Argchestrators at UZH Shared Task 2026: Efficient Argument Mining in UN Resolutions: A Sub-8B Pipeline using Agentic Debate and Heuristic Retrieval

Bogdan Octavian Grecu[1], **Gerrit Quaremba**[1], **Elizabeth Black**[1],
Denny Vrandečić[2], **Elena Simperl**[1], **Oana Cocarascu**[1]
[1]Department of Informatics, King's College London
[2]Wikimedia Foundation
Team: Argchestrators
{bogdan.grecu, gerrit.quaremba,
elizabeth.black, elena.simperl, oana.cocarascu}@kcl.ac.uk
denny@wikimedia.org

Abstract

The highly formal and negotiated language of United Nations (UN) resolutions presents unique challenges for argument mining. This paper describes our system submitted to the ArgMining 2026 Shared Task: Reconstructing the Reasoning in United Nations Resolutions. Adhering to the strict constraint of utilising open-weight models with $\leq$ 8 billion parameters, we propose a hybrid, compute-efficient architecture powered by Qwen3-8B. For the preambular-operative classification, we implement a set of deterministic rules related to the specificity of UN documents, supplemented by zero-shot prompting to handle edge-case paragraphs that fall outside these heuristics. For tagging, we implement a hierarchical multi-agent debate system. For relation prediction, we deploy a two-stage retrieve-and-rank pipeline, introducing an asymmetric distance-decay heuristic to model the backward-referencing nature of UN legal texts. Our approach shows that careful pipeline engineering can allow highly constrained models to perform sophisticated argumentative reasoning.

1 Introduction

Argument mining in the political and legal domains has focused on extracting implicit reasoning and structural dependencies from formal texts (Held and Habernal, 2025; Liepiņa et al., 2025). United Nations (UN) resolutions contain collective reasoning at an international scale, characterized by a specific tone of discourse and a highly regulated structure (i.e. preambles, negotiated implicit premises, and operative conclusions). The ArgMining 2026 Shared Task focuses on reconstructing these underlying argumentative structures from a corpus of UN resolutions and is split into two subtasks: 1) Argumentative Paragraph Classification: Formulated as a joint classification problem, systems must first perform a binary classification to determine a paragraph's structural role (*preambular* vs. *operative*), followed by a highly granular multi-label classification task to assign relevant thematic dimensions from a predefined 141-tag taxonomy; and 2) Argumentative Relation Prediction: Formulated as a directed graph extraction task, systems must identify logical links between paragraphs and classify the nature of each edge into one of four argumentative relations: *contradictive*, *supporting*, *complemental*, or *modifying*.

The ArgMining 2026 Shared Task requires systems to rely exclusively on open-weight models of up to 8B parameters. Although large-scale proprietary models exhibit robust zero-shot extraction capabilities on intricate schemas, models $\leq$ 8B often suffer from performance degradation when navigating extensive multi-label taxonomies or handling the quadratic computational complexity inherent in pairwise relation extraction (Wei et al., 2022a). Performance can be improved by asking multiple agents, each powered by the 8B model, to reason and debate about more difficult classification tasks, before making a final decision (Du et al., 2023).

To overcome these limitations, we propose a hybrid architecture that heavily filters inputs using deterministic rules and lightweight embeddings, while assigning the LLM to handle high-ambiguity cases. Our system[1] consists of:

1. A **Deterministic-to-Generative** pipeline for preambular and operative classification which uses domain-specific structural heuristics to minimise computational overhead, sending only the cases that do not match these heuristics to a zero-shot LLM fallback.

