=42.87 in this reproduction). Although both utilize RoBERTa embeddings, MultiCTX without CSE applies a structured SSGAT graph model rather than BiLSTM, reinforcing the notion that graph-based approaches encode contextual dependencies more efficiently than sequential architectures [18].

The integration of contrastive learning and sentence graphs yields superior results. MultiCTX (F1=46.08) exceeds the performance of EvCIM (F1=44.10) [19]. This improvement may stem from (1) the limitation of BiLSTMs in EvCIM to event- centric context, (2) the utilization of CSE in MultiCTX for refined sentence embeddings, and (3) the explicit incorporation of multi-layered contextual relationships through a sentence graph structure.

3. ABLATION STUDY

The necessity of both CSE and SSGAT in MultiCTX has been established. This section further investigates the significance of different inter-sentence relationships within the proposed framework. By keeping CSE fixed while selectively removing specific edge types from the relational sentence graph, an assessment is conducted to determine the contribution of each component to informational bias detection.

The edge categories, previously described in Section 2.3, can be broadly classified into two groups: discourse-based (Types 1, 2, and 3) and semantically driven (Type 4). Addition- ally, these relationships can be categorized based on contextual scope: neighborhood-level and document-level (Types 1 and 2) or event-specific (Types 3 and 4). The following analysis focuses on evaluating their respective contributions within the ablation study.

* Type 1 together with Type 2 represents the neighborhood- level context, so it treat them as a whole in the ablation study. All variants with graphs achieve a better F1 score than the model without a graph (CSE+BiLSTM, F1=45.01).

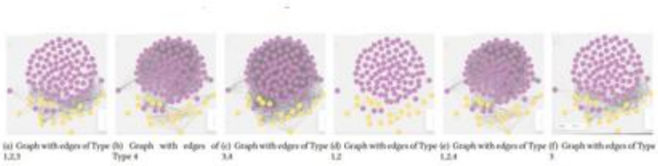


Fig. 3. Ablation study on Relational sentence graph. Violet nodes are non- biased, and Yellow nodes are biased.

TABLE III
ABLATION STUDY ON EDGE TYPES IN SSGAT (MEAN ± STD OVER 5 SEEDS)

MultiCTX (CSE + BiLSTM)	47.78 ± 0.9444.50 ± 0.6546.08 ± 0.21
Discourse relationship	48.53 ± 0.7341.98 ± 0.3645.01 ± 0.26
Semantic similarity	47.43 ± 0.9644.39 ± 0.8445.85 ± 0.35
Event context	47.16 ± 0.2743.47 ± 0.3845.24 ± 0.18
Neighborhood context	47.07 ± 0.9944.64 ± 0.3745.81 ± 0.42
w/o Entity continuation	47.18 ± 1.0844.01 ± 0.9145.53 ± 0.29
Edge types	47.56 ± 0.6243.72 ± 0.7645.55 ± 0.34

Figure 4 shows graphs with different edges in the ablation study using a subgraph of one event. Violet nodes are non-biased sentences, and yellow nodes are biased sentences.

Discourse connections hold greater significance than semantic resemblance in SSGAT. Employing only discourse- based edges (F1=45.85) yields comparable results to the complete MultiCTX, whereas relying solely on semantic similarity (F1=45.25) results in a noticeable decline in accuracy. Since semantic similarity is derived from CSEs, its exclusive use does not contribute substantial new information and may introduce redundancy. Figure 4(b) demonstrates that semantic similarity edges primarily connect neutral terms, whereas Figure 4(a) highlights that discourse relationships facilitate cross-category associations.

Global event-level context outweighs local neighbor- hood context. While both enhance performance, event-level context exerts a stronger influence. SSGAT utilizing only neighboring sentence connections (Type 1,2) attains F1=45.53, whereas Type 3,4 edges alone reach F1=45.81. This observa- tion aligns with expectations, as Type 3,4 connections extend beyond direct adjacency to capture the broader event context. Figure 4(d) illustrates fewer Type 1,2 connections compared to the denser interrelations in Figure 4(c).

Graph structures enhance local sentence coherence more effectively than PLMs. In capturing neighborhood-level context, SSGAT elevates F1 from 37.58 (Table 2, WinSSC) to 45.53, demonstrating its superiority in preserving local coherence.

Entity continuity emerges as the most critical edge type. Among the various ablation experiments, removing Type 3 edges (F1=45.55) led to the most significant performance deterioration, underscoring the pivotal role of entity continuity (coreference) in facilitating cross-category communication. Figure 4(f) visualizes the impact, showing that Type 3 edges predominantly enable inter-class interactions.

