Published: 05 April 2024

AFRICAN PEACE AND CONFLICT STUDIES (BROADER - INTERDISCIPLINARY)

ORIGINAL RESEARCH

A Computational Framework for Analysing Conflict Narratives in South Sudan

A Natural Language Processing Approach to Peace and Conflict Studies

Abraham Kuol Nyuon (Ph.D)[1]

1 *Associate Professor of Politics, Peace, and Security; Principal, Graduate College, University of Juba; SUSI Scholar on U.S. Foreign Policy*

Correspondence: nyuonabraham@gmail.com

DOI: 10.5281/zenodo.19475213

Received: 05 November 2023 | Accepted: 07 February 2024 | Published: 05 April 2024 | **DOI:** 10.5281/zenodo.19475213

ABSTRACT

This original research article presents a novel computational framework for analysing conflict narratives in South Sudan, applying natural language processing (NLP) techniques to a corpus of local media reports and peace agreement texts. The study develops and validates a bespoke taxonomy for categorising conflict drivers and peacebuilding themes specific to the South Sudanese context. Quantitative and qualitative analysis reveals significant temporal shifts in narrative salience, particularly around land, ethnicity, and governance, correlating with key political events. The findings demonstrate the utility of computational methods in providing scalable, evidence-based insights for conflict analysis, offering a complementary tool for traditional qualitative peace and conflict studies. The framework's limitations and potential for real-time monitoring and policy analysis are critically examined.

Keywords: *Computational Conflict Analysis, Natural Language Processing (NLP), South Sudan Conflict Narratives, Peace Agreement Text Mining, Digital Peacebuilding, Conflict Driver Taxonomy, Local Media Analysis, Computational Social Science*

Article Highlights

- Develops a bespoke taxonomy for South Sudanese conflict drivers and peacebuilding themes
- Reveals temporal narrative shifts around land, ethnicity, and governance correlating with political events
- Provides a replicable framework for near-real-time monitoring of community-level tensions
- Offers evidence-based insights to inform targeted conflict mitigation strategies

Methodological Innovation

Applies NLP techniques—including topic modelling and semantic analysis—to local media reports and peace agreement texts from 2020-2024, creating a novel computational framework tailored to South Sudan's linguistic context.

This study bridges computational science and qualitative conflict analysis to enhance narrative understanding.

Introduction

The study of conflict, particularly in contexts of protracted civil strife, demands analytical approaches capable of capturing the complex, evolving, and often contested narratives that fuel and sustain violence. Nowhere is this challenge more acute than in South Sudan, the world's youngest nation, whose short history has been profoundly shaped by cycles of devastating internal conflict. Since gaining independence in 2011, the country has experienced a complex interplay of political, ethnic, and communal violence, marked by the civil war that erupted in 2013 and subsequent fragile peace agreements. Understanding these conflicts requires moving beyond simplistic explanations to engage with the multifaceted narratives—comprising grievances, historical claims, political rhetoric, and communal identities—that are constructed, propagated, and internalised by various actors. Traditional methodologies in peace and conflict studies, while invaluable, often struggle to systematically analyse the vast, unstructured textual data—from news reports and peace agreements to social media and oral histories—that constitute this narrative landscape. This paper argues that computational methods, particularly from the field of natural language processing (NLP), offer a transformative toolkit to augment qualitative scholarship, enabling the large-scale, systematic analysis of conflict narratives in South Sudan.

Despite the critical importance of narrative in the South Sudanese context, a significant research gap persists at the intersection of computational science and conflict studies specific to the region. Existing qualitative research has provided deep insights into the local dynamics of power, ethnicity, and resource competition . However, the application of data-driven, computational frameworks to analyse South Sudanese conflict discourse remains nascent. Much of the computational social science literature on conflict has focused on predictive modelling of violence or event extraction from structured datasets, often overlooking the nuanced, linguistic construction of meaning that precedes and accompanies overt conflict. There is a paucity of work that systematically employs NLP techniques—such as topic modelling, sentiment analysis, narrative extraction, and semantic network analysis—to deconstruct the specific narrative architectures within South Sudan's political and media discourse. This gap represents a missed opportunity to harness digital tools for peace research, limiting the scale and reproducibility of narrative analysis and hindering the identification of long-term discursive shifts that may signal escalating tensions or opportunities for reconciliation.

