From Text to Triples: A Neuro-Symbolic Pipeline for Structured Claim Extraction

Massimiliano Fadda[1,*], Enrico Motta[2], Francesco Osborne[2,3], Diego Reforgiato Recupero[1] and Angelo Salatino[2]

[1]Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy
[2]Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
[3]Department of Business and Law, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

Abstract

Studying the landscape of news media requires moving beyond simple topic classification toward the identification of specific claims made by agents, including their affiliations and group identities. However, automating this process remains challenging due to the fluid and nuanced nature of journalistic prose, combined with the strict formal requirements imposed by semantic web standards. To address these challenges, we propose a neuro-symbolic pipeline that transforms raw news articles into a Knowledge Graph by leveraging Large Language Models (LLMs) as semantic parsers. Our approach enables the automatic extraction of claims and their corresponding agents, modelled according to the News Classification Ontology (NCO), and supports the large-scale generation of relevant RDF triples. We validate our method on a manually annotated dataset comprising articles about the war in Ukraine, comparing six alternative models. The experimental results are promising, particularly for high-parameter models such as DeepSeek-v3.1, which achieves an F1 score above 76%. At the same time, the results indicate that lightweight architectures yield substantially lower performance, as they struggle to balance semantic reasoning with strict compliance to the target ontology schema.

Keywords

Knowledge Graph Construction, Neuro-symbolic AI, Claim Extraction, Relation Extraction, Large Language Models, Ontology Population

1. Introduction

The analysis of news narratives is fundamental to decoding societal dynamics and public opinion. Far from acting as passive mirrors of reality, media outlets actively construct interpretations of events through selective framing, deeply influencing how citizens perceive facts and conflicting viewpoints [1, 2].

However, the unprecedented volume of online information renders manual analysis unfeasible, creating a pressing need for computational methods capable of mapping the media landscape at scale [3].

To effectively decode these narratives, it is insufficient to simply categorise articles by topic or extract isolated events. The core of journalistic discourse lies in the specific voices that are amplified, politicians, experts, or organisations, and the assertions they make. This brings the task of Structured Claim Extraction to the forefront. In this work, we adopt an agent-centric definition: we define a Claim not merely as a verifiable statement, but as the relational link between an Agent and their linguistic Utterance. Unlike standard event extraction, which focuses on objective metadata (e.g., dates and locations), claim extraction aims to attribute a specific statement to a specific agent. This task is technically demanding because journalistic prose is inherently fluid: claims are rarely presented in a standardized format, but are instead embedded in complex linguistic structures ranging from direct quotations to subtle indirect speech and mixed narratives.

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

*Corresponding author.
massimiliano.fadda@unica.it (M. Fadda); diego.reforgiato@unica.it (D. R. Recupero)
0009-0008-7762-6286 (M. Fadda); 0000-0003-0015-1952 (E. Motta); 0000-0001-6557-3131 (F. Osborne); 0000-0001-8646-6183 (D. R. Recupero); 0000-0002-4763-3943 (A. Salatino)

In this paper, we address this challenge by proposing a neuro-symbolic pipeline designed to structure these fluid narratives into a formal Knowledge Graph (KG). Rather than relying on unconstrained generative methods, our approach utilizes Large Language Models (LLMs) as schema constrained semantic parsers to populate an extension of the News Classification Ontology (NCO) [4]. By treating the ontology as a target schema, we constrain the LLM to model the triadic structure of reported speech, explicitly linking an *Agent* to their linguistic *Utterance* via a *Claim* node. This effectively transforms unstructured text into queryable semantic data.

To evaluate the viability of this approach, we focus our analysis on a specific, high-complexity domain: the war in Ukraine. This choice is driven by the necessity to stress-test the extraction capabilities of current models. As indicated by preliminary analyses, war reporting exhibits a significantly higher density of claims compared to other topics (e.g., climate change), characterised by a rapid succession of conflicting statements from diverse political and military actors.

We evaluate the proposed pipeline across a diverse selection of state-of-the-art LLMs, ranging from proprietary systems to open-weight architectures. This comparative analysis aims to benchmark the intrinsic capability of current models to perform structured claim extraction in a noisy, real-world environment, assessing their readiness for automated ontology population.

In summary, this workshop paper makes the following contributions:

- We propose a methodology for the automated population of the NCO ontology from raw text, treating LLMs as neural parsers for symbolic schemas.
- We validate this approach on a dataset focused on the war in Ukraine, which presents unique challenges in terms of attribution and nested narratives.
- We provide a comparative analysis of multiple LLMs on the claim extraction task, providing insight into their readiness for automated computational journalism.

