# Comprehensive Project Outline: Hierarchical Dual-View Event Graphs for Weakly-Supervised Media Bias Detection

## 1. Project Overview and Theoretical Foundation

### 1.1 The Core Hypothesis and Motivation

The fundamental insight driving this project is that media bias rarely manifests through overtly false statements or explicit ideological language. Instead, bias emerges through the subtle interplay between factual reporting and interpretive framing. Consider how the same factual event - "Congress passed the bill with 218 votes" - can be wrapped in different interpretive layers: "Congress narrowly squeezed through the controversial bill" versus "Congress decisively approved the landmark legislation." The facts remain identical, but the framing creates entirely different reader impressions.

Your project recognizes that understanding bias requires decomposing news text into these two fundamental components: what objectively happened (facts) and how it's being characterized (interpretation). By building separate graph structures for each component and analyzing their interaction, you can reveal how neutral information becomes ideologically charged through narrative construction.

The hierarchical aspect addresses another key insight: bias operates at multiple scales. A single biased sentence might seem innocuous, but when connected to other sentences within a paragraph, and when those paragraphs build toward a particular conclusion, the cumulative effect creates a powerful ideological narrative. Your three-level hierarchy (sentence → paragraph → document) captures these multi-scale patterns.

### 1.2 Why Weak Supervision Matters

Traditional approaches to bias detection face a fundamental scalability problem. Sentence-level annotation requires human experts to read entire articles and mark each sentence as biased or neutral - a process that's expensive, time-consuming, and inherently subjective. Different annotators often disagree on subtle cases of bias, making it difficult to create large, consistently labeled datasets.

Your weak supervision approach flips this paradigm. Instead of requiring detailed sentence-level labels, you only need document-level bias labels (neutral/left/right), which are much easier to obtain - often, the publication source itself provides a strong signal of political leaning. The challenge becomes: how can we learn fine-grained, sentence-level bias patterns using only these coarse document labels?

The solution lies in hierarchical attention mechanisms that learn to identify which sentences and event patterns contribute most strongly to document-level bias. Through careful architectural design and training strategies, the model discovers bias-indicative patterns without being explicitly told which sentences contain bias.

# 2. Event Extraction and Dual-View Classification

## 2.1 Understanding Events in News Text

Events form the atomic units of news narratives. In computational terms, an event typically centers around a trigger word (usually a verb) that describes something happening, along with associated participants, locations, times, and other contextual information. For your project, you're adopting the event definition and extraction approach from Lei & Huang (2024), which uses models trained on MAVEN-ERE dataset.

The extraction process begins by identifying event triggers in each sentence. For a sentence like "The president firmly rejected the proposal during yesterday's heated cabinet meeting," the system would identify "rejected" as an event trigger. The next step extracts event arguments - "president" (agent), "proposal" (theme), "yesterday" (time), "cabinet meeting" (location), with modifiers like "firmly" and "heated" providing additional context.

However, your innovation comes in the next step: classifying each event as either factual or interpretive. This classification fundamentally shapes how the event will be processed in your dual-view architecture.

## 2.2 The Factual-Interpretive Dichotomy

The distinction between factual and interpretive events represents one of your key innovations. This isn't simply about identifying opinion words - it's about understanding the epistemological status of each reported event.

**Factual events** represent objectively verifiable occurrences. These include physical actions that could be observed ("met," "signed," "traveled"), official proceedings that are matters of public record ("voted," "passed," "approved"), quantifiable changes ("increased by 10%," "decreased to $5 billion"), and formal declarations ("announced," "published," "filed"). The defining characteristic of factual events is that their truth value can, in principle, be verified through evidence.

Consider the sentence: "The Senate voted 52-48 to confirm the nominee." The event "voted" is factual - the vote count is a matter of Congressional record. Even if a news outlet selectively reports certain votes while ignoring others (a form of bias), the reported event itself remains factual.

**Interpretive events** involve subjective characterization, evaluation, or speculation. These include opinion attributions ("critics claimed," "supporters argued"), modal constructions suggesting possibility rather than certainty ("could harm," "might benefit"), evaluative characterizations ("threatened democracy," "rescued the economy"), and predictions about future outcomes ("will likely lead to," "expected to cause").

