YNUzwt at SemEval-2025 Task 10: Tree-guided Stagewise Classifier for Entity Framing and Narrative Classification

Qiangyu Tan, Yuhang Cui, Zhiwen Tang
Yunnan Key Laboratory of Intelligent Systems and Computing,
Yunnan University, Kunming, China
School of Information Science and Engineering, Yunnan University, Kunming, China
tanqiangyu@stu.ynu.edu.cn, cyhstart@outlook.com, zhiwen.tang@ynu.edu.cn

Abstract

This paper introduces a hierarchical classification framework, termed the **Tree-guided Stagewise Classifier (TGSC)**, which adopts a Chain-of-Thought (CoT) reasoning paradigm to address multi-label and multi-class classification challenges in multilingual news article analysis, as part of SemEval-2025 Task 10. Our approach leverages the zero-shot capabilities of Large Language Models (LLMs) through a structured hierarchical reasoning process. The classification proceeds step-by-step through the hierarchy: beginning at the root node, the model iteratively traverses category branches, making decisions at each level, and ultimately identifying the appropriate leaf-level category. To enhance classification accuracy, we design a prompt engineering strategy that embeds hierarchical structural constraints to better guide the reasoning process. Experimental results demonstrate the effectiveness of TGSC, showing competitive performance across multiple languages in both Subtask 1 and Subtask 2. The code for our system is publicly available at: https://github.com/startuniverse/SemEval_2025_task10.

1 Introduction

SemEval-2025 Task 10[1] (Piskorski et al., 2025) focuses on the multilingual characterization and extraction of narratives from online news. This task is of strategic significance for both computational social science and multilingual natural language processing (NLP), as it addresses key aspects of contemporary information ecosystems—such as cross-cultural narrative modeling and resilience to disinformation. The task comprises three subtasks. Our team participated in Subtask 1 (Entity Framing) and Subtask 2 (Narrative Classification), evaluating our approach across three languages: English, Portuguese, and Russian.

To address both subtasks, we propose the Tree-guided Stagewise Classifier (TGSC), a hierarchical reasoning framework that systematically harnesses the zero-shot capabilities of LLMs through parent-child knowledge propagation.

The architecture of TGSC adopts a stagewise classification protocol in which parent-level predictions dynamically condition subsequent child-level decisions by restructuring the reasoning context provided to the LLM. At each level of the hierarchy, validated parent class labels are explicitly integrated into CoT prompts using constrained text generation templates. This design guides the LLM's zero-shot inference toward taxonomically valid subclasses. The top-down classification process improves accuracy: parent-level decisions eliminate irrelevant categories, while child-level classification leverages contextual cues to resolve ambiguities among finer-grained classes. Importantly, TGSC achieves hierarchical constraint propagation purely through prompt engineering, without relying on parametric gating mechanisms. This preserves full zero-shot flexibility while structurally preventing category violations. By recursively anchoring coarse-grained classifications to guide more specific decisions, TGSC enables efficient knowledge transfer across classification tiers, effectively balancing broad conceptual coverage with fine-grained precision.

Our participation in this task demonstrates that a zero-shot hierarchical reasoning framework can achieve competitive performance across languages. Our contributions include the following:

- **Cognitive Scaffolding Framework**: Our approach embodies the pedagogical "scaffolding theory" through hierarchical chain-of-thought reasoning. This architecture progressively reduces decision entropy by initially resolving parent-level categorical ambiguities, then recursively refining predictions through child-node analysis using dynamically updated con-

[1] https://propaganda.math.unipd.it/semeval2025task10

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 2133–2139

textual constraints. The multi-phase deliberation process intrinsically prevents semantic subspace contamination through early elimination of taxonomically incompatible candidates, mirroring human hierarchical reasoning patterns while maintaining computational tractability.

- **Parameter-Efficient Taxonomic Decomposition**:Our framework strategically decomposes hierarchical taxonomies into LLM-interpretable reasoning chains through structured prompts that cascade parent-level decisions into child inference contexts. This zero-shot approach leverages pre-trained models' inherent knowledge while preventing error propagation through contextual anchoring, achieving flexible yet precise classification. The taxonomy-aware mechanism balances conceptual breadth with granular differentiation without task-specific adaptations or labeled data.

2 Background

2.1 Problem Formulation

We participated in two subtasks of SemEval-2025 Task 10: Entity Framing and Narrative Classification.