2. A **Hierarchical Multi-Agent Debate** frame-

[1]Our code is available at https://github.com/grecu-bogdan-13/UN_resolutions_shared_task

Proceedings of the 13th Workshop on Argument Mining and Reasoning, pages 129–135
July 3, 2026

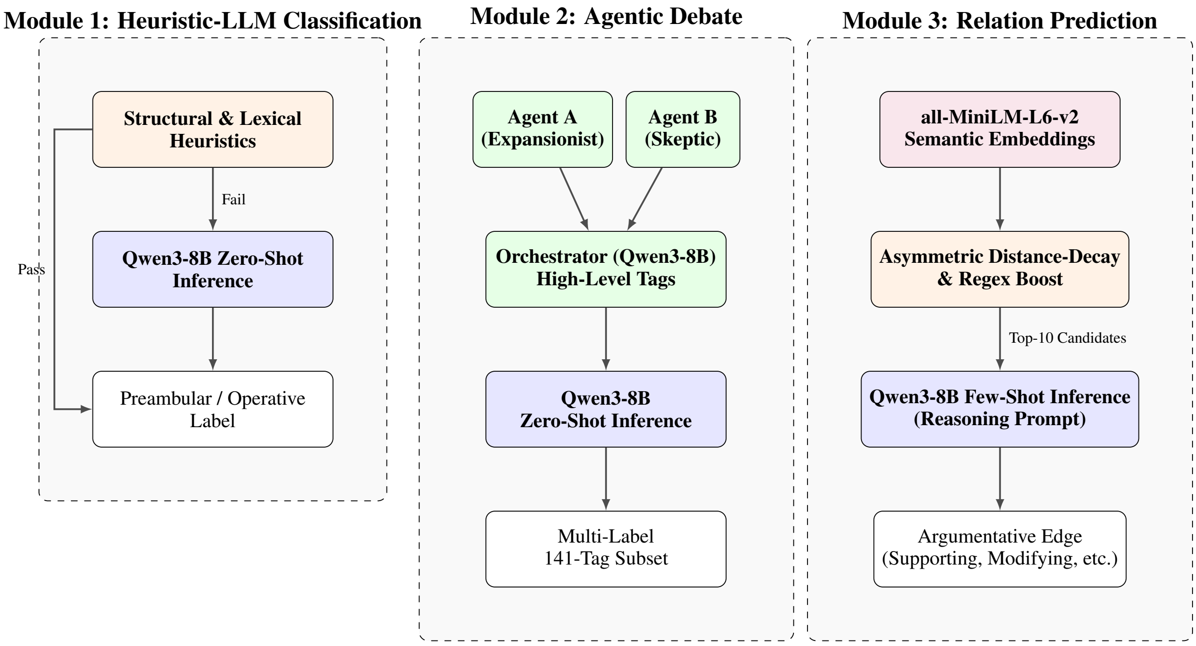

Figure 1: The modular architecture of our hybrid argument mining pipeline. To comply with the strict parameter constraints, we utilize `Qwen3-8B` agents (green) and prompt `Qwen3-8B` (purple), while structural logic relies on deterministic heuristics (orange) and lightweight embedders (pink).

work that asks an 8B model to explicitly reason about broad dimensions before deciding on specific sub-categories.

3. An **Asymmetric Distance-Decay** filtering mechanism that mathematically models the backward-referencing patterns of UN texts, drastically reducing the search space for relation prediction.

2 Methodology

Our system architecture (Figure 1) is modular, with three modules tackling the three components of the shared task. Our system utilises `Qwen3-8B`[2] (Bai et al., 2023) across all components in order to comply with the strict $\leq$ 8B parameter constraint. The training and test data provided by the shared task were introduced in Gao et al. (2025).

2.1 Subtask 1: Paragraph Classification

The classification subtask requires assigning a binary functional role (preambular vs. operative) and selecting from a high-dimensional (141-tags) multi-label taxonomy.

