Do Knowledge Graphs Influence LLMs in Entity Framing Analysis?

Luana Bulla[1], Beatrice Fiumanò[1,*], Nicolas Lazzari[1,2] and Valentina Presutti[1]

[1]University of Bologna, Bologna 40124, Italy
[2]University of Pisa, Pisa 56127, Italy

Abstract

Recent advances in NLP have shown that analysing documents using LLMs in pragma-discursive tasks benefits from the integration of explicitly modelled knowledge. In this paper, we investigate this approach on the Entity Framing task, which aims to characterise how an entity is portrayed in a document through linguistic and narrative choices. We test six LLMs of varying sizes under zero-shot and KG-enhanced settings, and demonstrate that a straightforward integration of structured knowledge does not necessarily improve performance, but that it can significantly impact the decision process of some models, depending on the knowledge integration strategy and model size.

Keywords

Entity Framing, Knowledge Graph, Large Language Models

1. Introduction

Media narratives shape how the general public perceives events and social actors, depending on how these are discursively construed. This process of constructing and disseminating a particular perspective on reality through language is known as *framing*. In particular, we refer to *entity framing* as the way social actors are portrayed within a text through specific linguistic and narrative choices. Framing is a particularly complex phenomenon to detect and analyse computationally [1], as it is instantiated through subtle discursive devices, such as word choice and topic selection.

In Fiumanò et al. [2], we investigate how framing-relevant signals can be extracted from text and structured as a Knowledge Graph (KG) representation, enabling query-driven analysis of framing patterns. A qualitative case study showed that the knowledge encoded in the KG successfully supports both event-centric and entity-centric framing analyses. However, this SPARQL-driven approach requires considerable manual effort, both in query design and in the interpretation of results, limiting its scalability.

In this work, we address this issue by evaluating the capabilities of Large Language Models (LLMs) for multilingual entity framing analysis, leveraging their broad textual and extra-textual knowledge alongside their multilingual capabilities. Our hypothesis is that, if the frame-semantic knowledge encoded in the KGs supports framing analysis when manually inspected by a domain expert, an LLM provided with the same knowledge should benefit from it and perform this analysis automatically. Our hypothesis is consistent with prior work showing that injecting ad-hoc, structured knowledge can guide LLM predictions toward more sound and theory-grounded outputs [3, 4]. Building on this insight, we evaluate the role that triple injection plays in the task of fine-grained entity role classification across a five-language entity framing benchmark [5], using target knowledge encoding frame-semantic information. To this end, we compare LLM performance under two settings: a *baseline* setting, where

RAGE-KG 2026: The Third International Workshop on Retrieval-Augmented Generation Enabled by Knowledge Graphs, co-located with ISWC 2026, October 25–26, 2026, Bari, Italy

*Corresponding author.
luana.bulla@unibo.it (L. Bulla); beatrice.fiumano@unibo.it (B. Fiumanò); nicolas.lazzari3@unibo.it (N. Lazzari); valentina.presutti@unibo.it (V. Presutti)
0000-0003-1165-853X (L. Bulla); 0009-0007-6918-0803 (B. Fiumanò); 0000-0002-1601-7689 (N. Lazzari); 0000-0002-9421-8566 (V. Presutti)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-11-04

CEUR-WS.org/Vol-4312/paper39.pdf

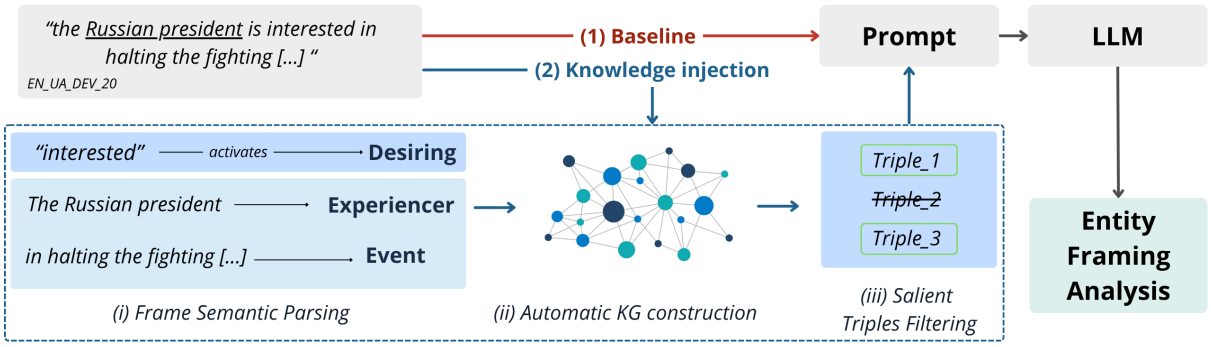


Figure 1: Example of the knowledge injection pipeline for a single entity mention (*the Russian president*). In the baseline (1) setting, only the raw text is provided to the LLM. In the KG-enhanced setting (2), the text is first parsed to extract semantic frames and frame elements (i), automatically structured into a KG (ii), and filtered to retain only salient, entity-centric triples (iii), before being injected into the prompt.

only the plain text is provided as input, and a *graph-enhanced* setting, where the text is accompanied by its KG representation, constructed following the DORIS ontology presented in Fiumanò et al. [2]. Since both the selection and the verbalization of structured knowledge are known to substantially affect model behaviour, we do not test a single KG-enhanced configurations: we compare three knowledge injection strategies of increasing specificity, from all salient triples to verbalised entity-centric ones. The pipeline is illustrated in Fig. 1. To thoroughly investigate the role of external knowledge, we test six LLMs (Gemma, Mistral and Qwen) from three families at two size regimes: small, under 8B parameters, and medium, under 30B parameters.

Through our experiments, we address the following research questions:

- **RQ1**: To what extent does injecting structured, frame-semantic knowledge improve LLM performance on multilingual entity framing, and does this vary across languages, injection strategies and model sizes?
- **RQ2**: How does knowledge injection affect model behaviour in terms of stability across runs and consistency with baseline predictions?

Our results demonstrate that knowledge injection yields broadly comparable performance overall, though with variation across model sizes and languages. In the KG-enhanced setting, performance improves for four out of six models on English and Bulgarian, while results are mixed for Hindi, Portuguese, and Russian.

To further probe how KG-injection affects model behaviour, we estimate the entropy of each model's role distribution given an entity in a text. Smaller models, such as Mistral 7B, exhibit higher answer uncertainty, i.e., less consistency across runs. We then compute the Hellinger distance between the probability distributions estimated from the baseline and KG-enhanced settings, to quantify how much structured knowledge influences model predictions, independently of their correctness with respect to the gold standard. Results show that smaller models are more strongly affected by triple injection, producing substantially different answers, whereas larger models show convergence between KG-augmented and baseline predictions. Finally, we conduct a qualitative analysis of the highest-entropy model to characterise in what way knowledge injection influences predictions, and to assess to what extent it affects interpretability and groundedness in input information.

We summarise our contribution as follows:

- We evaluate the impact of injecting ad-hoc, structured knowledge in LLMs for multilingual entity framing analysis across five languages and six models of varying sizes, comparing three knowledge injection strategies of increasing specificity;
- We perform an entropy analysis over repeated runs, together with a distributional distance measure between baseline and triple-augmented settings, showing how triple injection affects model uncertainty and prediction stability;

The rest of this paper is structured as follows: Section 2 reviews related work on computational approaches to framing analysis and entity role labeling. Section 3 describes the dataset used for the entity framing task. In Section 4, we describe how multilingual texts from the benchmark are structured into KGs. In Section 5, we present the task and our experimental setup, including prompts, evaluation settings, and knowledge injection techniques, and report results for role classification, prediction entropy and distributional distance between settings. We then discuss our findings in Section 6. Finally, Section 7 summarizes our contributions.

2. Related Work

Computational approaches to media framing have traditionally operated at the document level, formalising the task as single-label classification over a fixed frame inventory and relying on lexical or topical signals distributed across the article [6, 7, 8]. Because the unit of prediction is the article as a whole, this formulation does not model the participants through which a frame is realised.

Entity role labelling addresses this limitation by characterising entities according to their function in the narrative. For example, Gomez-Zara et al. [9] identify salient entities in news articles and classify them as heroes, villains, or victims using named-entity recognition and lexical and syntactic heuristics. Alleviating the need for effective heuristics, Frermann et al. [10] propose to frame the task as a multilabel classification problem solved through the use of Transformer-based models. Differently, Stammbach et al. [11] frame the task as zero-shot question answering, where an LLM is responsible for the classification process. Beyond relying on hero, villain, and victim as the set of possible archetypes, Hobson et al. [12] propose to frame the problem as an open-vocabulary classification task, where an LLM freely labels an entity using a role that can eventually be mapped back to either hero, villain, or victim.

Fine-grained entity framing, proposed by Mahmoud et al. [5], introduces additional challenges by framing the problem as a hierarchical multi-label classification task, where three coarse roles (protagonist, antagonist, and innocent) and 22 fine-grained archetypes are defined. In this paper, we follow this approach and rely on their multilingual dataset, described in Section 3, which has also been adopted for the SemEval-2025 Task 10 challenge [13].

