Integrating LLMs and the PrOntoNarr Ontology for Automated KG Construction in Multi-level Propaganda Narrative Analysis

Francesco Orciuoli[1,*,†], Antonella Pascuzzo[1,†] and Sabrina Senatore[2,†]

[1]*Dipartimento di Scienze Aziendali - Management & Innovation Systems (DISA-MIS), Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy*

[2]*Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica Applicata (DIEM), Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy*

Abstract

Knowledge Graphs (KGs) are essential for advancing semantic web technologies, yet their construction from unstructured text, particularly in domains like political discourse, remains a significant challenge. Traditional rule-based and machine learning methods often fail to capture the structural and persuasive dimensions of underlying narratives. To address this gap, this study proposes a comprehensive framework for identifying, modeling, and querying narratives within propaganda across three levels: text, paragraph, and campaign. Central to this approach is the PrOntoNarr ontology, a formal semantic model that integrates narrative theory with propaganda analysis to facilitate the construction of a KG. Leveraging the generative capabilities of Large Language Models (LLMs), the framework automates KG construction by transforming unstructured propaganda texts into formal RDF triples through prompt engineering. The resulting KG provides a structured representation of narrative elements in political speeches and enables complex, multi-level analysis via SPARQL queries. The framework is validated using a dataset of political propaganda speeches from the 2025 Gaza ceasefire negotiations, demonstrating its capacity to support researchers in uncovering patterns and persuasive strategies employed in modern persuasive discourse.

Keywords

Propaganda analysis, Knowledge Graph Construction, Large Language Models, Narrative Theory

1. Introduction

Propaganda has been a part of human society for centuries, gradually evolving from its religious roots to become a potent political and social tool. Over time, it has moved from word of mouth to mass media and, more recently, to digital platforms and social networks. Based on the work of [1], Propaganda is here defined as the deliberate and systematic attempt to shape perceptions, cognition, and behaviors in order to achieve the intentions of the propagandist. In the political domain, propaganda is central to the construction of discourse. Politicians use speeches to recount events, frame issues, and impose interpretations that aim to influence rather than merely inform. Such discourse projects a political identity tailored to resonate with an audience that either already shares these views or can be easily persuaded to adopt them. At the core of the speeches, there are narratives and their structure (e.g., schemas, characters, roles) which are strategically selected to maximize clarity, resonance, and persuasive power, making these recognizable as a form of propaganda [2]. While the immediate impact of an isolated speech may be limited to the occasion of its delivery, its integration into a broader campaign significantly amplifies its influence. Within such campaigns, individual speeches function as interconnected narrative units that mutually reinforce one another, generating cumulative and long-term effect on public opinion. Despite the centrality of narrative structures in propaganda, the

ESWC'26: 5th Workshop on LLM-Integrated Knowledge Graph Generation from Text (TEXT2KG), May 10-14, 2026 | Dubrovnik, Croatia

*Corresponding author.

†These authors contributed equally.

forciuoli@unisa.it (F. Orciuoli); apascuzzo@unisa.it (A. Pascuzzo); ssenatore@unisa.it (S. Senatore)

0000-0001-6899-4396 (F. Orciuoli); 0009-0005-2271-6338 (A. Pascuzzo); 0000-0002-7127-4290 (S. Senatore)

extraction and modeling of this knowledge from unstructured text remains a significant challenge. Existing methods (e.g., manually crafted rules, machine learning classifiers) struggle to capture the underlying schemas or the actors, events, and relations through which persuasion operates.

This paper addresses this gap by posing the following research question**: Can LLMs be effectively integrated with formal ontologies to automate the construction of Knowledge Graphs (KGs) for propaganda narrative analysis?**. Therefore, the primary objective is to assess the feasibility of this integration, and demonstrate its capacity to support propaganda analysis across multiple levels: text, paragraph, and campaign.

The core contributions of this work are i) the formalization of concepts from narrative theory into a dedicated ontology, the Propaganda Ontology Narrative Representation (PrOntoNarr), grounded in OntoNarr ontology [3]; ii) the automation of KG construction, demonstrating how LLMs via prompt engineering transforms unstructured texts into RDF triples; and iii) validation of the resulting KG, demonstrating how SPARQL queries enables a comprehensive analysis of propaganda patterns, that are otherwise invisible to traditional analysis. The framework is validated on a corpus of propaganda texts related to Gaza ceasefire negotiations. The paper is organized as follows: Section 2 reviews the literature, Section 3 details the proposed framework; Section 4 presents its validation; Section 5 highlights the added value of the framework, while Section 6 concludes the paper.

2. Related Works

Ontologies are increasingly employed to capture complex domains, with narrative theory emerging as a particularly promising area of investigation. A primary line of research formalizes narratives as ontologies grounded in traditional narratology. For instance, the authors of [4] conceptualize narratives in terms of fabula, narration, and plot. They introduce NOnt as both a mathematical specification and a Semantic Web implementation, in which events, timelines, and their relations are treated as first-class queryable objects within digital libraries. Subsequent extensions have captured the geospatial dimensions of narratives (NOnt+Space), enabling GeoSPARQL queries over narrative KG [5]. Similarly, [6] model literary narratives through ontologies and lexica to support SPARQL querying over characters, events, and temporal structures. However, while these models successfully handle literary domains, they do not address the structure of political discourse and propaganda narratives. The authors of [7] presents POlitical DIscourse Ontology (PODIO) to provide a structured representation of political debates on various digital platforms; however, they do not explicitly capture the narrative dimension of propaganda.

A second line of work focuses on the computational analysis of propaganda, a domain that remains largely independent of the ontological approaches. For example, [8] predominantly employs machine learning for the fine-grained, span-level identification of propaganda techniques in text. While effective for localized detection, such approaches bypass the modeling of the underlying narrative structures.