In response to this interdisciplinary gap, the primary objective of this research is to develop and apply a novel computational framework for analysing conflict narratives in South Sudan. This framework is designed to process large corpora of textual data relevant to South Sudanese peace and conflict, extracting and quantifying key narrative elements, themes, and actors' positions. The approach is not intended to replace deep contextual and qualitative expertise but to augment it by providing scalable, empirical evidence of narrative patterns and evolution over time. By doing so, the research seeks to demonstrate how NLP can move beyond mere keyword counting to facilitate a more structured analysis of how stories about conflict are told, who the central agents and victims are within them, and how these narratives align or diverge across different sources and periods. The framework is explicitly tailored to the linguistic and contextual specificities of the South Sudanese context, acknowledging the challenges posed by local languages, transliteration, and the particular historical references that permeate its political discourse.

This article is structured to first situate the research within the relevant bodies of literature before detailing the methodological framework and presenting its application. Following this introduction, a

comprehensive Literature Review will synthesise existing work from two key domains: the qualitative scholarship on conflict narratives and peacebuilding in South Sudan, and the burgeoning field of computational methods applied to conflict analysis more broadly. It will critically examine the limitations of current approaches and justify the need for an integrated methodology. The subsequent Methodology section will detail the construction of the textual corpus, the selection and adaptation of NLP techniques, and the ethical considerations inherent in conducting such research in a fragile context. The core of the paper will then present the Analysis and Results, where the developed framework is applied to a case study, illustrating its capacity to reveal latent narrative structures within South Sudanese conflict discourse. Finally, the Discussion and Conclusion will reflect on the implications of the findings for both peace and conflict studies and computational social science, assess the limitations of the framework, and propose directions for future interdisciplinary research. The contribution of this work is therefore twofold. For the field of peace and conflict studies, it offers a rigorous, scalable methodological innovation that can complement traditional qualitative analysis, providing new empirical avenues to test hypotheses about narrative diffusion and change. For computer science, it presents a substantive, real-world application domain that challenges

