cljteam at SemEval-2025 Task 10: Finetuning SmolLM2 with Taxonomy-based Prompting for Explaining the Dominant Narrative in Propaganda Text

Anca Marginean
Technical University of Cluj-Napoca, Cluj-Napoca, Romania
anca.marginean@cs.utcluj.ro

Abstract

XAI has been a long-standing goal of AI. Explaining why a text can be considered to have a dominant narrative, where the narrative is known, is of great importance for dealing with propaganda in the news. This paper reports on the participation of the system *clujteam* in Subtask 3 of Task 10 of Semveal 2025. The system obtained 7th place with a value of 0.72464 for F1macro, at 0.026 distance from the 1st place. The key components of the solution are the given taxonomy for the narratives and supervised fine-tuning of SmolLM2.

1 Introduction

Unfortunately, propaganda detection has become a crucial task in today's world. Understanding the dominant narrative and subnarratives, as well as identification of the different roles an entity can play, are two key aspects of propaganda identification.

Task 10 of Semeval 2025 (Piskorski et al., 2025), *Multilingual characterization and extraction of narratives from online news*, introduces 3 main subtasks in five languages: Entity Framing, Narrative Classification, and Narrative Explanation. This paper reports the results obtained by the system *clujteam* in the official competition for Subtask 3 in English. Additionally, post-competition experiments are reported for Subtask 1.

2 Background

Subtask 2 and 3 of Task 10 propose a taxonomy of narratives and subnarratives for two important topics for the online news: Ukraine-Rusia War (URW) and Climate Change. For each narrative and subnarrative, the taxonomy includes a definition for the relevant statements, instructions for the annotators, and possible examples. Figure 1 presents the narratives and subnarratives for URW topic. It can be observed that some subnarratives have quite similar meanings.

Subtask 3 is framed as a text-generation task where the input is a news text and a dominant narrative, and in some cases, the dominant subnarrative. The expected text must capture the explanation for having this subnarrative/narrative as the dominant one from the taxonomy. The evaluation relies on *bertscore* and the primary metric is *F1macro*.

3 System overview

Our solution for Subtask 3 is based on supervised fine-tuning (SFT) of a small language model using the data provided in the competition. Specifically, we use SmolLM2 1.7B-Instruct and 360M-Instruct (Allal et al., 2025), which offer a favorable balance between model size and performance.

3.1 SFT parameters

For SFT, we use the library trl[1] (Transformer Reinforcement Learning) from huggingface. The SFT on the 360M model was done on GPU 3090 (24GB) with $batch\_size = 4$ and $gradient\ accumulation\ steps = 4$. The used optimizer is *adamw_torch*. Differently, the SFT for the 1.7M model was done on 4 Nvidia V100 (32GB), with $batch\_size = 1$ per device, $gradient\ accumulation\ step = 2$. To reduce the GPU memory required by the 1.7B model, we changed the optimizer to *adamw_8bit* from *bitsandbytes* library, and we used mixed precision $fp16$. For both models, the learning rate was $2e-5$ and the number of epochs was 3. The parameters for the text generation were maintained consistently throughout all the experiments.

In terms of training data, we used 203 examples from the competition *training* set, while the 30 examples from the competition dev set were used for validation. The official *test* set of the com-

[1] https://huggingface.co/docs/trl/en/index

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 2395–2399
July 31 - August 1, 2025 ©2025 Association for Computational Linguistics