To bridge the gap between raw textual analysis and structured reasoning, a third line of research has started integrating LLMs with KGs. This approach assumes that while LLMs possess powerful natural language capabilities, they remain opaque black boxes; KGs, conversely, offer the explicit, formal structure required for precise and verifiable analysis [3]. The authors of [9] propose a multi-agent system in which multiple LLMs, each specialized on specific ontology modules, collaboratively extract RDF triples from heterogeneous data sources. However, to date, these hybrid pipelines have not been applied to the domain of narrative propaganda.

Motivated by this gap, the present work synthesizes these three intersecting lines of research to propose a framework aimed at advancing the formal understanding and computational representation of propaganda narratives. In particular, the framework formally models their structures in the PrOntoNarr ontology, populates it using LLMs to facilitate the construction of KGs, and finally explores the resulting KGs through SPARQL queries, enabling a deeper formal understanding of propaganda narratives.

3. The Proposed Framework

This section outlines the proposed framework for representing propaganda through its underlying narrative structures. This framework addresses the limitations of purely machine learning-based detection by enabling complex, structured reasoning over political discourse. As illustrated in Fig. 1, the workflow bridges textual analysis and formal semantic representation through three stages: i) pre-processing the input (propaganda texts); ii) using LLMs to populate the PrOntoNarr ontology and constructing KGs, and iii) systematically analyzing the resulting KGs via SPARQL queries. Together, these stages facilitate a comprehensive exploration of persuasive patterns and strategies across the text, paragraph, and campaign levels.

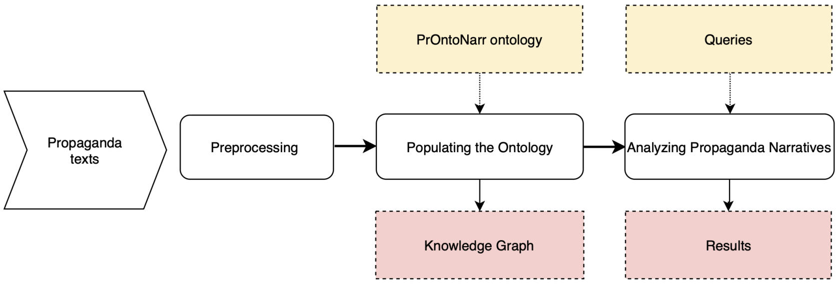

Figure 1: Overall Framework

The workflow begins with the acquisition of propaganda texts, which serve as the primary input for the framework. The way texts are processed reflects the analytical objective to be executed. Once processed, the texts support the ontology population process, guided by the PrOntoNarr ontology, a formal semantic model carefully designed to capture the structural nuances of propaganda narratives. The resulting KG enables the systematic analysis of propaganda phenomena through SPARQL queries. Consequently, the entire pipeline, from the initial segmentation of the text to the formulation of the final queries, requires a rigorous definition of the overarching analytical goals.

3.1. Preprocessing of the texts

The workflow (Fig. 1) begins with the systematic preprocessing of propaganda texts. This preprocessing must align strictly with the study's analytical objectives, making their definition essential (e.g., analyzing a propaganda campaign across different media and coming from several sources, analyzing a specific set of speeches delivered by the same speaker, analyzing multiple paragraphs of the same speech). To meet these different objectives, the framework implements three distinct pre-processing strategies:

- **A. Individual text level:** Each propaganda text is treated as a single unit, allowing for a detailed examination of the narrative structure and propaganda elements it contains.
- **B. Paragraph level:** Texts are segmented into paragraphs to allow for a more fine-grained analysis, making it possible to track the progression of narratives or propaganda elements throughout the text and compare their deployment across different sections. In this case, segmentation is performed using NLP toolkits[1] or structural line breaks.
- **C. Campaign level:** Multiple texts are analyzed collectively as components of a single propaganda campaign, facilitating the identification of recurring patterns, strategies, and schemas across texts.

While strategies A and C treat the full text as the unit of analysis, strategy B necessitates a segmentation step. This multi-level approach provides the flexibility required to model propaganda narratives comprehensively, ensuring the input data is systematically structured for downstream tasks, such as ontology population and KG construction. These three levels serve as reference cases to which more complex analytical needs can be mapped.

[1]https://www.nltk.org/

3.2. PrOntoNarr

The development of PrOntoNarr[2] follows MethOntology framework [10], adopting a structured lifecycle composed of four distinct phases: i) specification, which defines the ontology's scope and requirements; ii) conceptualization and formalization, where requirements are translated into a Glossary of Terms (GoT) and formal concepts, leveraging the reuse of existing models (in particular, OntoNarr); iii) implementation, coding the formal model within the Protégé[3] environment; and iv) evaluation, verifying the consistency of the resulting knowledge base using the HermiT[4] Reasoner.

To effectively model the narrative structures within political speech, PrOntoNarr integrates and extends the core classes of OntoNarr: `NarrativeText`, `NarrativeSchema`, `NarrativeComponent`, `NarrativeElement`, `Agent`, and `Archetype`. The `NarrativeText` class represents the portions of political speech in which narratives are present. Conversely, the `NarrativeSchema` class encompasses seven direct subclasses, each corresponding to a distinct schema, as detailed in Table 1.

Table 1
The function of each narrative schema in structuring narratives

Narrative Schema	Function of the Narrative Schema
AmericanDream	Emphasizes ambitious goals, national ideals, and the pursuit of a better future
HerosJourney	Highlights personal challenges, transformation, and the final return
PetalStructure	Organizes narratives around a main story supported by secondary or complementary stories
WaveStructure	Frames narratives around rising and falling emotional intensity
Redemption	Focuses on moral or personal redemption, recovery, and reconciliation
UsVsThem	Frames narratives as conflicts between an in-group and an out-group, emphasizing opposition and contrast
ThreeActStructure	Follows the traditional three-act division, including exposition, conflict, and resolution.

