# NarrativeNexus at SemEval-2025 Task 10: Entity Framing and Narrative Extraction

**Hareem Siraj, Kushal Chandani, Dua e Sameen** and **Ayesha Enayat**
Habib University, Karachi, Pakistan

hs07488@st.habib.edu.pk
kc07535@st.habib.edu.pk
ds07138@st.habib.edu.pk
ayesha.enayat@sse.habib.edu.pk

## Abstract

This paper presents NarrativeNexus' participation in SemEval-2025 Task 10, which focuses on fine-grained entity framing and narrative extraction. Our approach leverages BART, a transformer-based encoder-decoder model, fine-tuned for sequence classification and text generation. We participated in Subtask 1 (Entity Framing) and Subtask 3 (Narrative Extraction) on the English dataset.

For Subtask 1, we employed a BART-based classifier to identify and categorize named entities within news articles, mapping them to predefined roles such as protagonists, antagonists, and innocents. For Subtask 3, we used a generative BART model to produce justifications for dominant narratives.

Our framework incorporated data augmentation through paraphrasing, confidence thresholding for post-processing, and a hallucination filtering module. While the system demonstrated strong narrative coherence, distinguishing between similar roles (e.g., protagonist vs. innocent) proved challenging. NarrativeNexus secured 17th place in Subtask 1 and 5th place in Subtask 3. We highlight effective modeling strategies and discuss concrete directions for future improvements.

## 1 Introduction

Entity framing and narrative extraction are essential tasks in natural language processing (NLP), enabling applications in media bias detection, sentiment tracking, and sociopolitical discourse analysis.

SemEval-2025 Task 10 comprises three subtasks, of which we addressed two: (1) Entity Framing and (3) Narrative Extraction. Subtask 1 required classification of named entities into predefined roles, while Subtask 3 focused on generating textual justifications for dominant narratives in news articles.

We used BART, a pre-trained transformer model, fine-tuned separately for sequence classification and text generation. Despite BART's strong baseline capabilities, additional techniques such as data augmentation, hallucination filtering, and post-processing were necessary to address domain-specific challenges.

## 2 Related Work

We reviewed and analyzed previous research on entity framing and narrative extraction, categorizing studies based on their methodologies and objectives.

Sinelnik and Hovy (Sinelnik and Hovy, 2024) explored multilingual disinformation framing using XLM-RoBERTa, effectively detecting thematic frames across multiple languages. Their work highlights the challenges of multilingual preprocessing and the importance of aligning embeddings with linguistic differences.

Xu et al. (Xu et al., 2024) introduced NARCO, a graph-based Transformer-XL model designed for narrative coherence. Their approach improved causal and temporal dependencies in long-form texts, significantly enhancing narrative consistency in text generation tasks.

Papalampidi et al. (Papalampidi et al., 2022) proposed a dynamic entity memory mechanism within a Transformer-XL framework. By tracking entity attributes throughout a narrative, their model ensured coherence and consistency in generated texts, making it valuable for long-form content generation.

Schäfer et al. (Schäfer et al., 2024) applied BERTopic for fake news analysis, demonstrating its superiority over traditional topic modeling methods like LDA and NMF. Their study uncovered nuanced themes within misinformation campaigns, improving content classification and thematic anal-



*Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025)*, pages 1411–1414
July 31 - August 1, 2025 

ysis.

Sinelnik and Hovy (Sinelnik and Hovy, 2024) also examined lexicon-based approaches for frame detection. While these methods offer interpretability and lightweight computation, they struggle to capture rare or evolving thematic frames, limiting their effectiveness for complex analyses.

These studies collectively contribute to advancing NLP methodologies for entity framing, misinformation analysis, and narrative generation. Our work builds upon these insights, leveraging transformer-based models for structured content analysis in SemEval-2025 Task 10.

## 3 Background

SemEval-2025 Task 10 consists of three subtasks. We participated in:

- **Entity Framing (Subtask 1):** Classify named entities into protagonist, antagonist, or innocent roles using surrounding context.
- **Narrative Extraction (Subtask 3):** Generate justifications for dominant/subnarrative pairs given full article text.

The English dataset was provided by the task organizers, along with gold-standard annotations. Detailed task descriptions are available in the official task paper (Piskorski et al., 2025).

## 4 Dataset and Preprocessing

The English dataset included approximately:

- Subtask 1: 1242 training instances with entities and context passages (with augmentation).
- Subtask 3: 88 article entries containing dominant/subnarratives (no augmentation was done).

For Subtask 1, we used pre-extracted context passages accompanying each entity in the dataset. These were paired with the corresponding entity and formatted as 'Entity: entity. Context: context' for BART input.

For Subtask 3, documents were truncated to 1,024 tokens using the Hugging Face tokenizer's built-in truncation mechanism, with narrative and subnarrative prepended to the article.

## 5 System Overview

### 5.1 Subtask 1: Entity Framing

We developed a BART-based sequence classifier to categorize named entities into *protagonists, antagonists, and innocents*. The dataset contained news articles with named entity mentions, contextual passages, and gold-standard labels. Each training instance was formatted as follows:

```
"Entity: {named entity}. Context:
{text snippet}.  Classification:
{label}"
```

We fine-tuned `BART-large` using the BART tokenizer with a maximum sequence length of 512 tokens. The training hyperparameters, including batch size, epochs, and learning rate, are presented in Table 1. During training, we used cross-entropy loss and the AdamW optimizer with a linear decay learning rate scheduler.

To enhance generalization, we applied data augmentation techniques, particularly paraphrasing. We used a mix of GPT-based and Gemini-based paraphrasing APIs to rewrite approximately 15% of context passages in the training set. These paraphrased examples retained the original entity-role labels and were added back into the dataset, effectively doubling the training size.