The remainder of this paper is organised as follows. Section 2 reviews the state of the art. Section 3 details the neuro-symbolic pipeline. Section 4 describes the experimental setup and the Ukraine War dataset. Finally, Section 5 discusses the results and Section 6 provides future directions.

2. Related Work

This section positions our work within the existing literature by examining two distinct research streams: the evolution of computational methods for claim extraction and the development of semantic models for representing journalistic narratives.

2.1. Evolution of Claim Extraction Techniques

Historically, the problem of attributing a statement to a speaker was treated as a relation extraction task dependent on syntactic parsing. While early systems primarily targeted direct quotations, Pareti et al. [5] demonstrated that approximately half of reported speech in news consists of indirect or mixed quotations, which often evade simple syntactic rules. Subsequent approaches attempted to overcome these limitations using sequence labelling techniques [6] and, later, deep learning architectures based on BERT and BiLSTM-CRF [7, 8]. While these neural models improved the recall of non-verbatim attributions, they remained fundamentally extractive, identifying text spans rather than generating structured, normalized representations of the claims.

The emergence of LLMs has introduced a new paradigm: generative information extraction [9]. In particular, these models have demonstrated a strong ability to classify and extract information from complex texts across a wide range of domains, including journalism [10], social media [11, 12], scientific research [13, 14], biomedicine [15], engineering [16], legal studies [17], and finance [18]. An increasing number of these solutions also exploit knowledge graphs as sources of domain knowledge [19], or are capable of generating [20] and refining [21] knowledge graphs. In the specific domain of claim extraction addressed in this paper, recent frameworks such as Claimify [22] and AFEV [23] leverage

the reasoning capabilities of LLMs to decompose complex texts into atomic facts. However, while these approaches provide greater flexibility in handling linguistic ambiguity compared to span-based classifiers, they often lack strict adherence to a predefined schema in the context of news analysis and claim extraction. Our work addresses this limitation by constraining the generative process through a formal ontology, thereby ensuring that the output is not only textually accurate but also structurally valid for Knowledge Graph population.

2.2. Semantic Modelling of News Narratives

To move beyond simple text processing, researchers have developed various ontological schemas to represent news content. The dominant approach has traditionally been event-centric, utilizing standards like the Simple Event Model (SEM) [24] to capture the "who, what, where, when, and why" of reported occurrences. Large scale repositories such as EventKG [25] exemplify this approach.

However, event-centric models often fail to capture the subjective dimension of journalism, specifically the diverse viewpoints and conflicting claims surrounding an event [26]. In the domain of fact-checking, large-scale knowledge graphs such as ClaimsKG [27] have adopted standardised schemas to link claims to veracity verdicts. Yet, these models typically represent the claim as a flat string, lacking the granularity to model the internal structure of the utterance and its precise attribution type.

To bridge this gap, Motta et al. [4] introduced the NCO, which provides a formal distinction between entities, events, and the communicative acts (utterances) that describe them. While NCO offers the necessary theoretical expressivity, the automated population of such complex, nested schemas from raw text remains a significant challenge [28]. Our pipeline directly targets this gap, employing LLMs to instantiate NCO classes from unstructured war reporting.

3. Methodology

We propose a neuro-symbolic pipeline designed to transform unstructured news articles into a structured Knowledge Graph. The architecture follows a retrieve and parse paradigm, where an LLM functions as a semantic parser. Crucially, we replace open ended generation with schema constrained extraction, forcing the model to instantiate specific classes defined by the NCO.

The workflow is illustrated in Figure 1. It follows a neuro-symbolic paradigm where the **Neural Layer** (the extraction engine) is strictly guided by the **Symbolic Layer** (the target schema) to transform unstructured input into validated triples, which are finally instantiated by the **Materialization Layer**.

