Unlocking Human-Like Visible Logic: How Logic Diagrams Boost Logic Reasoning in Large Language Models?

Xin Wu[1,2,3], **Yuqi Bu**[4], **Mengchen Zhao**[1,2,3], **Qingbao Huang**[5], **Yi Cai**[1,2,3]*
[1]School of Software Engineering, South China University of Technology
[2] Key Laboratory of Big Data and intelligent Robot
(South China University of Technoloay), Ministry of Education
[3] Joint Guangdong-Hong Kong-Macao Research Laboratory of
Big Data and Robotic Intelligenceg, Ministry of Education
[4] School of Artificial Intelligence, Shenzhen Polytechnic University
[5] School of Electrical Engineering, Guangxi University

Abstract

Recent advancements in large language models (LLMs) have demonstrated their remarkable capabilities in natural language understanding and generation. However, these models still struggle with formal logical reasoning, often producing coherent yet invalid conclusions due to limitations in representing boundaries and relational structures through text alone. Human cognition frequently relies on visual representations to clarify logical structures involving category membership, inclusion, and relational hierarchies. Inspired by this, we investigate whether incorporating visual logic diagrams into LLMs' reasoning workflows can enhance their performance on formal logic tasks. We study this question in a controlled setting using syllogistic and conditional reasoning with programmatically generated Venn, Euler, and Linear diagrams. Across three Vision Language Models (VLMs) families, diagrams help in some settings but can also hurt performance, especially on logically invalid cases where models may over-rely on a single static visual instantiation. We therefore present this work as a reproducible evaluation framework and empirical analysis of when logic diagrams help or hinder language-conditioned reasoning.

1 Introduction

Recent advancements in artificial intelligence (AI) and natural language processing (NLP) have led to the development of Large Language Models (LLMs) capable of performing complex reasoning tasks. Despite their impressive capabilities, these models often struggle with formal logical reasoning (Ando et al., 2023; Lampinen et al., 2024; Eisape et al., 2024). Currently, LLMs primarily rely on textual input to conduct logical reasoning, which may inherently constrain their ability to represent abstract relationships and structures accurately (Ando et al., 2024). Language, though powerful, often falls short in expressing complex logical relationships, especially those involving boundaries and relational structures (Kosslyn and Pylyshyn, 1994). It frequently lacks the precision needed to clearly convey such concepts, leading to ambiguity and misinterpretation. For instance, when answering a syllogistic reasoning question, Qwen2-VL-7B-Instruct misinterprets the relational boundaries between sets (e.g., students, readers, essay writers). As shown in Figure 1, the model's reasoning appears syntactically sound but mishandles logical containment and intersection, drawing conclusions not supported by the premises. This highlights how LLMs can misrepresent structured relationships—such as set inclusion or overlap—when language alone is insufficient to encode precise boundaries and relational mappings.

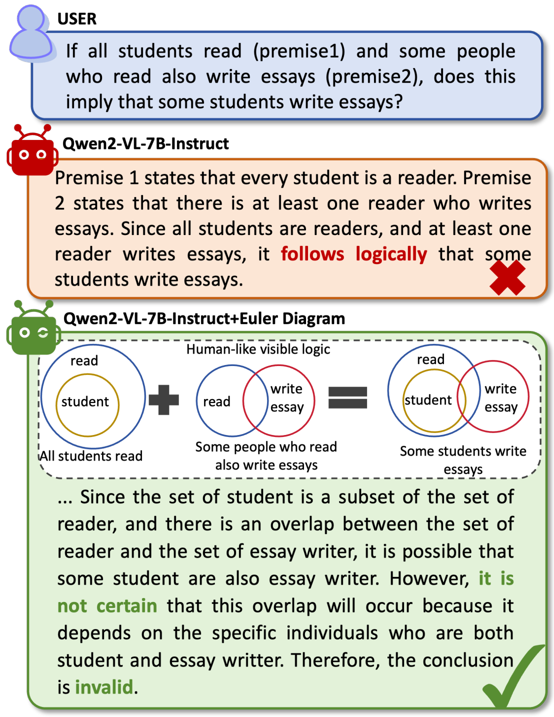

Figure 1: Using logic diagrams can help LLMs perform correct logical reasoning.

*Corresponding author. ycai@scut.edu.cn

In contrast, humans often rely on visual representations—such as diagrams and charts—to supplement or even replace textual descriptions in logical tasks. Visual tools like Venn diagrams, Euler diagrams, and flowcharts have long been used to clarify logical relations and abstract structures (Barwise and Etchemendy, 1996; Sato and Mineshima, 2015; Hooshyar, 2016; Bu et al., 2025). These diagrams provide a spatial and interpretable medium to represent inclusion, exclusion, overlap, and hierarchy, effectively addressing language's inefficiencies in expressing these dimensions (Sato et al., 2010, 2015). This raises a natural question: Can logic diagrams be integrated into LLMs inference processes to improve their logical reasoning performance?

As illustrated in Figure 1, when presented with a logic diagram that visually depicts the relationships between "students," "people who read," and "people who write essays," Qwen2-VL-7B-Instruct can recognize the logical structure and derive the correct conclusion. This suggests that logic diagrams can serve as explicit, interpretable evidence to assist LLMs in grasping complex relational patterns more effectively than through language alone. In this paper, we present a controlled empirical study of whether logic diagrams can help vision-language models with language-conditioned formal reasoning. We focus on two canonical reasoning types, syllogistic reasoning and conditional reasoning, and construct three diagram families: Venn, Euler, and Linear diagrams. We also study text-diagram alignment strategies, including color and symbolic cues, to better understand how different diagram designs affect model behavior.