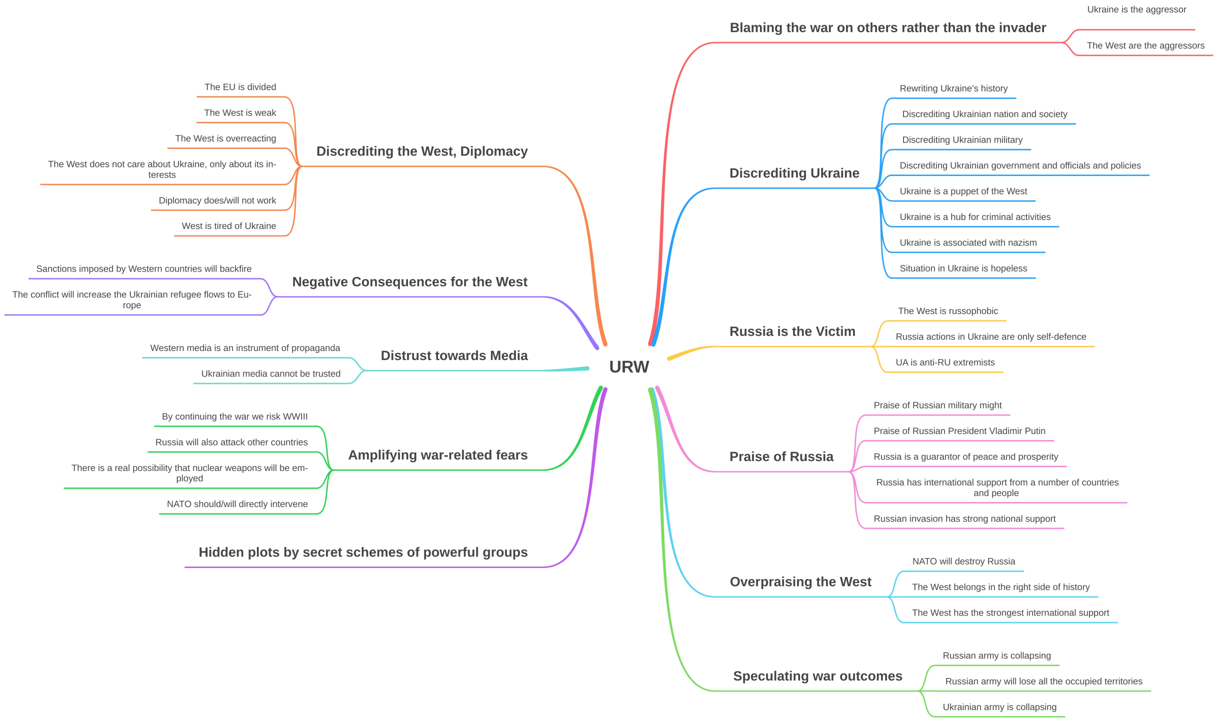

Figure 1: Ukraine War label taxonomy

petition includes 68 examples. We can observe that the training dataset is extremely small, yet the SFT works mainly because the SmolLM2-instruct models support tasks such as text rewriting and summarization.

3.2 SFT dataset preparation

Our SFT for Subtask 3 is done with distinct system and user prompts. In the prompts, square brackets [] indicate optional elements, while curly braces represent variables.

The user prompt includes the given text, the main narrative, and, if present, the subnarrative:

```
User prompt:

###News Text:
{content}

###Main narrative:
{main narrative}
[###Subnarrative
{subnarrative}]
```

For the system prompt, we experimented with two versions. One that is general for all the examples, respectively, one that is customized according to the given subnarrative (narrative). For the latter, we extracted the *Definition* of the subnarrative/narrative from the given Narrative-taxonomy. If the subnarrative is given, its definition is added at the end of the system prompt, otherwise, the *Definition* is taken from the narrative. Table 1 includes examples of such definitions for some narratives from URW and CC topics, respectively, for some URW subnarratives.

```
System prompt:

You understand propaganda in news about
Ukraina-Russia War and Climate Change,
and you know how to explain why a text
has as its main narrative a given one.

You get a news text and its main narra-
tive. In some cases, you also get a
subnarrative.
Generate a very brief explanation of the
 main narrative. The explanation MUST be
 grounded in text fragments that provide
 evidence for the given narrative's
claims. The evidence can only include
facts and opinions from the given text.

[Taxonomy-dependent part
In order to build the explanation,
analyze the {taxo[subnarrative] if
subnarrative else taxo[narrative]}].
```

4 Results

In the official leaderboard of Subtask 3 of Task 10, our system *clujteam* is positioned in the 7th place with a value of 0.72464 for F1macro. The best system obtained a value of 0.75040.

As expected, the SmolLM2 1.7B obtained the best results, even though mixed precision is used

Domain	Narrative/Subnarrative	Definition included in the taxonomy
	Narrative Definitions	
URW	Blaming the war on others rather than the invader	statements attributing responsibility or fault to entities other than Russia in the context of Russia's invasion of Ukraine.
URW	Discrediting Ukraine	statements that undermine the legitimacy, actions, or intentions of Ukraine or Ukrainians as a nation.
CC	Scientific community is unreliable	statements discrediting scientists, the scientific community and their actions.
CC	Controversy about green technologies	statements that express skepticism or criticism of environmentally friendly technologies.
	Subnarrative Definitions	
URW	Ukraine is the aggressor	statements that shift the responsibility of the aggression to Ukraine instead of Russia and portray Ukraine as the attacker.
URW	The West are the aggressors	statements that shift the responsibility for the conflict and escalation to the Western block.
URW	Ukraine is a puppet of the West	statements that claim that Ukraine is controlled or heavily influenced by Western powers, particularly the United States and European Union.