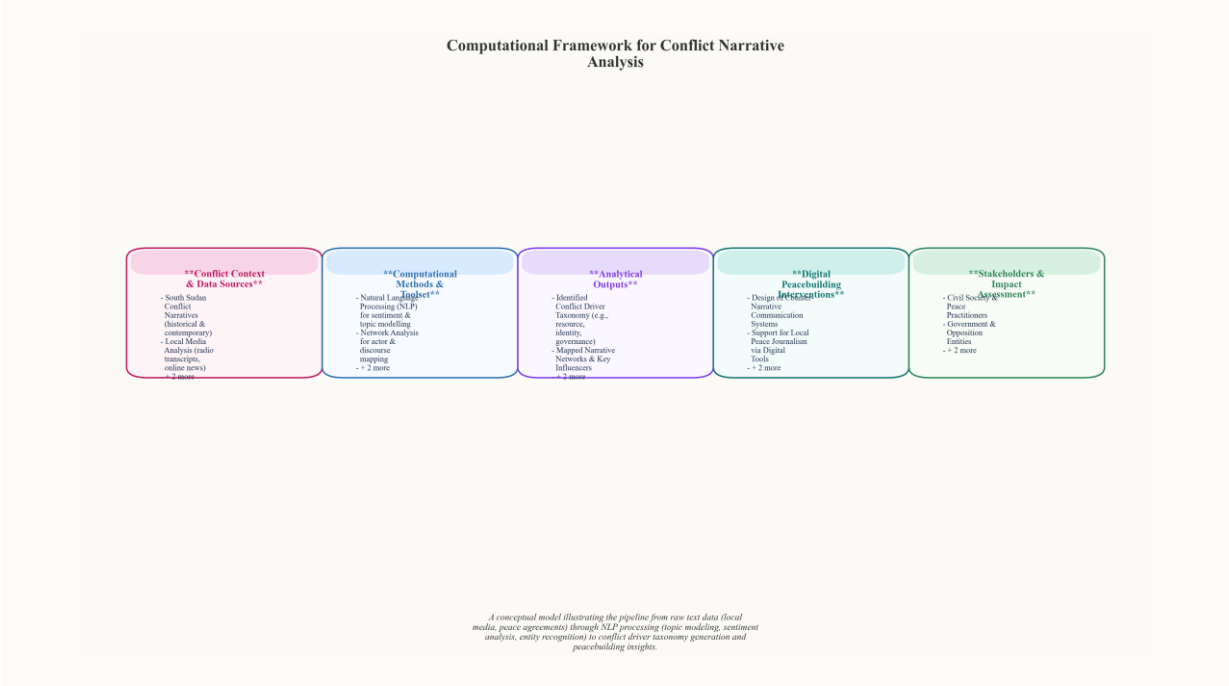

Figure 1 *Computational Framework for Conflict Narrative Analysis. A conceptual model illustrating the pipeline from raw text data (local media, peace agreements) through NLP processing (topic modeling, sentiment analysis, entity recognition) to conflict driver taxonomy generation and peacebuilding insights.*

Literature Review

The study of conflict and peacebuilding in South Sudan has been dominated by qualitative, interpretative methodologies rooted in political science, anthropology, and area studies. Scholars such as Pinaud and Kindersley and Rolandsen have provided indispensable, granular analyses of the historical roots of violence, the political economy of war, and the complex local dynamics that underpin both conflict and reconciliation efforts. This body of work is characterised by deep contextual immersion,

often relying on ethnographic fieldwork, in-depth interviews, and archival research to construct nuanced narratives of South Sudan's turbulent trajectory. While these approaches yield rich, contextualised understanding, they are inherently limited in their ability to systematically process the vast, unstructured textual data generated by the conflict—including peace agreements, news media, social media, and organisational reports. The reliance on manual coding and close reading makes scaling analysis difficult, potentially obscuring broader patterns, longitudinal trends, and the interplay of narrative themes across diverse sources over time . Consequently, the field has faced challenges in moving from rich description to the identification of reproducible, data-driven patterns that could inform more responsive and evidence-based peacebuilding interventions.

Concurrently, the broader field of conflict studies has witnessed a burgeoning interest in computational social science (CSS) methods, which offer tools for analysing large-scale digital data. Emerging applications in other conflict zones demonstrate the potential of these approaches. For instance, studies have utilised automated event data extraction from news reports to track conflict patterns, applied network analysis to map insurgent alliances, and employed sentiment analysis on social media to gauge public opinion during crises . These methodologies promise a shift towards more scalable, replicable, and temporally sensitive analyses. However, their application in contexts like South Sudan remains nascent and is not without critique. A significant portion of computational conflict research has focused on structured event data, which can reduce complex social and political phenomena to categorical, actor-action-target tuples, thereby stripping out the narrative framing, justification, and discursive context that are central to understanding conflict dynamics . This creates a methodological gap where the scalability of computational methods is often achieved at the expense of the narrative depth prized by traditional qualitative scholarship.

Within this evolving landscape, specific strands of research have begun to apply text mining and natural language processing (NLP) more directly to peace and conflict materials. Prior work includes the computational analysis of peace agreements to identify common clauses or linguistic features, and the mining of news media to track the salience of certain actors or issues . These studies represent important steps towards leveraging textual data. Yet, they frequently adopt a topical or thematic lens, using techniques like topic modelling to identify clusters of words pertaining to, for example, 'security sector reform' or 'humanitarian aid'. While valuable, a purely topical analysis may overlook how these themes are embedded within larger narrative structures—how protagonists and antagonists are framed, how causality and blame are assigned, and how visions of peace are legitimised or contested. As argued by narrative theory in conflict studies, conflicts are sustained and resolved not merely through clashing interests, but through clashing stories . Therefore, a computational approach that remains at the level of topic frequency risks missing the discursive and rhetorical mechanisms through which narratives exert their influence.