A multitude of submissions tackled the challenge, with approaches that can be conceptually divided into three groups: solving the task through prompt engineering [14, 15, 16, 17, 18], fine-tuning pre-trained language models [19, 20, 21, 22, 23], or combining both approaches [24, 25, 26, 27, 28].

In general, approaches that rely on prompt engineering to preprocess a document (for instance, by summarizing it) and then fine-tune a pre-trained language model to perform the final classification obtained the best results.

As introduced in Section 1, we rely on prompt engineering to predict the role of an entity. We observe, however, that our knowledge injection approach is similar in spirit to a combined approach: rather than preprocessing the document using an LLM, we transform it into a KG. This results in an intermediate representation that is easier to inspect and does not require task-specific fine-tuning, thereby reducing dependence on annotated data and the empirical label distribution.

3. Data

For our experiments, we rely on the multilingual entity framing benchmark introduced by Mahmoud et al. [5]. The dataset formulates entity framing as a fine-grained, multi-label span classification task. Given a news article and a set of entity mentions, the goal is to assign each mention one or more roles drawn from a hierarchical taxonomy of 22 archetypical roles nested under three coarse-grained categories, namely protagonist, antagonist, and innocent. Table 6 in the Appendix reports the full list of 22 fine-grained roles.

The corpus comprises 1,378 news articles in five languages (Bulgarian, English, Hindi, Portuguese, and Russian), covering two domains: the Ukraine–Russia War and Climate Change. Over 5,800 entity

	BG	EN	HI	PT	RU	All
Train	165 (389)	133 (440)	203 (1347)	206 (833)	89 (252)	796 (3261)
Dev	94 (237)	69 (245)	139 (983)	100 (417)	44 (114)	446 (1996)
Test	15 (30)	27 (90)	35 (279)	31 (115)	28 (85)	136 (599)
Total	274 (656)	229 (775)	377 (2609)	337 (1365)	161 (451)	1378 (5856)

Table 1
Distribution of articles and entity mentions (in parentheses) by language and split, from Mahmoud et al. [5].

mentions were annotated with role labels by language-specific teams of expert annotators, with each mention assigned one coarse-grained role and one or two corresponding fine-grained roles. Annotation was performed at the level of individual mentions rather than entities as a whole: the same entity may therefore be assigned different roles at different points within the same article, depending on how it is portrayed in each specific mention. For our experiments, we reuse the original splits released by the authors, as seen in Table 1. In particular, we rely on the the Dev split for LLM-based experiments, since the test set is not publicly available.

While the fine-grained role distribution is significantly imbalanced across the whole dataset, with 83.6% of instances labeled as *Victim*, this imbalance is mitigated in the dev set, where *Victim* still prevails but with a lower share (18.17%), followed by *Guardian* (13%) and *Foreign Adversary* (12.37%).

In general, the task of fine-grained entity framing is a challenging one, due to several factors. First, the taxonomy comprises 22 fine-grained archetypes, several of which are semantically close (e.g., *Victim* vs. *Exploited*, or *Instigator* vs. *Saboteur*), requiring models to draw subtle distinctions in narrative function. Second, the benchmark is multilingual and covers languages with significantly different resource levels and linguistic structures. Third, the task is highly context-dependent, as the correct label depends on the surrounding discourse rather than on the entity itself. For instance, the same entity can be assigned multiple co-occurring fine-grained roles, as its portrayal can shift across different mentions within the same article. Additionally, articles are drawn from media discourse on politically sensitive domains, where framing is often subtle and ideologically loaded, making roles harder to detect even for human annotators. This last point is reflected in the moderate inter-annotator agreement (IAA) scores reported by Mahmoud et al. [5], with Krippendorff's α between 0.436 (Russian) and 0.733 (Bulgarian). Besides the difficulty of the task, low IAA also indicates the subjectivity of the task.

4. From Text to Knowledge Graph

To provide LLMs with additional structured knowledge about entity construal, we represent each article from the benchmark presented in Section 3 as a KG, which we automatically construct using the pipeline presented in Fiumanò et al. [2]. Both the pipeline and the resulting KG are designed to support framing analysis. In particular, the KG is structured according to the DORIS ontology, which models the relationship between a textual artifact, such as a news article, the situation it describes, and the discursive interpretation of that situation (i.e., its framing).

While DORIS is not bound to any specific interpretative framework, we have shown that adopting Fillmore's Frame Semantics [29] is well suited to framing analysis. According to Frame Semantics, a text evokes abstract conceptual structures, or *Frames*, in which participating entities are assigned semantic roles, or *Frame Elements*. The choice of what frames are activated in a text and what roles each entity is assigned reflects the specific perspective adopted by the author in representing a real-world situation. As such, analysing a text in terms of semantic frames becomes a proxy for framing analysis.

Moreover, other studies have shown that the conceptual structure of Frames can be leveraged to analyse further semantic and pragmatic aspects, such as conceptual metaphors [30] or the diverse perspectives from which a text can be interpreted [31].

More importantly for this work, multilingual models that automatically annotate a document with Frames and Frame Elements, such as LOME, [32] have been proposed in literature. In Fiumanò et al. [2],

we rely on LOME to extract frame-semantic information from text, and then produce a theory-grounded KG through the use of the SPARQL Anything tool [33].

We apply this pipeline to the Dev split documents introduced in Section 3, obtaining a KG containing 669,519 triples, with an average of 128 frames annotated per document. Specifically, we rely on named graphs to explicitly maintain the relationship between a document and its corresponding KG [34]. This approach allows us to simplify queries in the KG, allowing users to perform document-bound analyses as well as analyses across different documents. We observe that `Political_locales`, `Leadership`, and `Hostile_encounter` are among the most commonly annotated frames, which aligns with the benchmark's focus on politics and conflict.

In Fig. 1 we report an excerpt of a document extracted from the dataset of Section 3, with frame semantic annotations extracted using LOME. In particular, the frame `Desiring` is evoked by the lexical unit "interested" with "The Russian president" being classified by the frame element `Experiencer`, and "in halting the fighting" by `Event`.

In Fiumanò et al. [2], we showed that DORIS-based KGs can be used to analyse entity framing by manually inspecting the roles taken by an entity in one or more texts. In the example before, a domain expert could follow this approach and hypothesize that the Russian president is framed as a peacemaker since, according to Frame Semantics, it is interpreted as the entity whose desire is to halt a fight. Building on this, in this paper we investigate how to scale this analysis by testing whether LLMs can leverage the frame-semantic information encoded in the KG to correctly assign fine-grained roles to entities.

5. From Knowledge Graph to Role Portrayal Analysis

In this section, we describe our methodology and experimental setup for LLM-driven entity framing analysis via entity role classification on the benchmark presented in Section 3.

Following the same setting as Mahmoud et al. [5], we frame the task as a multi-label classification problem. Given a text, an entity mention, and its corresponding span, the model has to select one or more entity roles for that entity. As introduced in Section 1, we evaluate LLMs under two main settings: a baseline setting, where the model is tasked to classify an entity based solely on the input text, and a KG-enhanced setting, where the model is additionally provided with information from the KG presented in Section 4.

For the latter, we experiment with three knowledge injection strategies, which we describe in detail in Section 5.1, alongside the rest of our experiments, which aim to answer **RQ1** and **RQ2** from Section 1.

5.1. Experimental Setup

The aim of our experiments is to evaluate the impact of triple injection on the task of fine-grained entity role classification, along three dimensions: overall model performance, prediction stability across repeated runs, and the extent to which predictions diverge between the baseline and KG-enhanced settings.

For a comprehensive evaluation, we experiment with 6 LLMs of different sizes: Gemma 4 E4b and 26B [35], Mistral small 3.2 24b [36], Mistral 7B [37], Qwen 3 30B and Qwen 3 4b [38]. Each model is executed on an NVIDIA RTX3080 GPU with 24GB of VRAM using the default configuration provided by Ollama[1]. We rely on model quantization for Gemma 4 26B, Mistral Small 3.2 and Qwen 3 30B[2].

Following Mahmoud et al. [5], we measure performance using precision, recall, micro- and macro-averaged F1 scores and exact match ratio. To mitigate the influence of the stochastic nature of LLMs on our results, we report all metrics averaged over three independent runs.

[1] https://github.com/ollama/ollama

[2] In particular, we use https://ollama.com/library/gemma4:26B-a4b-it-q4_K_M, https://ollama.com/library/mistral-small3.2:24b-instruct-2506-q4_K_M and https://ollama.com/library/qwen3:30B-a3b-instruct-2507-q4_K_M

Instruction	You are a powerful multi-label classifier. Given a text and an entity within the text, you have to classify the entity into one or two of the following fine-grained roles: Victim, Guardian, Foreign_Adversary, Virtuous, Instigator, Peacemaker, Incompetent, Tyrant, Conspirator, Rebel, Deceiver, Terrorist, Underdog, Corrupt, Exploited, Saboteur, Bigot, Traitor, Forgotten, Scapegoat, Martyr, Spy.
Task definition	This is the text: "..." this is the entity in the text: ...
Structured knowledge	Refine your decision based on these relevant semantic frames in the text: <structured data>
Instruction	Reply ONLY with the correct narrative role(s). Do not add any comment or text.

Figure 2: Structure of the prompt used to perform entity role classification. The *Instruction* and *Task definition* parts are shared across all approaches. The *Structured knowledge* part is specific to KG-enhanced approaches.