The contributions of this paper are three-fold:

- We present a reproducible framework for constructing programmatic logic diagrams for canonical syllogistic and conditional reasoning instances and pairing them with textual premises for VLM evaluation.
- We systematically compare diagram families and alignment strategies, revealing model-specific preferences and trade-offs in diagram use.
- We show that logic diagrams can improve performance in some settings but can also reduce accuracy, especially on invalid cases, and we analyze this over-reliance failure mode.

2 Related Work

With the rapid development of data-driven models, significant attention has been directed toward evaluating the logic reasoning capabilities of deep learning models. Early work primarily focus on assessing various types of logical reasoning, including formal reasoning (Clark et al., 2021; Tafjord et al., 2021; Han et al., 2022; Wu et al., 2024), mathematical reasoning (Cobbe et al., 2021; Patel et al., 2021), relational reasoning (Sinha et al., 2019), informal reasoning (Chen et al., 2022; Yang et al., 2024), and mixed reasoning (Yu et al., 2020; Liu et al., 2021; Wu et al., 2026). These studies typically involved evaluating models' performance on tasks where models receives textual input and produce textual output. However, pure textual reasoning fails to align with human cognitive processes, where humans often rely on visual aids to aid their reasoning.

Recent advancements (Wang et al., 2018; Ando et al., 2024) demonstrated that Euler diagrams could enhance the reasoning process of deep learning models when performing syllogistic reasoning tasks. Wang et al. (2018) introdce Euler-Net to learn vector representations of diagrams. Ando et al. (2024) incoperate Euler diagrams into GPT-4 for syllogism reasoning. Building on this foundation, our work extends the investigation by exploring a wider variety of diagram types and alignment methods to improve logical reasoning in LLMs. While Ando et al. (2024) focus primarily on Euler diagrams and syllogistic reasoning, our approach spans additional diagrammatic forms and evaluates their effectiveness across multiple types of logical tasks, including syllogism and conditional reasoning. Furthermore, we extend the evaluation to three different LLMs, providing a broader perspective on the generalizability of diagram-based reasoning improvements across different models.

3 Logic Diagrams Construction

We begin by introducing the background knowledge of logic diagrams, focusing on three main types: Venn diagrams, Euler diagrams, and linear diagrams. We then outline a three-step process for constructing logic diagrams in the context of syllogistic and conditional reasoning: (1) type extraction, (2) diagram construction, and (3) diagram combination.

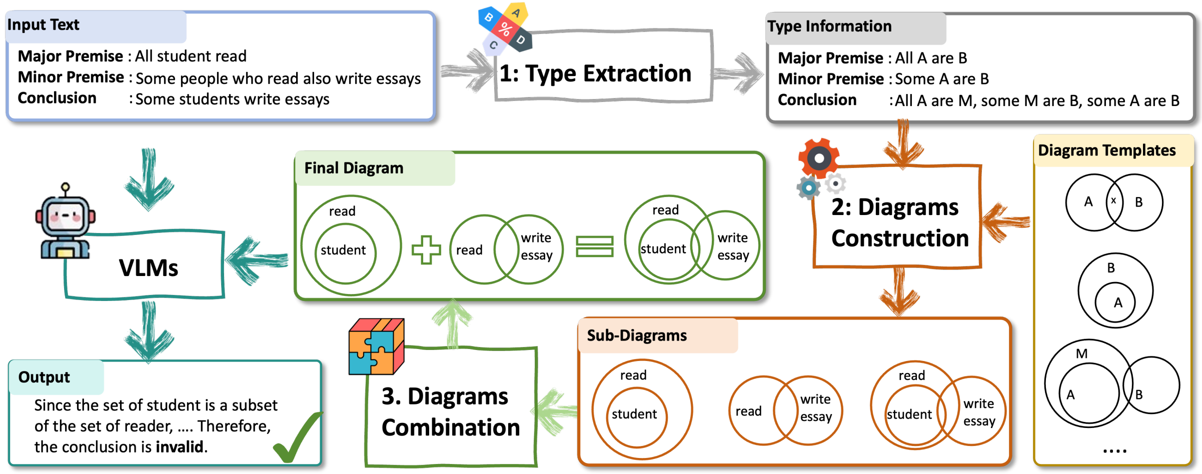

Figure 2: The proposed logic diagram construction framework, which includes three steps: (1) type extraction, (2) diagram construction, and (3) diagram combination.

3.1 Background

Logic diagrams serve as visual representations of logical relationships, aiding in the comprehension, analysis, and application of logical reasoning. These diagrams help illustrate the interplay between different logical elements and have been widely used in fields such as mathematics, computer science, philosophy, and artificial intelligence. Among the most commonly used logic diagrams are Venn diagrams, Euler diagrams, and Linear diagrams, each serving distinct purposes in representing logical structures and relationships. In this paper, we focus on these three main types of logic diagrams.

3.1.1 Venn Diagram

Venn diagrams (Venn, 1881) are widely used to depict set relationships and logical interactions. They consist of overlapping circles, where each circle represents a set, and the overlapping regions indicate shared elements between sets. Venn diagrams effectively illustrate concepts such as intersections, unions, and complements, making them essential tools for reasoning about categorical and propositional logic. Examples are shown in Figure 6. Each set is represented by a circle. Overlapping areas show common elements, while non-overlapping areas show differences. Shading represents an empty region (no elements exist). X marks show the presence of at least one element.