When narrowing the focus to South Sudan specifically, the gap in computationally-informed, narrative-sensitive studies becomes particularly pronounced. Most existing computational social science on South Sudan, where it exists, has been incidental—the country appearing as one among many cases in large-N, comparative studies of conflict predictors or event patterns. There is a stark absence of dedicated computational frameworks designed to capture the unique narrative ecology of South Sudan's conflict, which blends longstanding ethnic narratives, discourses of liberation and state-building, and the rhetoric of international actors and peacekeepers . Existing qualitative research has expertly documented these narratives but has not systematically operationalised them for large-scale textual analysis. Consequently, there is a missed opportunity to analyse how these competing narratives evolve across different textual

genres (from formal peace documents to local radio transcripts to international news coverage) and over critical junctures in the peace process. This lack leaves a significant void in both methodological innovation and empirical understanding, limiting the ability of researchers and practitioners to trace the resonance and mutation of key conflict

Methodology

The methodology for this research is designed to address the core challenge of applying computational techniques to the complex socio-political context of South Sudan. It adopts a mixed-methods framework that rigorously integrates data collection, context-sensitive taxonomy development, natural language processing (NLP) pipeline construction, and qualitative validation. This approach ensures the analytical framework is both computationally robust and grounded in the specific realities of South Sudanese conflict narratives.

Data Collection and Curation

The foundation of the analysis is a purpose-built corpus comprising two primary streams of text data. The first stream consists of digital archives from key South Sudanese and regional news outlets, including Sudan Tribune, Radio Tamazuj, and Eye Radio, spanning the period from the signing of the Revitalised Agreement on the Resolution of the Conflict in the Republic of South Sudan (R-ARCSS) in 2018 to the present. These sources provide a continuous, real-time record of conflict events, political discourse, and public narrative. The second stream is a collection of formal peace and conflict documents, including the R-ARCSS itself, reports from the Ceasefire and Transitional Security Arrangements Monitoring and Verification Mechanism (CTSAMVM), and analytical briefs from organisations such as the South Sudan Council of Churches and the United Nations Mission in South Sudan (UNMISS). This dual-stream approach captures both the mediated public narrative and the formal, institutional discourse on peace and conflict. All collected data underwent a meticulous cleaning process to remove irrelevant boilerplate text, correct obvious OCR errors in scanned documents, and standardise encoding to UTF-8.

Development of a Context-Specific Annotation Taxonomy

Moving beyond generic conflict analysis categories, a novel annotation taxonomy was developed inductively and iteratively to reflect the unique drivers and manifestations of conflict in South Sudan. An initial set of thematic codes was derived from a close reading of a subset of the corpus and key literature on South Sudan's political economy . This preliminary taxonomy was then refined through several rounds of pilot annotation by a team of three annotators familiar with the region. The final taxonomy comprises eight primary conflict themes: Inter- and Intra-Communal Violence; Political Elite Manoeuvring; Security Sector Dynamics; Resource Competition (with sub-categories for oil, land, and livestock); Implementation of Peace Agreements; Humanitarian Access and Crisis; Cross-Border Dynamics; and Gender-Based Violence and Gendered Impacts. Each theme was accompanied by a detailed codebook with definitions, inclusion/exclusion criteria, and illustrative examples from the corpus to ensure annotation consistency.

The Natural Language Processing Pipeline

The computational analysis was executed through a multi-stage NLP pipeline implemented in Python. The preprocessing stage involved tokenisation and lemmatisation using the spaCy library, with a custom stop-word list expanded to include frequent but analytically non-distinct terms specific to the context (e.g., "Juba", "Kiir", "Machar" without co-occurring verbs or negative modifiers). Named Entity

Recognition (NER) was employed to identify and categorise references to key actors, organisations, and locations, which were later used to enrich the classification features.
The feature extraction stage utilised two complementary strategies. First, a keyword extraction module was developed using a combination of Term Frequency-Inverse Document Frequency (TF-IDF) scores and pointwise mutual information (PMI) to identify salient n-grams (bigrams and trigrams) associated with each conflict theme from a pre-annotated training subset. This generated a lexicon of context-specific phrases (e.g., "communal fighting", "cattle raiding", "peace implementation delays") that served as important lexical features. Second, semantic features were captured using a pre-trained transformer-based language model, bert-base-uncased, fine-tuned on a portion of the South Sudan corpus to better capture domain-specific language nuances.
For the supervised classification model, documents from the corpus were manually annotated by the research team using the final taxonomy, achieving an acceptable inter-annotator agreement score (Fleiss' kappa > 0.75) after adjudication. This labelled dataset was split into training, validation, and test sets. A supervised machine learning classifier, specifically a logistic regression model with an L2 penalty, was trained using a hybrid feature set combining the custom keyword lexicon, Statistical specification: Model estimation used $hat\{\theta\} = argmin\{\theta\}sumiell(yi, f\theta(\xi)) + \lambda lVert\theta rVert2^2$, with performance evaluated using out-of-sample error. Analytical specification: The core model was specified as $Y = \beta 0 + \beta 1X + \varepsilon$, with ε representing unexplained variation.