Table 1: Examples of Definitions included in the taxonomy-dependent system prompt. Observation: only the Definition for the given narrative/subnarrative is included in the prompt.

dataset	model version	system prompt	Precision	Recall	F1-macro
dev	360M	taxo-dependent	0.713	0.706	0.709
test	360M	taxo-dependent	0.703	0.709	0.706
dev	360M	taxo-independent	0.712	0.710	0.712
test	360M	taxo-independent	0.707	0.716	0.711
dev	1.7B	taxo-dependent	0.731	0.718	0.725
test	1.7B	taxo-dependent	**0.723**	**0.726**	**0.725**
dev	1.7B	taxo-independent	0.721	0.715	0.718
test	1.7B	taxo-independent	0.709	0.715	0.712

Table 2: Results for Subtask3 in Semeva2025 Task 10

compared to the 360M model, and an 8bit optimizer is used. Due to hardware limitations, SFT of 1.7B model was not possible with 32bit precision. Table 2 presents Precision, Recall, and F1Macro for different experiments. We report both the values on the *dev* and *test* set.

The first observation is that the 360M model performance is not significantly lower than that of the 1.7B model. When comparing the performance of the 360M model to that of the 1.7B model, it's important to note that the SFT for the larger model was conducted using mixed precision and an 8-bit optimizer.

The second observation is that the 360M model achieves acceptable performance despite the limited size of both the training dataset and the model itself. This is justified by the fact that all versions of SmolLM2 are prepared for text summarization and rewriting.

The third observation is that the taxonomy-dependent prompt does not improve the performance of the 360M model, but rather slightly the opposite. We can just assume that this is due to the size of the model and the increased length of the system prompt. The 360M model obtained better performance when the taxonomy-independent prompt is used, at least according to the used metrics that rely on Bert score. Nonetheless, the text generated by the models that use the taxonomy-dependent prompt is more meaningful.

The fourth observation, and the most important in our view, is that the taxonomy-dependent prompt increases the performance for the 1.7B model. The model using taxonomy-dependent prompts is significantly more precise than the one with a general prompt (0.723 vs 0.709 on the *test* set).

The fifth observation is that even though the metrics do not show a significant difference between the quality of the SFT with the taxonomy-dependent prompt compared to SFT with the independent one, the generated text tends to be more meaningful when the Definition of the narrative statements is included. For example, the following text is generated by the taxonomy-dependent model: `The text suggests that the climate agenda has hidden motives, such as depopulation and population control. The author argues that the climate agenda is a tool for globalists to control the population and implement their plans for a one-world government. The text also criticizes the media.` The Taxonomy-independent model explains the dominant narrative of the same text with: `The text discusses the climate agenda and its hidden motives. The text also talks about the agenda of powerful groups and their hidden plots. The text also talks about the agenda of the powerful groups and their hidden motives.`

5 Post-competition experiments

In this section we describe our post-competition experiments for Subtask 1 using the same SFT on SmolLM2. The input for Subtask 1 is a news article and a list of entity mentions; it is required to classify the role and the subrole of the given entities using a predefined taxonomy of fine-grained roles. The subrole is not unique, so it is a multilabel classification. The same entity mention can occur several times in one text, not necessarily with the same role or subrole.

5.1 Single-step classification of all entity mentions

One approach for this subtask was to use SmolLM2 to generate the fine-grained roles for all the entity mentions in a single step. In our user prompt, we inserted special tags $\langle start\_entity\_i \rangle$ before and $\langle end\_entity\_i \rangle$ after each given entity mention to clearly delineate them. The expected output is a json object structured as a list of objects made of $entity\ number$, $entity$, and $entity\ role$. The first part of Figure 2 includes the used system prompt.

For clarity, we give here an example of the user prompt for the single-step classification of all entities:

```
Pence calls viral clip suggesting he
cares more about Ukraine than US 'fake
news'
Former Vice President  <start_entity_1>
Mike Pence <end_entity_1> fired back on
social media after ....
In the video, former Fox News host
Tucker Carlson questions the 2024
Republican presidential candidate about
where his priorities lie after Pence
criticizes the length of time it has
taken the <start_entity_2> Biden
administration <end_entity_2>  to provide
 <start_entity_3> Ukraine <end_entity_3>
 with weapons to fend off Russia's
invasion of the former Soviet state.
```

5.2 Additional task of independent classification of each entity mention based on a chatGPT extracted summary

Since the resulting training set is small, we considered an additional task for SFT of SmolLM2: given a list of actions associated with only one entity, the model needs to return a list of fine-grained roles. The system prompt used in SFT for this additional task is included in the second half of Figure 2.

