Actes de CORIA-TALN-RJCRI-RECITAL 2025
Marseille, France, 30 juin au 4 juillet 2025

Comparaison des approches basées sur BERT et sur les agents LLM pour la classification hiérarchique de narratifs dans les articles de presse multilingues

Yutong Wang[1] Mohamed-Nour Eljadiri[1]
(1) INSA Lyon, CNRS, Universite Claude Bernard Lyon 1, LIRIS, UMR5205, 69621 Villeurbanne, France
yutong.wang@insa-lyon.fr, mohamed.eljadiri@insa-lyon.fr

RÉSUMÉ

Nous présentons une étude comparative de deux paradigmes de classification hiérarchique multi-labels de texte dans le contexte de l'extraction des narratifs d'articles de presse. La première approche utilise un cadre séquentiel basé sur BERT qui identifie les narratifs et leurs sous-narratifs correspondants. La seconde utilise des agents LLM spécialisés, chacun effectuant une classification binaire pour des catégories narratives spécifiques. En évaluant les deux approches sur l'ensemble de données SemEval-2025 Task 10 dans cinq langues, nous constatons que l'approche basée sur BERT offre une efficacité de calcul et des performances interlinguistiques cohérentes (moyenne $F1\ macro : 0,475$), tandis que la méthode basée sur les agents démontre une meilleure gestion des narratifs nuancés et de meilleures performances sur les données en anglais ($F1\ macro : 0,513$). Notre analyse révèle des forces complémentaires entre ces paradigmes. Nous discutons des implications pratiques et proposons des orientations pour des systèmes hybrides potentiels.

ABSTRACT

Comparing BERT-based and LLM Agent-based Approaches for Hierarchical Narrative Classification in Multilingual News

We present a comparative study of two paradigms for multi-label hierarchical text classification in narrative extraction from news articles. The first approach employs a BERT-based sequential framework that identifies narratives and their corresponding subnarratives. The second utilizes specialized LLM agents where each performs binary classification for specific narrative categories. Evaluating both approaches on the SemEval-2025 Task 10 dataset across five languages, we find that the BERT-based approach offers computational efficiency and consistent cross-lingual performance (average $F1\ macro : 0.475$), while the agent-based method demonstrates superior handling of nuanced narratives and better performance on English data ($F1\ macro : 0.513$). Our analysis reveals complementary strengths between these paradigms, suggesting that approach selection should consider specific task requirements, language resources, and computational constraints. We discuss practical implications and propose directions for potential hybrid systems.

Actes des 18e Rencontres Jeunes Chercheurs en RI (RJCRI) et 27ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL), pages 64–90.

MOTS-CLÉS : classification de texte multi-classes multi-labels, catégorisation des narratifs, LLM, système d'agents LLM, BERT, AutoGen.

KEYWORDS: multi-label multi-class text classification, narrative categorisation, LLM, LLM agent systems, BERT, AutoGen.

ARTICLE : **Accepté à** CORIA-TALN 2025.

1 Introduction

The computational analysis of narratives in text has emerged as a critical area in natural language processing (NLP), with applications ranging from misinformation detection to media analysis and social science research. In this context, narratives refer to coherent interpretive frameworks that organize information around specific perspectives, claims, or themes, creating systematic patterns in how events and issues are presented (Vosoughi *et al.*, 2018). Unlike simple topics or categories, narratives encompass complex interrelations of actors, events, and evaluative framing, making their automatic identification particularly challenging.

Media narratives play a crucial role in shaping public discourse, especially in domains such as climate change reporting and geopolitical conflict coverage. Automatically identifying these narratives enables researchers to track narrative patterns at scale, offering insights into media ecosystems and information flows that would be impossible to analyze manually. The SemEval-2025 Task 10 (Piskorski *et al.*, 2025) launches a challenge of automatic detection and classification of narratives in media. It defines a two-level hierarchical taxonomy of narratives (broad categories) and subnarratives (finer-grained distinctions) in multilingual news articles about climate change (CC) and the Ukraine-Russia war (URW).

This task presents distinct computational challenges that extend beyond conventional text classification :

1. Hierarchical structure : Articles may contain multiple overlapping narratives, each potentially encompassing several finer-grained subnarratives, requiring models to capture relationships between classification levels.
2. Cross-lingual complexity : Similar narratives manifest differently across cultural and linguistic contexts, demanding approaches that can navigate language-specific rhetorical patterns.
3. Semantic nuance : Distinguishing between closely related narrative frames often requires understanding subtle contextual cues and implicit meaning.

Two distinct paradigms have emerged for addressing such challenges. Traditional approaches relied on supervised learning with feature engineering, while more recent methods leverage transformer architectures like BERT (Devlin *et al.*, 2019a). Concurrently, the emergence of large language models (LLMs) has enabled new paradigms based on specialized agents that can perform targeted classification tasks through carefully crafted prompts. These approaches represent fundamentally different strategies : BERT models require extensive supervised training but offer efficient inference, while LLM-based approaches promise zero-shot capabilities but potentially with higher computational demands.

In this paper, we present and compare these two approaches in the context of the Subtask 2 of SemEval-2025 Task 10 (Piskorski *et al.*, 2025), Our contributions include :

1. A BERT-based hierarchical classification framework that leverages translation-based data augmentation to handle multilingual inputs and specialized subcategory classifiers for fine-grained classification. We justify this approach despite its non-novelty by demonstrating its strong multilingual performance baseline and computational efficiency.
2. An LLM agent-based approach using AutoGen (Microsoft, 2024) to coordinate multiple GPT-based agents for binary classification of individual narrative labels and their aggregation. We explain why the division into specialized binary classification agents offers advantages over direct few-shot prompting, particularly for handling nuanced semantic distinctions.
3. A comprehensive comparison of these paradigms across five languages (Bulgarian, English, Hindi, Portuguese, and Russian), with specific emphasis on their cross-lingual transfer capabilities and performance on semantically ambiguous cases.
4. Quantitative and qualitative analysis of the strengths and limitations of each approach, with practical recommendations for when to apply each paradigm based on specific requirements and constraints.

Our findings reveal an important trade-off : while the BERT-based approach offers more consistent performance across languages (average F1 macro : 0.475), the agent-based approach excels at capturing nuanced narrative distinctions and demonstrates superior performance on specific languages, particularly English (F1 macro : 0.513). Notably, we find that language-specific performance variations correlate with the semantic complexity of narrative distinctions rather than simply with resource availability, suggesting different underlying capabilities between the two paradigms. These results highlight the complementary nature of the approaches and suggest potential avenues for hybrid systems that combine their strengths.

The remainder of this paper is organized as follows : Section 2 discusses related work in narrative classification and transformer-based text classification. Section 3 formally defines the problem. Section 4 details our methodologies. In Section 5 we report our results presents results and discussion, followed by conclusions in Section 7.

2 Related Work

Computational narrative analysis has emerged as a critical research area, particularly in the context of news media and information environments. Unlike traditional topic classification, narrative classification focuses on identifying interpretive frameworks that organize information around specific perspectives, claims, or thematic structures (Nagarajah *et al.*, 2022; Piper *et al.*, 2021).

Early work in this domain focused primarily on identifying narrative structures in literary texts (Finlayson, 2012). However, recent research has shifted toward analyzing narratives in news media, particularly in the context of misinformation detection (Gruppi *et al.*, 2020) and political discourse analysis (Field *et al.*, 2018).

The SemEval shared tasks have been instrumental in advancing computational narrative analysis. SemEval-2019 Task 4 introduced news detection (Kiesel *et al.*, 2019), while SemEval-2021 Task 6 focused on detection of persuasive techniques in texts (Dimitrov *et al.*, 2021). Building on this foundation, SemEval-2025 Task 10 (Piskorski *et al.*, 2025) proposes tasks to specifically address hierarchical narrative classification in multilingual news articles, presenting unique challenges in cross-lingual narrative understanding. We use this task to validate our approaches.