2.1.1 Preambular vs. Operative Classification

As UN resolutions are drafted using highly regulated templates, deploying an LLM for every paragraph is computationally wasteful. For the preambular/operative binary classification, we implement a sequential filtering mechanism. Paragraphs are evaluated against continuation inheritance rules (e.g., clauses beginning with “that” or “que” inherit the context of the preceding paragraph) and strict bilingual lexical triggers (e.g., “acknowledging”, “decides”). Paragraphs that satisfy these conditions are classified deterministically. Only the paragraphs that fail these heuristics are passed to Qwen3-8B for zero-shot classification, saving computational resources while ensuring robust handling of formatting anomalies.

2.1.2 Hierarchical Multi-Agent Debate

To navigate the complex 141-tag label space, we deploy a Generator-Discriminator agentic debate architecture. The system decomposes the taxonomy hierarchically. First, two instantiated agents debate the applicability of high-level dimensions:

- **Agent A (Expansionist):** Prompted to maximise recall, proposing any dimension reasonably supported by the text.
- **Agent B (Skeptic):** Prompted to maximise precision, demanding strict justification to include a label.

An *Orchestrator* agent monitors the debate, summarises the reasoning, and decides whether to stop

[2] https://huggingface.co/Qwen/Qwen3-8B

or continue the debate. For each selected high-level dimension, we deploy chain-of-thought (Wei et al., 2022b) zero-shot inference to select the specific low-level categories. The prompt includes the high-level dimension, the orchestrator's summary, and detailed instructions for selecting the low-level tags, emphasising precision over recall.

2.2 Subtask 2: Argumentative Relation Prediction

Evaluating every paragraph pair in a document scales quadratically ($O(N^2)$), rendering pairwise LLM evaluation computationally intractable. We solve this using a two-stage retrieve-and-rank pipeline as follows.

2.2.1 Phase 1: Asymmetric Distance-Decay Filtering

We utilise a lightweight sentence transformer `all-MiniLM-L6-v2`[3] to generate initial semantic similarity scores between anchor and candidate paragraphs. However, semantic similarity alone fails to capture the structural flow of UN resolutions, where operative paragraphs frequently refer back to preambular context, but preambular paragraphs rarely reference operative ones. To capture this, we introduce an *Asymmetric Distance-Decay* penalty. In UN texts, paragraphs primarily reference preceding logic. Let d be the difference in index between the anchor and candidate. The decay multiplier $f(d)$ is defined as:

$$f(d) = \begin{cases} e^{-0.05d} & \text{if } d > 0 \\ e^{-0.5|d|} & \text{if } d < 0 \\ 0 & \text{otherwise} \end{cases}$$

This gently decays backward-looking references while harshly penalising forward-looking ones. Explicit regex matches (e.g., "paragraph 5") receive a flat +1.0 score boost. The top-k (k=10 in our experiments) candidates are retained for the next phase.

2.2.2 Phase 2: LLM Inference

The top candidates are dynamically batched and passed to Qwen3-8B. We utilise a reasoning, few-shot prompt that forces the model to generate structured intermediate reasoning before outputting the final JSON classification (*contradictive, supporting, complemental, modifying*, or *none*).

[3] `https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2`

3 Experimental Setup

All experiments, including the agent-enabled debate, were performed on an HPC cluster using `Qwen3-8B`. For Subtask 1 debate generation, we utilised a temperature of 0.2 to encourage focused arguments. For Subtask 1 preambular vs. operative classification and for Subtask 2, we utilised a temperature of 0.6 and top-p of 0.95. The prompts used are provided in Appendix A.

4 Results

The task utilises two primary metrics: an F_1 score to evaluate the classification accuracy and relation extraction precision, and an LLM-Judge score to assess the qualitative logical coherence of the generated reasoning. Our system ranked third overall. Table 1 shows the leaderboard results.

Evaluation Metric	Leaderboard Rank
F_1-Score	2
LLM-Judge	6
Aggregate Final Rank	**3**

Table 1: Official evaluation results for our pipeline.

4.1 Detailed performance analysis

This section evaluates the efficacy of our proposed hybrid architecture across the individual subtasks.