In the Entity Framing task, the objective is to assign one or more roles to named entities mentioned in a news article, based on a predefined taxonomy. This task can be formulated as a multi-label, multi-class classification problem over text spans.

The Narrative Classification task involves assigning subnarrative labels to a news article according to a two-level taxonomy. Similar to Entity Framing, this is a multi-label, multi-class task, in which each article may be associated with multiple relevant subnarratives.

The annotation guidelines, including taxonomy definitions and operational instructions, are thoroughly documented in the official SemEval-2025 Task 10 technical report (Stefanovitch et al., 2025).

2.2 Related Work

Prior research in hierarchical classification has generally followed two main directions: hierarchical modeling approaches and reasoning-enhanced approaches.

Hierarchical Modeling Approaches. Dumais and Chen proposed hierarchical decomposition frameworks for efficient top-down categorization, though these methods typically require fully annotated datasets (Dumais and Chen, 2000). Wehrmann et al. introduced HMCN, a hybrid neural architecture that encodes hierarchical structures via specialized loss functions (Wehrmann et al., 2018). Zhu et al. developed HiTIN, a model that transforms label hierarchies into coding trees and incorporates structural encoders to encode hierarchical dependencies (Zhu et al., 2023). Pizarro et al. presented BA-CNN, an attention-based hierarchical model that enables dynamic feature flow between branches to enhance classification performance (Pizarro et al., 2023).

Reasoning-Enhanced Approaches. Cheng et al. proposed an end-to-end neural architecture search framework for hierarchical tasks, integrating domain-specific priors to guide model design (Cheng et al., 2020). In the realm of LLMs, Wei et al. introduced CoT prompting to support complex reasoning tasks (Wei et al., 2023), while Yao et al. explored tree-structured reasoning for more flexible and compositional inference (Yao et al., 2023). Zhang et al. proposed hierarchical knowledge integration to enrich LLM reasoning capabilities (Zhang et al., 2023). The AoR framework by Yin et al. uses dynamic sampling to select high-quality inference chains (Yin et al., 2024), and Goren et al. introduced Hierarchical Selective Classification (HSC), which refines inference by optimizing rule-based thresholds (Goren et al., 2025).

Unlike prior hierarchical models that rely heavily on annotated data or model modifications, our TGSC achieves hierarchical constraint propagation entirely through prompt engineering. TGSC harnesses the zero-shot reasoning ability of LLMs via structured Chain-of-Thought prompts, enabling taxonomically grounded multi-label classification without task-specific fine-tuning.

3 System Overview

3.1 Hierarchical Reasoning Algorithm

We introduce a Hierarchical Reasoning Algorithm (Algorithm 1) designed to overcome the limitations of traditional zero-shot classifiers by incorporating hierarchical constraints that reduce the prediction space and eliminate logical inconsistencies. This method directly addresses the challenges posed by flat label representations in complex taxonomies, where conventional approaches often fail to cap-

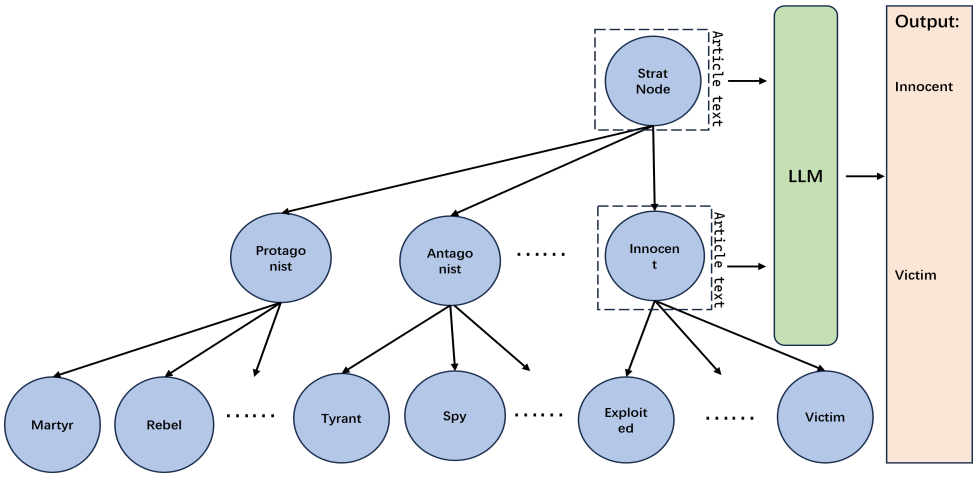

Figure 1: Overview of TGSC

ture the underlying hierarchical semantics. Inspired by the pedagogical theory of scaffolding, our algorithm mimics a staged, progressive learning process using a tree-structured reasoning framework.