3.1.2 Euler Diagram

Euler diagrams (Euler, 1768) are similar to Venn diagrams but offer a more flexible and intuitive way of representing logical relationships. Unlike Venn diagrams, which always display all possible set intersections, Euler diagrams depict only the actual relationships that exist between sets. This makes them more efficient for representing real-world logical structures, as they avoid unnecessary complexity. Examples are shown in Figure 6. Circles represent different sets or categories. Their relationships (overlapping, containing, or separate) show logical connections.

3.1.3 Linear Diagram

A Linear diagram (Lambert and Arndt, 1764) is a variant of an Euler diagram. In an Euler diagram, two-dimensional circles are used to represent sets, whereas in a Linear diagram, one-dimensional line segments are used to represent sets. Each line segment in the diagram represents a set. Their relationships (overlapping, containing, or separate) show logical connections.

3.2 Syllogism Reasoning Diagrams Construction

A syllogism is a fundamental form of deductive reasoning that consists of three statements: two premises and one conclusion. The premises provide the foundational truths, and the conclusion logically follows from them. An example of syllogism is shown in Figure 2. Our diagram generation pipeline is rule-based and template-driven for the controlled reasoning forms studied here: it extracts logical types, matches them to a predefined template library, and instantiates the final diagram from those templates.

Type Extraction. To extract type information from syllogistic reasoning, we first identify the structure of each statement in terms of logical quantifiers and subject-predicate form. The major premise typically presents a universal relationship, such as "All students read," which fits the type pattern "All A are B". The minor premise then provides a particular relation, for example, "Some people who read also write essays," which corresponds to the type "Some A are B". By identifying the shared middle term (in this case, "read"), the conclusion is drawn: "Some students write essays." The type pattern for the full syllogism can thus be described as: "All A are M", "Some M are B", "Some A are B".

Diagrams Construction We build a diagram template library that contains all possible diagram templates. Depending on the type corresponding to the premise, templates are extracted from the library and filled with content.

For the sub-diagram of the conclusion, including the information from the premises is helpful for understanding the logical relationship. Therefore, we also construct such diagram templates for all possible syllogistic structures. For example, the syllogism in Figure 2 corresponds to the structure "All A are M, some M are B, some A are B." We construct a diagram template with three circles for this structure.

Diagrams Combination Once the sub-diagrams for individual premises are constructed, they are combined into a unified diagram using logical operations: The "+" symbol represents the integration of one sub-diagram with another. The final diagram on the right denotes the candidate conclusion being evaluated against the premises. It should not be interpreted as asserting logical equivalence between the combined premise diagram and the conclusion diagram.

3.3 Conditional Reasoning Diagrams Construction

Conditional reasoning involves the logical relationship between two premises—a rule and a specific case—and a hypothesis. The rule expresses a general principle or condition, typically in the form "If A, then B." The specific case provides a real-world scenario or instance in which the rule may apply. The hypothesis is a conclusion that is tested based on the application of the rule to the specific case.

Type Extraction Firstly, we abstract the information of the rule to obtain a form similar to "if P then Q". For example, given a rule "If it rains, the ground will be wet." We extract P as "raining" and Q as "ground will be wet".

Next, we identify the specific case's relationship to the rule. Given a specific case s, its relationship with the rule may be one of the following four: $s \in \{P\}$, $s \in \neg P$, $s \in \{Q\}$, $s \in \neg Q$.

Diagrams Construction Based on this relationship, we can plot the position of the specific case in the rule diagram. We use a star symbol to represent the specific case. If the relationship is $s \in \{P\}$, then the star symbol is placed inside the set represented by P (for example, inside the circle in an Euler diagram). If the relationship is $s \in \neg P$, then the star symbol is placed outside the range represented by P. Thus, we can create a diagram that represents the specific case. Similarly, we can also construct a diagram for the hypothesis.

Diagrams Combination Finally, to visually represent the entire reasoning process, we combine the three sub-diagrams (rule, specific case, and hypothesis) into one diagram with logical symbols.

4 Experiments

4.1 Datasets

We evaluate LLMs' logical reasoning using two tasks: syllogism reasoning and conditional reasoning. We measure accuracy to assess how logic diagrams improve LLMs' reasoning performance on these tasks.

Syllogism reaasoning dataset: We follow the construction method of Lampinen et al. (2024)'s, and based on 12 sets of candidate entity tuples, we fill them into a predefined syllogism form to build a syllogism reasoning dataset. All candidate entity tuples are shown in Table 5. All predefined syllogism forms are shown in Table 6. Table 7 presents 10 constructed samples. This dataset contains 616 samples (308 valid, 308 invalid). Each instance consists of two premises and a conclusion, testing whether the conclusion logically follows.