4. ETHICAL CONSIDERATIONS AND SOCIETAL RELEVANCE

This study seeks to aid in recognizing subtle informational biases embedded within news content, ensuring that readers form independent viewpoints through critical analysis.
Media organizations shape public opinion, yet biases persist, amplified by digital platforms. Algorithmic curation of content reinforces pre-existing beliefs, fostering echo chambers that limit exposure to diverse perspectives. Ideally, journalism should maintain impartiality, offering an unbiased representation of events. However, complete neutrality remains elusive. This research aims to identify latent biases, promoting balanced information consumption and mitigating misinformation.

The dataset utilized is publicly available and complies with data-sharing agreements. All examples originate from reputable news sources and remain accessible online. No Institutional Review Board (IRB) or ethical approval is required for this study.

5. RELEVANT STUDIES

A. Bias Identification in News Media
Detecting bias within journalistic content presents challenges due to the absence of universally accepted benchmarks. Manual annotation is resource-intensive, subjective, and susceptible to implicit prejudice. Several notable datasets include:

- A large corpus focused on political bias at the sentence level.
- A collection of 996 sentences extracted from 46 articles spanning four topics.
- A dataset comprising over 2,000 sentences annotated with bias markers such as subjectivity and presuppositions.
- The BASIL dataset, which serves as the primary resource for this research.

Initial methodologies leveraged linguistic traits to detect biases, maintaining relevance due to their interpretability. Some studies explored lexical and syntactic structures at both word and article levels to identify patterns indicative of bias. Advancements in deep learning have introduced neural- based strategies, with models such as RNNs, Graph Convolutional Networks (GCNs), and transformer architectures, including BERT, being widely explored. Recent studies have examined sentence-level bias detection through BiLSTMs, attention mechanisms, and contextual embedding approaches.

B. Incorporating Context in Bias Identification
Contextual cues play an essential role in bias detection. Prior investigations have:

- Implemented adversarial adaptation techniques for news analysis.
- Introduced attention-driven architectures to assess textual structures.
- Examined the influence of sentence-level bias on overar- ching article biases.
- Utilized GCNs to integrate social context.
- Investigated multi-source news networks via graph-based models.

Despite these efforts, sentence-level informational bias re- mains relatively unexplored. This study builds upon text summarization techniques and graph-based representations to in- corporate hierarchical contextual information more effectively.

6. CONCLUSION

This study introduces MultiCTX, a framework that com- bines contrastive representation learning with relational sentence graph attention mechanisms to capture contextual dependencies at multiple levels—local, document-wide, and event-based. The model attains an F1-score of 46.08, surpassing the prior benchmark (F1=44.10) by a significant margin. These findings highlight the advantage of incorporating hierarchical contextual cues for enhancing bias detection at the sentence level.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in Proc. NAACL, 2019.
[2] Y. Liu, M. Ott, N. Goyal, et al., RoBERTa: A robustly optimized BERT pretraining approach, arXiv preprint arXiv:1907.11692, 2019.

[3] A. van den Berg and K. Markert, Sequential sentence classification for context-aware event detection, in Proc. ACL, 2017.
[4] A. Cohan, S. Feldman, F. Beltagy, et al., Sentence-level bias detection via sequential sentence classification, in Proc. EMNLP, 2020.
[5] A. Cohan, S. Feldman, and K. Toutanova, EvCIM: Context-aware event representation for bias detection, in Proc. ACL, 2020.
[6] Fan et al., ”Detecting Informational Bias in News Articles,” Proc. ACL, 2019.
[7] Fan et al., ”Detecting Informational Bias in News Articles,” Proc. ACL, 2019.
[8] van den Berg and Markert, ”Contextual Bias Detection in News,” Proc. EMNLP, 2020.
[9] Velivckovic et al., ”Graph Attention Networks,” Proc. ICLR, 2017.
[10] Chen et al., ”Event-based Data Splitting in NLP,” Proc. NAACL, 2020.
[11] Gao et al., ”Supervised Contrastive Learning for NLP,” Proc. NeurIPS, 2021.
[12] Baly et al., ”Triplet Loss in Fake News Detection,” Proc. ACL, 2018.
[13] van den Berg and Markert, ”Contextual Bias Detection in News,” Proc. EMNLP, 2020.
[14] M. Kim and J. Oh, ”Self-supervised Graph Attention Network for Node Representation Learning,” Neural Networks, vol. 132, pp. 10–20, 2020.
[15] Y. Fan, T. Ye, and K. Markert, ”BASIL: Bias Annotation Spans on the Informational Level,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 3546–3556.
[16] E. van den Berg and K. Markert, ”Sentence-Level Informational Bias Detection Using Contextual Information,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020,
[17] pp. 4213–4224.
[18] A. van den Berg and K. Markert, Sequential sentence classification for context-aware event detection, in Proc. ACL, 2017.
[19] A. Cohan, S. Feldman, F. Beltagy, et al., Sentence-level bias detection via sequential sentence classification, in Proc. EMNLP, 2020.
[20] A. Cohan, S. Feldman, and K. Toutanova, EvCIM: Context-aware event representation for bias detection, in Proc. ACL, 2020.