Fig. 2 shows the prompt used for classification, which we organize into three conceptual sections: *instruction* and *task definition*, which provide the LLM with the operative definition of the task to be performed, and *structured knowledge*, which provides the information extracted from the KG.

We note that, in both settings, we ask the model to reply with only one or two roles, reflecting human annotation behaviour observed on the dataset. Indeed, in the benchmark, only 8% of the total number of entities are labelled with more than one role, and only 0.3% with more than two roles. This design choice therefore has minimal impact on the best performance attainable by the model.

As for the `<structured data>` slot, which indicates knowledge extracted from the KG, we populate it differently depending on the knowledge injection strategy being tested, which we describe below.

We openly share the code for our experiments at https://github.com/beatrice-f/DORIAN.

5.1.1. Knowledge injection strategies

Deciding how to query the KG and how to incorporate the result into the prompt has a considerable influence on model performance. On one hand, injecting irrelevant or unrelated information might misguide the LLM, biasing its decision process [39, 40]. On the other hand, even when assuming access to the optimal structured knowledge, different verbalisation strategies can substantially influence model output [41].

To account for these two factors, we test three knowledge injection strategies: (i) injecting all salient triples; (ii) injecting only the salient triples where the entity fills one or more frame semantic roles; and (iii) injecting entity-centric salient triples in a verbalised form.

Salient triples Concerning the injection of non-relevant knowledge, during the testing phase we observed that a significant number of frames in our KGs are not directly relevant to framing analysis, such as the `Calendric_unit` or `Cardinal_numbers` frames, which capture generic temporal or numeric mentions rather than domain-specific content. This is due to the automated nature of the frame parsing process, which is designed to detect all frames activated in a text, regardless of their relevance to the domain.

To address this issue, we follow the approach of Minnema et al. [31], who identify a set of *salient frames*, defined as frames relevant to the domain at hand. While manual identification is feasible in narrow-scope settings, our broad domain coverage makes this effort impractical. We hence propose to automatically identify the set of salient frames based on the semantic type of their Frame Elements. The semantic type of a Frame Element specifies the category of the entity that fills that role. For instance, a Frame Element of semantic type `Human` constrains its filler entity to be a person. Other examples of semantic types include `Physical_entity`, `Event`, and `Goal`. By filtering frames according

to framing-relevant semantic types, we exclude less informative frames, such as the ones mentioned above.

```
PREFIX fstbox: <https://w3id.org/framester/framenet/tbox/>
PREFIX stype: <https://w3id.org/framester/framenet/abox/semType/>
SELECT DISTINCT ?frame
WHERE{
  ?fe a fstbox:FrameElement ;
      fstbox:hasSemType ?semtype ;
      fstbox:frameElementOf ?frame .
  VALUES ?semtype { stype:Sentient stype:Artifact stype:Human stype:Source stype:Animate_being
    stype:Group stype:Living_thing stype:Event stype:Organization } .
}
```

Query 1: SPARQL query used to extract the salient set of frames from Framester.

This allows us to extract a set of 617 salient frames via a SPARQL query on Framester [42], reported in Listing 1, selecting semantic types that denote animate or event-like entities, as these are more likely to correspond to the actors and happenings central to framing analysis.

Entity-centric salient triples To further refine the injected knowledge to include only key information on the entity, in strategy (ii) we only keep salient triples where the entity to be categorized fills one or more frame elements. This way, we ensure that no non-informative knowledge is provided to the LLM, minimizing distractions.

```
PREFIX  dul: <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#>
PREFIX  :   <http://w3id.org/doris/schema/>
SELECT ?eventId ?eventFrame ?frameAct ?argument ?argumentRole WHERE {
GRAPH <document_id> {
    { SELECT ?eventId WHERE {
        ?eventId a :SituationEncoding ; dul:hasPart ?encodedPart .
        ?encodedPart :encodes [ :hasRealization [ rdfs:label ?entityLabel ] ] .
        FILTER(CONTAINS(?entityLabel, <entity_mention>))
    }}
    ?eventId dul:hasPart ?encodedPart ;
        :encodes [ :hasRealization [ rdfs:label ?frameAct ] ] ;
        :hasDecoding [ :decodedAs ?eventFrame ] .
    ?encodedPart :encodes [ :hasRealization [ rdfs:label ?argument ]] ;
        :hasDecoding [ :decodedAs ?argumentRole ] .
    }
}
```

Query 2: SPARQL query used to extract entity-centric triples given the mention `<entity_mention>`.

Similarly to salient triples, we rely on the SPARQL query in Listing 2 to extract entity-centric knowledge based on the DORIS ontology.

Verbalized entity-centric salient triples Lastly, we hypothesise that injecting triples in a structured, machine-readable format, such as Turtle, might affect the model's performance. For this reason, we experiment with verbalizing the previously described entity-centric triples. To do so, we verbalise a semantic frame F anchored to the text t_f with its frame element FE_1 (resp. $FE_2, \cdots FE_n$) classifying the text t_1 (resp. $t_2, \cdots, t_n$) as

$$[F]"t_f" \rightarrow "t_1"[FE_1] \mid "t_2"[FE_2] \mid \cdots$$

For example, the annotations from Fig. 1 are verbalised as

$$\begin{aligned}[\texttt{Desiring}]\text{"interested"} \rightarrow &\text{"The Russian president"}[\texttt{Experiencer}] \mid \\ &\text{"in halting the fighting [...]"}[\texttt{Event}]\end{aligned}$$

5.2. Results

5.2.1. Does structured knowledge improve entity framing analysis?

Model	Prompt	Precision ↑	Recall ↑	Micro F1 ↑	Macro F1 ↑	EMR ↑
Fane et al. [15]		–	–	–	–	0.44
Caballero et al. [16]		–	–	–	–	0.33
Gemma 4 26B	Zero Shot	0.43 ±0.00	**0.58** ±0.00	**0.49** ±0.00	0.36 ±0.00	0.34 ±0.00
	Triples	**0.43** ±0.00	0.56 ±0.01	0.49 ±0.00	0.35 ±0.01	**0.35** ±0.01
	Triples Filt.	0.42 ±0.00	0.57 ±0.00	0.48 ±0.00	0.35 ±0.00	0.33 ±0.01
	Triples Filt. Verb.	0.42 ±0.01	0.58 ±0.01	0.49 ±0.01	0.34 ±0.02	0.32 ±0.01
Gemma 4 E4b	Zero Shot	0.34 ±0.00	0.53 ±0.01	0.41 ±0.00	0.31 ±0.01	0.18 ±0.01
	Triples	0.34 ±0.01	0.54 ±0.01	0.42 ±0.01	0.31 ±0.01	0.20 ±0.01
	Triples Filt.	0.35 ±0.01	0.55 ±0.00	0.43 ±0.01	0.30 ±0.01	0.20 ±0.01
	Triples Filt. Verb.	0.35 ±0.01	0.54 ±0.01	0.42 ±0.01	0.30 ±0.02	0.21 ±0.00
Mistral Small 3.2 24b	Zero Shot	0.31 ±0.00	0.52 ±0.01	0.39 ±0.00	0.28 ±0.01	0.13 ±0.00
	Triples	0.32 ±0.00	0.51 ±0.00	0.39 ±0.00	0.26 ±0.00	0.15 ±0.00
	Triples Filt.	0.33 ±0.00	0.54 ±0.00	0.41 ±0.00	0.29 ±0.00	0.15 ±0.01
	Triples Filt. Verb.	0.35 ±0.00	0.55 ±0.00	0.43 ±0.00	0.30 ±0.01	0.16 ±0.01
Mistral 7B	Zero Shot	0.18 ±0.00	0.34 ±0.00	0.23 ±0.00	0.15 ±0.01	0.02 ±0.01
	Triples	0.17 ±0.02	0.32 ±0.02	0.22 ±0.02	0.13 ±0.03	0.03 ±0.01
	Triples Filt.	0.19 ±0.01	0.34 ±0.01	0.24 ±0.01	0.16 ±0.01	0.03 ±0.01
	Triples Filt. Verb.	0.17 ±0.01	0.33 ±0.02	0.23 ±0.01	0.14 ±0.01	0.02 ±0.00
Qwen3 30B	Zero Shot	0.36 ±0.01	0.55 ±0.01	0.43 ±0.01	0.31 ±0.01	0.22 ±0.01
	Triples	0.35 ±0.00	0.55 ±0.01	0.43 ±0.00	0.30 ±0.01	0.21 ±0.00
	Triples Filt.	0.35 ±0.01	0.55 ±0.01	0.43 ±0.01	0.29 ±0.01	0.20 ±0.01
	Triples Filt. Verb.	0.36 ±0.00	0.56 ±0.01	0.44 ±0.00	0.30 ±0.01	0.22 ±0.01
Qwen 3 4b	Zero Shot	0.36 ±0.01	0.47 ±0.01	0.41 ±0.01	0.27 ±0.01	0.29 ±0.01
	Triples	0.33 ±0.00	0.46 ±0.01	0.39 ±0.00	0.25 ±0.01	0.23 ±0.01
	Triples Filt.	0.35 ±0.01	0.48 ±0.02	0.40 ±0.02	0.25 ±0.02	0.25 ±0.02
	Triples Filt. Verb.	0.34 ±0.00	0.47 ±0.00	0.39 ±0.00	0.24 ±0.01	0.24 ±0.01

Table 2
Results obtained by averaging three distinct runs of each LLM by relying on zero shot prompting and injecting the triples in the prompt, optionally filtering them and verbalizing the filtered subset. Best overall results are reported in bold. Best results for each model are underlined. Exact Match Ratio (EMR) for Fane et al. [15] and Caballero et al. [16] is taken from their respective papers.