Conditional reasoning dataset: We extract 20 samples from the "first_order_logic/modus_tollens" category in LogicBench (Parmar et al., 2024) and expand them into 160 data points (including 40 valid and 120 invalid) based on the 8 scenarios in conditional reasoning. The specific cases are

	Syllogism Reasoning			Conditional Reasoning		
	Valid	Invalid	Total	Valid	Invalid	Total
Llava-1.5-7b	72.40	25.65	49.03	55.00	80.00	73.75
+Venn Diagram	49.35	**58.44**	53.90	47.50	**85.83**	76.25
+Euler Diagram	68.18	40.58	54.38	52.50	82.50	75.00
+Linear Diagram	75.97	29.87	52.92	47.50	79.17	71.25
Qwen2-VL-7B-Instruct	91.23	31.49	61.36	47.50	86.67	76.88
+Venn Diagram	93.83	28.25	61.04	80.00	70.00	72.50
+Euler Diagram	93.18	41.88	**67.53**	**90.00**	74.17	**78.12**
+Linear Diagram	**95.78**	32.79	64.29	77.50	72.50	73.75
LLama3.2-vision-11B	59.74	54.87	57.31	77.50	70.83	72.50
+Venn Diagram	80.52	25.32	52.92	82.50	72.50	75.00
+Euler Diagram	75.00	27.27	51.14	67.50	70.00	69.38
+Linear Diagram	77.92	39.31	58.12	77.50	74.17	75.00

Table 1: Incorporating logic diagrams into LLMs' syllogism reasoning and conditional reasoning. The green marking indicates an improvement compared to the baseline. The red marking indicates an "over-reliance" phenomenon, meaning a significant decline in the invalid category.

shown in Table 8. Each instance follows an "if-then" structure, requiring inference based on a given condition.

4.2 Baselines

We select several widely used open-source vision LLMs to evaluate the effectiveness of logic diagram. Including Llava (Llava-1.5-7b (Liu et al., 2024)), Llama Vision (Llama-3.2-11B-Vision-Instruct (Dubey et al., 2024)), Qwen (Qwen2-VL-7B-Instruct (Wang et al., 2024)).

4.3 Implementation Details

All experiments are implemented based on Hugging Face, with the generation parameters set to the default values for each model. All experiments are conducted on NVIDIA GeForce RTX 4090. Since the responses generated by LLMs contain both the reasoning chain and the final answer, we use a Llama-3.1-8B model to extract the answer information from the responses.

4.4 Overall Performance

The experimental results are shown in Table 1. We analyze them from the perspectives of overall improvement, sub-tasks and models, as well as the observed over-reliance phenomenon.

Overall Improvement. The effect of diagrams is heterogeneous rather than uniformly positive. Some model-task combinations benefit substantially from particular diagram types (34 out of 54, as indicated by the green markings), while other settings show a trade-off in which gains on valid cases are offset by drops on invalid cases. These improvements are observed across both syllogism reasoning and conditional reasoning tasks, suggesting that logic diagrams, such as Venn, Euler, and Linear diagrams, can enhance the ability of LLMs to process and reason through logical scenarios. However, different types of diagrams show varying improvements for different models. For example, for the Llava model, the improvement from the diagram mainly focuses on recognizing invalid samples. But for the Qwen model, the improvement from the diagram is mainly seen in valid samples.

	Valid	Invalid	Total
Only text	91.23	31.49	61.36
Venn Diagram			
Only Diagram	94.81	22.40	58.60
Text + Diagram	93.83	28.25	61.04
Euler Diagram			
Only Diagram	90.58	33.12	61.85
Text + Diagram	93.18	41.88	67.53
Linear Diagram			
Only Diagram	93.18	33.12	63.15
Text + Diagram	95.78	32.79	64.29

Table 2: Ablation experiment results of Qwen LLMs on the syllogism reasoning task.

Syllogism Reasoning. Qwen2-VL-7B-Instruct excels with the highest overall total accuracy across syllogism reasoning tasks, achieving 67.53% when Euler diagrams. Morever, Qwen2-VL-7B-Instruct

Premise 1: All old people are famous actors.
Premise 2: No wealthy people are famous actors.
Hypothesis: No wealthy people are old people.

Qwen2-VL-7B-Instruct: Premise 1 states that all old people are famous actors. This means that if someone is old, they must be a famous actor. Premise 2 states that no wealthy people are famous actors. This means that if someone is wealthy, they cannot be a famous actor. However, the hypothesis states that no wealthy people are old people. This is not necessarily true based on the given premises. It is possible for a wealthy person to be old and still be a famous actor, as long as they are not the same person. Therefore, the hypothesis does not follow logically from the premises.

Qwen2-VL-7B-Instruct+Euler Diagram: ... Since old people are a subset of famous actors and wealthy people are not a subset of famous actors, it follows that wealthy people cannot be a subset of old people. Therefore, the hypothesis ""No wealthy people are old people"" is a valid conclusion based on the given premises.

Premise 1: Someone attending college is expected to put in a lot of effort when it comes to their studies.
Premise 2: Jenny is not attending college.
Hypothesis: Jenny does not put in a lot of effort when it comes to her studies.

LLaMA 3.2-vision-11B: The premises logically support the hypothesis. Premise 1 establishes a general expectation for college students, and Premise 2 states that Jenny is not a college student. Therefore, it can be logically concluded that Jenny does not put in a lot of effort when it comes to her studies, as the expectation mentioned in Premise 1 does not apply to her.

LLaMA 3.2-vision-11B+Venn Diagram: ... Since Jenny is not attending college (Premise 2), she is part of the region outside the ""attending college"" set. Therefore, if she does not put in a lot of effort when it comes to her studies, she would be part of the intersection of the ""effort to studies"" set with the region outside the ""attending college"" set. However, there is no information provided in the premises that directly links not attending college to not putting in a lot of effort... The hypothesis does not logically follow from the premises.