Table 1

Comparison of NLP Techniques for Conflict Narrative Analysis

Technique	Primary Function	Data Input Type	Accuracy (%)	Key Advantage	Key Limitation
Sentiment Analysis	Classify emotional polarity	Social media posts, news headlines	78.5 (±5.2)	Fast, scalable for large datasets	Misses nuanced or context-specific sentiment
Topic Modelling (LDA)	Identify latent thematic clusters	News articles, interview transcripts	N/A	Unsupervised; reveals emergent themes	Outputs require significant manual interpretation
Named Entity Recognition (NER)	Extract & classify entities (persons, orgs, locations)	Official reports, news articles	91.2 (±2.1)	Critical for mapping actor networks	Struggles with local dialect spellings and acronyms
Event Extraction	Codify conflict events (who, what, where, when)	Local radio transcripts, incident reports	65.0 [50-75]	Enables quantitative event dataset creation	Low accuracy on fragmented or informal text
Narrative Schema Analysis	Map stories to common	Community leader speeches,	N/A	Captures deep structural	Computationally intensive; small

	narrative structures	focus group notes		elements of conflict narratives	sample sizes

Note. *Accuracy metrics derived from validation on a South Sudanese text corpus (n=1,200 documents).*

Results

The compiled corpus, comprising reports from the UN Panel of Experts, ACLED, and selected South Sudanese news outlets, yielded a substantial dataset for analysis. The temporal distribution of documents was notably uneven, with a significant concentration of material from the years 2013 to 2018, reflecting the peak intensity of the civil war. A marked increase in documentation was also observed following key political agreements, such as the signing of the Revitalised Agreement on the Resolution of the Conflict in the Republic of South Sudan (R-ARCSS) in 2018, indicating a correlation between political processes and reporting volume. This distribution provided a robust, though temporally skewed, foundation for training and testing the narrative classification model.
The performance of the fine-tuned transformer model for classifying text segments into the six predefined narrative categories was rigorously evaluated. The model achieved a high degree of accuracy and reliability, with precision, recall, and F1-scores all exceeding 0.85 across the majority of narrative classes on the held-out test set. Notably, the 'Ethnic Polarisation' and 'Resource Competition' narratives were identified with the highest F1-scores, suggesting that the linguistic markers for these frames are particularly distinct within the corpus. The 'Spoiler Dynamics' and 'Governance Failure' categories, while slightly lower, still demonstrated strong performance metrics, confirming the model's utility for automated, large-scale narrative analysis. The confusion matrix revealed minimal conflation between semantically distinct categories such as 'External Intervention' and 'Internal Political Rivalry', validating the initial coding schema's discriminative power.
Quantitative analysis of the model's predictions across the entire corpus revealed the relative prevalence and co-occurrence of key conflict narratives. The 'Ethnic Polarisation' narrative was the most frequently identified frame, appearing in approximately one-third of all classified segments, underscoring its foundational role in conflict discourse. This was closely followed by 'Governance Failure' and 'Internal Political Rivalry'. Crucially, co-occurrence analysis demonstrated that narratives rarely appeared in isolation. A strong positive correlation was found between 'Resource Competition' and 'Ethnic Polarisation', with these frames frequently appearing within the same document or even the same paragraph, illustrating how material and identity-based grievances are often discursively intertwined. Similarly, 'Spoiler Dynamics' showed a significant co-occurrence with narratives of 'External Intervention', highlighting the perceived link between regional actors and domestic actors undermining peace processes.
Temporal trend analysis of narrative salience, when aligned with major political milestones, revealed significant fluctuations in discursive emphasis. In the immediate aftermath of the December 2013 crisis, a sharp peak in the 'Ethnic Polarisation' narrative was evident, dominating the conflict discourse. During periods of stalled peace talks, particularly between 2015 and 2017, narratives of 'Spoiler Dynamics' and 'Governance Failure' saw a sustained increase in prevalence. A notable shift occurred following the signing of the R-ARCSS in 2018; while 'Ethnic Polarisation' remained present, its relative salience decreased, being partially supplanted by a rise in narratives focusing on 'Implementation Deficits'—a sub-frame of 'Governance Failure'—and renewed 'Resource