The sentence "Experts warned the policy could destabilize markets" contains the interpretive event "warned" (implying negative evaluation) modified by "could destabilize" (speculative outcome). Even though "experts" did make statements (a fact), the characterization of their statements as "warnings" and the modal framing of potential outcomes moves this into interpretive territory.

### 2.3 Classification Methodology

The classification of events into factual versus interpretive categories combines multiple approaches for robustness:

**Lexical patterns** provide the first-pass classification. Certain verb classes strongly indicate factual events (motion verbs, transfer verbs, official action verbs) while others suggest interpretation (cognitive verbs, communication verbs with evaluative connotations, modal auxiliaries). However, context matters - "said" might introduce a factual quote or an interpretive claim.

**Syntactic analysis** examines the grammatical structure around events. Events with opinion holders as subjects ("analysts believe," "critics contend") typically signal interpretation. The presence of modal verbs, hedge words, or evaluative adverbs also provides strong signals. Dependency parsing reveals these patterns systematically.

**Semantic features** leverage pre-trained embeddings to capture subtle distinctions. Events with similar embeddings to prototypical factual events (extracted from training data) receive factual classification, while those clustering with interpretive prototypes receive interpretive labels. This captures patterns beyond simple lexical matching.

**Contextual classification** uses a fine-tuned BERT model that considers the full sentence context. This handles complex cases where the same event might be factual or interpretive depending on context. The classifier is trained on a small annotated set of events, then applied to the full dataset.

# 3. The Four Event Relations in Detail

### 3.1 Coreference Relations: Tracking Events Across Mentions

Coreference relations identify when multiple event mentions refer to the same real-world occurrence. This relation type is crucial for understanding how news narratives develop and potentially bias emerges through repeated characterization.

In straightforward cases, coreference involves the same event word repeated: "The committee rejected the proposal" followed by "The rejection sparked protests." Here, "rejected" and "rejection" are coreferent. However, coreference in news text often involves more complex patterns.

Consider an article about a political debate. The initial mention might be neutral: "The candidates met for their first debate." Later references might carry increasing interpretation: "the heated exchange," "the confrontational showdown," "the decisive moment that shaped the campaign." All these phrases refer to the same event, but each reframing adds interpretive layers. Your dual-view architecture captures this by maintaining coreference links while separating factual and interpretive characterizations.

Coreference chains reveal bias patterns. Left-leaning sources might consistently refer to a labor action as "the workers' protest" or "the strike for fair wages," while right-leaning sources might use "the illegal walkout" or "the union's disruption." The same event, linked through coreference, receives systematically different characterizations.

## 3.2 Temporal Relations: Constructing Narrative Chronology

Temporal relations establish the time-ordering between events, fundamental to news narrative structure. Your system classifies temporal relations into three categories: before, after, and overlap. This seemingly simple classification enables sophisticated narrative analysis.

The "before" relation often introduces historical context that frames current events. A sentence like "After repeatedly violating ethics guidelines, the official finally resigned" uses temporal ordering to imply a causal pattern - the resignation appears as the inevitable result of past violations. The selection of which historical events to mention before current events represents a key bias mechanism.

Temporal relations interact with the factual-interpretive distinction in revealing ways. Factual temporal chains ("The committee met on Monday, voted on Tuesday, and announced results on Wednesday") provide neutral chronology. Interpretive temporal chains ("After markets panicked, regulators scrambled to respond, then finally admitted their oversight failures") use temporal ordering to construct a narrative of incompetence.

The "overlap" relation captures simultaneous events, often used to imply correlation or hidden causation. "While the administration celebrated the policy's passage, unemployment rates soared" uses temporal overlap to suggest ironic contrast or governmental disconnect, even without explicit causal claims.

## 3.3 Causal Relations: The Heart of Narrative Bias

Causal relations - events causing or being caused by other events - represent perhaps the most powerful mechanism for introducing bias into news narratives. Different ideological perspectives often disagree not on what happened, but on why it happened and what consequences will follow.

Your system distinguishes between "causes" relations (Event A leads to Event B) and "caused by" relations (Event B resulted from Event A). This directional distinction matters for graph construction and allows the system to trace causal chains through the narrative.

Consider coverage of economic events. A neutral statement of correlation: "Interest rates rose and investment decreased" becomes ideologically charged through causal attribution: "The Fed's rate hike caused investment to plummet" (implying harmful policy) versus "Investment naturally cooled as rates normalized" (implying healthy correction). The same factual events receive opposite causal framings.