However, as noted by [11], classifying texts into these schemas is an interpretive process, as a single text frequently presents features of multiple overlapping schemas. To address this complexity, PrOntoNarr includes rigorous, and literature-grounded definitions (such as those in Table 1) directly into the ontology as `rdfs:comment` annotations to clarify their scope and characteristics. The structural logic of schemas is enforced through `NarrativeComponent`; each component is linked to its schema via subproperties of `belongsTo` (e.g., `belongsTo_AmericanDream`), preserving domain constraints. The logical sequence of components is modeled using the transitive object property `precedes`. For example, within the `UsVsThem` schema, the structural logic dictates that the definition of "Us" precedes the definition of "Them," which in turn precedes "Conflict." Due to transitivity, the reasoner infers that "Us" also precedes "Conflict".

Narrative content itself is represented by subclasses of `NarrativeElement` (e.g., Purpose, Solution, Challenge, Topic, Message, Object, Organization, and Person). These are linked to their respective NarrativeSchema using the `refersTo` property. However, Objects, Organizations, and Persons are also defined as subclasses of `Agent`. These are further enriched by assigning them traditional roles (e.g., Hero, Villain) through the `Archetype` class, which specifies the narrative functions these entities perform within the discourse.

While narrative structure is central, PrOntoNarr also captures the specific dimensions of propaganda (Fig. 2). This extension centers on the `PropagandaText` class, which serves as a hub linking the text to its persuasive dimensions and its metadata, while also encompass narrative(s) (NarrativeText). The persuasive dimensions are captured by two classes: `PersuasiveEffect`, which represents the intended psychological or emotional impact on the audience and `PropagandaTechnique`, which taxonomy (e.g., the selection of specific techniques and their definition) is derived from [12]. Both classes are linked to the PropagandaText via the object properties `evokesEffect` and `usePropagandaTechnique`, respectively.

[2]Available at: https://github.com/apascuzzo/PrOntoNarr
[3]https://protege.stanford.edu/
[4]http://www.hermit-reasoner.com/

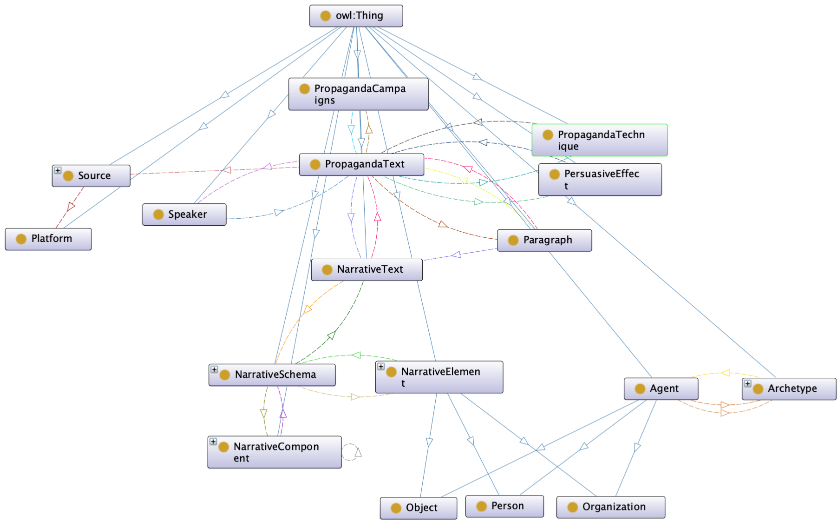

Figure 2: Overview of the PrOntoNarr schema extension.

The contextual metadata are captured through the `Speaker` and `Source` classes. A PropagandaText is attributed to a `Speaker` via the `deliveredBy` property, with the speaker further detailed through data properties. Similarly, the origin of the text is modeled by the `Source` class and linked via the `retrievedFrom` property. The `Source` class includes subclasses for diverse media types (e.g., social media posts, archival records, publications) and is annotated with data properties. Since modern propaganda predominantly operates within digital ecosystems, `Platform` class is also integrated to capture specific hosting environments.

Finally, as PrOntoNarr is designed to support multi-level analysis, other two classes are introduced: `PropagandisticCampaign` and `Paragraph`. At the macro level, the PropagandisticCampaign class aggregates all texts associated with a unified campaign. These texts are linked to the campaign using the `comprisesText` property, while the overarching strategic objectives are captured through the `hasOverallGoal` data property. At the micro level, a single PropagandaText can be segmented into smaller units; therefore, it is linked to instances of the Paragraph class via the `hasParagraph` property. It is worth reminding the reader that both the entire text (PropagandaText) and each individual Paragraph can present their own narrative structures. As a consequence, both are linked to NarrativeText using the `hasNarrative` property.

The resulting PrOntoNarr T-Box comprises 478 axioms, consisting of 64 classes, 39 object properties, and 11 data properties. To ensure logical robustness, 30 disjointness constraints were applied to schema components, and property characteristics (e.g., inverse, transitive) were rigorously defined. As the final step of the MethOntology lifecycle, the ontology was evaluated using the HermiT reasoner [13] via both the Protégé plugin[5] and the Owlready2 Python library[6]. The evaluation confirmed that the ontology is consistent and free of contradictions.

3.3. Populating the Ontology

The processed input serves as the basis for ontology population, a task primarily delegated to LLMs to ensure scalability. The rationale for this approach lies in the time-intensive nature of manually mapping unstructured textual content into formal semantic instances. To illustrate the depth of reasoning required, and thus the necessity of an automated solution, a walkthrough of the manual instantiation process is presented, using an excerpt from a speech by Donald Trump on the Gaza ceasefire (October 13, 2025). This example demonstrates the flexibility of the proposed framework in modeling the speech simultaneously across three levels.