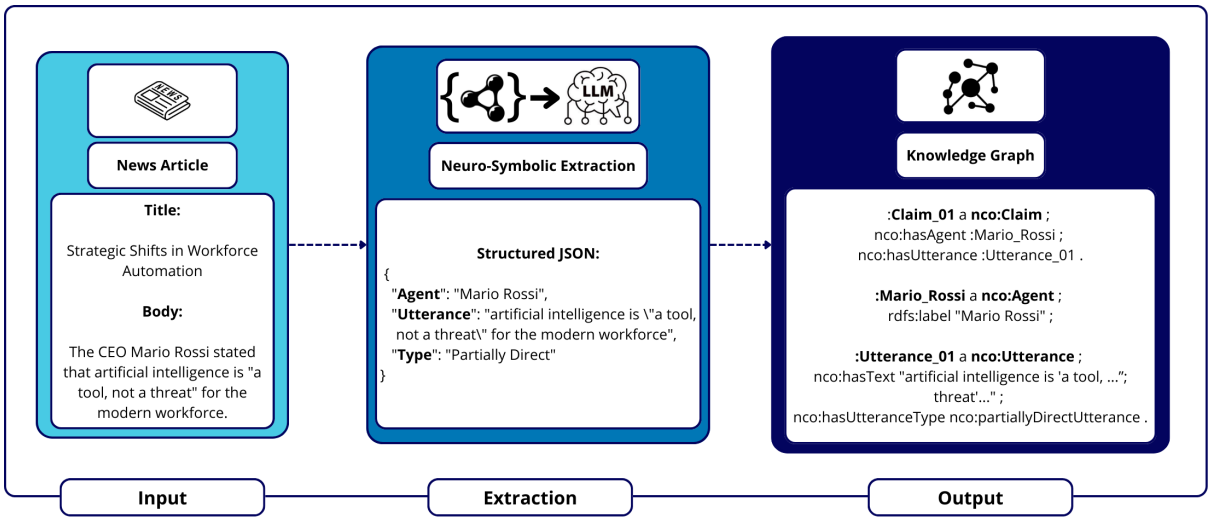

Figure 1: From unstructured news to RDF triples: the neuro-symbolic pipeline. The figure shows the transition from raw article body to a queryable Knowledge Graph. The LLM ensures semantic extraction via a strict JSON schema, enabling the automated instantiation of NCO ontological classes (nco:Agent, nco:Claim, nco:Utterance).

Problem Definition. We formulate the claim extraction task as the identification of a tuple $C = \langle a, u, \tau \rangle$, where:

- a is the **Agent** (the entity producing the statement);
- u is the **Utterance** (the normalized content of the statement);
- τ is the **Type** of attribution (Direct, Indirect, or Partially Direct).

Within the NCO ontology, the *Claim* is explicitly modelled not as the text itself, but as the event node connecting a and u.

3.1. Ontological Schema

The foundation of our pipeline is the NCO [4], which provides the semantic scaffolding for the extraction. Rather than treating a claim as a simple string, NCO decomposes the attribution into a set of interconnected entities. We target four primary classes:

- **nco:NewsItem**: The source document containing the metadata (e.g., publication date, title).
- **nco:Agent**: The entity (Person or Organization) to whom the statement is attributed.
- **nco:Claim**: The abstract node representing the act of stating something. It serves as the bridge between the Agent and the Utterance.
- **nco:Utterance**: The linguistic content of the statement. This class includes the property `nco:hasUtteranceType`, which categorizes the speech act into one of three taxonomic forms:
  - ***Direct***: Verbatim quotations (e.g., "We will not surrender.").
  - ***Indirect***: Paraphrased statements (e.g., The President stated that they would not surrender.).
  - ***Partially Direct***: Hybrid constructions where specific fragments are quoted within a narrative sentence (e.g., The spokesperson described the attack as a "significant escalation" of the conflict).

3.2. LLM-based Extraction

The core extraction engine employs an LLM to map raw text to the symbolic classes defined above. Unlike traditional supervised classifiers that require extensive training data, we utilise a few-shot prompting strategy that enforces *schema compliance*.

Input and Prompting Strategy. The input to this stage is the full body of the news article. The LLM is guided by a structured decomposition prompt rather than a standard open-ended query. The prompt explicitly defines the NCO concepts (Agent, Utterance, Claim) and enforces a three-step internal logic: (i) *Atomic Scan*, to split compound sentences into distinct claims; (ii) *Agent Resolution*, forcing the model to resolve pronouns (e.g., "he") to the full entity name mentioned earlier in the text; and (iii) *Utterance Normalization*, which rewrites indirect speech into grammatically complete sentences. Crucially, the prompt imposes a strict output format: the model must return a JSON object where each field corresponds to an ontological property. This forces the neural model to perform implicit "type checking" during generation, bridging the gap between unstructured text and the rigid NCO schema.

Output Specification. For each identified claim, the Neural Layer produces a structured object containing:

- `Agent Name`: The canonical string identifying the speaker.
- `Utterance Text`: The content of the claim, normalized to be self-contained.
- `Utterance Type`: The classification (Direct, Indirect, Partially Direct).
- `Source Evidence`: The verbatim sentence(s) from the text, used for traceability.

3.3. Knowledge Graph Construction

The final stage converts the intermediate JSON representations into a formalized Knowledge Graph. This process is deterministic and fully automated. Each entity is assigned a Uniform Resource Identifier (URI). To ensure consistency without external entity linking, URIs for Agents are generated based on normalized name strings, while URIs for Claims and Utterances are generated via MD5 hashing of their content and context. The resulting data is serialized as RDF triples (Turtle format), effectively grounding the unstructured news stream into a queryable semantic network.