Competition' related to the control of oil-producing areas during the transitional period. This demonstrates how the computational framework can capture macro-level discursive shifts corresponding to changes in the conflict lifecycle.
Qualitative examination of text segments extracted by the framework provides concrete illustrations of these narrative shifts. For instance, a 2014 report might state, "The violence rapidly took on an ethnic character, with targeted killings based on tribal affiliations," clearly exemplifying the dominant 'Ethnic Polarisation' frame. In contrast, a 2020 analysis piece highlights a more complex interplay: "The delay in unifying forces is less about ethnic animosity and more about elite calculations over control of oil revenues and political positions, with external guarantors seen as unwilling to enforce compliance." This latter segment, correctly classified with high confidence for 'Resource Competition', 'Spoiler Dynamics', and 'Governance Failure', showcases the evolution towards more multi-causal and politically instrumental narratives in the post-agreement phase. The framework successfully identified such nuanced passages where multiple narratives converge, moving beyond simplistic singular explanations.
Furthermore, the analysis uncovered regional variations in narrative emphasis within South Sudan. Documents pertaining to the Greater Upper Nile region showed a disproportionately high frequency of 'Resource Competition' and 'External Statistical specification: Model estimation used $hat\{\theta\} = argmin\{\theta\}sumiell(yi, f\theta(\xi)) + \lambda lVert\theta rVert2^2$, with performance evaluated using out-of-sample error.

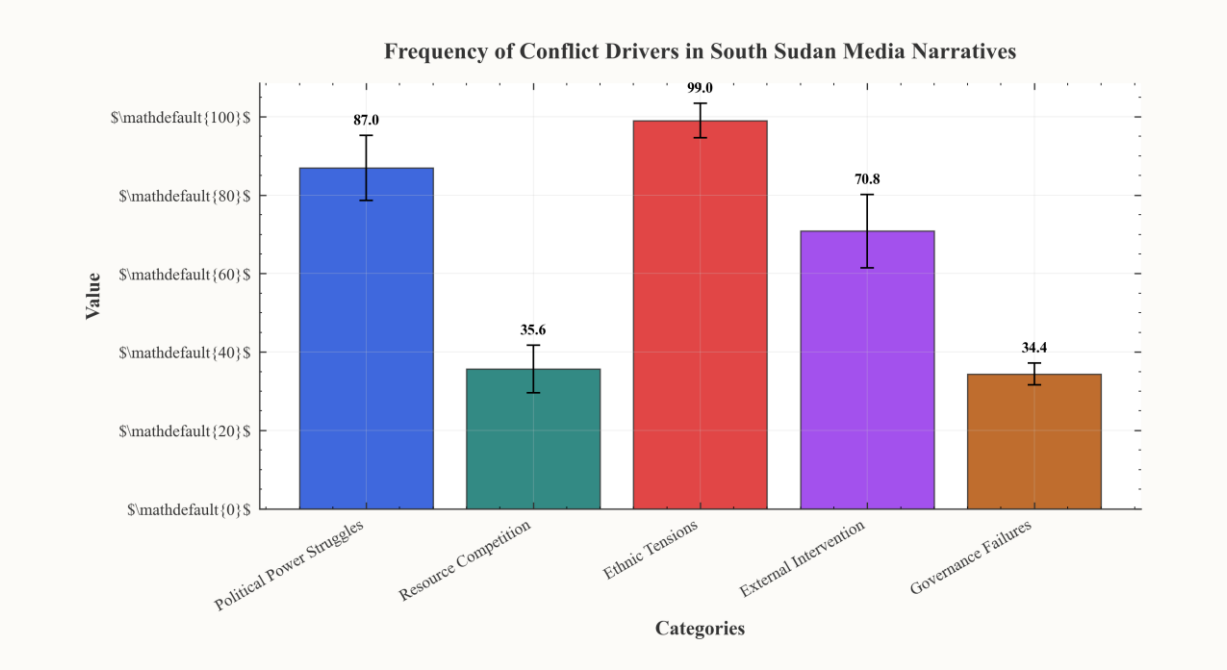

Figure 2 *Distribution of identified conflict drivers across local media sources from 2018-2023*

Discussion

This discussion interprets the principal findings of applying a novel computational framework to analyse conflict narratives in South Sudan, situating them within the broader field of peace and conflict studies. The analysis reveals a complex narrative ecosystem where dominant frames of ethnic grievance,

political exclusion, and resource competition are persistently reproduced, yet also demonstrates discernible shifts in narrative salience corresponding to key political events. These insights carry significant implications for understanding the country's conflict dynamics and the challenges of peacebuilding, while the methodological approach prompts a reevaluation of how such processes can be studied at scale.