[5]https://protege.stanford.edu
[6]https://owlready2.readthedocs.io/en/v0.48/

"We have peace in the Middle East, and it's a very simple expression, peace in the Middle East. After years of suffering and bloodshed, the war in Gaza is over. Humanitarian aid is now pouring in, including hundreds of truckloads of food, medical equipment, and other supplies. Civilians are returning to their homes. Together, we've achieved what everybody said was impossible at long last. We all know how to rebuild and we know how to build better than anybody in the world. I wanna express my tremendous gratitude to the Arab and Muslim nations who helped make this incredible breakthrough possible."

The manual instantiation (Fig. 3) is executed by a single human annotator tasked with coordinating three levels of analysis by i) identifying the macro-level context (e.g., the Campaign and all its constituent speeches); ii) analyzing each individual speech as a single unit to model its narrative structure and iii) decomposing the speech into Paragraphs to model their narrative structures as well.

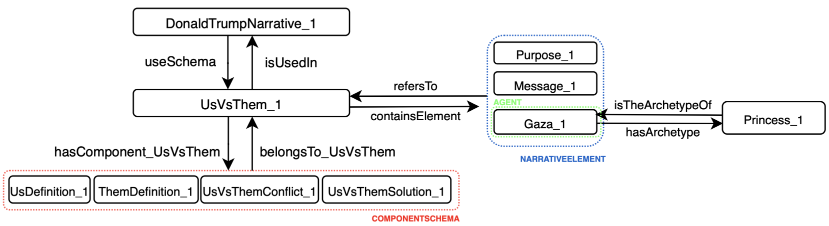

Figure 3: Manual creation of instances for Donald Trump's political speech

Listing 1 details the result of the manual instantiation, demonstrating the capacity of the ontology to model at three different levels of analysis. At the campaign level, the instance `GazaCampaign_1` captures the macro-context, articulating the general objective ("negotiation of a ceasefire in Gaza") and including the specific discourse `DonaldTrumpSpeech_1`. This text instance is further enriched with metadata, linking it to the speaker (`Speaker_1`), the source (`Source_1`), and propaganda technique (`PropagandaTechnique_1`). When treated as a single unit (individual level), the speech presents a narrative schema (`UsVsThem_1`) with its associated components. When segmented into paragraphs, `Paragraph_1` is linked to the text using the properties *isTheParagraphOf*, and linked to its own narrative (`NarrativeText_2`), distinct from the previous one, using the property *hasNarrative*.

```
:GazaCampaign1
   a :PropagandaCampaign ;
   :hasOverallGoal "Gaza␣ceasefire␣negotiation" ;
   :comprisesText :DonaldTrumpSpeech1 .

:DonaldTrumpSpeech1
   a :PropagandaText ;
   :isComprisedIn :GazaCampaign1 ;
   :hasParagraph :Paragraph_1 ;
   :hasNarrative :NarrativeText_1 ;
   :usePropagandaTechnique :LoadedLanguage_1 ;
   :deliveredBy :Speaker_1 ;
   :retrievedFrom :Source_1 .

:LoadedLanguage_1
   a :LoadedLanguage .

:Paragraph_1
   a :Paragraph ;
   :isTheParagraphOf :DonaldTrumpSpeech1 ;
   :hasNarrative :NarrativeText_2 .
```

Listing 1: Example of manual instantiation

While manual instantiation is reliable, this process is time-consuming and difficult to scale to large datasets. To overcome this bottleneck, an automated method based on LLMs for ontology population, and consequently, KG construction is introduced. LLMs demonstrate remarkable performances in complex tasks when guided by well-crafted prompts [14]; therefore, a template (Fig. 4), structured into functional blocks, is designed.

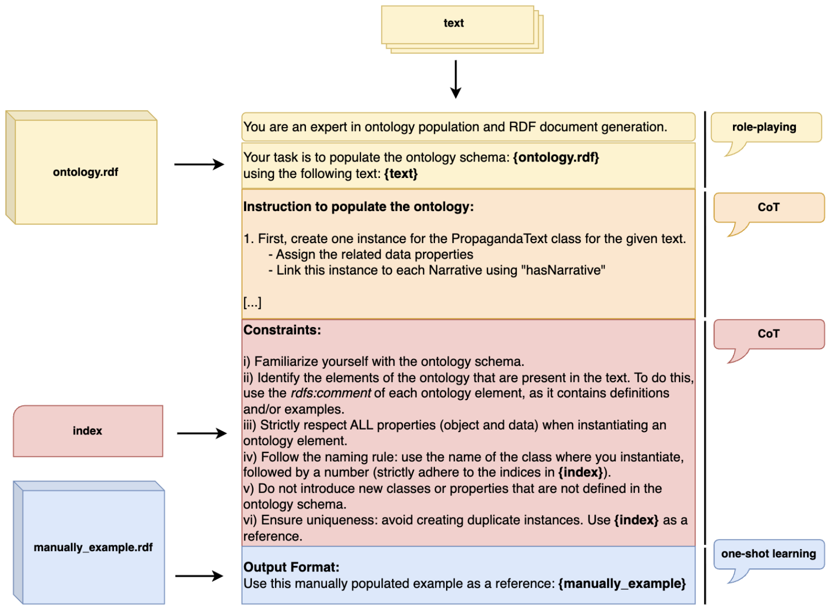

Figure 4: Prompt template.