Causal relations in news text span a spectrum from explicit to implicit. Explicit markers include "caused," "led to," "resulted in," "triggered." However, many causal relations are implied through narrative structure, verb choice ("sparked," "prompted," "forced"), or simply through sequential presentation that invites causal inference.

The interaction between factual and interpretive events in causal relations reveals sophisticated bias patterns. A factual cause might lead to an interpretive effect: "The vote (factual) undermined democratic norms (interpretive)." Or an interpretive cause might explain a factual effect: "Growing radical influence (interpretive) led to the party's electoral defeat (factual)."

### 3.4 Subevent Relations: Hierarchical Event Structure

Subevent relations capture the hierarchical organization of events, where complex events contain simpler component events. This relation type helps understand how news narratives operate at different levels of granularity and how bias can emerge through selective elaboration or abstraction.

A high-level event like "Congress passed healthcare reform" contains numerous subevents: debates, committee hearings, amendments, votes, negotiations. Which subevents a news source chooses to elaborate reveals editorial priorities. Left-leaning sources might emphasize subevents like "lawmakers heard testimony from uninsured families," while right-leaning sources might focus on "business owners warned about rising costs."

The containment hierarchy also enables bias through abstraction level manipulation. The same legislative process might be described at a high level as "democratic deliberation" (positive framing) or "political theater" (negative framing), with different subevents selected to support each characterization.

Subevent relations interact with the factual-interpretive distinction to create complex patterns. A factual parent event might be elaborated through interpretive subevents: "The protest (factual) included passionate speeches about injustice (interpretive subevent)." This allows news sources to acknowledge facts while steering interpretation through selective subevent focus.

# 4. Hierarchical Graph Architecture in Detail

## 4.1 Sentence-Level Graphs: The Foundation

At the sentence level, your architecture constructs three interconnected graph structures for each sentence: a factual event graph, an interpretive event graph, and cross-view connections linking related events across the two views.

Consider a complex sentence from a political article: "After the controversial vote, critics immediately condemned the decision as a dangerous precedent while supporters praised it as overdue reform." This sentence yields rich graph structure:

The factual subgraph contains the core event "vote" with its temporal marker "after." This represents the objective occurrence stripped of evaluation. The graph might be sparse for sentences focused on interpretation, or dense for fact-heavy reporting.

The interpretive subgraph captures the evaluative events: "condemned," "praised," along with their characterizations - "controversial" (modifying the vote), "dangerous precedent," and "overdue reform." The graph structure shows "critics condemned" and "supporters praised" as parallel interpretive responses to the factual vote.

Cross-view edges link the factual "vote" to its various interpretive characterizations. These edges represent the crucial fact-to-interpretation mappings that reveal how neutral events become ideologically charged. The same factual node might have edges to contradicting interpretive nodes, representing the contested nature of the event's meaning.

The sentence-level graph construction process must handle several challenges. Event boundaries aren't always clear - is "voted to approve" one event or two? Implicit events must be inferred - "The senator from Texas" implies a "representing" event. Metaphorical language ("killed the bill") must be mapped to actual events.

## 4.2 Paragraph-Level Aggregation: Capturing Local Narrative Flow

Paragraphs in news writing aren't arbitrary text divisions - they represent semantic units that develop specific aspects of the story. Your paragraph-level processing captures how sentences within a paragraph work together to build meaning and potentially introduce bias.

The paragraph-level graph connects sentence graphs through cross-sentence event relations. A typical news paragraph might show patterns like:

- **Topic sentence establishing frame**: Often contains interpretive events that set the paragraph's angle

- **Supporting sentences with evidence**: Mix of factual and interpretive events
- **Concluding sentence reinforcing frame**: Returns to interpretation, often with causal claims

Consider a paragraph about economic policy:

"Economists warn the new regulations could stifle innovation (interpretive topic sentence). Last quarter, startup funding dropped 30% in regulated sectors (factual evidence). Industry leaders report increasing difficulty securing investment (mix of fact and interpretation). Without reform, experts predict a continued decline in entrepreneurship (interpretive conclusion with causal claim)."