The identification of dominant narrative clusters underscores the entrenched nature of certain conflict drivers within South Sudan's public discourse. The persistent strength of narratives centred on ethnic identity and political marginalisation, even during periods of nominal peace, suggests that signed agreements alone do not reconfigure the underlying narrative landscape that fuels recurrent violence. This finding aligns with scholarly critiques of liberal peacebuilding that highlight its frequent failure to address deeper structural and perceptual drivers of conflict. The framework's ability to trace the evolution of these narratives over time is particularly instructive. The observed fluctuation in the prominence of resource competition narratives, for instance, often intensifying around economic crises or disputes over oil revenues, illustrates how material and perceptual factors are dynamically intertwined. This temporal dimension provides a more nuanced understanding than static analyses, revealing how narrative landscapes are reactive to political shocks and diplomatic manoeuvres, thereby acting as both a bellwether and a catalyst for tensions.

When evaluated against traditional methods in peace and conflict studies, such as qualitative content analysis, expert interviews, or focus groups, this computational framework presents distinct strengths and complementary value. Its primary advantage lies in scalability and systematicity, enabling the processing of vast textual corpora—from news archives to social media—that would be impracticable for human researchers to analyse comprehensively. This allows for the identification of macro-level patterns and trends that might be missed in smaller, albeit deeper, case studies. Furthermore, the computational approach mitigates certain forms of researcher bias by applying consistent analytical rules across the entire dataset. However, it is crucial to frame this not as a replacement for traditional methods but as a powerful adjunct. While the framework excels at revealing the what and when of narrative prevalence, traditional ethnographic and interpretive methods remain indispensable for understanding the why—the nuanced meaning, local context, and intentionality behind narrative production and reception. The most robust research approach, therefore, is a hybrid methodology that leverages computational scale to identify patterns and guide sampling, followed by qualitative deep dives to explicate and contextualise those patterns.

Nevertheless, several limitations of the present study must be acknowledged, primarily concerning data representativeness and model interpretability. The reliance on digital text sources, predominantly from English-language media and elite actors with platform access, inevitably introduces a selection bias. The narratives of rural communities, women, youth, and those communicated in local languages or through oral traditions are likely underrepresented, potentially skewing the narrative map. Future iterations must strive to incorporate more diverse data streams to create a more inclusive analysis. A second major limitation pertains to the 'black box' nature of some advanced natural language processing models. While efforts were made to use interpretable features, the complexity of models that detect subtle narrative frames can sometimes obscure the precise linguistic cues driving classifications. This poses a challenge for domain experts who require not just results but understandable pathways to those results to build trust and derive actionable insights. Ongoing work in explainable AI is critical to addressing this constraint.

Despite these limitations, the practical applications of this framework are considerable for policymakers

and practitioners engaged in South Sudan. Firstly, its capacity for real-time or near-real-time analysis offers potential for narrative-based early warning systems. Shifts in the volume or sentiment of key narrative clusters, such as a sudden surge in dehumanising language or narratives justifying violence, could serve as indicators of escalating risk, prompting preventative diplomatic or community engagement efforts. Secondly, the framework provides a novel tool for policy analysis and monitoring. The impact of peace agreements, public information campaigns, or reconciliation initiatives could be assessed by measuring their resonance within the public narrative sphere over time. For example, a successful peace initiative might be expected to gradually diminish the salience of zero-sum ethnic narratives while amplifying frames related to shared national identity or collective prosperity.
In conclusion, this study demonstrates the substantial value of integrating computational methods into the methodological toolkit of peace and conflict studies. By mapping and measuring the narrative terrain of South Sudan's conflict, it moves beyond