Such a template employs the role-playing technique [8], assigning to the model the role of expert in ontology population, followed by a clear definition of the task(s) to perform (yellow blocks). Next, detailed instructions on how performing the task is provided using Chain-of-Thought (CoT) [15] technique (orange block). To prevent the generation of invalid RDF triples, explicit constraints are defined (red blocks). In particular, the model is instructed to name each instance according to the format <instance-name_index>. To enforce this naming convention, the model is initialized with an index file where all counters for ontological classes and properties are set to 0. During each invocation (e.g., for each analyzed text), these indices are dynamically updated whenever a specific class or property is instantiated, and the incremented values are fed into subsequent calls. Finally, the model is grounded with contextual knowledge (blue blocks). This includes a manually curated instantiation example that serves as the primary reference for the model during output generation. Following a one-shot learning approach, this example acts as a key strategy to mitigate hallucinations. The proposed template was used to invoke the *gpt-oss-20b*[7] open-source model via Groq. Listing 2 illustrates the LLMs' output, revealing its capacity to successfully populate the ontology and construct KGs.

3.4. Analyzing Propaganda Narratives through SPARQL

The rigorous definition of analytical goals is required, as it not only guides the initial preprocessing of the textual input but also informs the formulation of queries. Once populated, the resulting KGs are deployed within a triple-store database. In this study, query formulation adheres to the heuristics proposed by [16]: i) relevant domain concepts, derived directly from the overarching analytical goals, are systematically mapped into semantic triples; and ii) the populated KG is navigated by explicitly leveraging the formal classes and properties established within the ontology. It is worth noting that SPARQL queries do not serve merely for simple data retrieval. Instead, they function as instrumental tools for empirically investigating complex narrative structures, allowing to extract structural insights and uncover the underlying mechanisms of real-world propaganda.

[7]https://console.groq.com/docs/model/openai/gpt-oss-20b

```
:DonaldTrumpSpeech1
   a :PropagandaText ;
   :deliveredFor "Gaza␣Peace␣Agreement" ;
   :deliveredIn "Washington,␣D.C." ;
   :deliveredAt "2025-11-11T08:00:00";
   :hasNarrative :NarrativeText_1 ;
   :deliveredBy :Speaker_1 ;
   :retrievedFrom :Source_1 .

:NarrativeText_1
   a :NarrativeText ;
   :useSchema :UsVsThem_1 ;
   :isTheNarrativeOf :DonaldTrumpSpeech1 .

:UsVsThem_1
   a :UsVsThem ;
   :isUsedIn :NarrativeText_1 ;
   :hasComponent :UsDefinition_1, :ThemDefinition_1, :UsVsThemConflict_1, :UsVsThemSolution_1 ;
   :containsElement Trump_1, :Gaza_1, Conflict_1.

:UsDefinition_1
   a :UsDefinition ;
   :belongsTo_UsVsThem :UsVsThem_1 ;
   rdfs:comment "President␣Donald␣Trump,␣and␣supportive␣Arab␣and␣Muslim␣nations." ;
   :precedes :ThemDefinition_1 .

:DonaldTrump_1
   a :Agent ;
   :hasArchetype "Hero" .
```

Listing 2: Example of LLM-based instantiation

4. Experimentation and Validation

The proposed framework is validated through a case study analyzing 25 political propaganda texts published between September and October 2025, focusing on the Gaza ceasefire negotiations. The corpus was collected from official institutional and media sources (e.g., the Scottish Government[8], The Times of Israel[9], and Al Jazeera[10]), and preprocessed to align with the analytical goals. Within this context, three analytical objectives are defined, each systematically mapped to a distinct level of analysis:

- **Individual Text Level (Strategy A):** Identification of propaganda techniques within a specific political address (e.g., Donald Trump's inaugural speech on the crisis).
- **Paragraph Level (Strategy B):** Examination of the evolution of narrative schemas and their rhetorical purposes across the paragraphs of a single speech.
- **Campaign Level (Strategy C):** Comparative analysis of the overarching narrative schemas leveraged by key political figures (Trump, Netanyahu, and Abbas) within the broader ceasefire debate.

The PrOntoNarr ontology was populated using template in Fig. 4. To evaluate the quality of the automated instantiation process (hereafter, LLM-based), a manual instantiation (hereafter, human-based) was also performed, serving as the ground truth. The metrics proposed in the work of [17] were applied: i) *schema metrics*, to assess whether the ontology design accurately models the domain knowledge and ii) *knowledge base metrics* to evaluate the effectiveness of the ontology design and the amount of real-world knowledge it represents. Following the official specifications, both metrics were computed in Python. The results, summarized in Table 2, highlight fundamental structural differences between the two methods.

[8] https://www.gov.scot/publications/gaza-israel-first-ministers-speech/
[9] https://www.timesofisrael.com/
[10] https://www.aljazeera.com

Table 2
Comparison of Schema and Knowledge Base metrics between Human- and LLM-based population methods.

		Human Method	LLM-based Method
Schema Metrics	Attribute richness	0.096	0.400
	Inheritance richness	1.269	1.800
	Relationship richness	0.445	0.318
	Axioms-class ratio	10.904	8.107
	Inverse relations ratio	0.417	0.306
	Class-relation ratio	0.437	0.379
KB Metrics	Average population	0.942	0.920
	Class richness	0.519	0.413

The human method yields higher *Relationship Richness* (0.44 vs. 0.32) and *Class Richness* (0.52 vs. 0.41), revealing a focus on establishing meaningful connections between instances. On the other hand, the LLM-based method demonstrates higher *Attribute Richness* (0.40 vs. 0.09) and *Inheritance Richness* (1.80 vs. 1.27), suggesting that the model tends to produce instances with a denser set of properties and a wider hierarchical distribution, while establishing fewer interconnections between them. It is important to note that the higher metrics for the human-based method are entirely expected, and the gap with the LLM-based method highlights that propaganda analysis requires a level of abstract reasoning and contextual understanding that current LLMs struggle to fully replicate. Nevertheless, the LLM-based results remain highly promising: the model captures a comprehensive and semantically meaningful portion of the underlying information in a fully automated setting. The human method ensures higher conceptual accuracy, while the LLM-based approach offers scalability and efficiency, successfully replicating domain knowledge (Average Population ≈ 0.93) while significantly reducing processing time.