In Table 2 we report the results obtained by the different models in the narrative classification task. Additionally, we report the Exact Match Ratio (EMR) of two baseline approaches for comparison: Fane et al. [15], which employs a multi-turn prompting strategy, and the approach of Caballero et al. [16], which employs a Chain-of-thought prompting strategy. Both approaches rely on GPT-4o, which has orders of magnitude more parameters than the open models we rely on [43]. Finally, in Table 3 we report the micro F1 scores for each language.

In general, we observe that the best performing model, Gemma 4 26B, achieves comparable results with and without the injection of knowledge extracted from the KG, across all languages. A similar pattern can be observed for the other models tested. These results provide a negative answer to **RQ1**, suggesting that, unlike tasks addressed by other works (e.g., Cuccarini et al. [4]), entity framing analysis does not straightforwardly benefit from structured knowledge injection. Performance remains comparable regardless of whether triples are filtered by the target entity or not, and regardless of whether they are verbalised or not.

We argue that this result stems from two factors. First, as observed earlier, entity framing analysis is a challenging task that requires nuanced, contextual knowledge that the open LLMs tested in Table 2

LLM	Prompt	BG 🇧🇬	EN 🇬🇧	HI 🇮🇳	PT 🇵🇹	RU 🇷🇺
Gemma 4 26B	Zero Shot	**0.28** ±0.01	**0.32** ±0.01	0.24 ±0.02	0.28 ±0.01	**0.33** ±0.04
	Triples	0.27 ±0.01	0.29 ±0.04	0.23 ±0.00	0.29 ±0.00	0.28 ±0.03
	Filt. Triples	0.27 ±0.03	0.31 ±0.01	0.23 ±0.00	0.28 ±0.01	0.27 ±0.01
	Verb. Filt. Triples	0.27 ±0.02	0.29 ±0.05	0.24 ±0.01	**0.31** ±0.01	0.26 ±0.03
Gemma 4 E4b	Zero Shot	0.26 ±0.02	0.24 ±0.03	0.23 ±0.01	0.22 ±0.02	0.29 ±0.04
	Triples	0.24 ±0.00	0.21 ±0.01	0.25 ±0.03	0.25 ±0.03	0.26 ±0.05
	Filt. Triples	0.27 ±0.02	0.25 ±0.09	**0.25** ±0.01	0.25 ±0.01	0.26 ±0.03
	Verb. Filt. Triples	0.26 ±0.02	0.23 ±0.02	0.23 ±0.03	0.27 ±0.05	0.27 ±0.05
Mistral Small 3.2 24b	Zero Shot	0.18 ±0.00	0.19 ±0.01	0.21 ±0.01	0.21 ±0.01	0.23 ±0.02
	Triples	0.25 ±0.01	0.17 ±0.01	0.20 ±0.00	0.19 ±0.00	0.20 ±0.01
	Filt. Triples	0.27 ±0.02	0.17 ±0.01	0.22 ±0.00	0.20 ±0.02	0.25 ±0.00
	Verb. Filt. Triples	0.25 ±0.02	0.22 ±0.01	0.23 ±0.00	0.21 ±0.00	0.23 ±0.01
Mistral 7B	Zero Shot	0.10 ±0.02	0.15 ±0.03	0.12 ±0.00	0.15 ±0.01	0.10 ±0.02
	Triples	0.11 ±0.04	0.11 ±0.01	0.10 ±0.02	0.09 ±0.01	0.09 ±0.01
	Filt. Triples	0.15 ±0.02	0.18 ±0.01	0.10 ±0.01	0.14 ±0.00	0.12 ±0.01
	Verb. Filt. Triples	0.13 ±0.05	0.15 ±0.03	0.09 ±0.00	0.14 ±0.02	0.13 ±0.04
Qwen3 30B	Zero Shot	0.25 ±0.01	0.24 ±0.04	0.24 ±0.01	0.24 ±0.03	0.26 ±0.02
	Triples	0.23 ±0.02	0.25 ±0.01	0.24 ±0.02	0.24 ±0.00	0.21 ±0.01
	Filt. Triples	0.23 ±0.04	0.25 ±0.03	0.25 ±0.01	0.23 ±0.02	0.21 ±0.02
	Verb. Filt. Triples	0.26 ±0.01	0.25 ±0.02	0.24 ±0.01	0.23 ±0.00	0.23 ±0.02
Qwen3 4b	Zero Shot	0.17 ±0.03	0.19 ±0.01	0.23 ±0.01	0.25 ±0.02	0.18 ±0.01
	Triples	0.20 ±0.02	0.18 ±0.02	0.23 ±0.00	0.21 ±0.01	0.15 ±0.02
	Filt. Triples	0.22 ±0.05	0.17 ±0.01	0.22 ±0.02	0.23 ±0.01	0.16 ±0.00
	Verb. Filt. Triples	0.20 ±0.05	0.19 ±0.02	0.21 ±0.02	0.20 ±0.03	0.15 ±0.02

Table 3
Language specific micro F1 results obtained by averaging three distinct runs of each LLM using the same setting in Table 2. Best overall results are report in bold. Best results for each model are underlined.

may not be able to fully leverage. Second, although our prior work showed that constructing a KG according to the pipeline presented in Section 4 enables manual framing analysis, it is possible that noisy predictions produced by LOME negatively impact the performance. We also note, however, that, beyond noise, the effectiveness of knowledge injection may also depend on the type and content of the knowledge that is provided.

While results in Tables 2 and 3 show the impact of triple injection on model performance, they do not inform us on how the KG-augmented prompt shapes LLMs' decision process. To answer **RQ2**, in the next section we analyse the models' predictions through a probabilistic lens.

5.2.2. How does structured knowledge affect LLM predictions?

To address **RQ2**, we start from a simple consideration: if an LLM is influenced by the KG-augmented setting, this effect should be observable in its predictions, beyond their correctness with respect to a gold standard. For instance, the LLM might shift its preference toward different entity roles when augmented with knowledge from the KG, or it might become more uncertain about which role best fits an entity, predicting different roles across repeated runs. In this section, we investigate both hypotheses: the first through a direct comparison between the predictive distributions of the baseline and KG-enhanced settings (using Hellinger distance), and the second through the entropy of each setting's predictive distribution.

To do so, let $\mathbf{R}$ denote the set of 22 fine-grained roles introduced by Mahmoud et al. [5], and assume that an LLM has a latent probability distribution $\mathbf{P}(\mathbf{R}|t)$ over entity roles for a given entity in an input text t. It follows that the roles predicted for that entity can be considered as a sample drawn from that

distribution, i.e., $Pred \sim \mathbf{P}(\mathbf{R}|t)$ with $Pred \subset \mathbf{R}$. If the distribution $\mathbf{P}$ has a low degree of uncertainty, we should observe consistent answers from the LLM across repeated runs. Conversely, if the distribution $\mathbf{P}$ has a high degree of uncertainty, we will observe a high variability between predictions across repeated runs.

We follow this observation to quantitatively measure whether the model's predictions are influenced by the KG-augmented prompt.

We estimate the probability that $\mathbf{P}(\mathbf{R}|t)$ by counting the number of times that an LLM predicts the entity role r when prompted with text t, normalised by the total number of (possibly repeated) roles predicted by the LLM across three repeated runs[3].

We then measure the degree of uncertainty of $\mathbf{P}$ by computing its entropy $H(\mathbf{P})$. This allows us to understand whether injecting triples into the prompt influences the variability of the LLM's output and, more generally, which LLMs display a higher degree of uncertainty.

LLM	Prompt	Avg	BG 🇧🇬	EN 🇬🇧	HI 🇮🇳	PT 🇵🇹	RU 🇷🇺
Gemma 4 26B	Zero Shot	0.496	0.559	0.709	0.425	0.308	0.478
	Triples	0.481	0.528	0.683	0.416	0.265	0.512
	Filt. Triples	0.496	0.578	0.710	0.459	0.262	0.472
	Verb. Filt. Triples	0.501	0.582	0.697	0.475	0.291	0.458
Gemma 4 E4b	Zero Shot	0.852	0.915	1.031	0.798	0.701	0.814
	Triples	0.831	0.890	1.022	0.779	0.709	0.756
	Filt. Triples	0.810	0.857	1.030	0.794	0.653	0.716
	Verb. Filt. Triples	0.795	0.797	1.013	0.796	0.650	0.719
Mistral Small 3.2 24b	Zero Shot	0.628	0.622	0.687	0.573	0.633	0.623
	Triples	0.594	0.575	0.730	0.566	0.519	0.581
	Filt. Triples	0.585	0.637	0.705	0.575	0.534	0.476
	Verb. Filt. Triples	0.565	0.582	0.697	0.558	0.543	0.447
Mistral 7B	Zero Shot	1.105	1.037	1.166	1.123	1.088	1.110
	Triples	1.251	1.140	1.304	1.247	1.228	1.335
	Filt. Triples	1.042	0.926	1.070	1.082	1.005	1.125
	Verb. Filt. Triples	1.096	0.961	1.161	1.116	1.086	1.158
Qwen 3 30B	Zero Shot	0.750	0.783	0.877	0.715	0.581	0.793
	Triples	0.739	0.801	0.843	0.749	0.554	0.749
	Filt. Triples	0.764	0.831	0.853	0.767	0.604	0.763
	Verb. Filt. Triples	0.749	0.738	0.893	0.754	0.565	0.793
Qwen 3 4b	Zero Shot	0.566	0.566	0.702	0.631	0.406	0.527
	Triples	0.666	0.625	0.899	0.721	0.491	0.596
	Filt. Triples	0.652	0.674	0.807	0.737	0.589	0.452
	Verb. Filt. Triples	0.638	0.635	0.809	0.716	0.555	0.475

Table 4
Entropy computed from the probability distribution $\mathbf{P}$. Higher values indicate a higher uncertainty of the answer.