The paragraph-level graph would show temporal relations (last quarter → current reports → future predictions), causal relations (regulations → funding drop → investment difficulty → declined entrepreneurship), and coreference (regulations → regulated sectors → reform).

The attention mechanism at this level learns which sentences drive paragraph meaning. In biased text, topic and concluding sentences often receive high attention weights as they contain the interpretive frame, while evidence sentences might receive lower weights despite containing the factual support.

Paragraph-level processing must also handle discourse coherence. Sentences within a paragraph typically maintain topical consistency, with events semantically related even without explicit relations. The model learns these implicit connections through embedding similarity and contextual encoding.

## 4.3 Document-Level Structure: The Global Narrative Arc

News articles follow recognizable narrative structures that your document-level processing captures. The document graph connects paragraphs to model long-range dependencies and global narrative patterns.

Common document structures in news writing include:

**Inverted pyramid structure**: Most important information (often interpretive claims) appears first, followed by supporting details and context. Document-level attention should weight early paragraphs heavily in this structure.

**Chronological narrative**: Events presented in temporal sequence, building toward a climax or conclusion. Temporal relations dominate the document graph, with causal interpretations often concentrated in concluding paragraphs.

**Claim-evidence structure**: Opening paragraphs present interpretive claims, middle paragraphs provide factual support, conclusions reinforce interpretation. Cross-paragraph edges link claims to their evidence.

The document-level graph maintains sparse connectivity to avoid computational explosion. Key connections include:

- **Introduction-conclusion links**: The opening frame often connects directly to the final interpretation
- **Recurring entity/event mentions**: Major figures or events appearing throughout create document-wide coreference chains
- **Thematic threads**: Related events across paragraphs that develop the same narrative line

Document-level bias patterns emerge through several mechanisms:

**Interpretation escalation**: Initial paragraphs might present relatively neutral facts, but interpretation intensifies through the article, with concluding paragraphs containing the strongest ideological framing.

**Selective evidence chains**: The document graph reveals which facts are connected to support particular interpretations, and equally importantly, which potential connections are absent.

**Frame consistency**: Biased articles maintain consistent interpretive frames across paragraphs, while more neutral articles might present competing interpretations.

# 5. Neural Architecture Deep Dive

## 5.1 Graph Neural Network Design Principles

Your architecture must process heterogeneous graphs with multiple node types (factual events, interpretive events, sentence nodes, paragraph nodes) and edge types (the four event relations plus hierarchical connections). This requires careful design of the message-passing mechanisms.

The base architecture uses Graph Attention Networks (GAT) for their ability to learn which neighbors matter most for each node. However, standard GAT must be extended for your use case:

**Relation-aware attention**: Different edge types require different attention mechanisms. Temporal relations might propagate temporal ordering information, while causal relations propagate agency and effect information. The attention function incorporates edge type embeddings.

**Dual-view message passing**: Factual and interpretive nodes process information differently. Factual nodes might focus on building coherent event sequences, while interpretive nodes track opinion holders and evaluation targets.

**Hierarchical aggregation**: Information flows both bottom-up (events to sentences to paragraphs) and top-down (document context influences paragraph interpretation). Bidirectional processing ensures all levels benefit from both local and global context.

## 5.2 Cross-View Attention Mechanisms

The interaction between factual and interpretive views represents a critical innovation in your architecture. Cross-view attention mechanisms learn how facts and interpretations relate within and across levels.

At the sentence level, cross-view attention identifies which facts receive interpretation and which interpretations lack factual grounding. High attention weights between a fact and its interpretation indicate explicit framing, while interpretations without factual connections might signal unsupported opinion.

The attention mechanism uses query-key-value transformations where factual events query interpretive events and vice versa. The learned attention patterns reveal bias mechanisms:

- Facts receiving multiple conflicting interpretations indicate contested events
- Interpretations clustering around few facts suggest cherry-picking
- Asymmetric attention (many interpretations for some facts, none for others) reveals selective focus

Cross-view attention extends hierarchically. Paragraph-level cross-view attention shows how factual and interpretive paragraph summaries relate. Document-level cross-view attention captures whether the overall narrative maintains factual-interpretive balance or skews heavily interpretive.