The LLM's performance in extracting narrative is further assessed using the evaluation framework proposed in [18], employing string-matching (Levenshtein, Jaro-Winkler) and set-based similarity metrics (Cosine). In this setup, the LLM's output serve as candidates against the manually populated one which serve as ground truth. The findings are detailed in Table 3.

Table 3
Summary statistics of LLM narrative extraction performance across 25 documents.

Metric	Mean	Max	Min	Std
levenshtein	3.94004	4.727	3.100	0.443297
avg_jaro_winkler	0.94188	0.950	0.931	0.006173
avg_cosine	0.46792	0.533	0.378	0.048449
matches_count	42.16000	76.000	27.000	15.447977
instance_coverage	64.55560	88.100	52.940	9.641290
properties_coverage	100.00000	100.000	100.000	0.000000

The high mean Jaro-Winkler similarity (0.942) and the low mean Levenshtein distance (3.94) demonstrate strong string-level accuracy, suggesting that the model reproduces ground-truth labels with minimal variation (e.g., discrepancies are largely limited to the index, such as *ThreeActStructure_1* versus *ThreeActStructure_2*). This precision, coupled with a Properties Coverage of 1.00 confirms that the LLM perfectly internalizes the T-Box structure. The model's ability to identify and correctly label every property defined in the ontology explains the high *Attribute Richness* observed in Table 2.

However, a divergence emerges at the set-based level (A-Box). The Cosine similarity score (0.468) and the Instance Coverage (64.55) indicate that, while individual labels are well-formed, the model does not consistently retrieve the complete set of expected instances. This drop in coverage is primarily due to the LLM's tendency to aggregate multiple occurrences of the same technique into a single instance. For example, if the model identifies the Bandwagon technique in multiple segments of a speech, it often creates a single instance and concatenates all corresponding text snippets in the `rdfs:comment` data property, rather than generating distinct URIs for each occurrence. In contrast, the human-based method correctly instantiates separate URIs for each individual occurrence. Despite this limitation, the

```
SELECT ?text1 ?techniqueComment
WHERE {
   ?text1 a PrOntoNarr:PropagandaText ;
        PrOntoNarr:isDeliveredBy ?speaker1 ;
        PrOntoNarr:usePropagandaTechnique ?technique .

   ?speaker1 a PrOntoNarr:Speaker ;
          rdfs:comment "Donald␣Trump" .

   ?technique a PrOntoNarr:PropagandaTechnique .
        rdfs:comment ?techniqueComment .
}
```

Listing 3: SPARQL query for the first analytical objective

LLM can still be used to populate the ontology comprehensively and construct KGs, as the observed gaps are due to structural instance aggregation rather than labeling errors or conceptual misalignment. These limitations may be mitigated in future work by enforcing stricter constraints in the prompt or employing models with larger context windows. To ensure a consistent sequence of indices and avoid hallucinations, a file storing the updated counters for each ontological element is maintained and passed to the model during each invocation (see Fig. 4).

Moreover, the disambiguation of rhetorical figures is also handled by employing the CoT technique and instructing the model to leverage the annotations in the `rdfs:comment` when populating the ontology, as suggested by [19]. The resulting KGs were stored in a GraphDB[11] triple-store database. To demonstrate the efficacy of the proposed framework in facilitating multi-level propaganda analysis, three SPARQL queries were formulated, each mapped to the analytical objectives (Strategies A, B, and C).

The first objective (Strategy A) evaluates the identification of propaganda techniques within Donald Trump's inaugural address on the Gaza ceasefire, treating the speech as a single unit. Listing 3 presents the formulated SPARQL query, while Table 4 summarizes the extracted results. The query successfully identified the presence of two techniques: Loaded Language and Bandwagon, which were also identified by human annotators. This confirms the framework's capacity to formalize and capture propaganda strategies at the individual text level.

Table 4
Result of the first analytical objective

text	techniqueComment
PropagandaText_1	Loaded Language
	Bandwagon

The second objective (Strategy B) examines the progression of narrative schemas within the same speech through a paragraph-level analysis. This approach involves segmenting the discourse into its paragraphs to track structural and rhetorical shifts. Listing 4 presents the SPARQL query designed to systematically retrieve the Purpose associated with the underlying narrative schema of each paragraph. The results, summarized in Table 5, reveal a clear rhetorical trajectory. Trump initiates the discourse by highlighting civilian suffering (`Paragraph_1`), thereby capturing the audience's attention and establishing a sense of urgency. He subsequently transitions to context-setting (`Paragraph_2`) to help the audience understanding the conflict's complexity, before communicating them the proposed resolution (`Paragraph_3`), emphasizing the need for diplomatic efforts. Finally, in the fourth paragraph (`Paragraph_4`), Trump calls upon the public to support humanitarian aid and advocate for lasting peace. These findings demonstrate the framework's capacity to model and analyze the strategic layering of political discourse, uncovering structural insights that would be entirely lost if the text were treated as a single unit.

The third objective (Strategy C) focuses on a comparative analysis of narrative schemas employed

[11] https://graphdb.ontotext.com

```
SELECT ?text ?paragraph ?purposeComment
WHERE {
   ?text a PrOntoNarr:PropagandaText ;
       PrOntoNarr:isDeliveredBy ?speaker ;
       PrOntoNarr:hasParagraph ?paragraph .

   ?speaker a PrOntoNarr:Speaker ;
        rdfs:comment "Donald␣Trump" .

   ?paragraph a PrOntoNarr:Paragraph ;
          PrOntoNarr:isTheParagraphOf ?text ;
          PrOntoNarr:hasNarrative ?narrative .

   ?narrative a PrOntoNarr:NarrativeText ;
          PrOntoNarr:isTheNarrativeOf ?paragraph ;
          PrOntoNarr:useSchema ?schema .

   ?schema a PrOntoNarr:NarrativeSchema ;
        PrOntoNarr:isUsedIn ?narrative ;
        PrOntoNarr:containsElement ?purpose .

   ?purpose a PrOntoNarr:Purpose ;
         PrOntoNarr:refersTo ?schema ;
         rdfs:comment ?purposeComment .
}
ORDER BY ?paragraph
```