Table 4 reports the value of $H(\mathbf{P})$ for each LLM in different settings and across the five languages. Unlike the results in Tables 2 and 3, we observe diverse patterns on $H(\mathbf{P})$ depending on the model, the prompt and the language of the documents. For example, Gemma 4 26B, the best performing method in Table 2, does not display a significant change in the estimated entropy with or without KG-enhanced prompting. Differently, Qwen 3 30B displays a lower entropy when prompted using all the triples related to a document, indicating that theyinfluence its decision process. Finally, we highlight that Mistral 7B displays the highest entropy on the classified fine-grained roles. This provides an interpretation over it

[3]Note that this corresponds to estimating $\mathbf{P}(\mathbf{R}|t)$ using the Monte Carlo method with 3 samples by assuming conditional independence between entity roles, $\mathbf{P}(\mathbf{R}|t) = \prod_r \mathbf{P}(\mathbf{R} = r|t)$. We rely on Laplace smoothing to cope with the sparsity of the predictions.

being the worst performing model in Table 2: the model does not strongly commit to an answer when analysing the role of an entity.

The findings of Table 4 provide an initial answer to **RQ2**: the influence of structured knowledge on LLMs depends on the specific model at hand as well as the prompting strategy that is being used. What remains to understand is whether, assuming an LLM is influenced by the KG-augmented prompt, there is a noticeable difference in its decisions.

We propose to answer these questions by comparing the similarity of the probability distribution $\mathbf{P}$ of an LLM being prompted with or without structured knowledge. We measure this similarity by using the Hellinger distance. Intuitively, if an LLM is highly influenced by a KG-augmented prompt, its distribution $\mathbf{P}_{KG}$ will diverge (i.e., it is dissimilar) from its $\mathbf{P}_{ZS}$ when prompted by only using the textual document[4].

LLM	Triples	Filtered Triples	Verbalized Filtered Triples
Gemma 4 26B	0.198	0.189	0.188
Gemma 4 E4b	0.382	0.358	0.350
Mistral Small 3.2 24b	0.294	0.287	0.259
Mistral 7B	0.597	0.562	0.555
Qwen 3 30B	0.316	0.317	0.309
Qwen 3 4b	0.373	0.368	0.371

Table 5
Hellinger distance computed between the probability distributions induced by the zero shot and KG-augmented settings. Higher values indicate a larger shift in model's predictions, thus a higher influence of the triples injection.

Table 5 reports the pairwise Hellinger distance between the KG-augmented settings and the baseline approach. Similarly to the entropy in Table 4, we observe that not all LLMs are equally influenced by the use of KG-augmented prompting.

Notably, Mistral 7B, which displayed the highest entropy in Table 4, is the model whose distribution diverges the most when prompted using structured knowledge. This complements the previous answer to **RQ2**: although not all models are equally influenced by KG-augmented prompting, some models display a significantly different behaviour.

6. Discussion

In Section 5, we addressed **RQ1** to assess whether triple injection improves LLM performance on entity role classification, and **RQ2** to assess the impact of KG-based augmentation on prediction behaviour. Our results indicate that, overall, triple injection yields results comparable to the baseline, providing a negative answer to RQ1. However, the picture is more nuanced.

While knowledge injection does not yield a consistent improvement across models and languages (Tables 2 and 3), its effect is not uniform. Mistral Small 3.2 24b stands out as the model that mostly benefits from structured knowledge, improving across some of the metrics with respect to the baseline setting. Conversely, Qwen3 4b is harmed by injection, and the remaining models are largely indifferent. This heterogeneity suggests that the value of structured knowledge injection also depends on model size and architecture, as well as input language. For instance, the KG-augmented setting improves Gemma 4 E4b performance in every language except for Russian, a language which generally seems not to benefit from injection across most of the models. Conversely, English and Bulgarian are the most positively affected by the augmented setting. While these differences highlight that the benefit of injected knowledge is also affected by the model's multilingual competence, this may also depend on the quality of language-specific frame-semantic parsing.

[4]We define the similarity $sim(\mathbf{P}_{ZS}, \mathbf{P}_{KG}) = 1 - H(\mathbf{P}_{ZS}, \mathbf{P}_{KG})$ with $H(\mathbf{P}_{ZS}, \mathbf{P}_{KG})$ the Hellinger distance.

More generally, we observe that filtering out non-entity-relevant triples tends to help more than unfiltered injection, pointing to the importance of salience in RAG approaches, and indicating that noisy or redundant triples can misguide, rather than support, the LLM.

The entropy analysis (Table 4) offers a complementary lens on this heterogeneity. Gemma 4 26B, the best-performing model overall, is also the most confident, and its entropy is almost unaffected by knowledge injection, suggesting it neither relies on nor is distracted by the injected triples. A similar behaviour is observed for the largest model, Qwen 3 30B. Mistral Small 3.2 24b and Gemma 4 E4b, by contrast, become both more confident and more accurate under verbalised, filtered injection, indicating that in this case the model is able to productively exploit the additional knowledge. On the other hand, both smaller models Mistral 7B and Qwen 3 4b are more uncertain with unfiltered triple injection and partially stabilized by entity-centric filtering. We note, however, that higher entropy should not be read as an inherently negative signal. As defined in Section 1, entity framing is a highly subjective task, as reflected in the moderate inter-annotator agreement reported in Section 3, and a degree of uncertainty in model predictions may simply mirror the genuine ambiguity of the task rather than a failure of the model itself.

The Hellinger distance results in Table 5 corroborate the entropy findings. Mistral 7B shows the largest divergence between its zero-shot and KG-augmented predictions across every knowledge injection strategy (0.56–0.60), followed by models Gemma 4 E4b and Qwen 3 4b, suggesting that knowledge injection has a greater impact on smaller models. Conversely, Gemma 4 26B is the least affected, with predictions shifting only marginally between the two evaluated settings.

Overall, our findings suggest that the benefits of knowledge injection previously reported for other, highly context-dependent tasks [4] do not straightforwardly transfer to entity framing analysis, and that realizing such benefits requires careful model selection, knowledge filtering, and injection strategy. Our entropy and Hellinger distance analyses show that larger models, such as Gemma 4 26B, are largely indifferent to structured knowledge injection, while smaller models are considerably more responsive to it. This makes smaller models a more promising target for future work exploring alternative knowledge types and injection strategies.

7. Conclusion

In this work, we investigated whether LLMs can effectively leverage structured frame-semantic knowledge to scale entity framing analysis beyond the manual, graph-based approach presented in Fiumanò et al. [2]. Using a multilingual benchmark on fine-grained role portrayal and entity framing, we compared six LLMs of varying sizes under a text-only baseline and three KG-enhanced settings, differing in the scope and verbalization of the injected knowledge.

Our results provide a partially negative answer to **RQ1**: structured knowledge injection does not yield a consistent performance improvement across models or languages, with Mistral Small 3.2 24b as a notable exception. Addressing **RQ2**, our entropy and Hellinger distance analyses reveal that this negative result at the accuracy level masks substantial differences in model behaviour: larger models remain mostly indifferent to injected knowledge, while smaller models are more sensitive to it.

Taken together, our findings indicate that the benefits of structured knowledge injection for entity framing analysis are not uniform, but depend on an interplay between model size, architecture, and verbalization strategy. Smaller models are the most responsive to injected knowledge, and thus the most promising direction for future work exploring alternative knowledge representations and injection strategies.

Contribution

According to the CRediT taxonomy[5], BF and NL contributed to Conceptualization, Investigation, Methodology, Software, Visualization. BF, NL and LB contributed to Writing – original draft, Writing –

[5] https://credit.niso.org/

review & editing. VP contributed to Project administration and Funding acquisition.

Acknowledgments

This research is supported by INFINITY: an EU Horizon Europe project under Grant Agreement No 101233051. Beatrice Fiumanò is funded by the National Recovery and Resilience Plan (NRRP), funded by the European Union – NextGenerationEU - Mission 4 "Education and Research", Component 1 "Enhancement of the offer of educational services: from nurseries to universities" - Investment 4.1 "Extension of the number of research doctorates and innovative doctorates for public administration and cultural heritage".(DM 118/2023).