During classification, the root node acts as the initial scaffold, offering a broad conceptual foundation to guide the model's reasoning. Internal nodes introduce intermediate constraints that systematically narrow the prediction space, steering the model through increasingly specific decision stages. This hierarchical guidance culminates at the leaf nodes, where the model makes fine-grained, context-aware predictions. The hierarchical structure of root and internal nodes serves not only as a constraint mechanism but also as a semantic guide that progressively structures the model's understanding from general to specific categories.

This top-down traversal—from root to internal nodes, and ultimately to leaf nodes—mirrors human decision-making patterns, where an initial high-level comprehension is incrementally refined into detailed, nuanced judgments. At each level, predictions are informed by the contextual signals propagated from higher tiers in the taxonomy. This staged reasoning process enhances prediction accuracy and coherence, aligning closely with cognitive strategies observed in human learning and decision-making.

3.2 Context-Aware Hierarchical Prompt Routing

We introduce a dynamic prompt engineering framework that leverages a novel context propagation mechanism through hierarchical prompt routing within a tree-structured classification process. In this framework, each node in the classification tree corresponds to a predefined prompt template, specifically designed for its respective classification level. As the model traverses the tree, each node integrates its prompt template with the article text to construct a level-specific input. This input serves as the final prompt for the LLM, guiding it to generate an informed and context-aware prediction.

The hierarchical nature of the tree ensures that context is progressively refined and propagated at each stage of reasoning. As the model moves from root to leaf nodes, the accumulated context from earlier stages constrains and informs subsequent predictions, enabling the LLM to perform increasingly fine-grained classification in alignment with the underlying taxonomy. This structured approach ensures semantic coherence across classification tiers and enhances the model's ability to distinguish

Algorithm 1 Hierarchical Reasoning Algorithm

Input: Article text T, Tree structure $T = \{N_1, N_2, ..., N_m\}$
Output: Label sequence $L = [L_1, L_2, ..., L_k]$

Initialize $L = \emptyset$
Set current node $N \leftarrow N_1$
while N is not a leaf node **do**
    Construct input: input $\leftarrow P_N(T)$
    Predict label: $L_N \leftarrow$ LLM(input)
    Append L_N to L
    Set current node $N \leftarrow$ child(N)
end while
Construct input: input$_N \leftarrow P_N(T)$
Predict label: $L_N \leftarrow$ LLM(input$_N$)
Append L_N to L
return L

among closely related subcategories.

An example of the prompt generation process for Subtask 2 is illustrated in Figure 2. Since the task structures of Subtask 1 and Subtask 2 are largely similar, we only provide the prompt template for Subtask 2. Additional implementation details and the full code are available in our GitHub repository.

4 Experimental Setup

4.1 Datasets and Evaluation Metrics

The experimental datasets span three languages: English, Portuguese, and Russian. We utilized the official test sets provided for evaluation. For Subtask 1, the dataset includes 63 English samples, 57 Russian samples, and 71 Portuguese samples, totaling 191 instances. For Subtask 2, there are 101 English samples, 60 Russian samples, and 100 Portuguese samples, resulting in 261 instances overall.

The official metrics used in both subtasks include exact match ratio (EMR) and Sample F1. EMR evaluates the prediction accuracy at the level of leaf nodes. Sample F1 is the average of F1 scores across all test samples.

4.2 Implementation Details

For our experiments, we utilized the OpenAI API for token-based inference. Additionally, we integrated the Phi-3.5 model (Abdin et al., 2024) into our framework. Inference with the Phi-3.5 4B model was conducted on a single NVIDIA RTX 4090 GPU.

5 Results

5.1 Main Result

In Subtask 1, our method achieved rankings of 15th, 13th, and 12th for English, Portuguese, and Russian, respectively. In Subtask 2, we attained rankings of 8th, 8th, and 5th, demonstrating competitive performance across all three languages. Tables 1 and 2 compare our results with those of leading competitors and baseline models for Subtasks 1 and 2, respectively.