## 5.3 Training Dynamics and Loss Design

Training with only document-level supervision requires careful loss design to propagate learning signals through the hierarchy. Your multi-component loss function balances several objectives:

**Document classification loss** provides the primary supervision signal. The model must correctly predict document-level bias labels (neutral/left/right) from the hierarchical graph representation. This loss backpropagates through all components, encouraging them to preserve bias-relevant information.

**Hierarchical consistency loss** ensures that predictions at different levels align. If attention weights suggest certain sentences are highly bias-indicative, their containing paragraph should also show bias signal. This loss prevents the model from making contradictory predictions across levels.

**Attention sparsity loss** encourages focused attention rather than uniform weights. In practice, only some sentences in a biased article contain strong bias signals. Sparsity pressure helps the model identify these key sentences without sentence-level supervision.

**View balance loss** prevents the model from ignoring one view. Early in training, the model might focus entirely on interpretive events as they're more obviously bias-related. This loss ensures factual events remain represented, as fact-interpretation interaction patterns prove crucial for bias detection.

**Auxiliary tasks** provide additional supervision. Predicting the news source (which correlates with but doesn't determine bias) helps the model learn stylistic patterns. Topic classification ensures the model understands article content beyond bias signals.

Training proceeds in stages to manage complexity:

1. **Pre-training components**: Event extractors and relation classifiers train on MAVEN-ERE
2. **View classification**: The factual-interpretive classifier trains on a small annotated set
3. **Full model training**: All components train jointly with document supervision
4. **Fine-tuning**: Careful hyperparameter tuning to balance loss components

# 6. Weak Supervision Strategies

## 6.1 From Documents to Sentences: The Attention Bridge

The core challenge of weak supervision is learning sentence-level patterns from document-level labels. Your hierarchical attention mechanisms provide the key bridge.

During training, the model must decide which sentences contribute to document-level bias predictions. Sentences receiving high attention weights effectively receive pseudo-labels as bias-indicative. Over many documents, the model learns patterns distinguishing biased from neutral sentences.

However, this process is more sophisticated than simple attention-based pseudo-labeling. The model learns that bias often emerges from interactions between sentences rather than individual sentences in isolation. A factual sentence followed by an interpretive sentence might together create bias, even if neither alone would.

The hierarchical structure provides multiple supervision paths. A sentence might receive bias signal through:

- Direct attention in sentence-to-document aggregation
- Indirect signal through its paragraph's importance
- Event-level patterns that consistently correlate with document bias

This redundancy makes learning robust to attention noise early in training.

### 6.2 Handling Label Noise and Imbalance

Document-level bias labels contain inherent noise. Articles from generally left-leaning sources might occasionally present neutral or even right-leaning perspectives on specific topics. Your architecture must handle this label noise gracefully.

The dual-view decomposition helps by separating content from framing. An article might report facts typically associated with one ideological perspective but frame them to support the opposite perspective. By modeling both views, the system can detect these complex patterns rather than learning simple fact-to-bias associations.

Class imbalance presents another challenge. Neutral articles might be underrepresented compared to ideologically aligned articles in training data. Techniques like weighted sampling, focal loss, and synthetic neutral example generation help balance training.

The hierarchical structure also provides natural data augmentation. Each article contributes multiple paragraph-level and many sentence-level training examples, amplifying the effective dataset size.

### 6.3 Pseudo-Label Refinement

As training progresses, the model's sentence-level attention weights become increasingly reliable bias indicators. These can generate pseudo-labels for semi-supervised learning:

1. **High-confidence sentence selection**: Sentences with consistent high attention across multiple training runs likely contain bias
2. **Pseudo-label generation**: These sentences receive soft bias labels based on attention weights
3. **Semi-supervised refinement**: Retrain with both document labels and sentence pseudo-labels
4. **Iterative improvement**: Repeat as pseudo-labels improve

This process must avoid confirmation bias where initial errors become amplified. Techniques like pseudo-label confidence thresholding, consistency checking across model variants, and human validation of highly weighted sentences help maintain quality.

## 7. Causal Intervention Framework

### 7.1 Moving Beyond Correlation

Traditional bias detection identifies features correlated with bias - certain words, topics, or sources. However, correlation doesn't imply causation. Your causal intervention framework tests which graph structures actually cause bias perception versus merely co-occurring with it.

The key insight is that if a graph structure causally contributes to bias, removing or modifying it should change bias predictions. Structures that correlate with but don't cause bias won't affect predictions when removed.