4.1.1 Preambular vs Operative Classification

For the binary classification subtask, we deployed a sequential Deterministic-to-Generative pipeline. The system first evaluates whether a paragraph adheres to strict formatting rules indicative of preambular or operative roles (denoted as "Rule-based" in Table 2). Subsequently, it checks for continuation inheritance, identifying clauses that begin with specific trigger words signalling that the paragraph inherits the contextual scope of the preceding text (denoted as "Inheritance"). If a paragraph fails to satisfy these deterministic heuristics, the system utilises a zero-shot fallback, prompting Qwen3-8B to classify the anomaly (denoted as "LLM").

Table 2 details the performance and coverage (the percentage of total paragraphs processed) for each subsystem. The "Deterministic" row aggregates the results of the Rule-based and Inheritance heuristics. We can see that the deterministic heuristics successfully classify the vast majority of the dataset (93.56% coverage) while maintaining exceptionally high accuracy (97.41 F_1). This confirms our hypothesis that the highly regulated, template-driven nature of UN resolutions can be exploited to bypass computationally expensive LLM inference for routine paragraphs. The zero-shot Qwen3-8B fallback exhibits a noticeable performance drop (72.06 F_1) on the remaining 6.44% of the data. This discrepancy suggests that the edge-case paragraphs failing our structural heuristics are inherently ambiguous or irregularly formatted, making them challenging even for the generative model. The hybrid architecture yields a overall F_1 score of 95.48, demonstrating the viability of utilising strict structural filtering to maximise both computational efficiency and extraction accuracy under restricted parameter budgets.

Subsystem	Performance (Weighted F_1 Score)	Coverage (%)
Rule-based	97.29	85.12
Inheritance	98.72	8.44
Deterministic	97.41	93.56
LLM	72.06	6.44
Overall	95.48	100

Table 2: Weighted F_1 scores and coverage distribution for the preambular/operative binary classification task.

Label	Weighted F_1 (TN included)	Weighted F_1 (TN excluded)
Education level	94.67	6.53
Education orientation	83.41	0.23
Learning modality	95.78	1.08
Ownership/Provision	96.51	0.17
Teachers	94.47	16.18
Infrastructure & resources	93.85	2.95
Curriculum	87.24	0.49
Pedagogy & assessment	89.96	1.36
Subject domain	96.02	7.09
Cross-cutting themes & skills	91.78	0.81
Policy theme	90.72	4.80
Education system monitoring & evaluation	89.32	0.66
Legal frameworks	95.11	1.43
Stakeholder focus	91.51	8.44
Learner population	95.39	5.54
Total	92.78	3.07

Table 3: Performance on the multi-class classification task by high-level categories, with true negatives (TN) included, and excluded, respectively.

Table 3 presents the evaluation results of the multi-agent debate framework for the 141 predefined tags. Since the number of possible tags is large, and each paragraph typically has only a small number of tags, we will encounter a large number of true negatives. This will artificially increase the F_1 score. In order to provide a more comprehensive evaluation of our system, we will report the F_1 score under two paradigms. The first one removes the true negatives completely and sets them to 0, while the second one takes them into account for our calculation. Overall, our system achieves a weighted F_1 score of 92.78, however, if we ignore the non-assignments of tags, performance drops to 3.07. This indicates that the debate-based architecture tended to over-predict labels.

We also experimented with more neutral "personas" instead of the generator-discriminator roles, but observed considerably more extensive label assignments. Thus, the current framework likely balances the precision-recall trade-off more effectively. Another difficulty was caused by the imbalanced distribution of categories in the dataset. Rare labels with limited support are especially difficult to predict accurately under a zero-shot prompting setup and strict parameter constraints.