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude Sonnet 5 to improve writing style, grammar and spelling check, and for coding assistance. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] Y. Otmakhova, S. Khanehzar, L. Frermann, Media framing: A typology and survey of computational approaches across disciplines, in: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 15407–15428. doi:10.18653/v1/2024.acl-long.822.

[2] B. Fiumanò, N. Lazzari, S. P. Ponzetto, V. Presutti, Victim or assailant? exploring narratives through knowledge graph queries, in: Proceedings of the 10th Workshop on Linked Data in Linguistics (LDL-2026)@ LREC 2026, 2026, pp. 40–49.

[3] S. De Giorgis, A. Gangemi, A. Russo, Neurosymbolic graph enrichment for grounded world models, Information Processing & Management 62 (2025) 104127. URL: https://www.sciencedirect.com/science/article/pii/S030645732500069X. doi:https://doi.org/10.1016/j.ipm.2025.104127.

[4] M. Cuccarini, L. Draetta, B. Fiumanò, S. Bistarelli, R. Damiano, V. Presutti, Unveiling stereotypes: Combining knowledge graphs and LLMs for implied stereotype generation, in: C. Bosco, E. Jezek, M. Polignano, M. Sanguinetti (Eds.), Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025), CEUR Workshop Proceedings, Cagliari, Italy, 2025, pp. 314–326. URL: https://aclanthology.org/2025.clicit-1.33/.

[5] T. Mahmoud, Z. Xie, D. I. Dimitrov, N. Nikolaidis, P. Silvano, R. Yangarber, S. Sharma, E. Sartori, N. Stefanovitch, G. Da San Martino, J. Piskorski, P. Nakov, Entity framing and role portrayal in the news, in: W. Che, J. Nabende, E. Shutova, M. T. Pilehvar (Eds.), Findings of the Association for Computational Linguistics: ACL 2025, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 302–326. URL: https://aclanthology.org/2025.findings-acl.17/. doi:10.18653/v1/2025.findings-acl.17.

[6] D. Card, A. E. Boydstun, J. H. Gross, P. Resnik, N. A. Smith, The media frames corpus: Annotations of frames across issues, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 438–444. doi:10.3115/v1/P15-2072.

[7] S. Liu, L. Guo, K. Mays, M. Betke, D. T. Wijaya, Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence, in: Proceedings of the 23rd Conference on Computational Natural Language Learning, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 504–514. doi:10.18653/v1/K19-1047.

[8] J. Piskorski, N. Stefanovitch, N. Nikolaidis, G. Da San Martino, P. Nakov, Multilingual multifaceted understanding of online news in terms of genre, framing, and persuasion techniques, in: Pro-

ceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 3001–3022. doi:10.18653/v1/2023.acl-long.169.

[9] D. Gomez-Zara, M. Boon, L. Birnbaum, Who is the hero, the villain, and the victim? detection of roles in news articles using natural language techniques, in: Proceedings of the 23rd International Conference on Intelligent User Interfaces, 2018, pp. 311–315.

[10] L. Frermann, J. Li, S. Khanehzar, G. Mikolajczak, Conflicts, villains, resolutions: Towards models of narrative media framing, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 8712–8732. doi:10.18653/v1/2023.acl-long.486.

[11] D. Stammbach, M. Antoniak, E. Ash, Heroes, villains, and victims, and GPT-3: Automated extraction of character roles without training data, in: Proceedings of the 4th Workshop of Narrative Understanding, Association for Computational Linguistics, Seattle, United States, 2022, pp. 47–56. doi:10.18653/v1/2022.wnu-1.5.

[12] D. G. Hobson, D. Ruths, A. Piper, Evaluating taxonomy-free character role labeling in news stories using large language models, in: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2025. doi:10.18653/v1/2025.emnlp-main.750.

[13] J. Piskorski, T. Mahmoud, N. Nikolaidis, R. Campos, A. M. Jorge, D. Dimitrov, P. Silvano, R. Yangarber, S. Sharma, T. Chakraborty, N. Guimaraes, E. Sartori, N. Stefanovitch, Z. Xie, P. Nakov, G. Da San Martino, SemEval-2025 task 10: Multilingual characterization and extraction of narratives from online news, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 2610–2643. doi:10.18653/v1/2025.semeval-1.331.

[14] T. Mahmoud, Z. Xie, P. Nakov, Bertastic at semeval-2025 task 10: State-of-the-art accuracy in coarse-grained entity framing for hindi news, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 386–396.

[15] E. Fane, M. Surdeanu, E. Blanco, S. Corman, Fane at SemEval-2025 task 10: Zero-shot entity framing with large language models, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 1924–1934. doi:10.18653/v1/2025.semeval-1.251.

[16] A. Caballero, A. Rodrigo, R. Centeno, NlpUned at SemEval-2025 task 10: Beyond training: A taxonomy-guided approach to role classification using LLMs, in: S. Rosenthal, A. Rosá, D. Ghosh, M. Zampieri (Eds.), Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), Association for Computational Linguistics, Vienna, Austria, 2025, pp. 296–301. URL: https://aclanthology.org/2025.semeval-1.42/.

[17] S. Aryal, P. Dhungana, Howard university-ai4pc at semeval-2025 task 10: Ensembling llms for multi-lingual multi-label and multi-class meta-classification, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 1585–1592.

[18] B. Wang, R. Song, X. Wang, G. Shi, L. Hu, H. Huang, C. Feng, gowithnlp at SemEval-2025 task 10: Leveraging entity-centric chain of thought and iterative prompt refinement for multi-label classification, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 373–380. doi:10.18653/v1/2025.semeval-1.53.

[19] V. Saravanan, S. Wilson, cocoa at semeval-2025 task 10: Prompting vs. fine-tuning: A multilevel approach to propaganda classification, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 590–594.

[20] E. Rønningstad, G. Negi, Ltg at semeval-2025 task 10: Optimizing context for classification of narrative roles, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 440–447.

[21] N. Li, Y. Zhang, J. Wang, D. Xu, X. Zhang, Ynu-hpcc at semeval-2025 task 10: A two-stage approach to solving multi-label and multi-class role classification based on deberta, in: Proceedings of the

19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 1993–1999.
[22] Ö. Karabulut, S. Eslami, A. Gharaee, M. Andrews, Tuebingen at semeval-2025 task 10: Class weighting, external knowledge and data augmentation in bert models, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 1248–1254.
[23] J. Liu, Y. Liu, T. Wang, X. Xu, Y. Wang, Y. Jiang, Team qust at semeval-2025 task 10: Evaluating large language models in multiclass multi-label classification of news entity framing, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 1052–1056.
[24] L. Sun, X. Wan, Y. Lin, F. Su, P. Chen, PATeam at SemEval-2025 task 10: Two-stage news analytical framework: Target-oriented semantic segmentation and sequence generation LLMs for cross-lingual entity and narrative analysis, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 2121–2132. doi:10.18653/v1/2025.semeval-1.276.
[25] A. Rajeev, R. Mamidi, adithjrajeev at SemEval-2025 task 10: Sequential learning for role classification using entity-centric news summaries, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 241–246. doi:10.18653/v1/2025.semeval-1.36.
[26] I. Diaz, F. Vázquez, C. Luna, A. Conde, G. Sierra, H. Gómez-Adorno, G. Bel-Enguix, LATE-GIL-nlp at SemEval-2025 task 10: Exploring LLMs and transformers for characterization and extraction of narratives from online news, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 657–665. doi:10.18653/v1/2025.semeval-1.92.
[27] R. Raghav, A. P. Vemali, D. Aswal, R. Ramesh, P. Tusham, P. Rishi, TartanTritons at SemEval-2025 task 10: Multilingual hierarchical entity classification and narrative reasoning using instruct-tuned LLMs, in: Proceedings of the 19th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 1964–1973. doi:10.18653/v1/2025.semeval-1.255.
[28] T. Lv, J. Li, C. Liu, Y. Kang, L. Luo, Y. Sun, H. Lin, Dutir at semeval-2025 task 10: A large language model-based approach for entity framing in online news, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 174–179.
[29] C. J. Fillmore, Frame Semantics and the Nature of Language, Annals of the New York Academy of Sciences 280 (1976) 20–32. URL: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1976.tb25467.x. doi:https://doi.org/10.1111/j.1749-6632.1976.tb25467.x. arXiv:https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1976.tb25467.x.
[30] Y. Otmakhova, M. Guida, L. Frermann, Not all ANIMALs are equal: metaphorical framing through source domains and semantic frames, in: M. Liakata, V. P. Moreira, J. Zhang, D. Jurgens (Eds.), Findings of the Association for Computational Linguistics: ACL 2026, Association for Computational Linguistics, San Diego, California, United States, 2026, pp. 38334–38355. URL: https://aclanthology.org/2026.findings-acl.1911/. doi:10.18653/v1/2026.findings-acl.1911.
[31] G. Minnema, S. Gemelli, C. Zanchi, T. Caselli, M. Nissim, SocioFillmore: A tool for discovering perspectives, in: V. Basile, Z. Kozareva, S. Stajner (Eds.), Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 240–250. URL: https://aclanthology.org/2022.acl-demo.24/. doi:10.18653/v1/2022.acl-demo.24.
[32] P. Xia, G. Qin, S. Vashishtha, Y. Chen, T. Chen, C. May, C. Harman, K. Rawlins, A. S. White, B. Van Durme, LOME: Large ontology multilingual extraction, in: D. Gkatzia, D. Seddah (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Online, 2021, pp. 149–159. URL: https://aclanthology.org/2021.eacl-demos.19/. doi:10.18653/v1/2021.eacl-demos.19.
[33] L. Asprino, E. Daga, A. Gangemi, P. Mulholland, Knowledge graph construction with a façade: A unified method to access heterogeneous data sources on the web, ACM Transactions on Internet

Technology 23 (2023) 1–31. URL: https://doi.org/10.1145/3555312. doi:10.1145/3555312.