Both subtasks are framed as multi-label, multi-class classification problems. However, a key distinction lies in their focus: Subtask 1 emphasizes local information—particularly the semantic attributes of named entities—whereas Subtask 2 centers on capturing broader, global narrative structures. The performance differences observed between the subtasks suggest that LLMs are particularly effective at modeling global semantic patterns but may struggle with fine-grained, localized reasoning due to noise or complexity in the input.

System	en	pt	ru
DUTIR	0.413(1)	0.593(1)	0.565(1)
PATeam	0.383(2)	0.492(2)	0.444(6)
DEMON	0.375(3)	0.367(6)	0.467(4)
TGSC	0.200(15)	0.162(13)	0.266(12)
HowardUniversity AI4PC	0.081(24)	0.131(14)	0.126(14)
Baseline	0.038	0.047	0.051

Table 1: Official Evaluation on Subtask 1

System	en	pt	ru
GATENLP	0.438(1)	0.480(1)	0.518(1)
PATeam	0.339(7)	0.409(2)	0.434(2)
iLostTheCode	0.320(9)	0.293(5)	0.411(3)
TGSC	0.321(8)	0.266(8)	0.335(5)
DUTtask10	0.165(23)	0.026(13)	0.033(13)
Baseline	0.013	0.014	0.008

Table 2: Official Evaluation on Subtask 2

This observation highlights an important direction for future work: enhancing the ability of LLMs to accurately extract and reason over local information, while preserving their strength in understanding global context. Developing methods to mitigate local noise and better isolate entity-level semantics could further improve classification performance in tasks requiring detailed linguistic precision.

5.2 Ablation Study

To evaluate the contribution of the hierarchical stagewise classification mechanism, we conduct an ablation study comparing our full model with a simplified variant that performs classification in a single stage, without hierarchical reasoning. The results are summarized in Table 3.

The hierarchical stagewise approach consistently outperforms the single-stage variant across both subtasks, highlighting its effectiveness in enhancing classification performance. These findings demonstrate that decomposing the task into multiple reasoning stages not only improves accuracy but also promotes consistency in multi-label, multi-class classification.

By first identifying a coarse-grained parent category, the stagewise process effectively constrains the subsequent prediction space, allowing downstream stages to focus on finer-grained distinctions within narrower semantic scopes. In contrast, the

Root node level prompt	*"input":* I want you to complete a narrative classification task. I will give you a news article. After you read the article, you have two options, 'Ukraine War' and 'Climate Change'. If you think the article is related to 'Ukraine War', output 'URW'. If you think the article is related to 'Climate Change', output 'CC'. Your output result should be only 'URW' or 'CC'. ***{news article}*** *"output":* CC	
Internal node level prompt	*"input":* There are subcategories under "CC". Please select the appropriate subcategory that is relevant to the article. The subcategories are as follows: ***{news article}*** ***{narrative_1;...;narrative_N}*** *"output":* CC: Criticism of climate policies	*"input":* There are subcategories under "CC". Please select the appropriate subcategory that is relevant to the article. The subcategories are as follows: ***{news article}*** ***{narrative_1;...;narrative_N}*** *"output":* CC: Controversy about green technologies
leaf node level prompt	*"input":* Please further categorize the specific content of "Criticism of climate policies" The following are more detailed classification options: ***{news article}*** ***{subnarrative_1;...;subnarrative_N}*** *"output":* CC: Criticism of climate policies: Climate policies are ineffective	*"input":* Please further categorize the specific content of "Controversy about green technologies" The following are more detailed classification options: ***{news article}*** ***{subnarrative_1;...;subnarrative_N}*** *"output":* CC: Controversy about green technologies: Nuclear energy is not climate friendly

Figure 2: Template of subtask 2

single-stage approach lacks this structured guidance, making it more susceptible to ambiguity and reduced precision. The observed performance gap reinforces the value of integrating hierarchical reasoning into the classification pipeline.

While the hierarchical structure is the key innovation of TGSC, we also investigated whether specific prompt template designs contributed significantly to the performance gains. To this end, we experimented with alternative prompt templates by varying the instruction styles, contextual phrasing, and information ordering across hierarchical levels.

The experimental results revealed that the overall performance remained largely stable across different prompt variants. This suggests that the hierarchical stagewise reasoning framework, rather than prompt template alone, plays the primary role in enhancing classification accuracy.