Figure 3: Representative qualitative examples on syllogistic and conditional reasoning with and without logic diagrams. The examples illustrate how diagrams can change model attention to relevant set or condition structure.

stands out with the highest valid accuracy at 95.78% when using Linear diagrams. Llava-1.5-7b performs best in terms of invalid accuracy, particularly with Venn diagrams at 58.44%. This is the lowest error rate for invalid reasoning, suggesting the model is better at identifying invalid cases. For LLama3.2-vision-11B, although incorporating diagrams improves valid accuracy, it also causes a decrease in invalid accuracy, resulting in an overall increase in total accuracy only for Linear diagrams.

Conditional Reasoning. Qwen2-VL-7B-Instruct excels in total accuracy, achieving 78.12% when using Euler diagrams. However, we also observe a decrease in invalid accuracy for Qwen2-VL-7B-Instruct with each diagram. Similarly with syllogism reasoning, Llava-1.5-7b has the best performance for invalid reasoning, particularly with Venn diagrams, showing 85.83% invalid reasoning. This suggests that the model is more adept at avoiding errors in identifying invalid cases. For LLaMA 3.2-vision-11B, there is no decrease in invalid accuracy for conditional reasoning, and combining the Venn diagram and Linear diagram both led to improvements in total accuracy.

Diagrams Comparison. The experimental results show that different models have distinct preferences for the three types of diagrams. Specifically, Llava-1.5-7b performs well with all three diagrams, especially when combined with the Venn diagram, where it achieves the highest invalid accuracy in both syllogism and conditional reasoning tasks. It is also evident that Llava-1.5-7b primarily uses the diagrams to identify invalid samples. Similarly, Qwen2-VL-7B-Instruct also performs well with all three diagrams, particularly with the Euler diagram, where it achieves the highest total accuracy in both syllogism and conditional reasoning tasks. In contrast, LLama3.2-vision-11B shows significant variability in its use of different diagrams. Specifically, its use of the Euler diagram is noticeably weaker than with other diagrams, but it performs better with the Linear diagram, showing an improvement in total accuracy across both tasks. In summary, the three types of diagrams all contribute to enhancing the reasoning abilities of LLMs. The Venn diagram helps Llava-1.5-7b identify invalid samples, the Euler diagram combined with Qwen2-VL-7B-Instruct achieves the highest performance in two tasks, and the Linear diagram is the most comprehensible form for LLama3.2-vision-11B.

Over-reliance. Besides the performance improvement brought about in most cases, we also observe that incorporating diagrams leads to a significant decrease in invalid accuracy for certain models on specific tasks. For example, in conditional reasoning, the invalid accuracy of Qwen2-VL-7B-Instruct decreases. We name this phenomenon "over-reliance," and the reason behind it is that LLMs, when reasoning with a given diagram, treat the diagram as the correct answer without performing logical reasoning. However, the information depicted in the diagram is not necessarily correct, especially for invalid samples, where the diagram itself contains errors. LLMs need to use reasoning to detect these errors. However, due to over-reliance on visual information, LLMs end up treating incorrect information as correct, which leads to a significant drop in invalid accuracy. This suggests that current LLMs still have limitations when it comes to detecting erroneous information.

4.5 Ablation Study

In the ablation study, we mainly compare the impact of combining text and diagrams. Specifically, we compare the accuracy of Qwen2-VL-7B-Instruct on syllogism reasoning under different settings: only text, only diagram, and text+diagram. The experimental results are shown in Table 2. When incorporating Euler and Venn diagrams, the performance in the text+diagram setting is better than both the only text and only diagram settings. Additionally, Qwen2-VL-7B-Instruct, when reasoning with only diagrams, also achieves better performance than with only text, indicating that Qwen2-VL-7B-Instruct can understand and utilize Euler and Venn diagrams effectively. However, for Venn diagrams, Qwen2-VL-7B-Instruct does not show any performance improvement in either the only diagram or text+diagram settings. A possible reason for this is that Venn diagrams, compared to Euler and Linear diagrams, have a more complex representation of information, which hinders the understanding of LLMs. For example, in the expression of "All A are B" shown in Figure 6, a Venn diagram uses shading to represent the portion where "A" does not exist, while in Euler and Linear diagrams, the set representing "A" is fully contained within the set representing "B". This simpler representation is more conducive to understand.

Text-only LLMs. Our objective is to investigate whether LLMs are capable of understanding and utilizing logic diagrams. To this end, we primarily conduct experiments using vision-language models (VLMs) that can process image inputs. The experimental results support our hypothesis. Furthermore, we compare the performance of the base LLM used within the VLM against that of the VLM itself when handling only textual input. For Qwen2-VL-7B, its base LLM, Qwen-7B, achieves an accuracy of 57.14% on the syllogism reasoning task, which is lower than the 61.36% accuracy of Qwen2-VL-7B. This demonstrates the rationale for using VLMs as the baseline.

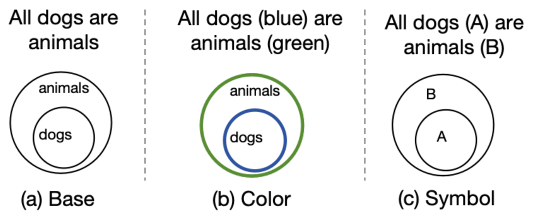

Figure 4: Three different text-diagram alignment methods: base, color and symbol.