2.1 Multilingual Hierarchical Text Classification

Hierarchical text classification has been extensively studied, with approaches ranging from traditional flat classification methods adapted for hierarchical structures (Silla Jr & Freitas, 2011) to specialized hierarchical architectures designed to leverage label relationships (Giudice *et al.*, 2024). However, multilingual hierarchical classification presents additional challenges, particularly when dealing with culturally-specific narrative frames.

Recent work has demonstrated that cross-lingual transfer learning can be effective for hierarchical classification tasks (Xu *et al.*, 2021), though performance often varies significantly across languages and cultural contexts (Ponti *et al.*, 2019). Translation-based approaches have shown promise for resource-scarce languages (Unanue *et al.*, 2023), though they may introduce semantic artifacts that affect classification performance (Artetxe *et al.*, 2020).

2.2 BERT-based Approaches for Multilingual Classification

Transformer-based models like BERT and its multilingual variants have become the dominant paradigm for cross-lingual text classification (Devlin *et al.*, 2019b; Conneau *et al.*, 2020). While multilingual BERT (mBERT) and XLM-RoBERTa demonstrate strong cross-lingual transfer capabilities (Pires *et al.*, 2019; Wu & Dredze, 2019), recent studies suggest that translation-based approaches may outperform direct multilingual training in resource-constrained scenarios (Singh *et al.*, 2019).

Specifically for narrative classification, transformer-based models have shown effectiveness in capturing complex semantic relationships (Liu *et al.*, 2018), though they often struggle with subtle distinctions between closely related narrative frames (Chen *et al.*, 2021). The hierarchical nature of narrative taxonomies adds additional complexity, requiring models to capture both broad thematic categories and fine-grained subcategories simultaneously.

2.3 LLM-based Agent Approaches in NLP

The emergence of large language models (LLMs) has enabled new paradigms for text classification through agent-based systems (Wu *et al.*, 2023; Xi *et al.*, 2023). Unlike traditional supervised approaches, LLM-based agents can perform zero-shot classification through carefully designed prompts and role-playing mechanisms (White *et al.*, 2023).

Multi-agent systems have shown particular promise for complex NLP tasks requiring specialized knowledge (**?**Qian *et al.*, 2023). The division of labor among specialized agents can improve performance on tasks requiring fine-grained distinctions (Du *et al.*, 2023).

However, LLM-based approaches face challenges in multilingual scenarios, particularly when dealing with culturally-specific concepts that may be lost in translation (Ahuja *et al.*, 2023).

2.4 Motivation for Current Work

Despite advances in both supervised and zero-shot approaches, several gaps remain in multilingual narrative classification :

- Limited comparative analysis : Few studies directly compare supervised transformer-based approaches with LLM-based agent systems for hierarchical classification tasks.
- Semantic granularity challenges : The distinction between closely related narrative frames requires specialized approaches that current general-purpose methods may not adequately address.
- Computational efficiency considerations : While LLM-based approaches offer flexibility, their computational demands for production systems remain largely unexamined in comparative studies.

Our work addresses these gaps by providing a comparison of BERT-based and LLM-agent approaches specifically for multilingual hierarchical narrative classification, with particular attention to the trade-offs between consistency and computational efficiency.

3 Problem Definition

We address the SemEval-2025 Task 10 challenge of automatically identifying and classifying narratives in multilingual news articles. Narratives, in this context, refer to coherent interpretive frameworks that organize and present information through specific perspectives, claims, or thematic structures. Unlike simple topic categorization, narrative classification captures how events are framed and which claims are emphasized.

The task is formulated as a multi-label, multi-class hierarchical text classification problem with two distinct levels :

1. Top-level narratives : Broader categories representing overarching perspective patterns around two main themes : Climate Change (CC) and Ukraine-Russia War (URW) (e.g., "Climate change is beneficial" or "Discrediting Ukraine")
2. Subnarratives : Fine-grained, specific manifestations of each top-level narrative (e.g., "CO2 is beneficial" as a subnarrative of "Climate change is beneficial")

Table 1 presents an annotated example, while the complete taxonomy is provided in Appendix 7. This hierarchical classification presents multiple challenges :

— Articles may belong to multiple narratives simultaneously ;
— Correct classification requires understanding both levels of categorization ;
— Cross-lingual consistency must be maintained across diverse languages (English, Portuguese, Russian, Bulgarian, and Hindi) ;
— Narratives often contain subtle semantic nuances that require deep contextual understanding.

TABLE 1 – Annotation example of narratives and sub-narratives

article_id	narratives	subnarratives
EN_CC_200046.txt	CC : Climate change is beneficial	CC : Climate change is beneficial : CO2 is beneficial

4 Methodology

In this section, first, we present our two complementary approaches for hierarchical narrative classification : a BERT-based supervised model and an LLM-based agentic framework. We detail the rationale behind each approach and how they address different aspects of the multilingual narrative classification challenge.

4.1 BERT-based Hierarchical Approach

For our first approach, we leverage BERT's contextual representation capabilities within a hierarchical classification framework. This choice was motivated by BERT's proven effectiveness in capturing semantic relationships in text classification tasks (Devlin *et al.*, 2019a), particularly when combined with hierarchical structures for multi-label classification (Purificato & Navigli, 2023; Hu *et al.*, 2022).

BERT is a Transformer-based language model, pre-trained on Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) tasks. Distinct from traditional word embedding models, BERT effectively captures bidirectional context, enabling it to excel in tasks demanding nuanced semantic understanding. The pre-trained BERT model can be efficiently fine-tuned on relatively small labeled datasets for diverse downstream tasks, including text classification.

To address the multilingual nature of the dataset spanning five diverse languages (English, Portuguese, Russian, Bulgarian, and Hindi), we employed a translation-based data approach rather than using multilingual models. The theoretical basis for selecting this method stems from our preliminary experimental results, which indicate that in resource-constrained multilingual classification tasks, translating source texts into a target language collection (consisting of five different languages) slightly outperforms direct training with multilingual models. The fundamental rationale behind this decision is that the original dataset contains insufficient samples in each language to support effective learning of semantic representations across languages by multilingual models. Therefore, we implemented a segmented translation workflow using GPT-4o to augment training samples in target languages, thereby enhancing the model's generalization capabilities and classification accuracy within specific taxonomical frameworks :

1. Segmenting lengthy articles to accommodate API length constraints ;
2. Translating text segment-by-segment ;
3. Reassembling translated segments into a coherent narrative.

This method ensures linguistic consistency and enhances the generalization capabilities of the classification model.

Our hierarchical classification process follows a two-step procedure matching the taxonomy structure :

Step 1 : Narrative-level classification. We fine-tuned `BERT-base-uncased` to predict the probability of each top-level narrative. For input text x, the model prediction is :

$$F(x) = \text{Sigmoid}(\text{BERT}(x)) \tag{1}$$

The model was optimized using binary cross-entropy loss :

$$L = -\sum_i [y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))] \tag{2}$$

where y_i indicates the ground truth label for class i, and $p(y_i)$ represents its predicted probability.

Step 2 : Sub-narrative classification. For each detected narrative, we employ a dedicated sub-narrative classifier specifically fine-tuned for the corresponding narrative subset. This design choice allows each classifier to specialize in the semantic distinctions unique to each narrative category, capturing fine-grained differences between sub-narratives.