[34] J. J. Carroll, C. Bizer, P. Hayes, P. Stickler, Named graphs, Journal of Web Semantics 3 (2005) 247–267. URL: https://www.sciencedirect.com/science/article/pii/S1570826805000235. doi:https://doi.org/10.1016/j.websem.2005.09.001, world Wide Web Conference 2005——Semantic Web Track.

[35] G. Team, S. E. Abd, V. Aggarwal, R. Algayres, A. Andreev, O. Bachem, I. Ballantyne, C. Brick, V. Cărbune, M. Casbon, M. Chaturvedi, V. Cotruta, A. Coucke, P. Culliton, R. Dadashi, L. Dixon, M. Elhawaty, U. Evci, C. Farabet, J. Ferret, F. Galgani, S. Girgin, J.-B. Grill, M. Grootendorst, J. Guo, C. Hardin, Y. He, S. M. Hernandez, O. Homburger, L. Hussenot, J. Ji, A. Joulin, A. Kamath, P. Kassraie, O. Lacombe, P. Lahoti, G. Liu, G. Martins, L. Martins, T. Matejovicova, R. Merhej, N. Momchev, S. Mondal, R. Mullins, S. R. Panyam, S. Pathak, S. Perrin, A. S. Pinto, E. Pot, A. Pouget, A. Ramé, S. Ramos, D. Reid, D. Rim, M. Rivière, K. Roth, L. Rouillard, O. Sanseviero, P. G. Sessa, S. Settle, D. Sinopalnikov, S. Smoot, P. Stanczyk, A. Steiner, L. Stewart, I. Tolstikhin, M. Tschannen, A. Tsitsulin, N. Vieillard, R. Wu, P. Xu, H. Yang, E. Yvinec, L. Zhang, J. Zou, N. Aagnes, A. Abdelhamed, S. Agrawal, S. Agrawal, I. Alabdulmohsin, J. B. Alayrac, U. Alon, C. Amarnath, A. Anand, C. Anastasiou, S. Ariafar, F.-X. Aubet, K. Axiotis, F. Barbero, J. Barral, A. Bendebury, U. Bergmann, S. Bileschi, K. Black, M. Blondel, S. Borgeaud, A. Bražinskas, R. Burnell, R. Busa-Fekete, M. Cai, G. Cameron, C. Caucheteux, G. Chadha, J. Chan, A. Chawla, B. J. Chen, J. Chen, L. Chen, X. Chen, D. Cheng, T. hsiang Chien, N. Chinaev, Y. Chou, Z. Chu, B. Coleman, P. Consul, S. Conway-Rahman, S. Crowell, D. Cutler, V. Dani, S. Daruki, A. Das, D. Deutsch, N. Dikkala, L. Ding, Q. Ding, S. Dodhia, K. Donhauser, T. Doshi, A. Dragan, A. Druinsky, S. Dua, Z. Egyed, D. Eisenbud, D. Eppens, C. Fan, B. Fatemi, Y. Fathullah, V. Feinberg, M. Ferev, T. Fujimoto, I. Galatzer-Levy, J. Gante, S. Geisler, S. Ghosal, A. M. Girgis, A. Go, A. Gokhale, A. Grills, Y. Gu, P. Gupta, G. Guruganesh, R. Hadsell, H. Harkous, J. Harlalka, D. Hassabis, A. Hauth, J. Heyward, A. Hosseini, C.-Y. Hsia, I.-H. Hsu, X. Huang, Y. Huang, K. Hui, A. Hutter, T. I, F. Iliopoulos, A. Jain, G. Jawahar, Z. Ji, Q. Jin, M. Johnson, K. Joshi, A. Kandoor, W.-C. Kang, K. Kavukcuoglu, M. Kazemi, K. Kenealy, A. Khalifa, P. Kirk, S. Kothawade, V. Kovalev, N. Kovelamudi, A. Kraft, R. Kumar, H. Kuppam, J. Lannin, C.-Y. Lee, S. Lee, D. Lepikhin, D. Li, Q. Li, V. Liévin, E. Lin, Z. Lin, C. Liu, T. Liu, T. Liu, X. Liu, M. Lunayach, M. Ma, G. Madan, A. Maksai, E. Malmi, M. Matuszak, D. McDuff, G. Menghani, D. Mirylenka, K. Misiunas, V. Misra, A. Mitran, K. Mohamed, M. Mukha, E. Noland, J. O'Donnell, K. Olszewska, B. Orlando, W. Pan, R. Panigrahy, U. Parekh, C. Park, E. Paskie, L. Peng, B. Petrini, S. Petrov, J. Pfeiffer, B. Piot, M. Plomecka, S. Poder, O. Ponce, A. Pramanik, D. Racz, A. Rajan, M. Ramanovich, A. Rao, M. Ritter, V. Rodrigues, E. Rosen, M. Rybiński, N. Sachdeva, M. E. Sander, R. Sathyanarayana, S. Savla, S. Schmidgall, T. Schuster, B. Seguin, A. Sellergren, A. Severyn, I. Shafran, D. Shah, Y. Shangguan, A. Shenoy, P. Shenoy, R. Shivanna, P. Sho, L. Spangher, W. Stokowiec, T. Strother, Y. Su, Y. Sun, M. Sundararajan, A. Tacchetti, M. H. Taege, P. Tafti, C. Tekur, R. Thapa, M. Traverse, L. Treven, T. Tu, C. T. Tung, P. Veličković, M. P. Venkat, S. G. Venkatesh, V. Venkiteswaran, F. Visin, A. Vitvitskyi, K. Vodrahalli, W. Wang, X. Wang, T. Warkentin, J. Wassenberg, J. Wieting, L. Xiao, H. Xu, Y. Xu, F. Xue, A. Yadav, J. Yan, A. Yang, L. Yang, M.-H. Yang, Z. Ying, J. H. Yoo, S. Zafar, F. Zhang, J. Zhang, J. Zhang, X. Zhang, C. Zhao, D. Zhou, C. Zou, Gemma 4 technical report, 2026. URL: https://arxiv.org/abs/2607.02770. arXiv:2607.02770.

[36] A. H. Liu, K. Khandelwal, S. Subramanian, V. Jouault, A. Rastogi, A. Sadé, A. Jeffares, A. Jiang, A. Cahill, A. Gavaudan, A. Sablayrolles, A. Héliou, A. You, A. Ehrenberg, A. Lo, A. Eliseev, A. Calvi, A. Sooriyarachchi, B. Bout, B. Rozière, B. D. Monicault, C. Lanfranchi, C. Barreau, C. Courtot, D. Grattarola, D. Dabert, D. de las Casas, E. Chane-Sane, F. Ahmed, G. Berrada, G. Ecrepont, G. Guinet, G. Novikov, G. Kunsch, G. Lample, G. Martin, G. Gupta, J. Ludziejewski, J. Rute, J. Studnia, J. Amar, J. Delas, J. S. Roberts, K. Yadav, K. Chandu, K. Jain, L. Aitchison, L. Fainsin, L. Blier, L. Zhao, L. Martin, L. Saulnier, L. Gao, M. Buyl, M. Jennings, M. Pellat, M. Prins, M. Poirée, M. Guillaumin, M. Dinot, M. Futeral, M. Darrin, M. Augustin, M. Chiquier, M. Schimpf, N. Grinsztajn, N. Gupta, N. Raghuraman, O. Bousquet, O. Duchenne, P. Wang, P. von Platen, P. Jacob, P. Wambergue, P. Kurylowicz, P. R. Muddireddy, P. Chagniot, P. Stock, P. Agrawal, Q. Torroba, R. Sauvestre, R. Soletskyi, R. Menneer, S. Vaze, S. Barry, S. Gandhi, S. Waghjale,

S. Gandhi, S. Ghosh, S. Mishra, S. Aithal, S. Antoniak, T. L. Scao, T. Cachet, T. S. Sorg, T. Lavril, T. N. Saada, T. Chabal, T. Foubert, T. Robert, T. Wang, T. Lawson, T. Bewley, T. Bewley, T. Edwards, U. Jamil, U. Tomasini, V. Nemychnikova, V. Phung, V. Maladière, V. Richard, W. Bouaziz, W.-D. Li, W. Marshall, X. Li, X. Yang, Y. E. Ouahidi, Y. Wang, Y. Tang, Z. Ramzi, Ministral 3, 2026. URL: https://arxiv.org/abs/2601.08584. arXiv:2601.08584.

[37] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mistral 7b, 2023. URL: https://arxiv.org/abs/2310.06825. arXiv:2310.06825.