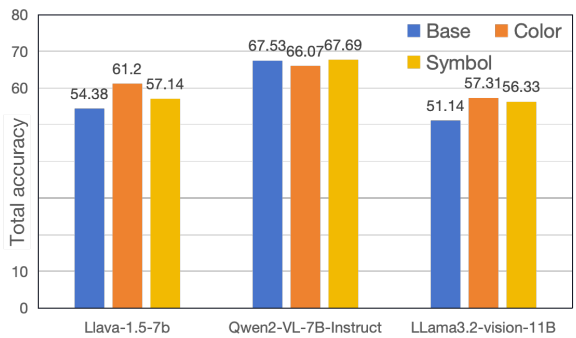

Figure 5: Different Alignment methods comparison on syllogism reasoning.

4.6 Case study

Figure 3 shows a comparison of Qwen2-VL-7B-Instruct and LLaMA 3.2-vision-11B, both with and without logic diagrams, in syllogism and conditional reasoning tasks. In the syllogism example, the text-only output proposes a conclusion that is inconsistent with the premise "No wealthy people are famous actors," whereas the diagram-assisted output more directly tracks the set relations and reaches the correct decision. In the conditional example, the text-only output exhibits the fallacy of denying the antecedent, while the diagram-assisted output better attends to the missing condition before accepting the hypothesis. We use these examples as qualitative illustrations of how diagrams can redirect model attention to relevant structure.

5 Discussion

5.1 Alignment Methods

We compare two other methods for aligning text and diagrams: color and symbol. Specifically, we add color and symbol markers after the original construction process to guide the LLMs in aligning the information between the text and the diagram. Figure 4 shows the three text-diagram alignment methods for the Euler diagram. In the color alignment method, we mark the circles with different colors and label them in the text. In the symbol

	Valid	Invalid	Total
Llava-1.5-7b			
w/o instruction	78.57	32.14	55.36
w/ instruction	68.18	40.58	54.38
Qwen2-VL-7B-Instruct			
w/o instruction	97.40	15.91	56.66
w/ instruction	93.18	41.88	67.53
LLama3.2-vision-11B			
w/o instruction	66.56	36.04	51.30
w/ instruction	75.00	27.27	51.14

Table 3: Comparison of proactivity of LLMs in utilizing visual information under the condition of whether instructions on how to use the diagram are provided.

alignment method, we do not fill in the specific content in the diagram but retain the symbols, with the corresponding content noted in the text.

The results in Figure 5 show that the alignment methods of color and symbol can further assist LLMs in improving their logical reasoning abilities. For the symbol alignment method, it results in performance improvements of 2.76%, 0.16%, and 5.19% for Llava-1.5-7b, Qwen2-VL-7B-Instruct, and LLama3.2-vision-11B, respectively. The color alignment method lead to performance improvements of 6.82% and 6.17% for Llava-1.5-7b and LLama3.2-vision-11B, respectively, which are greater than the improvements brought by symbol alignment, and ultimately resulted in higher accuracy. However, color alignment does not bring any improvement for Qwen2-VL-7B-Instruct. This indicates that LLMs have preferences for alignment methods.

5.2 Proactivity Analysis

We compare whether LLMs can autonomously and accurately understand information in diagrams without being provided instructions on how to read them. The results are shown in Table 3. When instructions are removed, the total accuracy of Llava-1.5-7b and LLama3.2-vision-11B does not decrease. This suggests that these two LLMs have certain background knowledge about how to use logic diagrams. However, for Qwen2-VL-7B-Instruct, its total accuracy drops by 10.87% when instructions are not provided, mainly due to a significant decline in its invalid accuracy. This indicates that without instructions, Qwen2-VL-7B-Instruct exhibits over-reliance, treating the given diagram as absolute correct information.

	Venn	Euler	Linear
Llava-1.5-7b	7/14	4/4	10/20
Qwen2-VL-7B-Instruct	26/48	32/48	36/50
LLama3.2-vision-11B	23/35	8/16	26/44

Table 4: Results of the human evaluation. "x/y" represents that "y" samples clearly use diagram-assisted reasoning, and "x" of those have correct reasoning.

5.3 Human Evaluation

We conduct a manual evaluation of the responses generated by LLMs to assess whether they use diagrams to assist in logical reasoning and whether they use the diagrams accurately.

Qwen. The results in Table 4 show that Qwen2-VL-7B-Instruct has the highest utilization rate of diagrams, reaching 96%, 96%, and 100% for Venn, Euler, and Linear diagrams, respectively. Additionally, Qwen2-VL-7B-Instruct demonstrates relatively high accuracy in understanding diagrams, with accuracy rates of 54%, 66%, and 72%, respectively.

Llava. In contrast, Llava-1.5-7b shows lower utilization rates of diagrams, at 28%, 8%, and 40%. However, this does not imply that logic diagrams are not helpful for Llava-1.5-7b. The statistical results only reflect samples where the diagram information is explicitly mentioned in the response. From the experimental results in Table 1, it appears that logic diagrams provide a relatively stable improvement for Llava-1.5-7b, suggesting that this improvement may be implicitly influencing its reasoning process.

Llama. For LLama3.2-vision-11B, the utilization rates for logic diagrams are 70%, 32%, and 88%, respectively. Interestingly, both Llava-1.5-7b and LLama3.2-vision-11B show very low utilization of Euler diagrams, which is in stark contrast to Qwen2-VL-7B-Instruct. This indicates that LLMs have preferences for different types of diagrams.