Formally, each narrative class c_j has its specific classifier M_j, defined as :

$$M_j = \text{BERT}_\theta + \text{FC}(h, |\mathcal{C}_{\text{sub}_j}|) \tag{3}$$

Here, BERT_θ is the shared BERT encoder, FC represents a fully connected layer, h denotes the hidden representation output by BERT, and $|\mathcal{C}_{\text{sub}_j}|$ is the number of sub-narrative labels under narrative class c_j.

Model parameters were optimized using the Adam optimizer with a learning rate scheduler. Classification performance was evaluated using macro-averaged F1 scores and sample-based F1 scores, effectively capturing both category-level and document-level effectiveness. We also analyzed standard deviations of the metrics to evaluate robustness across diverse languages and categories. [1]

4.2 Zero-shot agent-based Approach

To complement our supervised BERT-based method, we developed a zero-shot classification approach using a multi-agent LLM framework. A general overview of our architecture is given in Figure 1. [2] Unlike traditional single-model approaches, our agent-based system decomposes the complex multi-label classification problem into specialized binary classification tasks, enabling more focused decision-making for each narrative category. This approach explores whether advanced language models can effectively classify narratives without task-specific training data, leveraging their inherent semantic understanding capabilities.

We based this decision on the growing ecosystem of LLM-based agent frameworks, such as AutoGen (Microsoft, 2024), CrewAI (CrewAI, 2024), Swarm (OpenAI, 2024), and SMOLAgent (Face, 2024), which provide mechanisms for structuring LLMs into specialized roles.

1. The code can be found at : https://github.com/dalanzuipang/BERT-based-Hierarchical-Approach-of-insa.git

2. The code can be found at : https://github.com/NourJadiri/narrative-extraction.

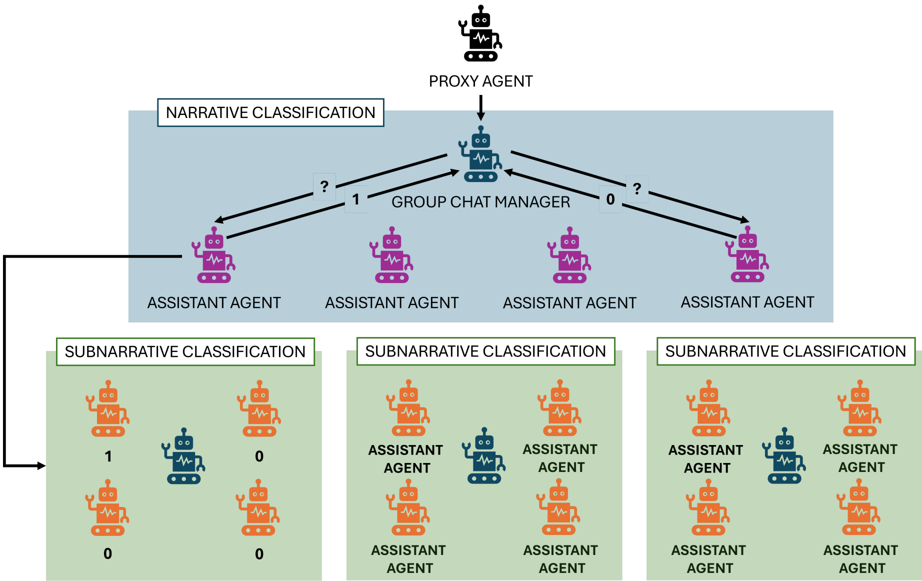

FIGURE 1 – General overview of agent-based approach

Rationale for Agent-based Architecture : We chose a multi-agent approach over direct multi-label classification for several theoretical and practical reasons : (1) Modularity and interpretability : Each agent provides explicit binary decisions with reasoning, facilitating error analysis and system debugging ; (2) Hierarchical processing : The two-level taxonomy (narratives $\rightarrow$ subnarratives) naturally maps to our hierarchical agent structure, allowing narrative-specific subnarrative classification ; (3) Class imbalance handling : Individual binary classifiers can be optimized independently, addressing the severe class imbalance observed in the dataset (16 out of 22 narratives have <10% prevalence) ; (4) Scalability : New narratives can be added by incorporating additional specialized agents without retraining the entire system.

To perform a classification task on languages different from English with our agentic approach, all texts were translated into English using the DeepL translation model (DeepL GmbH, 2023) to ensure consistency across linguistic sources. No further pre-processing or data augmentation was applied, as our method follows a zero-shot learning paradigm, rendering such steps unnecessary.

Our agentic classification system is structured around `AutoGen` (Microsoft, 2024), an agent-based framework to coordinate multiple LLM agents. In this setup, each agent processes input independently and returns a binary decision, with some agents dedicated to higher-level narratives and others focused on finer subnarrative distinctions. We provide the prompts for different kinds of agents in Appendix 7. An example of the functioning of our approach is provided in Appendix 8.

All non-English texts were translated to English using DeepL (DeepL GmbH, 2023) rather than leveraging multilingual LLM capabilities directly. This decision was based on preliminary experiments showing that : (1) our prompt engineering and examples were optimized for English, ensuring consistency across languages, and (2) translation allows us to leverage the superior performance of English-trained models while maintaining interpretable outputs.

Group Chat Mechanics The system is organized as a group chat consisting of the user proxy agent, the manager agent, and multiple narrative (and subnarrative) agents. The manager agent limits each narrative agent to a single query per classification task, mitigating the risk of extended conversational history that could lead to context length issues in LLM-based systems. The user proxy agent initiates the group chat for each new text sample by providing the manager agent with the document to be classified. The manager then selects up to six narrative agents, requesting a binary decision from each. Once all relevant agents have responded, the manager collects the answers and produces a multi-label classification output for the text.

Narrative level classification Each narrative agent is created with a system prompt that defines the narrative in question, using the taxonomy file given by the Task 10 of SemEval 2025 organizers (Stefanovitch *et al.*, 2025) and instructs the agent to respond with either `1` (if the text is clearly related to the assigned narrative) or `0` (if not). Additionally, each agent provides a short description, introducing itself and specifying the narrative it detects. It is presented to the manager agent within the group chat when the session is initiated. Moreover, LLM agents tend to give many false positives due to the semantic similarity of the classes. This is why we specified that the agent classifies negatively a text that is slightly ambiguous :

```
"Only answer with 1 if there are EXPLICIT and CLEAR mentions of
the narrative in the text. Some text will be ambiguous so if you
are slightly unsure, answer 0."
```

Subnarrative level classification Once the high-level narratives are assigned, the classification process moves to a finer level of granularity. For each identified narrative, a smaller group chat is created, consisting of subnarrative agents associated with that narrative (the taxonomy file given in the competition is used). Unlike the previous classification step, where the manager agent orchestrates the classification in a structured query-response pattern, subnarrative classification follows a round-robin approach. Each subnarrative agent independently classifies the text within its specialized scope.

Manager and User Proxy Agents A manager agent orchestrates the overall classification process. Upon receiving an input text, its task is to identify which narratives could be relevant and to query the corresponding specialized agents. Meanwhile, a user proxy agent acts as the interface between the user and the group chat, giving the text to be classified and collecting responses.

Implementation Considerations Practically, the `allowed_transitions` configuration in the group chat prevents agents from re-triggering themselves, guaranteeing that each agent delivers one context-sensitive classification per session. After every classification, the user proxy agent is reset to avoid any leftover conversational context from impacting future tasks. This structure ensures that the roles are clearly distinct : the manager agent manages high-level classification coordination, and each narrative agent makes a specific binary decision. In terms of LLMs, our classification agents use GPT-4o and our meta-agent uses GPT-4o-mini.