[38] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, C. Zheng, D. Liu, F. Zhou, F. Huang, F. Hu, H. Ge, H. Wei, H. Lin, J. Tang, J. Yang, J. Tu, J. Zhang, J. Yang, J. Yang, J. Zhou, J. Zhou, J. Lin, K. Dang, K. Bao, K. Yang, L. Yu, L. Deng, M. Li, M. Xue, M. Li, P. Zhang, P. Wang, Q. Zhu, R. Men, R. Gao, S. Liu, S. Luo, T. Li, T. Tang, W. Yin, X. Ren, X. Wang, X. Zhang, X. Ren, Y. Fan, Y. Su, Y. Zhang, Y. Zhang, Y. Wan, Y. Liu, Z. Wang, Z. Cui, Z. Zhang, Z. Zhou, Z. Qiu, Qwen3 technical report, 2025. URL: https://arxiv.org/abs/2505.09388. arXiv:2505.09388.

[39] F. Cuconasu, G. Trappolini, F. Siciliano, S. Filice, C. Campagnano, Y. Maarek, N. Tonellotto, F. Silvestri, The power of noise: Redefining retrieval for rag systems, SIGIR '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 719–729. URL: https://doi-org.ezproxy.unibo.it/10.1145/3626772.3657834. doi:10.1145/3626772.3657834.

[40] W. Shi, H. Chen, J. Li, Y. Zhao, Y. Zhang, Q. Chen, J. Zhang, R. Zhang, J. Zhu, J. Xu, X. Zhou, Making RALM robust to irrelevant contexts via layer knowledge guided attention, in: W. Che, J. Nabende, E. Shutova, M. T. Pilehvar (Eds.), Findings of the Association for Computational Linguistics: ACL 2025, Association for Computational Linguistics, Vienna, Austria, 2025, pp. 3652–3668. URL: https://aclanthology.org/2025.findings-acl.188/. doi:10.18653/v1/2025.findings-acl.188.

[41] C. Ringwald, F. Gandon, C. Faron, F. Michel, H. A. Akl, 12 shades of RDF: impact of syntaxes on data extraction with language models, in: A. Meroño-Peñuela, Ó. Corcho, P. Groth, E. Simperl, V. Tamma, A. G. Nuzzolese, M. Poveda-Villalón, M. Sabou, V. Presutti, I. Celino, A. Revenko, J. Raad, B. Sartini, P. Lisena (Eds.), The Semantic Web: ESWC 2024 Satellite Events - Hersonissos, Crete, Greece, May 26-30, 2024, Proceedings, Part I, volume 15344 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 81–91. URL: https://doi.org/10.1007/978-3-031-78952-6_8. doi:10.1007/978-3-031-78952-6_8.

[42] A. Gangemi, M. Alam, L. Asprino, V. Presutti, D. R. Recupero, Framester: A wide coverage linguistic linked data hub, in: E. Blomqvist, P. Ciancarini, F. Poggi, F. Vitali (Eds.), Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, volume 10024 of *Lecture Notes in Computer Science*, 2016, pp. 239–254. URL: https://doi.org/10.1007/978-3-319-49004-5_16. doi:10.1007/978-3-319-49004-5_16.

[43] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, R. Campbell, A. Cann, B. Carey, C. Carlson, R. Carmichael, B. Chan, C. Chang, F. Chantzis, D. Chen, S. Chen, R. Chen, J. Chen, M. Chen, B. Chess, C. Cho, C. Chu, H. W. Chung, D. Cummings, J. Currier, Y. Dai, C. Decareaux, T. Degry, N. Deutsch, D. Deville, A. Dhar, D. Dohan, S. Dowling, S. Dunning, A. Ecoffet, A. Eleti, T. Eloundou, D. Farhi, L. Fedus, N. Felix, S. P. Fishman, J. Forte, I. Fulford, L. Gao, E. Georges, C. Gibson, V. Goel, T. Gogineni, G. Goh, R. Gontijo-Lopes, J. Gordon, M. Grafstein, S. Gray, R. Greene, J. Gross, S. S. Gu, Y. Guo, C. Hallacy, J. Han, J. Harris, Y. He, M. Heaton, J. Heidecke, C. Hesse, A. Hickey, W. Hickey, P. Hoeschele, B. Houghton, K. Hsu, S. Hu, X. Hu, J. Huizinga, S. Jain, S. Jain, J. Jang, A. Jiang, R. Jiang, H. Jin, D. Jin, S. Jomoto, B. Jonn, H. Jun, T. Kaftan, Łukasz Kaiser, A. Kamali, I. Kanitscheider, N. S. Keskar, T. Khan, L. Kilpatrick, J. W. Kim, C. Kim, Y. Kim, J. H. Kirchner, J. Kiros, M. Knight, D. Kokotajlo, Łukasz Kondraciuk, A. Kondrich, A. Konstantinidis, K. Kosic, G. Krueger, V. Kuo, M. Lampe, I. Lan, T. Lee, J. Leike, J. Leung, D. Levy, C. M. Li, R. Lim, M. Lin, S. Lin, M. Litwin, T. Lopez, R. Lowe, P. Lue, A. Makanju, K. Malfacini, S. Manning,

T. Markov, Y. Markovski, B. Martin, K. Mayer, A. Mayne, B. McGrew, S. M. McKinney, C. McLeavey, P. McMillan, J. McNeil, D. Medina, A. Mehta, J. Menick, L. Metz, A. Mishchenko, P. Mishkin, V. Monaco, E. Morikawa, D. Mossing, T. Mu, M. Murati, O. Murk, D. Mély, A. Nair, R. Nakano, R. Nayak, A. Neelakantan, R. Ngo, H. Noh, L. Ouyang, C. O'Keefe, J. Pachocki, A. Paino, J. Palermo, A. Pantuliano, G. Parascandolo, J. Parish, E. Parparita, A. Passos, M. Pavlov, A. Peng, A. Perelman, F. de Avila Belbute Peres, M. Petrov, H. P. de Oliveira Pinto, Michael, Pokorny, M. Pokrass, V. H. Pong, T. Powell, A. Power, B. Power, E. Proehl, R. Puri, A. Radford, J. Rae, A. Ramesh, C. Raymond, F. Real, K. Rimbach, C. Ross, B. Rotsted, H. Roussez, N. Ryder, M. Saltarelli, T. Sanders, S. Santurkar, G. Sastry, H. Schmidt, D. Schnurr, J. Schulman, D. Selsam, K. Sheppard, T. Sherbakov, J. Shieh, S. Shoker, P. Shyam, S. Sidor, E. Sigler, M. Simens, J. Sitkin, K. Slama, I. Sohl, B. Sokolowsky, Y. Song, N. Staudacher, F. P. Such, N. Summers, I. Sutskever, J. Tang, N. Tezak, M. B. Thompson, P. Tillet, A. Tootoonchian, E. Tseng, P. Tuggle, N. Turley, J. Tworek, J. F. C. Uribe, A. Vallone, A. Vijayvergiya, C. Voss, C. Wainwright, J. J. Wang, A. Wang, B. Wang, J. Ward, J. Wei, C. Weinmann, A. Welihinda, P. Welinder, J. Weng, L. Weng, M. Wiethoff, D. Willner, C. Winter, S. Wolrich, H. Wong, L. Workman, S. Wu, J. Wu, M. Wu, K. Xiao, T. Xu, S. Yoo, K. Yu, Q. Yuan, W. Zaremba, R. Zellers, C. Zhang, M. Zhang, S. Zhao, T. Zheng, J. Zhuang, W. Zhuk, B. Zoph, Gpt-4 technical report, 2024. URL: https://arxiv.org/abs/2303.08774. arXiv:2303.08774.

Protagonist	
Guardian	Heroes who protect values or communities, ensuring safety and upholding justice.
Martyr	Individuals who sacrifice their well-being, or even their lives, for a greater good.
Peacemaker	Individuals who advocate for harmony, resolving conflicts and bringing about peace.
Rebel	Revolutionaries who challenge the status quo and fight for significant change.
Underdog	Entities who, despite a disadvantaged position, strive against greater forces.
Virtuous	Individuals portrayed as righteous, fair, and upholding high moral standards.
Antagonist	
Instigator	Those who initiate conflict and provoke violence or unrest.
Conspirator	Individuals involved in plots and covert activities to undermine or deceive others.
Tyrant	Leaders who abuse their power, ruling unjustly and oppressing others.
Foreign Adversary	Entities from other nations creating geopolitical tension.
Traitor	Individuals who betray a cause or country, seen as disloyal.
Spy	Individuals engaged in espionage, gathering and transmitting sensitive information.
Saboteur	Those who deliberately damage or obstruct systems to cause disruption.
Corrupt	Individuals or entities engaging in unethical or illegal activities for personal gain.
Incompetent	Entities causing harm through ignorance, lack of skill, or poor judgment.
Terrorist	Individuals who engage in violence and terror to further ideological ends.
Deceiver	Manipulators who twist the truth and spread misinformation.
Bigot	Individuals accused of hostility or discrimination against specific groups.
Innocent	
Forgotten	Marginalized groups who are overlooked and ignored by society.
Exploited	Individuals or groups used for others' gain, often without consent.
Victim	People suffering harm due to circumstances beyond their control.
Scapegoat	Entities unjustly blamed for problems or failures to divert attention.

Table 6
The taxonomy of entity roles, taken from Mahmoud et al. [5].