4. Experimental Setup

We evaluate the proposed pipeline on the domain of conflict reporting, specifically the war in Ukraine. Selected for its high claim density and narrative complexity, this domain serves as a rigorous stress test for automated extraction. In this section, we detail the dataset characteristics, the set of LLMs compared, and the semantic matching protocol.

4.1. Dataset: The Ukraine War Corpus

We constructed a Gold Standard dataset focused on the War in Ukraine. This domain was selected as a stress test for the extraction pipeline due to its narrative complexity, characterized by conflicting propaganda, official military statements, and nested diplomatic statements that challenge standard NLP parsers.

To ensure ground truth reliability, the corpus was manually annotated by domain experts following a strict protocol to resolve ambiguous attributions and coreferences. The dataset is composed of 24 long-form articles retrieved from a diverse range of outlets (including *Daily Mail*, *BBC*, *The Guardian*, *Associated Press*). Crucially, this corpus shifts the evaluation focus from document count to extraction density: as shown in Table 1, the dataset contains a total of 367 atomic claims (avg. 15.29 per article) with a significant presence of indirect and partially direct speech (approx. 50%), making it a rigorous benchmark for testing LLM capabilities in capturing non-verbatim attributions.

Table 1
Statistics of the Ukraine War Gold Standard. The dataset is characterized by a high density of claims and a balanced distribution of utterance types.

Metric	Value
Total Articles	24
Total Claims	367
Claims per Article (Mean ± SD)	15.29 ± 5.56
Utterance Types distribution:	
Direct Quotations	183 (49.9%)
Indirect Speech	80 (21.8%)
Partially Direct (Mixed)	104 (28.3%)

4.2. Selected Models

To assess the intrinsic extraction capabilities of current generation LLMs, we selected six models representing a diverse spectrum of architectures, sizes, and access paradigms (open-weights vs. proprietary). The selected models are listed in Table 2. All models were queried using the same structured prompt with a temperature setting of 0 to ensure deterministic reproducibility.

Table 2
The six LLMs selected for the evaluation. For Mixture of Expert (MoE) models, active parameters are shown in parentheses next to the total count.

Model Identifier	Size (Active Params)	Context (k)	Open Weights
`openai/gpt-oss-20b`	21B (3.6B)	131	Yes
`google/gemini-2.5-flash-lite`	N/A	1048	No
`openai/gpt-4.1-nano`	N/A	1047	No
`deepseek/deepseek-chat-v3.1`	671B (37B)	163	Yes
`meta/llama-4-maverick`	400B (17B)	1048	Yes
`qwen/qwen3-235b-a22b-2507`	235B (22B)	262	Yes

4.3. Evaluation Protocol

Evaluating generative extraction requires accounting for semantic equivalence rather than exact string matching. We employ a two-stage alignment protocol to identify True Positives (TP), False Positives (FP), and False Negatives (FN):

1. **Agent Alignment:** Extracted agents are matched against the Gold Standard (GS) using a hierarchical funnel: (i) exact string match, (ii) name containment (e.g., "Zelenskyy" in "President Zelenskyy"), and (iii) semantic similarity using vector embeddings (threshold > 0.7).
2. **Claim Verification:** To handle structural variations, such as an LLM splitting a single GS claim into multiple sentences or merging distinct ones, we employ a greedy combinatorial matching algorithm. For each aligned agent, the algorithm identifies the optimal alignment that maximizes the aggregate semantic similarity (computed via `en_core_web_lg`[1] vectors) between extracted and GS utterances. The threshold $\tau = 0.75$ acts strictly as a lower bound, any candidate alignment below this value is discarded, while the highest scoring match among the remaining candidates is selected as a True Positive.

Metrics Calculation. To quantify the extraction quality, we define the following outcomes based on the alignment results:

- **True Positive (TP):** A claim extracted by the model that successfully aligns with a Gold Standard (GS) entry.
- **False Positive (FP):** An extracted claim that fails to align with any GS entry. This category includes hallucinations, misattributions, or the extraction of neutral factual statements (e.g., reporting an event) that do not constitute an argumentative claim.
- **False Negative (FN):** A claim present in the GS that the model failed to identify or correctly attribute.

Based on these definitions, we report Micro-averaged Precision, Recall, and F1-Score. This approach aggregates TPs, FPs, and FNs across the entire corpus, ensuring that each of the 367 claims in the Gold Standard is weighted equally in the final assessment.

5. Results and Discussion

Table 3 presents the performance of the six evaluated models on the Ukraine War Gold Standard, ranked by Micro-F1 score. The results validate the feasibility of the proposed neuro-symbolic pipeline but highlight significant disparities in how different architectures handle the dual constraint of semantic extraction and schema compliance.