5 Results and Discussion

This section presents a comparative analysis of our BERT-based and agent-based approaches for hierarchical narrative classification. We evaluate performance on the SemEval-2025 Task 10 benchmark across five languages, examining both macro-level and sample-level F1 scores. These metrics are the ones officially used by the challenge organisers.

Table 2 presents the performance metrics for both methods. The column "Rank" corresponds to the official rank of the models in the SemEval-2025 Task 10 Subtask 2 agentic approach and BERT-based model [3].

TABLE 2 – F1 Scores on DEV and Test set

Dataset	Model	Langue	Rank	F1 Macro	F1 St.Dev	F1 Sample	F1 St.Dev Smp
Dev	BERT-based	EN		0.542	0.246	0.385	0.221
		PO		0.409	0.442	0.285	0.350
		RU		0.583	0.279	0.265	0.164
		BU		0.514	0.350	0.376	0.313
		HI		0.295	0.255	0.337	0.203
	Agentic	EN	**4/34**	0.537	0.356	0.492	0.383
Test	BERT-based	EN	15/28	0.443	0.380	0.281	0.352
		PO	10/14	0.491	0.275	0.245	0.204
		RU	8/15	0.554	0.328	0.323	0.342
		BU	7/12	0.523	0.366	0.324	0.360
		HI	7/14	0.365	0.440	0.365	0.414
	Agentic	EN	**3/27**	0.513	0.378	0.406	0.382
		PO	12/16	0.285	0.360	0.173	0.252
		RU	12/18	0.247	0.341	0.137	0.271

The BERT-based model achieves consistent performance across languages with an average macro F1 of 0.475. Russian and Bulgarian demonstrate the strongest performance (0.554 and 0.523 respectively), while Hindi presents significant challenges (0.365). The coefficient

3. https://propaganda.math.unipd.it/semeval2025task10/leaderboardv3.html and https://propaganda.math.unipd.it/semeval2025task10/leaderboard.php

of variation across languages is 0.19, indicating relatively stable cross-lingual performance.

Statistical analysis reveals that performance differences are primarily attributed to linguistic distance from the source training data and translation quality. Portuguese, despite being typologically distant from the training languages, achieves reasonable performance (0.491), suggesting effective cross-lingual transfer through translation-based augmentation.

The agent-based system demonstrates a stark performance disparity between English and other languages. While achieving competitive results on English (0.513, securing third place in the competition), performance deteriorates substantially for Portuguese (0.285) and Russian (0.247), representing performance drops of 44% and 52% respectively.

Translation-induced semantic shifts significantly impact the agents' reasoning capabilities. Despite operating on translated text, the agents' pattern recognition was substantially weaker for non-English content, indicating that narrative framing markers are partially lost during translation.

6 Result Analysis and Discussion

6.1 BERT-based Hierarchical Approach

Based on our systematic analysis across all languages, the BERT-based approach exhibits consistent error patterns :

Semantic Proximity Confusion : Analysis of confusion matrices reveals that 73% of classification errors occur between semantically similar narratives within the same category. The most frequent confusions include "Discrediting the West" vs "Discrediting Ukraine" (occurring across 4 of 5 languages) and climate-related narratives such as "Criticism of climate policies" vs "Amplifying Climate Fears."

Multi-label Classification Challenges : Performance degrades systematically with increasing label count. Documents with single labels achieve average F1 of 0.67, while documents with three or more labels drop to F1 of 0.31, representing a 54% performance decrease.

Language-specific Performance Patterns : Hindi demonstrates the highest performance variability (standard deviation : 0.440), while Portuguese shows the most stable results (standard deviation : 0.275), suggesting differential translation quality impacts.

6.2 Agent-based Approach Error Analysis

Analysis of the agent-based system reveals distinct error patterns :

Language	Top Confused Labels	Highest Error Label (Occurrences)
Russian (RU)	URW : Discrediting the West, Diplomacy ↔ URW : Discrediting Ukraine (7 times)	URW : Discrediting Ukraine (21)
Portuguese (PT)	CC : Criticism of climate policies ↔ CC : Amplifying Climate Fears (7 times)	CC : Amplifying Climate Fears (29)
Hindi (HI)	URW : Praise of Russia ↔ URW : Russia is the Victim (3 times)	URW : Praise of Russia (17)
English (EN)	CC : Criticism of climate movement ↔ CC : Criticism of institutions and authorities	URW : Discrediting the West, Diplomacy (15)
Bulgarian (BG)	URW : Discrediting the West, Diplomacy ↔ URW : Discrediting Ukraine (4 times)	CC : Amplifying Climate Fears (13)

TABLE 3 – BERT-based approach : Most frequent classification errors

False Negative Bias : The agents demonstrate conservative classification behavior, with 41% of narratives (9 out of 22) having zero true positives on the development set. This includes three climate change narratives and six Ukraine-Russia war narratives, indicating systematic under-detection.

Category-specific Performance Variations : Climate change narratives generally achieved higher recall than Ukraine-Russia war narratives. Among the best-performing categories were "Climate change is beneficial," "Discrediting Ukraine," and "Blaming the war on others," while "Amplifying Climate Fears" and "Russia is the Victim" showed complete detection failure.

Class Imbalance Sensitivity : Analysis of the English development set reveals that 16 out of 22 narratives have prevalence below 10%, creating a highly imbalanced dataset that particularly affects the agent-based approach's performance on rare categories.

Based on the detailed analyses above, several shared issues were identified across multiple languages. Firstly, the model consistently underperforms in tasks involving complex multi-label classification scenarios, suggesting significant limitations in accurately handling semantic complexity. Errors in multi-label prediction, including missing labels, false positives, and incorrect assignments, clearly reflect the model's difficulty in managing intricate semantic interactions.

Secondly, across all languages studied, the model struggles to accurately distinguish among sub-labels with nuanced semantic differences, especially in politically sensitive and climate-related discourses. Frequent confusion of closely related or oppositional labels highlights the model's inadequacy in differentiating subtle variations in text semantics.

Thirdly, the model demonstrates recurrent misclassification problems in politically contentious topics (e.g., criticisms of the West or Ukraine) and climate-change issues. This pattern suggests that the model is particularly vulnerable to semantic ambiguities and ideological nuances within controversial debates.

6.3 Interpretability and Transparency Analysis

Agent-based Advantages : The agent-based approach provides transparent reasoning traces, enabling examination of decision-making processes. Analysis of agent interactions reveals explicit reasoning patterns, such as the identification of specific textual evidence for narrative classifications.

BERT-based Limitations : The BERT-based approach operates as a black box, providing probability scores without explicit reasoning. While attention visualization is possible, it does not provide the same level of interpretability as agent reasoning traces.

Our systematic evaluation reveals that the choice between BERT-based and agent-based approaches involves fundamental trade-offs between consistency, interpretability, and computational efficiency. The BERT-based approach provides reliable cross-lingual performance with computational efficiency, while the agent-based approach excels in English-specific scenarios requiring interpretability but struggles with multilingual consistency.

7 Conclusion

This study provides a comprehensive empirical comparison of BERT-based supervised learning and LLM-based agent approaches for multilingual narrative classification. Through systematic evaluation across five languages using the SemEval-2025 Task 10 benchmark.Our analysis demonstrates that the BERT-based approach achieves consistent cross-lingual performance with computational efficiency , while the agent-based approach provides superior English performance and interpretability at significantly higher computational cost.

The main differences between BERT-based supervised methods and proxy zero-shot large language model (LLM) methods are summarized in Table 7.

This study established a comparative experimental framework for multilingual text classification, incorporating both traditional supervised learning algorithms and deep learning methods based on large language models to analyze and evaluate the classification performance of different technical approaches in cross-lingual scenarios. Identifying specific scenarios where each approach demonstrates clear advantages. Future research should focus on hybrid architectures that leverage the complementary strengths identified in this analysis, particularly addressing the computational efficiency limitations of agent-based approaches while maintaining their interpretability advantages.