5.3 Experiments on Open-Source Models

We further evaluate TGSC on open-source LLMs to assess its generalizability beyond proprietary APIs. As shown in Table 4, introducing stagewise classification leads to a substantial performance improvement. Notably, the baseline model without stagewise classification performs at or near zero, underscoring the critical role of hierarchical, staged reasoning in achieving meaningful results on multi-label, multi-class classification tasks.

System	en	pt	ru
Subtask 1			
TGSC	**0.200**	**0.161**	**0.266**
TGSC w/o stagewise classification	0.187	0.128	0.238
Subtask 2			
TGSC	**0.321**	**0.266**	**0.335**
TGSC w/o stagewise classification	0.196	0.227	0.148

Table 3: Ablation Study

The stagewise approach enables the model to incrementally refine its predictions by traversing from coarse-grained parent categories to fine-grained child categories. This structured progression significantly narrows the decision space at each stage, reducing ambiguity and promoting more accurate, contextually grounded classifications. In contrast, single-stage inference lacks this hierarchical guidance, often resulting in vague or incorrect outputs.

These findings reaffirm the effectiveness of stagewise classification in enhancing model precision and consistency, particularly in complex classification scenarios where both global context and

System	en
Subtask 1	
TGSC(Phi 3.5)	**0.077**
TGSC(Phi 3.5) w/o stagewise classification	0.000
Subtask 2	
TGSC(Phi 3.5)	**0.049**
TGSC(Phi 3.5) w/o stagewise classification	0.000

Table 4: Experiments on Open Source Models

local semantic distinctions must be captured.

6 Conclusion

This paper introduces the Tree-guided Stagewise Classifier (TGSC), a hierarchical framework that leverages Chain-of-Thought reasoning for multi-label news categorization. TGSC progressively narrows the classification space through stagewise refinement, transitioning from coarse-grained parent-category hypotheses to fine-grained child-category predictions via structured reasoning paths. Experimental results validate the effectiveness of TGSC in zero-shot multilingual settings, demonstrating consistent accuracy improvements over baseline models across English, Portuguese, and Russian—without requiring model retraining.

Nevertheless, TGSC also has limitations. Specifically, the stagewise hierarchical structure introduces a potential risk of error propagation: misclassifications at parent nodes may constrain or mislead subsequent child-level predictions. In future work, we plan to explore mechanisms such as confidence-based node re-evaluation, multi-path traversal, or uncertainty-aware prompting strategies to further enhance robustness against hierarchical prediction errors.

Acknowledgments

This work is supported by the Open Project Program of Yunnan Key Laboratory of Intelligent Systems and Computing (ISC24Y03), and Yunnan Fundamental Research Project (202501AT070231).

References

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, Weizhu Chen, Yen-Chun Chen, Yi-Ling Chen, Hao Cheng, Parul Chopra, Xiyang Dai, Matthew Dixon, Ronen Eldan, Victor Fragoso, Jianfeng Gao, Mei Gao, Min Gao, Amit Garg, Allie Del Giorno, Abhishek Goswami, Suriya Gunasekar, Emman Haider, Junheng Hao, Russell J. Hewett, Wenxiang Hu, Jamie Huynh, Dan Iter, Sam Ade Jacobs, Mojan Javaheripi, Xin Jin, Nikos Karampatziakis, Piero Kauffmann, Mahoud Khademi, Dongwoo Kim, Young Jin Kim, Lev Kurilenko, James R. Lee, Yin Tat Lee, Yuanzhi Li, Yunsheng Li, Chen Liang, Lars Liden, Xihui Lin, Zeqi Lin, Ce Liu, Liyuan Liu, Mengchen Liu, Weishung Liu, Xiaodong Liu, Chong Luo, Piyush Madan, Ali Mahmoudzadeh, David Majercak, Matt Mazzola, Caio César Teodoro Mendes, Arindam Mitra, Hardik Modi, Anh Nguyen, Brandon Norick, Barun Patra, Daniel Perez-Becker, Thomas Portet, Reid Pryzant, Heyang Qin, Marko Radmilac, Liliang Ren, Gustavo de Rosa, Corby Rosset, Sambudha Roy, Olatunji Ruwase, Olli Saarikivi, Amin Saied, Adil Salim, Michael Santacroce, Shital Shah, Ning Shang, Hiteshi Sharma, Yelong Shen, Swadheen Shukla, Xia Song, Masahiro Tanaka, Andrea Tupini, Praneetha Vaddamanu, Chunyu Wang, Guanhua Wang, Lijuan Wang, Shuohang Wang, Xin Wang, Yu Wang, Rachel Ward, Wen Wen, Philipp Witte, Haiping Wu, Xiaoxia Wu, Michael Wyatt, Bin Xiao, Can Xu, Jiahang Xu, Weijian Xu, Jilong Xue, Sonali Yadav, Fan Yang, Jianwei Yang, Yifan Yang, Ziyi Yang, Donghan Yu, Lu Yuan, Chenruidong Zhang, Cyril Zhang, Jianwen Zhang, Li Lyna Zhang, Yi Zhang, Yue Zhang, Yunan Zhang, and Xiren Zhou. 2024. Phi-3 technical report: A highly capable language model locally on your phone.