`DeepSeek-v3.1` emerges as the most robust model, achieving the top F1-score (76.57) with a perfect balance between Precision and Recall. This indicates a superior ability to interpret the nuances of war

[1] en_core_web_lg — https://spacy.io/models/en#en_core_web_lg

Table 3
Performance comparison on the Ukraine War dataset. Models are sorted by Micro-F1. *DeepSeek-v3.1* offers the best stability, while *gpt-oss-20b* maximizes Recall at the expense of Precision.

Model	Precision	Recall	F1	TP	FP	FN
deepseek/deepseek-chat-v3.1	**76.57**	76.57	**76.57**	281	86	86
qwen/qwen3-235b-a22b-2507	71.54	71.93	71.74	264	105	103
google/gemini-2.5-flash-lite	67.31	75.75	71.28	278	135	89
openai/gpt-oss-20b	59.14	**82.02**	68.72	301	208	66
meta/llama-4-maverick	70.83	55.59	62.29	204	84	163
openai/gpt-4.1-nano	43.06	40.75	41.87	152	201	215

All metrics are Micro-averaged over the 367 Gold Standard claims.

reporting without deviating from the NCO ontology structure. It is closely followed by qwen3-235b (F1 71.74), confirming that high-parameter open-weight models are now competitive with proprietary solutions for structured information extraction.

Conversely, the remaining models exhibit distinct trade-offs. openai/gpt-oss-20b acts as an "over extractor," achieving the highest Recall in the benchmark (82.02) but suffering from low Precision (59.14), likely due to the hallucination of non argumentative sentences as claims. On the other spectrum, meta/llama-4-maverick proves overly conservative, missing nearly half of the relevant claims (Recall 55.59). Finally, the poor performance of gpt-4.1-nano (F1 41.87) establishes a clear lower bound for this task: lightweight models appear to lack the reasoning capacity required to simultaneously parse complex narratives and satisfy strict JSON schema constraints.

6. Conclusion and Future Work

In this work, we demonstrated that LLMs can effectively function as schema-constrained semantic parsers, bridging the gap between unstructured conflict reporting and formal Knowledge Graphs. By enforcing the NCO as a strict generation target, we showed that high-parameter models like DeepSeek-v3.1 can extract claims with a good precision. However, our benchmark also reveals a critical reasoning threshold: while large architectures successfully balance semantic extraction with schema compliance, lightweight models struggle to maintain this dual focus, resulting in significant degradation of output quality.

These findings suggest that the future of automated computational journalism lies not in single model optimisation, but in composite neuro-symbolic architectures. Our immediate research agenda moves beyond the single-prompt paradigm towards an orchestration layer, where distinct models are leveraged for extraction, verification, and deduplication. Furthermore, to transform these isolated graphs into a truly interoperable web of data, we plan to integrate an explicit Entity Linking module, grounding textual agents to canonical URIs. Ultimately, this pipeline represents a foundational step towards real-time, transparent mapping of global media narratives, capable of scaling across diverse domains beyond the specific complexities of war reporting.

Declaration on Generative AI

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

References

[1] S. Zaklama, Exploring the foundations of media framing theory, European Modern Studies Journal 9 (2025) 75–89. doi:10.59573/emsj.9(1).2025.7.

[2] M. Djerf-Pierre, A. Shehata, B. Johansson, Media salience shifts and the public's perceptions about reality: How fluctuations in news media attention influence the strength of citizens' sociotropic beliefs, Mass Communication and Society 28 (2025) 459–484. URL: https://doi.org/10.1080/15205436.2023.2299209. doi:10.1080/15205436.2023.2299209. arXiv:https://doi.org/10.1080/15205436.2023.2299209.

[3] E. Motta, F. Osborne, M. M. L. Pulici, A. Salatino, I. Naja, Capturing the viewpoint dynamics in the news domain, in: M. Alam, M. Rospocher, M. van Erp, L. Hollink, G. A. Gesese (Eds.), Knowledge Engineering and Knowledge Management, Springer Nature Switzerland, Cham, 2025, pp. 18–34.

[4] E. Motta, E. Daga, A. Gangemi, M. L. Gjelsvik, F. Osborne, A. Salatino, The epistemology of fine-grained news classification, Semantic Web 16 (2025) 22104968251344461. URL: https://doi.org/10.1177/22104968251344461. doi:10.1177/22104968251344461. arXiv:https://doi.org/10.1177/22104968251344461.

[5] S. Pareti, T. O'Keefe, I. Konstas, J. R. Curran, I. Koprinska, Automatically detecting and attributing indirect quotations, in: D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, S. Bethard (Eds.), Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle, Washington, USA, 2013, pp. 989–999. URL: https://aclanthology.org/D13-1101/.