6 Conclusion

In this paper, we explore the impact of incorporating Venn, Euler, and Linear diagrams on enhancing the reasoning abilities of LLMs. The results show that in more than 60% of cases, visual logic diagrams can lead to performance improvements. Additionally, through comparative experiments, we reveal that different LLMs exhibit distinct preferences for specific diagram types and alignment methods. At the same time, we present the limitations of current LLMs in understanding diagrams, indicating that there is still room for improving LLMs' comprehension of diagrams. Future work could explore incorporating diagrams into pretraining during the fine-tuning phase; or explore an dynamic diagram understanding method to avoid the problem of over-reliance.

Several directions remain for future analysis. First, stronger text-only reasoning baselines, such as chain-of-thought and self-consistency prompting, would help isolate when diagram gains reflect genuinely useful visual grounding rather than improvements that could also be obtained from better textual prompting alone. Second, future work should study mitigation strategies for the over-reliance phenomenon, for example by treating the diagram as a hypothesis to be verified against the premises, or by using multiple-world or uncertainty-aware diagram renderings for invalid cases. Third, since our current pipeline is template-driven and evaluated in controlled settings, an important next step is to test robustness under paraphrased, noisy, or linguistically complex premises, and to measure how errors in type extraction or diagram construction propagate to final reasoning accuracy. Finally, it would be valuable to introduce a diagram-to-structure fidelity analysis to determine whether models truly recover the intended logical relations from the image, and to explore dynamic or hybrid diagram representations that better capture counterexamples and alternative valid worlds.

Limitations

This work is a controlled empirical study of diagram-grounded reasoning in canonical syllogistic and conditional forms. The diagram generator is rule-based and relies on template-compatible inputs, so the current framework does not directly extend to arbitrary natural-language arguments or highly noisy premises. In addition, the diagram families studied here are most practical for a small number of entities, and scalability becomes more difficult as the number of sets and relations grows. Finally, a single static diagram can encourage over-reliance, especially on invalid cases where correctness depends on considering alternative possible worlds or counterexamples. These results should therefore be interpreted as evidence about the current behavior of vision-language models under structured visual reasoning support, rather than as a complete solution to general logical reasoning.

Acknowledgements

This research is supported by the National Natural ScienceFoundation of China (62476097, 62506133), the Guangdong S&T Programme (2025B0101120003), Guangdong Provincial Fund for Basic and Applied Basic Research Regional Joint Fund Project (Key Project) (2023B1515120078), the Fundamental Research Fundsfor the Central Universities, South China University of Technology (x2rjD2250190). Yuqi Bu was supported by Shenzhen Polytechnic University Research Fund (6026310031K). Mengchen Zhao was partially supported by the Guangdong Basic and Applied Basic Research Foundation under Grant 2025A1515010247. Qingbao Huang was supported by Guangxi Natural Science Foundation Key Project (2025GXNSFDA069017), the Bagui Scholar Program of Guangxi (2025).

References

Risako Ando, Takanobu Morishita, Hirohiko Abe, Koji Mineshima, and Mitsuhiro Okada. 2023. Evaluating large language models with neubaroco: Syllogistic reasoning ability and human-like biases. In *Proceedings of the 4th Natural Logic Meets Machine Learning Workshop*, pages 1–11.

Risako Ando, Kentaro Ozeki, Takanobu Morishita, Hirohiko Abe, Koji Mineshima, and Mitsuhiro Okada. 2024. Can euler diagrams improve syllogistic reasoning in large language models? In *International Conference on Theory and Application of Diagrams*, pages 232–248. Springer.

Jon Barwise and John Etchemendy. 1996. Visual information and valid reasoning. *Logical reasoning with diagrams*, 3:25.

Yuqi Bu, Xin Wu, Zirui Zhao, Yi Cai, David Hsu, and Qiong Liu. 2025. Walk in others' shoes with a single glance: Human-centric visual grounding with top-view perspective transformation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26904–26923.

Jiangjie Chen, Rui Xu, Ziquan Fu, Wei Shi, Zhongqiao Li, Xinbo Zhang, Changzhi Sun, Lei Li, Yanghua Xiao, and Hao Zhou. 2022. E-kar: A benchmark for rationalizing natural language analogical reasoning. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3941–3955.

Peter Clark, Oyvind Tafjord, and Kyle Richardson. 2021. Transformers as soft reasoners over language. In *Proceedings of the Twenty-Ninth International Confer-*

ence on International Joint Conferences on Artificial Intelligence, pages 3882–3890.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, and 1 others. 2021. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Tiwalayo Eisape, Michael Tessler, Ishita Dasgupta, Fei Sha, Sjoerd Steenkiste, and Tal Linzen. 2024. A systematic comparison of syllogistic reasoning in humans and language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8418–8437.

Leonhard Euler. 1768. Lettres à une princesse d'allemagne sur divers sujets de physique & de philosophie.

Simeng Han, Hailey Schoelkopf, Yilun Zhao, Zhenting Qi, Martin Riddell, Wenfei Zhou, James Coady, David Peng, Yujie Qiao, Luke Benson, and 1 others. 2022. Folio: Natural language reasoning with first-order logic. *arXiv preprint arXiv:2209.00840*.

Danial Hooshyar. 2016. *A Flowchart-Based Intelligent Tutoring System Model to Improve Students' Problemsolving Skills*. University of Malaya (Malaysia).