The samples for this additional subtask are obtained from the Subtask 1 data by using chatGPT 4o:

User prompt is similar to the single-step classification (each entity mention is marked out with special tags).

System prompt: *You are given a text where entities are enclosed between the tags <start_entity_number> and <end_entity_number>. For each entity, identify its actions and summarize the key information associated with it, focusing primarily on the surrounding context. Ensure the summary is concise, capturing the most relevant details related to the entity's role and actions in the text.*

Example of relevant types of actions: - Protects values, communities, or individuals from harm. - Upholds justice and ensures safety. - Takes on leadership roles (e.g., law enforcement, soldiers, community leaders). ....

Present the output in JSON format, with each entity as an object containing: - entity: The name of the entity. - entity number: the number of the entity. - summary: A list of the entity's actions, significance, or other important characterizations

```
System prompt for single step classification of all entities:
You are given a text where entities are enclosed between the tags <start_entity_num-
 ber> and <end_entity_number>.
For each entity, identify its role or roles.  The possible roles are
{list of fine-grained roles from the given taxonomy}.

Present the output in JSON format, with each entity as an object containing:
- entity: the name of the entity.
- entity number: the number of the entity.
- roles: one or more roles from the given list.
Ensure the response uses only information extracted from the given text.
```

```
System prompt for independent classification of one entity:
You are given an entity and a text describing the actions or views of that entity.
Your task is to classify the entity's behavior based on the following roles:
{list of fine-grained roles from the given taxonomy}.
Assign one or more of these roles to the entity based on the given text.
Return only the identified roles as a comma-separated list without any additional
text or explanation. If no role applies, return 'None'.
```

Figure 2: System prompts for Subtask 1: the single-step prompt was employed to classify all entity mentions at once, while the independent classification prompt considered each entity separately based on a summary of its actions.

	EMR	microP	macroR	microF1	Acc main role
v1	0.264	0.287	0.260	0.270	0.824
v2	0.319	0.382	0.340	0.360	0.890

Table 3: Subtask 1 results on *dev* set for SmolLM2 1.7B trained on: (*v1*) single-step classification vs (*v2*) single-step classification & independent

of the entity (e.g. pedophile, terrorist, martyr, ..).

Ensure the response remains precise and contextual and uses only information extracted from the given text.

5.3 Experiments and results

We run three types of experiments: *v1* - SmolLM2 is trained exclusively with samples of single-step classification; *v2* - SmolLM2 is trained on a combination of both single-step classification samples and independent classification; *v3* - SmolLM2 is trained on the data from the *v2* experiment, to which the samples from Subtask 3 are added.

The best result on the test set obtained by our system *clujteam* for Subtask 1 in the Post-competition Leaderboard is: $Exact\ Match\ Ratio = 0.2894$, $microP = 0.3447$, $macroR = 0.3057$, $microF1 = 0.3240$, $Accuracy\ main\ role = 0.8638$. It was obtained for SmolLM2 1.7M trained in mixed precision on v3. For this subtask, the differences in performance between SmolLM2 360M and 1.7B are more significant, even when we compare 32-bit precision for the small model to mixed precision for the larger model.

6 Conclusions

The paper describes the participation of *clujteam* system at Subtask 3 of Semeval Task 10. Our experiments indicate that taxonomies defined for the narratives and subnarratives play a key role in obtaining a model that generates precise explanations in resource-constrained settings.

Acknowledgements

This work was conducted as part of the project PN-IV-P6-6.3-SOL-2024-2-0312.

References

Loubna Ben Allal, Anton Lozhkov, Elie Bakouch, Gabriel Martín Blázquez, Guilherme Penedo, Lewis Tunstall, Andrés Marafioti, Hynek Kydlíček, Agustín Piqueres Lajarín, Vaibhav Srivastav, Joshua Lochner, Caleb Fahlgren, Xuan-Son Nguyen, Clémentine Fourrier, Ben Burtenshaw, Hugo Larcher, Haojun Zhao, Cyril Zakka, Mathieu Morlon, Colin Raffel, Leandro von Werra, and Thomas Wolf. 2025. SmolLM2: When Smol Goes Big – Data-Centric Training of a Small Language Model. *Preprint*, arXiv:2502.02737.

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.