[6] T. O'Keefe, S. Pareti, J. R. Curran, I. Koprinska, M. Honnibal, A sequence labelling approach to quote attribution, in: J. Tsujii, J. Henderson, M. Paşca (Eds.), Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, Jeju Island, Korea, 2012, pp. 790–799. URL: https://aclanthology.org/D12-1072/.

[7] Y. Chen, Z.-H. Ling, Q.-F. Liu, A neural-network-based approach to identifying speakers in novels, in: Interspeech 2021, 2021, pp. 4114–4118. doi:10.21437/Interspeech.2021-609.

[8] T. Ni, Q. Wang, G. Ferraro, Explore bilstm-crf-based models for open relation extraction, CoRR abs/2104.12333 (2021). URL: https://arxiv.org/abs/2104.12333. arXiv:2104.12333.

[9] D. Xu, W. Chen, W. Peng, C. Zhang, T. Xu, X. Zhao, X. Wu, Y. Zheng, Y. Wang, E. Chen, Large language models for generative information extraction: A survey, Frontiers of Computer Science 18 (2024) 186357.

[10] E. Motta, F. Osborne, M. M. Pulici, A. Salatino, I. Naja, Capturing the viewpoint dynamics in the news domain, in: International Conference on Knowledge Engineering and Knowledge Management, Springer, 2024, pp. 18–34.

[11] K. Yang, T. Zhang, Z. Kuang, Q. Xie, J. Huang, S. Ananiadou, Mentallama: interpretable mental health analysis on social media with large language models, in: Proceedings of the ACM Web Conference 2024, 2024, pp. 4489–4500.

[12] S. Angioni, S. Consoli, D. Dessì, F. Osborne, D. R. Recupero, A. Salatino, Exploring environmental, social, and governance (esg) discourse in news: An ai-powered investigation through knowledge graph analysis, IEEE Access 12 (2024) 77269–77283.

[13] A. Borrego, D. Dessì, D. Ayala, I. Hernández, F. Osborne, D. R. Recupero, D. Buscaldi, D. Ruiz, E. Motta, Research hypothesis generation over scientific knowledge graphs, Knowledge-Based Systems 315 (2025) 113280.

[14] T. Aggarwal, A. Salatino, F. Osborne, E. Motta, Large language models for scholarly ontology generation: An extensive analysis in the engineering field, Information Processing & Management 63 (2026) 104262.

[15] J. A. Omiye, H. Gui, S. J. Rezaei, J. Zou, R. Daneshjou, Large language models in medicine: the potentials and pitfalls: a narrative review, Annals of internal medicine 177 (2024) 210–220.

[16] G. Cascini, A. Fantechi, E. Spinicci, Natural language processing of patents and technical documentation, in: International Workshop on Document Analysis Systems, Springer, 2004, pp.

508–520.
[17] J. Savelka, K. D. Ashley, The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts, Frontiers in Artificial Intelligence 6 (2023) 1279794.
[18] M. Birti, A. Maurino, F. Osborne, Optimizing large language models for esg activity detection in financial texts, in: Proceedings of the 6th ACM International Conference on AI in Finance, 2025, pp. 856–863.
[19] A. Meloni, D. Reforgiato Recupero, F. Osborne, A. Salatino, E. Motta, S. Vahadati, J. Lehmann, Exploring large language models for scientific question answering via natural language to sparql translation, ACM Transactions on Intelligent Systems and Technology (2025).
[20] D. Dessí, F. Osborne, D. R. Recupero, D. Buscaldi, E. Motta, Scicero: A deep learning and nlp approach for generating scientific knowledge graphs in the computer science domain, Knowledge-Based Systems 258 (2022) 109945.
[21] S. Tsaneva, D. Dessì, F. Osborne, M. Sabou, Knowledge graph validation by integrating llms and human-in-the-loop, Information Processing & Management 62 (2025) 104145.
[22] D. Metropolitansky, J. Larson, Towards effective extraction and evaluation of factual claims, 2025. URL: https://arxiv.org/abs/2502.10855. arXiv:2502.10855.
[23] L. Zheng, C. Li, Z. Liu, F. Huang, H. Jia, Z. Ye, X. Zhang, Fact in fragments: Deconstructing complex claims via llm-based atomic fact extraction and verification, 2025. URL: https://arxiv.org/abs/2506.07446. arXiv:2506.07446.
[24] W. R. van Hage, V. Malaisé, R. Segers, L. Hollink, G. Schreiber, Design and use of the simple event model (sem), Journal of Web Semantics 9 (2011) 128–136. URL: https://www.sciencedirect.com/science/article/pii/S1570826811000199. doi:https://doi.org/10.1016/j.websem.2011.03.003, provenance in the Semantic Web.
[25] S. Gottschalk, E. Demidova, Eventkg: A multilingual event-centric temporal knowledge graph, 2018. URL: https://arxiv.org/abs/1804.04526. arXiv:1804.04526.
[26] F. Plötzky, W.-T. Balke, It's the same old story! enriching event-centric knowledge graphs by narrative aspects, in: 14th ACM Web Science Conference 2022, WebSci '22, ACM, New York, NY, USA, 2022, p. 34–43. URL: http://dx.doi.org/10.1145/3501247.3531565. doi:10.1145/3501247.3531565.
[27] A. Tchechmedjiev, P. Fafalios, K. Boland, M. Gasquet, M. Zloch, B. Zapilko, S. Dietze, K. Todorov, Claimskg: A knowledge graph of fact-checked claims, in: The Semantic Web – ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II, Springer-Verlag, Berlin, Heidelberg, 2019, p. 309–324. URL: https://doi.org/10.1007/978-3-030-30796-7_20. doi:10.1007/978-3-030-30796-7_20.
[28] H. Bian, Llm-empowered knowledge graph construction: A survey, 2025. URL: https://arxiv.org/abs/2510.20345. arXiv:2510.20345.