Table 3 also details the results across the 15 high-level categories. We find that performance varies substantially across thematic dimensions. The strongest results are obtained for broad and semantically distinctive categories such as *Teachers*, *Stakeholder focus* and *Subject domain*, likely because these categories are comparatively well represented in the corpus and contain strong lexical or policy-oriented cues. Performance is substantially weaker for narrow or low-support categories such as *Education orientation* and *Education system monitoring & evaluation*, suggesting that the hierarchical decomposition strategy is partially successful. By first debating high-level dimensions before selecting low-level tags, the system is able to maintain relatively strong recall despite the large label space (a 141-label multi-label problem). The architecture therefore succeeds at narrowing the search space and guiding the downstream tag prediction process. Nevertheless, the transition from high-level dimensions to specific sub-tags remains noisy, particularly when multiple semantically adjacent categories coexist within the same paragraph.

4.1.2 Argumentative Relation Prediction

Evaluating relation extraction in documents with high paragraph counts presents a unique challenge due to the quadratic ($\mathcal{O}(N^2)$) scaling of possible links. Moreover, the majority of paragraph pairs do not have an argumentative relationship, leading to extreme class imbalance. For a comprehensive evaluation of our system, we report the results using two distinct paradigms. The first focuses exclusively on the model's ability to extract and correctly classify active argumentative edges by ignoring the heavily dominant none class. The second is a Full

Matrix Evaluation, which assesses every possible paragraph combination in the document, factoring in true negative (none-none) agreements.

Relation Class	Precision	Recall	F_1 Score	Support
Complemental	2.40	52.09	4.58	597
Supporting	0.84	35.58	1.64	104
Contradictive	0	0	0	7
Modifying	4.85	3.10	3.78	969
Standard Aggregate Performance (Excluding 'none' class)				
Weighted Avg	3.71	22.54	3.92	1677
None	99.42	87.46	93.06	135701
Full Aggregate Performance (Including 'none' class)				
Weighted Avg	98.25	86.67	91.97	137378

Table 4: Performance metrics for Subtask 2. The table contrasts the Standard Evaluation (which isolates active relation classes) against the Full Matrix Evaluation (which includes the sparse none class).

The evaluation metrics highlight the severe difficulties associated with highly imbalanced relation extraction under constrained parameter limits. When including the none class, the system achieves a 91.97 weighted F_1 score. However, the Standard Evaluation reveals significant shortcomings in the Phase 2 LLM Inference. The system suffers from critically low precision across all active classes, indicating that while the model captures a moderate portion of valid links, it fundamentally over-predicts active relations, generating a massive volume of false positives. This signals that the Phase 1 Asymmetric Distance-Decay filtering should be more strict, minimising the search space even further, and that the model should be asked to be more "conservative" when considering whether a pair of paragraphs is related or not.

4.2 Performance Analysis and Discussion

The difference between our F_1 rank (2^{nd}) and the LLM-Judge rank (6^{th}) highlights the functional trade-offs inherent in a hybrid, parameter-constrained architecture.

Improving Precision through Heuristics: The second place F_1 rank suggests that utilising deterministic heuristics (such as the lexical triggers and asymmetric distance-decay) effectively constrained the search space. This strategy mitigated the tendency of LLMs to hallucinate argumentative components, a behaviour frequently observed in purely generative solutions.

Limitations in Explanation Quality: The sixth place ranking in the LLM-Judge evaluation suggests a limitation in the system's capacity for deep qualitative explanation. Although the 8B model successfully performed the classification, relying on it to explicitly reason and explain its choices for the first and last tasks produced justifications that likely fell short of the evaluator's linguistic and qualitative standards.

Evaluating the Hybrid Approach: These results demonstrate that for highly regulated domains like UN resolutions, hybrid architectures provide superior extraction accuracy (F_1) by utilising the document's structure to guide the LLM. However, a significant gap remains between efficient, accurate extraction, and the generation of high-quality reasoning under strict parameter limitations.

5 Conclusion

In this paper, we described a hybrid architecture designed to parse the argumentative structure of UN resolutions under strict parameter limitations. By combining deterministic heuristics, hierarchical agentic debate, and asymmetric semantic retrieval, we showed how smaller language models can perform complex, domain-specific reasoning tasks.