Xuelian Cheng, Yiran Zhong, Mehrtash Harandi, Yuchao Dai, Xiaojun Chang, Tom Drummond, Hongdong Li, and Zongyuan Ge. 2020. Hierarchical neural architecture search for deep stereo matching.

Susan Dumais and Hao Chen. 2000. Hierarchical classification of web content. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, page 256–263, New York, NY, USA. Association for Computing Machinery.

Shani Goren, Ido Galil, and Ran El-Yaniv. 2025. Hierarchical selective classification.

Jakub Piskorski, Tarek Mahmoud, Nikolaos Nikolaidis, Ricardo Campos, Alípio Jorge, Dimitar Dimitrov, Purificação Silvano, Roman Yangarber, Shivam Sharma, Tanmoy Chakraborty, Nuno Ricardo Guimarães, Elisa Sartori, Nicolas Stefanovitch, Zhuohan Xie, Preslav Nakov, and Giovanni Da San Martino. 2025. SemEval-2025 task 10: Multilingual characterization and extraction of narratives from online news. In *Proceedings of the 19th International Workshop on Semantic Evaluation*, SemEval 2025, Vienna, Austria.

Iván Pizarro, Ricardo Ñanculef, and Carlos Valle. 2023. An attention-based architecture for hierarchical classification with cnns. *IEEE Access*, 11:32972–32995.

Nicolas Stefanovitch, Tarek Mahmoud, Nikolaos Nikolaidis, Jorge Alípio, Ricardo Campos, Dimitar Dimitrov, Purificação Silvano, Shivam Sharma, Roman Yangarber, Nuno Guimarães, Elisa Sartori, Ana Filipa Pacheco, Cecília Ortiz, Cláudia Couto, Glória Reis de Oliveira, Ari Gonçalves, Ivan Koychev, Ivo Moravski, Nicolo Faggiani, Sopho Kharazi, Bonka Kotseva, Ion Androutsopoulos, John Pavlopoulos, Gayatri Oke, Kanupriya Pathak, Dhairya Suman, Sohini Mazumdar, Tanmoy Chakraborty, Zhuohan Xie, Denis Kvachev, Irina Gatsuk, Ksenia Semenova, Matilda Villanen, Aamos Waher, Daria Lyakhnovich, Giovanni Da San Martino, Preslav Nakov, and Jakub Piskorski. 2025. Multilingual characterization and extraction of narratives from online news: Annotation guidelines. Technical Report JRC141322, European Commission Joint Research Centre, Ispra (Italy).

Jonatas Wehrmann, Ricardo Cerri, and Rodrigo Barros. 2018. Hierarchical multi-label classification networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5075–5084. PMLR.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models.

Zhangyue Yin, Qiushi Sun, Qipeng Guo, Zhiyuan Zeng, Xiaonan Li, Tianxiang Sun, Cheng Chang, Qinyuan Cheng, Ding Wang, Xiaofeng Mou, Xipeng Qiu, and XuanJing Huang. 2024. Aggregation of reasoning: A hierarchical framework for enhancing answer selection in large language models.

Jiajie Zhang, Shulin Cao, Tingjian Zhang, Xin Lv, Juanzi Li, Lei Hou, Jiaxin Shi, and Qi Tian. 2023. Reasoning over hierarchical question decomposition tree for explainable question answering. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14556–14570, Toronto, Canada. Association for Computational Linguistics.

He Zhu, Chong Zhang, Junjie Huang, Junran Wu, and Ke Xu. 2023. HiTIN: Hierarchy-aware tree isomorphism network for hierarchical text classification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7809–7821, Toronto, Canada. Association for Computational Linguistics.