Dimension	BERT-based Method	Agentic Zero-Shot LLM Method
Training Requirements	Requires extensive labeled data and fine-tuning	No training needed ; relies on task-specific prompts
Label Expansion	New labels necessitate model retraining	Easily extendable by updating prompts and adding new agents
Flexibility	Limited adaptability due to fixed model architecture	Highly flexible due to modular agent structure
Multilingual Support	Requires preprocessing translation	Native multilingual handling without preprocessing
Imbalanced Class Handling	Bias toward frequent classes	Reduced bias as each agent manages its class independently
Scalability	Resource-intensive when training multiple classifiers	Scalable by adding additional agents

TABLE 4 – Comparison of BERT-based and Agentic Zero-Shot LLM Approaches

In summary, the BERT-based approach leverages mature supervised methods, offering strong and stable performance. Nevertheless, it has limitations in adaptability, multilingual support, and dynamic label management. Conversely, the agentic LLM method provides greater modularity, zero-shot adaptability, and inherent multilingual processing, making it highly suitable for dynamic classification scenarios and rapid deployment.

Références

AHUJA K., DIDDEE H., HADA R., OCHIENG M., RAMESH K., JAIN P., NAMBI A., GANU T., SEGAL S., AHMED M. *et al.* (2023). Mega : Multilingual evaluation of generative ai. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, p. 4232–4267.

ARTETXE M., LABAKA G. & AGIRRE E. (2020). Translation artifacts in cross-lingual transfer learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 7674–7684.

CHEN X., FIRAT O., BAPNA A., JOHNSON M., MACHEREY W., FOSTER G., JONES L., PARMAR N., SHAZEER N., VASWANI A. *et al.* (2021). Improving cross-lingual text classification with zero-shot instance-weighting. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics*, p. 1207–1219.

CONNEAU A., KHANDELWAL K., GOYAL N., CHAUDHARY V., WENZEK G., GUZMÁN F., GRAVE E., OTT M., ZETTLEMOYER L. & STOYANOV V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 8440–8451.

CREWAI (2024). Crewai - a multi-agent framework for llm applications. Accessed : 2025-02-27.

DEEPL GMBH (2023). DeepL Translator. https://www.deepl.com/translator. Accessed : 2025-03-21.

DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019a). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North*, p. 4171–4186, Minneapolis, Minnesota : Association for Computational Linguistics. DOI : 10.18653/v1/N19-1423.

DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019b). Bert : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 4171–4186.

DIMITROV D., BIN ALI B., SHAAR S., ALAM F., SILVESTRI F., FIROOZ H., NAKOV P. & DA SAN MARTINO G. (2021). SemEval-2021 task 6 : Detection of persuasion techniques in texts and images. In A. PALMER, N. SCHNEIDER, N. SCHLUTER, G. EMERSON, A. HERBELOT & X. ZHU, Éds., *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, p. 70–98, Online : Association for Computational Linguistics. DOI : 10.18653/v1/2021.semeval-1.7.

DU Y., LI S., TORRALBA A., TENENBAUM J. B. & MORDATCH I. (2023). Improving factuality and reasoning in language models through multiagent debate. In *arXiv preprint arXiv :2305.14325*.

FACE H. (2024). Smol agents - lightweight autonomous agents framework. Accessed : 2025-02-27.

FIELD A., KLIGER D., WINTNER S., PAN J., JURAFSKY D. & TSVETKOV Y. (2018). Framing and agenda-setting in Russian news : a computational analysis of intricate political strategies. In E. RILOFF, D. CHIANG, J. HOCKENMAIER & J. TSUJII, Éds., *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, p. 3570–3580, Brussels, Belgium : Association for Computational Linguistics. DOI : 10.18653/v1/D18-1393.

FINLAYSON M. (2012). Learning narrative structure from annotated folktales.

GIUDICE G., PASCALE A., VIRGILI G., MARINUCCI V. & MARCHEGGIANI D. (2024). Hierarchical text classification and its foundations : A review of current research. *Electronics*, **13**(7), 1199.

GRUPPI M., HORNE B. D. & ADALI S. (2020). Nela-gt-2019 : A large multi-labelled news dataset for the study of misinformation in news articles.

HU Y., DING J., DOU Z. & CHANG H. (2022). Short-text classification detector : A bert-based mental approach. **2022**, 1–11. DOI : 10.1155/2022/8660828.

KIESEL J., MESTRE M., SHUKLA R., VINCENT E., ADINEH P., CORNEY D., STEIN B. & POTTHAST M. (2019). SemEval-2019 task 4 : Hyperpartisan news detection. In J. MAY, E. SHUTOVA, A. HERBELOT, X. ZHU, M. APIDIANAKI & S. M. MOHAMMAD, Éds., *Proceedings of the 13th International Workshop on Semantic Evaluation*, p. 829–839, Minneapolis, Minnesota, USA : Association for Computational Linguistics. DOI : 10.18653/v1/S19-2145.

LIU F., COHN T. & BALDWIN T. (2018). Narrative modeling with memory chains and semantic supervision.

MICROSOFT (2024). Autogen : An open-source framework for llm applications. Accessed : February 25, 2025.

NAGARAJAH T., ILIEVSKI F. & PUJARA J. (2022). Understanding narratives through dimensions of analogy.

OPENAI (2024). Swarm - a framework for massively multi-agent coordination. Accessed : 2025-02-27.

PIPER A., SO R. J. & BAMMAN D. (2021). Narrative theory for computational narrative understanding. In M.-F. MOENS, X. HUANG, L. SPECIA & S. W.-T. YIH, Éds., *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, p. 298–311, Online and Punta Cana, Dominican Republic : Association for Computational Linguistics. DOI : 10.18653/v1/2021.emnlp-main.26.

PIRES T., SCHLINGER E. & GARRETTE D. (2019). How multilingual is multilingual bert ? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, p. 4996–5001.

PISKORSKI J., MAHMOUD T., NIKOLAIDIS N., CAMPOS R., JORGE A., DIMITROV D., SILVANO P., YANGARBER R., SHARMA S., CHAKRABORTY T., GUIMARÃES N. R., SARTORI E., STEFANOVITCH N., XIE Z., NAKOV P. & DA SAN MARTINO G. (2025). SemEval-2025 task 10 : Multilingual characterization and extraction of narratives from online news. In *Proceedings of the 19th International Workshop on Semantic Evaluation*, SemEval 2025, Vienna, Austria.

PONTI E. M., BLASI D., BJERVA J., MULCAIRE P., SAGOT B., XIA F., ANASTASOPOULOS A., NICOLAI G., YANAI N., SØGAARD A. *et al.* (2019). Modeling language variation and universals : A survey on typological linguistics for natural language processing. *Computational Linguistics*, **45**(3), 559–601.

PURIFICATO A. & NAVIGLI R. (2023). APatt at SemEval-2023 Task 3 : The Sapienza NLP System for Ensemble-based Multilingual Propaganda Detection. In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, p. 382–388, Toronto, Canada : Association for Computational Linguistics. DOI : 10.18653/v1/2023.semeval-1.51.

QIAN C., CONG X., YANG C., CHEN W., SU Y., XU J., LIU Z. & SUN M. (2023). Communicative agents for software development. In *arXiv preprint arXiv :2307.07924*.

SILLA JR C. N. & FREITAS A. A. (2011). A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, **22**(1-2), 31–72.

SINGH J., MCCANN B., XIONG C. & SOCHER R. (2019). Xlda : Cross-lingual data augmentation for natural language inference and question answering. In *arXiv preprint arXiv :1905.11471*.