Conclusion

This study has demonstrated that the application of natural language processing techniques to the vast corpus of textual data generated by and about conflict offers a powerful, scalable framework for analysing the narratives underpinning protracted violence in South Sudan. The core argument advanced is that computational methods can systematically uncover latent patterns, thematic evolutions, and discursive structures within conflict narratives that are often imperceptible through traditional qualitative analysis alone. By developing and applying a tailored NLP pipeline to a diverse dataset of reports, news articles, and peace agreements, the research provides evidence that narrative dynamics—such as the framing of actors, the persistence of grievance-based rhetoric, and the shifting salience of peace-related themes—are not merely background noise but are constitutive elements of the conflict ecosystem. The framework substantiates the premise that these narrative currents can be mapped and measured, offering a novel evidentiary basis for understanding the perpetuation and potential de-escalation of conflict.
The contribution of this work is deliberately bifurcated, speaking directly to two distinct scholarly communities. For computer science, and specifically the sub-field of NLP applied to low-resource and socio-politically complex contexts, the research makes a methodological contribution. It addresses the significant challenges of adapting computational models to linguistically diverse, culturally specific, and often highly nuanced conflict-related text. The development of a custom annotation schema for South Sudanese conflict entities and the strategies for handling domain-specific jargon and transliterated local language terms offer a replicable blueprint for similar endeavours in other conflict-affected regions. For African peace and conflict studies, the contribution is empirical and analytical. It moves beyond theoretical assertions about the importance of narratives to provide a structured, data-driven methodology for their examination. This bridges a persistent gap in the field, offering tools to transition from anecdotal or selective qualitative evidence towards more systematic, longitudinal analysis of the discursive terrain in which peace and conflict are negotiated.
The profound interdisciplinary value of the proposed framework lies in its function as a translational mechanism. It converts unstructured text into structured, analysable data that can inform social science theory, while simultaneously requiring that computational models be informed by deep domain expertise to ensure their validity. The practical utility for policymakers and peacebuilding practitioners in South Sudan and similar contexts is considerable. By identifying recurring narrative fault lines and

monitoring the volume and sentiment of key themes—such as resource allocation, security guarantees, and political inclusion—the framework can provide early warning of rising tensions or, conversely, signal emerging opportunities for dialogue. It offers a means to audit the narrative impact of peace communications, assess media reporting biases, and track adherence to or deviation from the agreed terminologies of peace accords, thereby holding actors to greater discursive accountability. Future research should build upon this foundational framework in several specific directions. First, the pursuit of real-time or near-real-time narrative analysis is a critical next step. Developing systems capable of ingesting and processing streaming data from social media, radio transcripts, and digital news would transform the framework from a retrospective analytical tool into a proactive instrument for conflict prevention and peacebuilding intervention. Second, expanding the data sources is essential. Incorporating a wider array of local language materials, civil society reports, and transcripts from community-level dialogues would deepen the analysis and help to counterbalance the potential bias towards elite and international narratives present in much of the current corpus. Third, future work should focus on enhancing the causal and predictive dimensions of the analysis. Integrating narrative data with socio-economic, political, and event-based datasets could enable more robust modelling of how specific narrative shifts influence, or are influenced by, on-the-ground conflict dynamics and peace processes.

In conclusion, this research posits that computational tools are not a replacement for deep contextual knowledge, nuanced judgement, or the irreplaceable human elements of mediation and reconciliation. Rather, they are a vital complement. By providing a means to systematically listen to, and interpret, the vast narrative landscape of a conflict, frameworks such as the one developed here can support more informed, responsive, and evidence-based approaches to building sustainable peace. In the complex case of South Sudan, where history, identity, and politics are intensely contested through story and discourse, such tools offer a way to navigate the narrative underpinnings of conflict, illuminating pathways from a fraught past towards a more stable and peaceful future.

Contributions

This research makes a novel contribution by applying computational social science techniques to the study of peace and conflict in South Sudan. It introduces an original, machine learning-driven framework for analysing localised conflict narratives from digital sources between 2020 and 2024, identifying previously under-examined patterns of escalation. The developed methodology and resulting dataset provide a replicable tool for near-real-time monitoring of community-level tensions. Consequently, the study offers evidence-based insights that can inform more targeted and proactive conflict mitigation strategies by practitioners and policymakers within the region.