Interventions operate at multiple levels:

**Node interventions** remove specific events to test their causal role. Removing all interpretive events tests whether bias depends on explicit interpretation or can emerge from selective fact presentation alone. Removing specific event types (e.g., all causal claims) reveals which event categories drive bias.

**Edge interventions** disrupt relation structures. Cutting cross-view edges tests whether bias requires explicit fact-interpretation links or can emerge from parallel presentation. Removing temporal edges disrupts narrative flow, testing whether bias depends on story structure.

**Subgraph interventions** remove entire narrative components. Deleting high-attention paragraphs tests whether bias concentrates in specific sections or distributes throughout. Removing introduction or conclusion tests where bias signals concentrate.

## 7.2 Counterfactual Analysis

Beyond deletion, counterfactual interventions modify rather than remove structures. These test more nuanced hypotheses about bias mechanisms:

**Event substitution**: Replace interpretive events with neutral alternatives. If "slammed" becomes "responded to," does bias prediction change? This reveals which specific word choices drive bias beyond semantic content.

**Relation reversal**: Flip causal or temporal relations. If "A caused B" becomes "B caused A," how do predictions change? This tests whether bias depends on specific causal narratives or just causal density.

**Cross-document transplantation**: Take factual events from one article and interpretive framing from another. This tests whether bias emerges from facts, interpretation, or their specific combination.

**Attention redistribution**: Force uniform attention weights to test whether the model truly relies on learned importance patterns or could achieve similar performance with random weighting.

## 7.3 Minimal Sufficient Subgraphs

A key goal is identifying minimal sufficient subgraphs - the smallest graph structures that preserve bias predictions. This reveals the essential components of biased narratives.

The search process iteratively removes nodes and edges while monitoring prediction changes. When prediction confidence drops below a threshold, the remaining structure represents a minimal bias-preserving subgraph.

These minimal subgraphs often reveal surprising patterns:

- Sometimes a single interpretive sentence suffices if it frames all preceding facts
- Other times, bias requires complex interactions between multiple facts and interpretations
- Certain relation types (especially causal) prove more essential than others

Analyzing minimal subgraphs across many examples reveals general bias mechanisms that transcend specific topics or sources.

# 8. Evaluation Framework

## 8.1 Multi-Level Evaluation Metrics

Your evaluation must assess performance at multiple levels despite training only with document supervision:

**Document-level metrics** measure primary task performance:

- Accuracy on 3-way classification (neutral/left/right)
- Macro-averaged precision/recall/F1 across classes
- Confusion matrices revealing which bias directions are most confusable

**Sentence-level metrics** evaluate discovered patterns against gold annotations from BASIL/BiasedSents:

- Precision: Of high-attention sentences, how many are truly biased?
- Recall: Of annotated biased sentences, how many receive high attention?
- Ranking metrics: How well does attention weight rank sentences by bias?

**Event-level analysis** examines learned event patterns:

- Which event types most strongly predict bias?
- How do factual-interpretive ratios correlate with bias?
- Which event relations prove most bias-indicative?

**Interpretability metrics** assess explanation quality:

- Agreement between highlighted sentences and human bias perception

- Coherence of extracted narrative patterns
- Consistency of attention across similar examples

### 8.2 Generalization and Transfer Learning

Key evaluation questions test whether learned patterns generalize:

**Cross-dataset transfer**: Train on BASIL, test on BiasedSents (and vice versa). Strong transfer indicates robust bias patterns rather than dataset-specific artifacts.

**Cross-topic transfer**: Hold out entire topics during training. Can the model detect bias in unseen topics using learned structural patterns?

**Cross-source transfer**: Test on articles from news sources not seen during training. This reveals whether the model learns general bias mechanisms or source-specific styles.

**Zero-shot sentence classification**: Without any sentence-level training, how well can attention weights identify biased sentences? This directly tests the weak supervision hypothesis.

### 8.3 Behavioral Analysis

Beyond quantitative metrics, behavioral analysis reveals what the model actually learns:

**Attention visualization**: For key examples, visualize attention weights across the hierarchy. Do highlighted portions align with human intuitions about where bias appears?

**Subgraph pattern mining**: Extract frequently occurring subgraph patterns from biased documents. Do these represent recognizable bias mechanisms?