Our modular architecture opens several promising avenues for future research across all components of the pipeline. First, the zero-shot LLM fallback could be replaced with either a specialised, distilled model trained exclusively on the edge-cases that fail the deterministic heuristics, or simply by replacing the zero-shot prompt with a few-shot prompt. Second, for the multi-label tagging, incorporating Retrieval-Augmented Generation (RAG) (Lewis et al., 2021) to ground a third "Domain Expert" agent in historical UN databases could help resolve highly technical taxonomy disputes between the Expansionist and Skeptic agents. Furthermore, while our asymmetric distance-decay formula effectively models UN document topology, future work should explore fine-tuning the embedding model via contrastive learning on UN-specific corpora, or replacing the heuristic decay entirely with a learned structural prior using Graph Neural Networks (GNNs). Finally, while our Phase 2 inference focuses on intra-document relations, we could extend our counterfactual reasoning prompts to handle inter-document dependencies, allowing the system to map argumentation graphs across decades of historical UN resolutions.

References

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, and 1 others. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2023. Improving factuality and reasoning in language models through multiagent debate. *Preprint*, arXiv:2305.14325.

Yingqiang Gao, Fabian Winiger, Patrick Montjourides, Anastassia Shaitarova, Nianlong Gu, Simon Peng-Keller, and Gerold Schneider. 2025. SpiritRAG: A Q&A System for Religion and Spirituality in the United Nations Archive. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 26–41.

Lena Held and Ivan Habernal. 2025. Contemporary LLMs struggle with extracting formal legal arguments. In *Proceedings of the Natural Legal Language Processing Workshop 2025*, pages 292–303, Suzhou, China. Association for Computational Linguistics.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Preprint*, arXiv:2005.11401.

Rūta Liepiņa, Francesca Galloni, Francesca Lagioia, Marco Lippi, Mariaceleste Musicco, Burcu Sayin, Andrea Passerini, and Giovanni Sartor. 2025. Legal argument mining: Recent trends and open challenges. In *Proceedings of the First Argument Mining and Empirical Legal Research Workshop (AMELR 2025) co-located with the 20th International Conference on Artificial Intelligence and Law (ICAIL 2025)*, volume 4089 of *CEUR Workshop Proceedings*, pages 1–14. CEUR-WS.org.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022a. Emergent abilities of large language models. *Preprint*, arXiv:2206.07682.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022b. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

A Prompt Templates

Below are the prompt templates utilized in our pipeline. Variables provided at runtime (such as the paragraph text or agent names) are denoted by bracketed placeholders (e.g. `{paragraph}`).

Preambular vs Operative LLM Inference

Classify the UN resolution paragraph as 'preambular' or 'operative'.
Provide reasoning and then the final JSON with the key "type".
Paragraph: {paragraph}

High-Level Debate: Agent A (Expansionist)

Debate role: Expansionist. Bias toward broader coverage and include multiple labels when reasonably supported by the UN resolution paragraph.
Turn: {turn}/4
Paragraph: {paragraph}
High-level labels: {$label_1$}, {$label_2$} ...
Current selection (may be empty): {current_selection}
Orchestrator summary so far: {summary}
Modify the current selection by adding and/or removing high-level labels so it best matches the paragraph. Output JSON array of label strings only.

High-Level Debate: Agent B (Skeptic)

Role: Skeptic. Bias toward parsimony and keep labels as few as possible, including a label only when clearly justified by the UN resolution paragraph.
Turn: {turn}/4
Paragraph: {paragraph}
High-level labels: {$label_1$}, {$label_2$} ...
Current selection (may be empty): {current_selection}
Orchestrator summary so far: {summary}
Modify the current selection by adding and/or removing high-level labels so it best matches the paragraph. Output JSON array of label strings only.