Listing 4: SPARQL query for the second analytical objective

Table 5
Result of the second analytical objective

text	paragraph	purposeComment
PropagandaText_1	Paragraph_1	Highlight the urgent humanitarian crisis in Gaza, emphasizing the suffering of civilians and the need for immediate international attention.
	Paragraph_2	Explain the recent escalation of hostilities between Israel and Gaza, providing background on the causes and key actors involved.
	Paragraph_3	Announce the agreed ceasefire, describe the terms for ending hostilities, and stress the importance of compliance by all parties.
	Paragraph_4	Urge international leaders, organizations, and citizens to support ongoing humanitarian aid and diplomatic efforts to maintain lasting peace.

across the broader Gaza Ceasefire Campaign by three key political actors: Donald Trump, Benjamin Netanyahu, and Mahmoud Abbas. Listing 5 presents the SPARQL query, while Table 6 illustrates the extracted results, highlighting distinct rhetorical profiles. Trump employs the widest range of schemas, reflecting his diverse communication style and political background. These include *AmericanDream*, that is consistent with his tendency to frame US involvement as a morally imperative mission, and *Hero's Journey* and *UsVsThem*, reflecting his strategy to mobilize support by creating clear distinctions between allies and adversaries. Netanyahu, on the other hand, focuses heavily on *UsVsThem* schema to emphasize a defensive and antagonistic stance toward Hamas and the broader context of Gaza, and on *Hero's Journey* to frame the conflict through the lens of national defense and moral justification. Abbas, compared with the other two leaders, employs fewer narrative schemas. As a Palestinian leader striving to maintain legitimacy and protect his people under difficult circumstances, he primarily relies on the *Redemption* schema, framing the Gaza situation as a historical injustice which require rectification. For this reason, he appeals to both domestic and international audiences in the ceasefire negotiations. These findings validate the framework's capacity to perform macro-political analysis, effectively identifying the specific strategies adopted by each speaker across multiple speeches within a unified campaign.

```
SELECT ?speaker (GROUP_CONCAT(DISTINCT ?schema; SEPARATOR=",␣") AS ?schemasUsed)
WHERE {
   ?PoliticalSpeechText a PrOntoNarr:PropagandaText ;
                 PrOntoNarr:hasNarrative ?NarrativeText ;
                 PrOntoNarr:isComprisedIn ?PropagandisticCampaign ;
                 PrOntoNarr:isDeliveredBy ?speaker .

   ?speaker a PrOntoNarr:Speaker ;
         rdfs:comment ?speakerName .

   FILTER(?speakerName IN ("Donald␣J.␣Trump", "Benjamin␣Netanyahu", "Mahmoud␣Abbas"))

   ?PropagandisticCampaign a PrOntoNarr:PropagandisticCampaign ;
                   PrOntoNarr:hasOverallGoal ?campaignGoal .
   FILTER(CONTAINS(LCASE(?campaignGoal), "Gaza␣ceasefire␣negotiations"))

   ?NarrativeText PrOntoNarr:useSchema ?schema .
   ?schema a PrOntoNarr:NarrativeSchema .
}
GROUP BY ?author
```

Listing 5: SPARQL query for the third analytical objective

Table 6
Result of the third analytical objective

Author	schemasUsed
Donald J. Trump	AmericanDream, HerosJourney, UsVsThem
Benjamin Netanyahu	HerosJourney, ThreeActStructure, UsVsThem, WavePlot
Mahmoud Abbas	Redemption

5. The added value of the proposed framework

The novelty of the proposed framework lies in its integration of formal ontologies, KGs, and LLMs, which effectively overcomes the limitations of traditional approaches to propaganda analysis. Traditional NLP methods and machine learning approaches, particularly those based on token-level classification or embedding, rely on flat, isolated textual relationships [20], [21]. In these models, propaganda is often treated as a sequence of stylistic patterns, frequently overlooking the narrative elements that drives persuasion [22]. In such black-box architectures, semantic associations remain implicit within high-dimensional vector spaces, making it difficult to follow the logical steps of a persuasive argument. In contrast, the proposed framework leverages the PrOntoNarr ontology to provide an explicit representation of knowledge. By moving from a flat representation to this one, unstructured political discourse is transformed into a graph of interconnected entities. This change allows a narrative schema, such as UsVsThem, to evolve from a simple label into a queryable path of logical relationships [23], ensuring full semantic traceability.

Furthermore, KGs address the challenge of document consistency; while traditional methods treat separate paragraphs as disjointed units, KGs explicitly link them, allowing to visualize how a single narrative is reinforced across a text or an entire campaign. This graph structure facilitates the comparison of patterns by using SPARQL queries, such as isolating all agents assigned a specific archetype or identifying the co-occurrence of propaganda techniques, tasks that are computationally unfeasible with flat classifications.

Equally important in the proposed framework is the use of LLM in the KG construction. By leveraging prompt engineering, LLMs can understand propaganda elements and map them to formal RDF triples with high precision. This ensures that the framework is not only theoretically grounded through ontology, but also scalable, enabling the systematic analysis of large propaganda datasets while maintaining explicit and queryable semantic connections.