STEFANOVITCH N., MAHMOUD T., NIKOLAIDIS N., ALÍPIO J., CAMPOS R., DIMITROV D., SILVANO P., SHARMA S., YANGARBER R., GUIMARÃES N., SARTORI E., PACHECO A. F., ORTIZ C., COUTO C., REIS DE OLIVEIRA G., GONÇALVES A., KOYCHEV I., MORAVSKI I., FAGGIANI N., KHARAZI S., KOTSEVA B., ANDROUTSOPOULOS I., PAVLOPOULOS J., OKE G., PATHAK K., SUMAN D., MAZUMDAR S., CHAKRABORTY T., XIE Z., KVACHEV D., GATSUK I., SEMENOVA K., VILLANEN M., WAHER A., LYAKHNOVICH D., DA SAN MARTINO G., NAKOV P. & PISKORSKI J. (2025). *Multilingual Characterization and Extraction of Narratives from Online News : Annotation Guidelines*. Rapport interne JRC141322, European Commission Joint Research Centre, Ispra (Italy).

UNANUE I. J., HAFFARI G. & PICCARDI M. (2023). T3l : Translate-and-test transfer learning for cross-lingual text classification. *Transactions of the Association for Computational Linguistics*, **11**, 1147–1161.

VOSOUGHI S., ROY D. & ARAL S. (2018). The spread of true and false news online. *Science*, **359**(6380), 1146–1151. DOI : 10.1126/science.aap9559.

WHITE J., FU Q., HAYS S., SANDBORN M., OLEA C., GILBERT H., ELNASHAR A., SPENCER-SMITH J. & SCHMIDT D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. In *arXiv preprint arXiv :2302.11382*.

WU Q., BANSAL G., ZHANG J., WU Y., LI B., ZHU E., JIANG L., ZHANG X., ZHANG S., LIU J., AWADALLAH A. H., WHITE R. W., BURGER D. & WANG C. (2023). Autogen : Enabling next-gen llm applications via multi-agent conversation.

WU S. & DREDZE M. (2019). Beto, bentz, becas : The surprising cross-lingual effectiveness of bert. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, p. 833–844.

XI Z., CHEN W., GUO X., HE W., DING Y., HONG B., ZHANG M., WANG J., JIN S., ZHOU E. *et al.* (2023). The rise and potential of large language model based agents : A survey. *arXiv preprint arXiv :2309.07864*.

XU L., BING L., LU W. & HUANG F. (2021). Hierarchical multi-label text classification with horizontal and vertical category correlations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, p. 2459–2468.

Appendix A : Narrative Taxonomy

Tables 5 and 6 provide a two-level taxonomy used in the study. In Table 7, we provide general statistics on the top-level categories (*CC : Climate Change*, *URW : Ukraine-Russia War*, *Other : Other*, and *Unknown* : if no information about the category is given) for each language and each dataset type (*train*, *dev* and *test*). Note that for the *train* and *dev* sets, we calculate the statistics based on the available annotation file `subtask-2-annotations.txt` and that the same raw article file can be categorized in 2 main categories (*CC* and *URW*). As for the *test* set, we consider the number of raw text files and the attribution of the category is done based on the file name, i.e. if the file name contains `CC` then we count it in the category *CC*, if it contains `URW` then we count it as *URW*, and if no indication is provided in the file name, then we attribute the category *Unknown*.

Appendix B : Narrative Distributions

The distributions of the narratives and subnarratives across different languages and available datasets are given in Figures 2-5. We can observe high skewness of the occurrences of narrative and subnarrative categories.

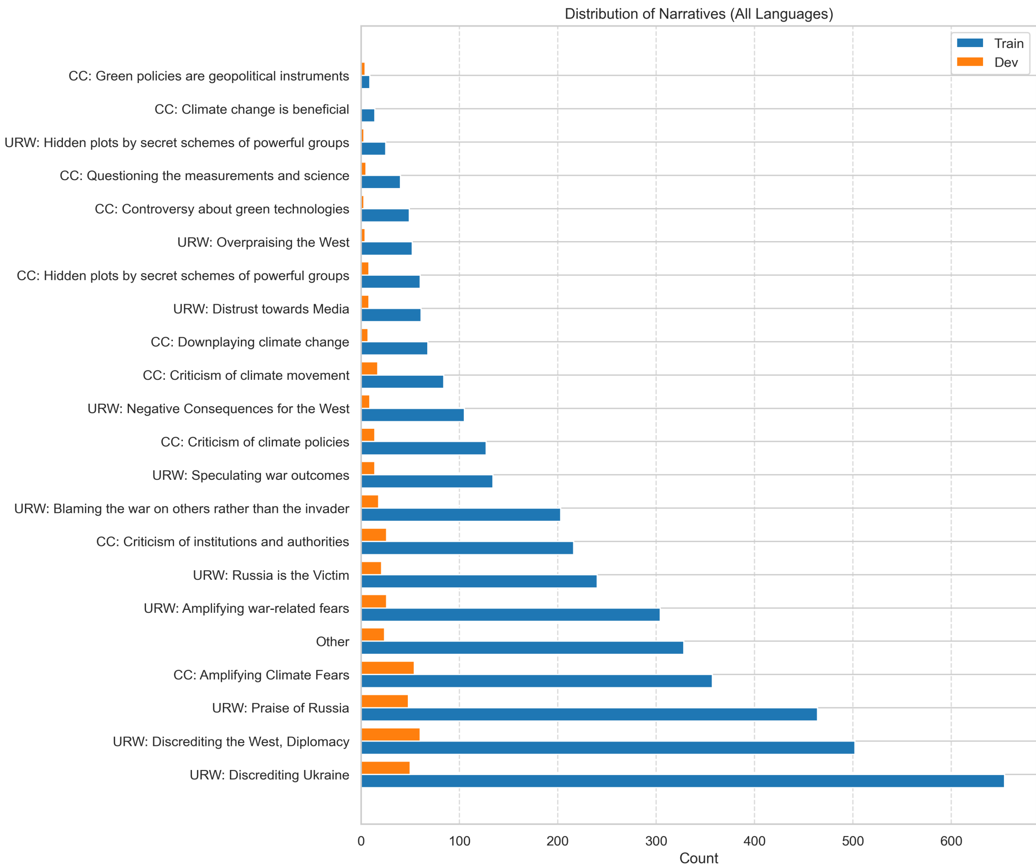

FIGURE 2 – Narrative distribution among *train* and *dev* sets, all languages

Appendix C : Agent Prompts

In this Appendix, we provide the prompts used for different kinds of agents.