A. System Prompt for Claim Extraction

The following is the complete system prompt used for the claim extraction task across all six evaluated models. The prompt was designed following an iterative refinement process and selected empirically based on extraction quality on a held out subset of articles. All models were queried using this identical prompt with temperature = 0 to ensure deterministic reproducibility.

The prompt enforces a three-phase internal reasoning workflow: (i) *Atomic Scan*, to identify and decompose compound attributions into individual claim units; (ii) *Agent Resolution*, to resolve pronouns and partial names to their canonical form; and (iii) *Utterance Classification & Transformation*, to normalize the utterance text and classify its type (*direct*, *indirect*, or *partially-direct*). Output is constrained to a strict JSON schema that mirrors the NCO ontological structure, effectively forcing the model to perform implicit type checking during generation.

Listing 1: System prompt used for structured claim extraction.

```
SYSTEM_PROMPT:
    You are an advanced AI model, hyper-specialized in extracting structured claims from news articles with maximum precision. Your mission is to identify,
        analyze, and format every single attributed claim into a perfect JSON output.

    ### Core Objective
    Your goal is to extract **every single atomic claim** in chronological order. A claim is a statement, opinion, or assertion made by a **person** or an
        **organization**, and it must be explicitly attributed to that agent.

    ### Internal Thought Process (Follow these steps)

    To ensure maximum accuracy, you must internally adopt the following three-phase workflow for every potential claim you identify:

    **Phase 1: Scan & Atomic Identification**
    1.  **Exhaustive Scan:** Read the entire article, sentence by sentence. Do not skip any part.
    2.  **Identify Attribution Patterns:** Look for verbs of assertion (e.g., 'said', 'claimed', 'stated', 'announced') that link a statement to an agent.
    3.  **Extract Atomically:** Your most critical rule is **One Claim per Attributed Clause**.
        *   If one sentence contains multiple attributions ("X said A, while Y argued B"), you MUST create two separate JSON objects.
        *   If one agent makes multiple points in consecutive sentences ("He stated A. He then added B."), create two separate JSON objects.
        *   The goal is atomic extraction, not summarization.

    **Phase 2: Agent Resolution & Canonicalization**
    For each identified claim, determine the agent and format the `agent_*` fields according to these rules:
    1.  **Resolve Pronouns/Partial Names:** If the initial agent is a pronoun ('he', 'she') or a partial name ('Smith'), scan the full article text to find the
            most complete mention (e.g., "The Secretary of State, John Smith"). Use this full information for the steps below.
    2.  **Specific Person:**
        *   **Text:** "The President of the ECB, Christine Lagarde, stated..."
        *   `agent_name`: "Christine Lagarde"
        *   `agent_description`: "President of the ECB"
        *   `agent_type`: "person"
    3.  **Spokesperson for a Specific Organization:**
        *   **Text:** "A spokesperson for the Ministry of Health confirmed..."
        *   `agent_name`: "Ministry of Health"
        *   `agent_description`: "A spokesperson for the Ministry of Health"
        *   `agent_type`: "organization"
    4.  **Specific Organization or Inanimate Source:**
        *   **Text:** "A new study from MIT indicated..."
        *   `agent_name`: "MIT Study"
        *   `agent_description`: "A new study from MIT"
        *   `agent_type`: "organization"

    **Phase 3: Utterance Classification & Transformation**
    For each claim, classify and transform the utterance text (`utterance_text`) to create a clear, standalone statement.
    1.  **Classify the Utterance Type (`utterance_type`):**
        *   **`direct`**: The statement is a verbatim quote, fully enclosed in quotation marks.
        *   **`partially-direct`**: The statement is mostly a paraphrase but includes quoted fragments.
        *   **`indirect`**: The statement is a complete paraphrase, with no quotation marks.

    2.  **Transform the Utterance Text (`utterance_text`):**
        *   **For `direct`**: Extract only the exact text inside the quotation marks.
        *   **For `indirect` and `partially-direct`**:
            a. **Isolate:** Remove all attributional phrases (e.g., "he said that...", "she described it as...").
            b. **Reconstruct:** Rewrite the remaining statement into a complete, grammatically coherent sentence. This may require changing verb tenses (e.g.,
                "was" to "is").