High-Level Debate: Agent A (Orchestrator)

ou are the orchestrator for a debate over labels for UN resolution paragraphs. You summarize the debate and decide whether more debate is needed. Return ONLY a JSON object with keys: summary (string), continue (boolean), selection (array). No extra text.

Tag LLM Inference

You are an expert education labeller selecting sub-level labels for ONE high-level label from UN resolution paragraphs. You must be concise, evidence-based, and context-aware of UN resolution language (normative statements, policy commitments, rights framing, implementation language). Follow these rules: 1) Select only from the allowed categories. 2) Use the paragraph as primary evidence and the orchestrator summary as supporting context. 3) Prefer precision over over-labeling, but include multiple categories when clearly supported. 4) Do not invent categories or rely on external facts. 5) Keep reasoning concise and tied to concrete phrases/themes in the paragraph. Return ONLY a JSON object with keys: selection (array of category strings), thinking (string). No extra text.

Relation LLM Inference (Phase 2)

You are mapping the internal logic of a single, isolated UN document.
You will be given an Anchor Paragraph and a Candidate Paragraph.
Determine the relationship between the Candidate and the Anchor.
Output ONLY a JSON object in this exact format: {"relation": "label"}.
The label must be one of: 'contradictive', 'supporting', 'complemental', 'modifying', or 'none'.
SUPPORTING: Candidate provides justification, evidence, or context making Anchor's directive valid.
COMPLEMENTAL: Candidate addresses the same theme as Anchor and adds additional info, without depending on each other.
MODIFYING: Candidate changes, qualifies, restricts, or expands the scope of the Anchor.
CONTRADICTIVE: Candidate asserts something conflicting with the Anchor.
NONE: The paragraphs are not related.
Counterfactual test: Does the existence of the Candidate amend, restrict, or expand the specific mandate established in the Anchor?
- If YES -> 'modifying'.
- If NO (it adds a related but separate action) -> 'complemental'.
Provide your thinking process in <think>...</think> tags, then output the final JSON.
If the value is none, no need to output it.
Here are some examples of paragraphs and their relation
Example 1
Paragraph A: Considering that a certain number of students admitted to secondary schools are not in a position to benefit effectively from the instruction provided therein;
Paragraph B: Deems it necessary, in order to avoid as much as possible errors in orientation and the discouragement that may result, to organize student guidance during the final regulated year of primary education, with the collaboration of the teacher, the physician, and the vocational guidance service, with the decision remaining the responsibility of the family.
Relation: supporting
Example 2
Paragraph A: Considers desirable greater coordination between primary education and secondary education in order to facilitate, especially during the initial years of study, the easy transition from one category of education to another.
Paragraph B: Deems it necessary, in order to avoid as much as possible errors in orientation and the discouragement that may result, to organize student guidance during the final regulated year of primary education, with the collaboration of the teacher, the physician, and the vocational guidance service, with the decision remaining the responsibility of the family.
Relation: complemental
Example 3
Paragraph A: Considers it desirable to improve the selection methods for admission to secondary schools proper. For this selection, the following elements should be taken into account: a) the primary school leaving certificate, as well as the individual report prepared by the primary school teachers, b) an examination conducted according to scientific methods aimed at identifying not only the knowledge acquired but also the candidate's aptitude to continue their studies.
Paragraph B: Draws the attention of educational authorities to the fact that, since any selection involves forced elimination, any student excluded from the secondary schools proper should be directed towards other studies or practical vocational training corresponding to their aptitudes.
Relation: modifying
Example 4
Paragraph A: Indigenous peoples have the right to self-determination. By virtue of that right they freely determine their political status and freely pursue their economic, social and cultural development.
Paragraph B: Nothing in this Declaration may be construed as authorizing or encouraging any action which would dismember or impair, totally or in part, the territorial integrity or political unity of sovereign and independent States.
Relation: contradictive