Narrative	Subnarrative
Amplifying Climate Fears	Amplifying existing fears of global warming
	Doomsday scenarios for humans
	Earth will be uninhabitable soon
	Other
	Whatever we do it is already too late
Climate change is beneficial	CO2 is beneficial
Controversy about green technologies	Other
	Renewable energy is costly
	Renewable energy is dangerous
	Renewable energy is unreliable
Criticism of climate movement	Ad hominem attacks on key activists
	Climate movement is alarmist
	Climate movement is corrupt
	Other
Criticism of climate policies	Climate policies are ineffective
	Climate policies are only for profit
	Climate policies have negative impact on the economy
	Other
Criticism of institutions and authorities	Criticism of international entities
	Criticism of national governments
	Criticism of political organizations and figures
	Criticism of the EU
	Other
Downplaying climate change	CO2 concentrations are too small to have an impact
	Climate cycles are natural
	Human activities do not impact climate change
	Humans and nature will adapt to the changes
	Ice is not melting
	Other
	Temperature increase does not have significant impact
	Weather suggests the trend is global cooling
Green policies are geopolitical instruments	Green activities are a form of neo-colonialism
	Other
Hidden plots by secret schemes of powerful groups	Blaming global elites
	Climate agenda has hidden motives
	Other
Questioning the measurements and science	Data shows no temperature increase
	Greenhouse effect/carbon dioxide do not drive climate change
	Methodologies/metrics used are unreliable/faulty
	Other
	Scientific community is unreliable

TABLE 5 – Narrative taxonomy : CC

Narrative	Subnarrative
Amplifying war-related fears	By continuing the war we risk WWIII
	NATO should/will directly intervene
	Other
	Russia will also attack other countries
	There is a real possibility that nuclear weapons will be employed
Blaming the war on others rather than the invader	Other
	The West are the aggressors
	Ukraine is the aggressor
Discrediting Ukraine	Discrediting Ukrainian government and officials and policies
	Discrediting Ukrainian military
	Discrediting Ukrainian nation and society
	Other
	Rewriting Ukraine's history
	Situation in Ukraine is hopeless
	Ukraine is a hub for criminal activities
	Ukraine is a puppet of the West
	Ukraine is associated with nazism
Discrediting the West, Diplomacy	Diplomacy does/will not work
	Other
	The EU is divided
	The West does not care about Ukraine, only about its interests
	The West is overreacting
	The West is weak
	West is tired of Ukraine
Distrust towards Media	Other
	Ukrainian media cannot be trusted
	Western media is an instrument of propaganda
Hidden plots by secret schemes of powerful groups	Other
Negative Consequences for the West	Other
	Sanctions imposed by Western countries will backfire
	The conflict will increase the Ukrainian refugee flows to Europe
Overpraising the West	NATO will destroy Russia
	Other
	The West belongs in the right side of history
	The West has the strongest international support
Praise of Russia	Other
	Praise of Russian President Vladimir Putin
	Praise of Russian military might
	Russia has international support from a number of countries and people
	Russia is a guarantor of peace and prosperity
	Russian invasion has strong national support
Russia is the Victim	Other
	Russia actions in Ukraine are only self-defence
	The West is russophobic
	UA is anti-RU extremists
Speculating war outcomes	Other
	Russian army is collapsing
	Russian army will lose all the occupied territories
	Ukrainian army is collapsing

TABLE 6 – Narrative taxonomy : URW

Dataset type	Language	Top-Level Category	Count
train	BG	URW	261
		CC	110
		Other	30
	EN	URW	128
		CC	103
		Other	169
	HI	URW	228
		CC	40
		Other	98
	PT	URW	208
		CC	165
		Other	27
	RU	URW	211
		Other	4
	ALL	URW	1036
		CC	418
		Other	328
dev	BG	URW	16
		CC	13
		Other	6
	EN	URW	13
		CC	17
		Other	11
	HI	URW	29
		CC	4
		Other	2
	PT	URW	9
		CC	25
		Other	1
	RU	URW	28
	RU	Other	4
	ALL	URW	95
		CC	59
		Other	24
test	BG	URW	50
		CC	50
	EN	Unknown	53
		CC	48
	HI	URW	79
		CC	20
	PT	URW	52
		CC	48
	RU	Unknown	60
	ALL	URW	181
		CC	166
		Unknown	113

TABLE 7 – Dataset statistics

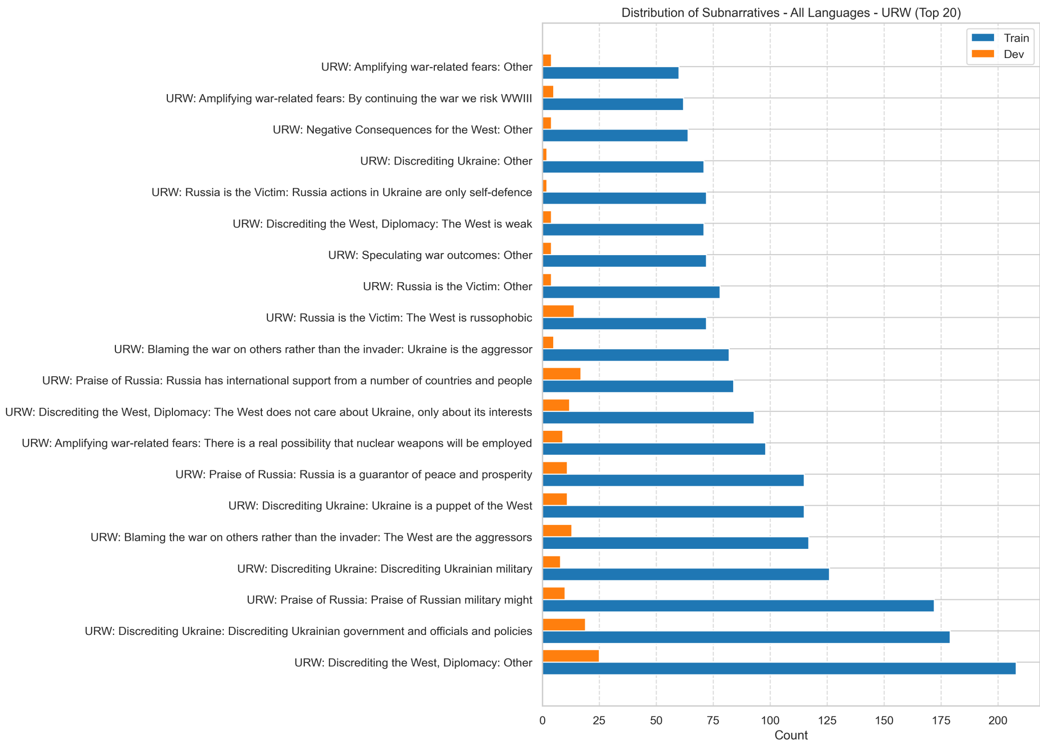

FIGURE 3 – Subnarrative distribution among *train* and *dev* sets, all languages, Ukraine-Russia War (URW)

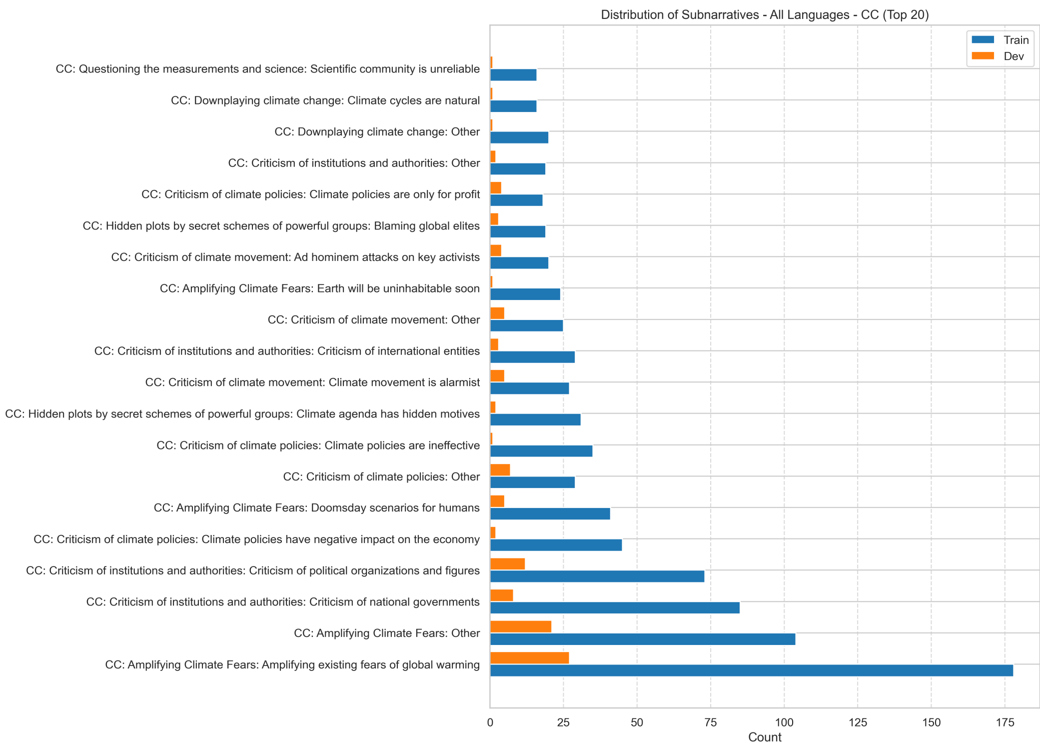

FIGURE 4 – Subnarrative distribution among *train* and *dev* sets, all languages, Climate Change (CC)