            c. **Transform to First-Person (if applicable):** Change the perspective to the first person ("his/her" -> "my") **only if** the statement reflects
                the agent's **own** actions, beliefs, or commitments.
                *   *Correct Example:* "The minister confirmed that **his** proposal was ready." -> `utterance_text`: "**My** proposal is ready."
                *   *Incorrect Example (DO NOT transform):* "The engineer described **the engine** as 'a marvel'." -> `utterance_text`: "**The engine** is 'a
                    marvel'." (The agent is describing an external object).

    ### JSON Output Format (Mandatory)
    Your entire output MUST be a single JSON code block. Do not include any text or explanations before or after the JSON block. The syntax must be perfect.

    ```json
    {{
      "results": [
        {{
          "agent_name": "canonical name of the agent",
          "agent_description": "short description of the agent's role or context",
          "agent_type": "person/organization",
          "utterance_type": "direct/partially-direct/indirect",
          "utterance_text": "the clean, transformed text of the claim",
          "source_context": "the full, original sentence from the article where the claim appears"
        }}
      ]
    }}
    ```

    ### Comprehensive Examples

    Example 1: Direct and Indirect Claim
    *   **Article Text:** The President of the European Commission, Ursula von der Leyen, stated that the new measures are essential. "We can no longer wait,"
            she later added.
    *   **JSON Output:**
        ```json
        {{
          "results": [
            {{
              "agent_name": "Ursula von der Leyen",
              "agent_description": "President of the European Commission",
              "agent_type": "person",
              "utterance_type": "indirect",
              "utterance_text": "The new measures are essential.",
              "source_context": "The President of the European Commission, Ursula von der Leyen, stated that the new measures are essential."
            }},
            {{
              "agent_name": "Ursula von der Leyen",
              "agent_description": "President of the European Commission",
              "agent_type": "person",
              "utterance_type": "direct",
              "utterance_text": "We can no longer wait",
              "source_context": "\"We can no longer wait,\" she later added."
            }}
          ]
        }}
        ```

    Example 2: Partially-Direct Claim
    *   **Article Text:** The Italian Prime Minister, Giorgia Meloni, said the reform would "strengthen national sovereignty" and improve citizen trust.
    *   **JSON Output:**
        ```json
        {{
          "results": [
            {{
              "agent_name": "Giorgia Meloni",
```
```

```
            "agent_description": "Italian Prime Minister",
            "agent_type": "person",
            "utterance_type": "partially-direct",
            "utterance_text": "The reform will strengthen national sovereignty and improve citizen trust.",
            "source_context": "The Italian Prime Minister, Giorgia Meloni, said the reform would \"strengthen national sovereignty\" and improve citizen
                  trust."
          }}
        ]
      }}
      ```

    Example 3: Organization Agent
    *   **Article Text:** The European Central Bank announced that interest rates would remain unchanged.
    *   **JSON Output:**
        ```json
        {{
          "results": [
            {{
              "agent_name": "European Central Bank",
              "agent_description": "European Central Bank",
              "agent_type": "organization",
              "utterance_type": "indirect",
              "utterance_text": "Interest rates will remain unchanged.",
              "source_context": "The European Central Bank announced that interest rates would remain unchanged."
            }}
          ]
        }}
        ```

    ### Final Checklist
    - **Explicit Attribution:** Is every claim clearly attributed to an agent?
    - **Atomic Extraction:** Have you created a separate object for each individual claim?
    - **Correct Agent:** Has the agent been resolved and formatted correctly according to the rules?
    - **Clean Utterance Text:** Is the `utterance_text` properly transformed, free of attributional phrases, and in the first person only when appropriate?
    - **Full Context:** Does the `source_context` contain the complete original sentence?
    - **Chronological Order:** Are the claims listed in the order they appeared in the article?
    - **Valid JSON Format:** Is the output a single, syntactically perfect JSON block?

    Now, process the provided article text.

USER_PROMPT:
    ## Article Text to Process:
    {input_article_text}
```
```