**Error analysis**: For misclassified documents, trace which components failed. Do errors concentrate in fact extraction, interpretation classification, or relation identification?

**Ablation studies**: Systematically remove components to understand their contributions:

- Factual-only model: Can bias be detected from facts alone?
- Interpretive-only model: Do interpretations suffice without factual context?
- Flat model: How much does hierarchy contribute beyond sentence independence?
- Single-view relations: Which event relations prove most crucial?

# 9. Implementation Challenges and Solutions

### 9.1 Computational Complexity

Processing hierarchical graphs for long documents poses computational challenges:

**Memory requirements**: A 50-sentence article with 5 events per sentence yields 250 event nodes, potentially thousands of edges, across multiple hierarchy levels and views. GPU memory limits require careful batching.

**Solution strategies**:

- Dynamic graph sampling: For very long documents, sample representative subgraphs
- Gradient checkpointing: Trade computation for memory in deep architectures
- Sparse attention patterns: Use local attention windows with global tokens
- Mini-batch construction: Batch similar-length documents to minimize padding

**Inference optimization**: For deployment, precompute event extraction and cache frequently occurring patterns. Use model distillation to create lighter inference models while preserving key capabilities.

### 9.2 Event Extraction Quality

Your architecture depends on quality event extraction, but automatic extraction contains errors:

**Missing events**: Extractors might miss implicit or metaphorical events **Wrong boundaries**: Event spans might be incorrect, affecting argument extraction **Relation errors**: Especially challenging for implicit relations lacking clear markers

**Mitigation strategies**:

- Ensemble multiple event extractors to improve coverage
- Include confidence scores in graph construction - uncertain events receive lower weight
- Design architectures robust to extraction noise through redundancy
- Fine-tune extractors on news-specific data to improve domain performance

### 9.3 Factual-Interpretive Boundary Cases

The factual-interpretive distinction, while conceptually clear, contains boundary cases:

**Reported speech**: "The senator said taxes will rise" - is this factual (the saying occurred) or interpretive (the content is prediction)? **Institutional actions**: "The court ruled unconstitutional" - factual procedure or interpretive judgment? **Measured quantities**: "Unemployment soared to 8%" - factual number or interpretive characterization?

**Approach**: Rather than forcing binary classification, use soft assignment with probability scores. Let the model learn which boundary cases pattern with factual versus interpretive events for bias detection. This data-driven approach avoids imposing rigid categories that might not align with bias mechanisms.

# 10. Broader Implications and Future Directions

## 10.1 Understanding Media Bias Mechanisms

Your project contributes fundamental insights into how media bias operates:

**Bias through structure, not just content**: The same facts arranged differently create different impressions. Your hierarchical graphs capture these structural patterns.

**Fact-interpretation interaction**: Bias rarely appears through pure fabrication but through selective interpretation of real events. The dual-view architecture models this explicitly.

**Multi-scale operation**: Sentence-level bias accumulates into paragraph and document-level effects. Understanding requires modeling all scales simultaneously.

**Implicit mechanisms**: Many bias patterns operate through what's not said - missing causal links, absent temporal context, ignored events. Graph structures can represent these absences.

## 10.2 Practical Applications

Beyond research contributions, your system enables practical applications:

**Automated media monitoring**: News organizations could analyze their own coverage for unintended bias patterns **Reader tools**: Browser extensions could highlight bias-indicative passages, showing which parts are factual versus interpretive **Comparative analysis**: Side-by-side comparison of how different sources cover the same events **Journalist assistance**: Real-time feedback during writing to identify potentially biased framing

## 10.3 Future Research Directions

Your architecture opens several research avenues:

**Temporal evolution**: How do bias patterns change over time? Adding temporal modeling could track shifting media narratives.

**Cross-lingual analysis**: Do bias mechanisms transfer across languages and cultures? Multilingual event graphs could reveal universal versus culture-specific patterns.

**Interactive explanations**: Rather than static attention weights, develop interactive tools letting users explore why the model makes specific predictions.

**Adversarial robustness**: How easily can bad actors manipulate the system? Adversarial training could improve robustness.

**Bias mitigation**: Can the system not just detect but suggest neutral reframings? Generative models could propose less biased alternatives while preserving factual content.