6. Conclusion and Future Works

Narratives are central to propaganda, as they rely on well-defined structures to shape and influence public opinion. To address the lack of formal methods for analyzing these structures in the propaganda domain, this study introduces a methodological framework centered on the PrOntoNarr ontology. The framework also bridges the gap between text and formal logic by integrating LLMs into the KG construction pipeline, thereby automating the creation of narrative-based KGs. The methodology consists of three phases: i) pre-processing unstructured text, ii) populating the ontology using LLMs, and iii) analyzing propaganda narratives by navigating the resulting KGs. A key feature of this approach is the alignment of analytical objectives with three levels of processing: individual text, paragraph, and campaign. This structure enables researchers to move from broad thematic overviews to fine-grained examinations of propaganda techniques. The framework's effectiveness is validated on a corpus of political speeches from the 2025 Gaza ceasefire negotiations, showing that LLMs can be effectively constrained to populate complex ontologies and make intricate narrative patterns accessible for systematic exploration through SPARQL queries. Despite these promising findings, the study is limited by the relatively small dataset, constrained by the number of speeches available during the negotiation phase. However, the generalizability of the framework is supported by the design of the PrOntoNarr ontology, which is grounded in universal principles of narrative theory and propaganda analysis rather than being tied to a specific domain. As a result, applying the framework to new contexts requires only adapting the contextual information in the LLM prompts, while the ontology itself remains unchanged. Future work will focus on expanding the dataset across diverse geopolitical contexts and incorporating a temporal dimension into the ontology. This extension will enable the formal study of how propaganda narratives adapt, evolve, and intensify over time within digital and political ecosystems.

Acknowledgement

This work was partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

Declaration on Generative AI

During the preparation of this work, the author(s) used Google Gemini in order to: refine the English language, correct grammar, and improve readability. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] G. S. Jowett, V. O'Donnell, Propaganda & persuasion, Sage publications, 2018.
[2] S. Groth, Political narratives/narrations of the political: An introduction, Narrative Culture 6 (2019) 1–18.
[3] E. Damiano, F. Orciuoli, A. Pascuzzo, S. Senatore, Interacting with political narratives through llms: An approach based on ontologies and graph embeddings, in: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2025, pp. 4918–4223.
[4] C. Meghini, V. Bartalesi, D. Metilli, Representing narratives in digital libraries: The narrative ontology, Semantic Web 12 (2021) 241–264.
[5] V. Bartalesi, N. Pratelli, Representing geospatial knowledge in narratives, ACM Journal on Computing and Cultural Heritage 18 (2025) 1–15.
[6] A. F. Khan, A. Bellandi, G. Benotto, F. Frontini, E. Giovannetti, M. Reboul, Leveraging a narrative ontology to query a literary text, in: 7th Workshop on Computational Models of Narrative (CMN 2016), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2016, pp. 10–1.

[7] I. Guillén-Pacho, A. Iglesias-Molina, C. Badenes-Olmedo, Ó. Corcho, Podio: A political discourse ontology., in: EKAW (Companion), 2024.
[8] A. Gaeta, V. Loia, A. Lorusso, F. Orciuoli, A. Pascuzzo, Towards a llm-based intelligent system for detecting propaganda within textual content, Computers and Electrical Engineering 128 (2025) 110765.
[9] I. Traore, A multi-llm agent system for modular ontology population: A case study on adhd, in: International Conference on Agents and Artificial Intelligence (ICAART 2026), 2026.
[10] M. Fernández-López, A. Gómez-Pérez, N. Juristo, Methontology: From ontological art towards ontological engineering, in: Spring Symposium on Ontological Engineering, AAAI, Stanford University, California, 1997, pp. 33–40.
[11] R. Wodak, Complex texts: Analysing, understanding, explaining and interpreting meanings, Discourse studies 13 (2011) 623–633.
[12] G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, P. Nakov, Semeval-2020 task 11: Detection of propaganda techniques in news articles, in: Proceedings of the fourteenth workshop on semantic evaluation, 2020, pp. 1377–1414.
[13] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, Hermit: an owl 2 reasoner, Journal of automated reasoning 53 (2014) 245–269.
[14] A. S. Lippolis, M. J. Saeedizade, R. Keskisärkkä, S. Zuppiroli, M. Ceriani, A. Gangemi, E. Blomqvist, A. G. Nuzzolese, Ontology generation using large language models, in: European Semantic Web Conference, Springer, 2025, pp. 321–341.
[15] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems 35 (2022) 24824–24837.
[16] A. Loizou, R. Angles, P. Groth, On the formulation of performant sparql queries, Journal of Web Semantics 31 (2015) 1–26.
[17] B. Lantow, Ontometrics: Putting metrics into use for ontology evaluation., in: KEOD, 2016, pp. 186–191.
[18] S. S. Norouzi, A. Barua, A. Christou, N. Gautam, A. Eells, P. Hitzler, C. Shimizu, Ontology population using llms, arXiv preprint arXiv:2411.01612 (2024).
[19] J. M. C. de Sá, M. Da Silveira, C. Pruski, Semantic change characterization with llms using rhetorics, arXiv preprint arXiv:2407.16624 (2024).
[20] M. S. I. Malik, T. Imran, J. M. Mamdouh, How to detect propaganda from social media? exploitation of semantic and fine-tuned language models, PeerJ Computer Science 9 (2023) e1248.
[21] V.-A. Oliinyk, V. Vysotska, Y. Burov, K. Mykich, V. Basto-Fernandes, Propaganda detection in text data based on nlp and machine learning., in: MoMLeT+ DS, 2020, pp. 132–144.
[22] P. N. Ahmad, J. Guo, N. M. AboElenein, Q. M. u. Haq, S. Ahmad, A. D. Algarni, A. A. Ateya, Hierarchical graph-based integration network for propaganda detection in textual news articles on social media, Scientific Reports 15 (2025) 1827.
[23] B. Abu-Salih, M. Al-Tawil, I. Aljarah, H. Faris, P. Wongthongtham, K. Y. Chan, A. Beheshti, Relational learning analysis of social politics using knowledge graph embedding, Data Mining and Knowledge Discovery 35 (2021) 1497–1536.