Post-processing involved confidence thresholding to improve classification reliability. A softmax threshold of 0.5 was used for filtering uncertain predictions. Entities with <50% max confidence were labeled as "Uncertain" and excluded during test-time prediction. Error analysis revealed that distinguishing between *protagonists and innocents* was particularly challenging due to contextual ambiguities in news articles. The model was evaluated on the basis of Accuracy and F1 score. Table 3 presents the evaluation metrics and the scores.

### 5.2 Subtask 3: Narrative Extraction

For narrative extraction, we fine-tuned BART-large-cnn using a text-to-text generative approach. The dataset contained dominant narratives, subnarratives, and full article texts, structured as follows:

```
"Narrative: {dominant narrative}.
Subnarrative:    {subnarrative}.
Article: {full text}"
```

The model was trained with a batch size of 4 for eight epochs, optimizing with cross-entropy loss and the AdamW optimizer. The hyperparameter settings for Subtask 3 are detailed in Table 2.

A key challenge was ensuring factual consistency between generated justifications and the original article. To mitigate hallucination, we introduced an additional filtering step where justifications with low confidence scores were discarded. The evaluation relied on BLEU and ROUGE scores to measure output fluency and relevance. The model's performance results are presented in Tables 3 and 4.

## 6 Experimental Setup

### 6.1 Hyperparameters

The experimental configuration for each subtask is detailed in Tables 1 and 2.

Table 1: Experimental setup for Subtask 1

| Configuration | Value |
|---|---|
| Pre-Trained Model | facebook/bart-large |
| Epochs | 5 |
| Batch Size | 8 |
| Learning Rate | 5e-5 |
| Data Splits | 80% Train, 20% Validation |

Table 2: Experimental setup for Subtask 3

| Configuration | Value |
|---|---|
| Pre-Trained Model | BART-large-cnn |
| Epochs | 8 |
| Batch Size | 4 |
| Learning Rate | 2e-5 |
| Data Splits | 80% Train, 20% Validation |

## 7 Results

The performance of our system on the official test sets is presented in Table 4.

Table 3: Training Performance metrics for Subtask 1 and Subtask 3

| Metric | Subtask 1 | Subtask 3 |
|---|---|---|
| Accuracy | 0.835498 | – |
| F1-score | 0.737705 | – |
| BLEU-4 | – | 0.104933 |
| ROUGE-L | – | 0.4138472 |

Table 4: Test Performance metrics for Subtask 1 and Subtask 3

| Metric | Subtask 1 | Subtask 3 |
|---|---|---|
| Exact Match Ratio | 0.18300 | – |
| Micro Precision | 0.20850 | – |
| Micro Recall | 0.18490 | – |
| Micro F1-score | 0.19600 | – |
| Accuracy | 0.71910 | – |
| Precision | – | 0.71991 |
| Recall | – | 0.74267 |
| F1 Macro | – | 0.73085 |

For Subtask 1, our system achieved an Exact Match Ratio of 0.18300, with micro precision, recall, and F1-scores of 0.20850, 0.18490, and 0.19600 respectively. The accuracy for identifying the main role of entities was 0.71910. These results indicate the difficulty in capturing fine-grained entity role distinctions, suggesting potential improvements through better feature engineering and model enhancements.

For Subtask 3, our system ranked 10th, achieving a Precision of 0.71991, Recall of 0.74267, and an F1 Macro score of 0.73085. The model demonstrated strong coherence in narrative justification generation, though further refinements in dataset curation and text conditioning techniques could boost performance further.

These results emphasize the effectiveness of transformer-based architectures in entity framing and narrative generation while highlighting areas where improvements in feature extraction and model fine-tuning could lead to higher accuracy.

## 8 Conclusion

We presented NarrativeNexus' approach to SemEval-2025 Task 10, focusing on entity framing and narrative extraction using BART. Our findings highlight the potential of pre-trained transformer models for structured content analysis, particularly in maintaining narrative coherence. However, challenges such as fine-grained role differentiation and ensuring factual consistency in justification generation remain areas for future improvement. We aim to explore alternative model architectures, data augmentation strategies, and more refined evaluation techniques in future work to enhance entity framing and narrative extraction capabilities. The insights gained from this work contribute to advancing automated media analysis and NLP applications in discourse understanding.

## References


Pinelopi Papalampidi, Kris Cao, and Tomas Kocisky. 2022. Towards coherent and consistent use of entities in narrative generation. In *International Conference on Machine Learning*, pages 17278–17294. PMLR.

Jakub Piskorski, Tarek Mahmoud, Nikolaos Nikolaidis, Ricardo Campos, Alípio Jorge, Dimitar Dimitrov, Purificação Silvano, Roman Yangarber, Shivam Sharma, Tanmoy Chakraborty, Nuno Guimarães, Elisa Sartori, Nicolas Stefanovitch, Zhuohan Xie, Preslav Nakov, and Giovanni Da San Martino. 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.

Karla Schäfer, Jeong-Eun Choi, Inna Vogel, and Martin Steinebach. 2024. Unveiling the potential of bertopic for multilingual fake news analysis-use case: Covid-19. In *International Conference on Web Search and Data Mining 2024*.

Antonina Sinelnik and Dirk Hovy. 2024. Narratives at conflict: Computational analysis of news framing in multilingual disinformation campaigns. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 225–237.

Liyan Xu, Jiangnan Li, Mo Yu, and Jie Zhou. 2024. Fine-grained modeling of narrative context: A coherence perspective via retrospective questions. *arXiv preprint arXiv:2402.13551*.