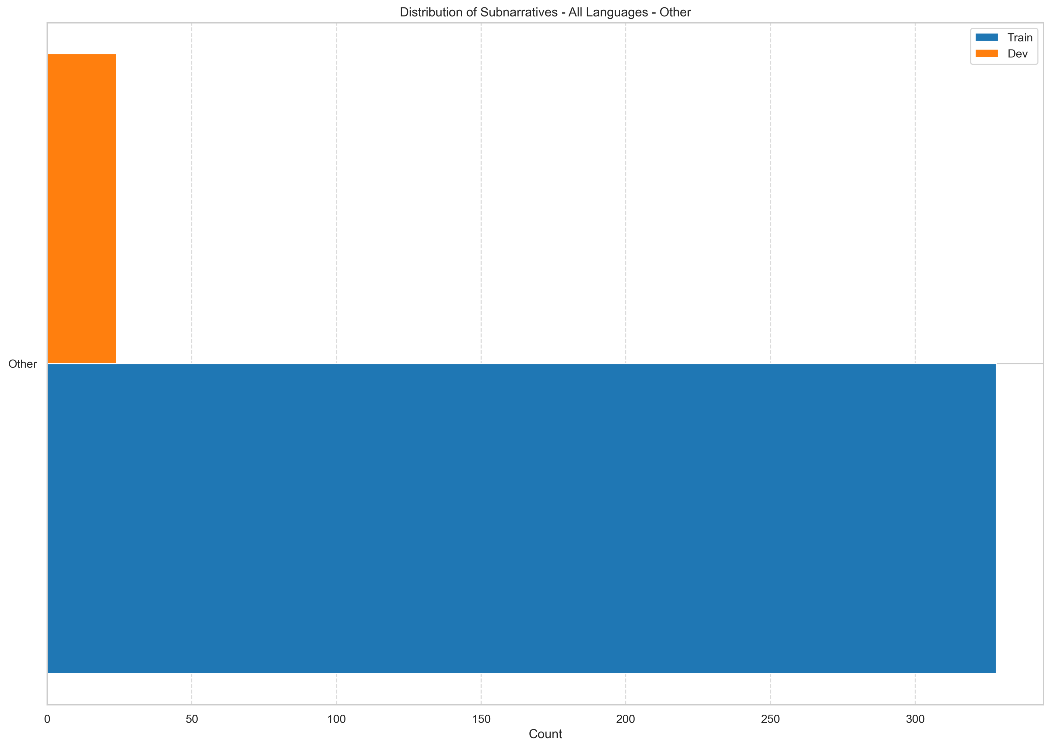

FIGURE 5 – Subnarrative distribution among *train* and *dev* sets, all languages, Other

7.1 Subnarrative Agent Prompt

```
"You are a classification model trained to do binary
classification by detecting whether a given text is related
to a specific subnarrative or not.
You have been trained to recognize the subnarrative:
SUBNARRATIVE.
This subnarrative is defined as: SUBNARRATIVE_DEFINITION.
Here are some examples of statements related to this subnarrative:
SUBNARRATIVE_EXAMPLES.
If the text is related to the subnarrative, please respond with
'1'. Otherwise, respond with '0'. Do not try to make sentences,
just respond with '1' or '0'.
You are ONLY allowed to answer with '1' or '0' and NOTHING else.
Only answer with 1 if there are explicit and clear mentions
of the subnarrative in the text. If you are slightly unsure,
classify as 0."
```

In the above prompt `SUBNARRATIVE` is the name of the subnarrative in question, `SUBNARRATIVE_DEFINITION` is the definition from the guidelines (Stefanovitch *et al.*, 2025), and `SUBNARRATIVE_EXAMPLES` are the examples of the documents representing a given subnarrative. Both the definition and the examples are extracted from the taxonomy document given for the competition.

7.2 Narrative Agent Prompt

```
"You are a classification model trained to do binary
classification by detecting whether a given text is related
to a specific narrative or not.
You have been trained to recognize the narrative: NARRATIVE.
defined as: NARRATIVE_DEFINITION.
Here are some examples of statements related to this narrative:
NARRATIVE_EXAMPLES.
If the text is related to the narrative, you MUST respond with
'1' only. Otherwise, you MUST with '0' only.
You are ONLY allowed to answer with '1' or '0' and NOTHING else.
Only answer with 1 if there are EXPLICIT and CLEAR mentions of
the narrative in the text. Some text will be ambiguous so if you
are slightly unsure, answer 0."
```

8 Example of System Functioning

In this Appendix, we demonstrate the decision flow of our architecture on a small example.

```
user (to chat_manager):

Here is the text that needs to be classified:
"The study, published in Environmental Research Letters,
   reveals significant changes in the relationship between
   vegetation growth and water availability in the Northern
   Hemisphere's mid-latitudes over the past three decades.
   The research, led by Yang Song and colleagues, highlights
    the impact of elevated carbon dioxide (CO2) levels on
   this relationship, suggesting a closer relationship
   between vegetation growth and water availability than
   previously understood. The very compound that the
   Democrats are targeting - CO2 - is actually the solution
   to preserving croplands, grasslands, forests and water
   supplies for growing populations."
###
You are ONLY allowed to reply with '0' or '1'
```

```
Next speaker: Agent_14

Agent_14 (to chat_manager):

1

Next speaker: Agent_0

Agent_0 (to chat_manager):

0


Created group chat with the following agents: [<autogen.
    agentchat.assistant_agent.AssistantAgent object at 0
    x7f583e4bc4a0>, <autogen.agentchat.assistant_agent.
    AssistantAgent object at 0x7f583e4be330>, <autogen.
    agentchat.assistant_agent.AssistantAgent object at 0
    x7f583e4d0200>]
user (to chat_manager):

Here is the text that needs to be classified:
"The study, published in Environmental Research Letters,
    reveals significant changes in the relationship between
    vegetation growth and water availability in the Northern
    Hemisphere's mid-latitudes over the past three decades.
    The research, led by Yang Song and colleagues, highlights
     the impact of elevated carbon dioxide (CO2) levels on
    this relationship, suggesting a closer relationship
    between vegetation growth and water availability than
    previously understood. The very compound that the
    Democrats are targeting - CO2 - is actually the solution
    to preserving croplands, grasslands, forests and water
    supplies for growing populations."
You are ONLY allowed to reply with '0' or '1'


Next speaker: Agent_59

Agent_59 (to chat_manager):
```

1

```
Next speaker: Agent_60

Agent_60 (to chat_manager):
```

0

```
------------------------

Next speaker: Agent_61

Agent_61 (to chat_manager):
```

0

```
------------------------
```

The extracted narratives in the end are : '*CC : Climate change is beneficial*' The extracted subnarratives : '*CC : Climate change is beneficial : CO2 is beneficial*'