Stephen M Kosslyn and Zenon Pylyshyn. 1994. Image and brain: The resolution of the imagery debate. *Nature*, 372(6503):289–289.

Johann Heinrich Lambert and Hans Werner Arndt. 1764. *Neues Organon*, volume 1. Wendler Leipzig.

Andrew K Lampinen, Ishita Dasgupta, Stephanie CY Chan, Hannah R Sheahan, Antonia Creswell, Dharshan Kumaran, James L McClelland, and Felix Hill. 2024. Language models, like humans, show content effects on reasoning tasks. *PNAS nexus*, 3(7):pgae233.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26296–26306.

Jian Liu, Leyang Cui, Hanmeng Liu, Dandan Huang, Yile Wang, and Yue Zhang. 2021. Logiqa: a challenge dataset for machine reading comprehension with logical reasoning. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 3622–3628.

Mihir Parmar, Nisarg Patel, Neeraj Varshney, Mutsumi Nakamura, Man Luo, Santosh Mashetty, Arindam Mitra, and Chitta Baral. 2024. Logicbench: Towards systematic evaluation of logical reasoning ability of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13679–13707.

Arkil Patel, Satwik Bhattamishra, and Navin Goyal. 2021. Are nlp models really able to solve simple math word problems? In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2080–2094.

Yuri Sato, Sayako Masuda, Yoshiaki Someya, Takeo Tsujii, and Shigeru Watanabe. 2015. An fmri analysis of the efficacy of euler diagrams in logical reasoning. In *2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 143–151. IEEE.

Yuri Sato and Koji Mineshima. 2015. How diagrams can support syllogistic reasoning: an experimental study. *Journal of Logic, Language and Information*, 24:409–455.

Yuri Sato, Koji Mineshima, and Ryo Takemura. 2010. The efficacy of euler and venn diagrams in deductive reasoning: Empirical findings. In *International Conference on Theory and Application of Diagrams*, pages 6–22. Springer.

Koustuv Sinha, Shagun Sodhani, Jin Dong, Joelle Pineau, and William L Hamilton. 2019. Clutrr: A diagnostic benchmark for inductive reasoning from text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4506–4515.

Oyvind Tafjord, Bhavana Dalvi, and Peter Clark. 2021. Proofwriter: Generating implications, proofs, and abductive statements over natural language. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3621–3634.

John Venn. 1881. Symbolic logic. *Nature*, 24(613):284–285.

Duo Wang, Mateja Jamnik, and Pietro Liò. 2018. Investigating diagrammatic reasoning with deep neural networks. In *International Conference on Theory and Application of Diagrams*, pages 390–398. Springer.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. 2024. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*.

Xin Wu, Yuqi Bu, Yi Cai, and Tao Wang. 2026. Triple-constrained and assumption-based zero-shot logical reasoning in reading comprehension. *Knowledge-Based Systems*, page 115702.

Xin Wu, Yi Cai, and Ho-fung Leung. 2024. Abstract-level deductive reasoning for pre-trained language models. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 70–76.

Zonglin Yang, Li Dong, Xinya Du, Hao Cheng, Erik Cambria, Xiaodong Liu, Jianfeng Gao, and Furu Wei. 2024. Language models as inductive reasoners. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 209–225.

Weihao Yu, Zihang Jiang, Yanfei Dong, and Jiashi Feng. 2020. Reclor: A reading comprehension dataset requiring logical reasoning. In *International Conference on Learning Representations*.

A Example Appendix

A.1 Over-reliance

The experimental results show that after incorporating diagrams, the invalid accuracy of some LLMs tends to decrease. The potential reason lies in the fact that an invalid diagram may appear to be correct at first glance, and if LLMs treat it as absolutely correct information, they can be misled. For example, the syllogism "All happy people are healthy people, Some healthy people are librarians, Some librarians are happy people" is invalid. The diagram drawn for this syllogism, shown in Figure 8, seems to have a correct structure, but in reality, the segment representing "librarians" doesn't necessarily need to intersect with the segment representing "happy people." As shown in Figure 7, at this point, the segments for "librarians" and "happy people" do not intersect, but this situation does not violate the premise. This indicates that the given premises can lead to multiple possible conclusions, meaning that "Some librarians are happy people" is not necessarily true, and thus the argument is invalid. When humans use diagrams, they often rely on imagination to explore potential changes in the elements of the diagram. However, current LLMs can only process a static image, making it difficult to perform this dynamic analysis. In the future, dynamic diagram displays or other logical representations could be explored to address this issue.

A.2 Form Analysis

We categorize and statistically analyze the accuracy of different forms of data in syllogism and conditional reasoning. The results are shown in Tables 9 and 10. An interesting phenomenon is that different diagrams have a noticeable impact on different forms. For example, the Euler diagram significantly improves the performance of Qwen2-VL-7B-Instruct in the AII, IAI, and AAI types of syllogisms. However, for the EAE and AEE types, the Venn diagram performs better. These trends also appear in the results for conditional reasoning. This suggests that the effectiveness of the diagram presentation is linked to the type of reasoning, similar to how humans apply different methods when performing different types of reasoning. This provides insight into designing hybrid diagram methods in the future.

A.3 More Diagrams

We present more cases of Euler, Venn, and Linear Diagrams in Figures 9-16. Our data and templates will also be made publicly available after acceptance.

A.4 Prompts

The prompts for only text, text+Euler, text+Venn, and text+Linear are shown in Figure 17, 18, 19, 20, respectively.

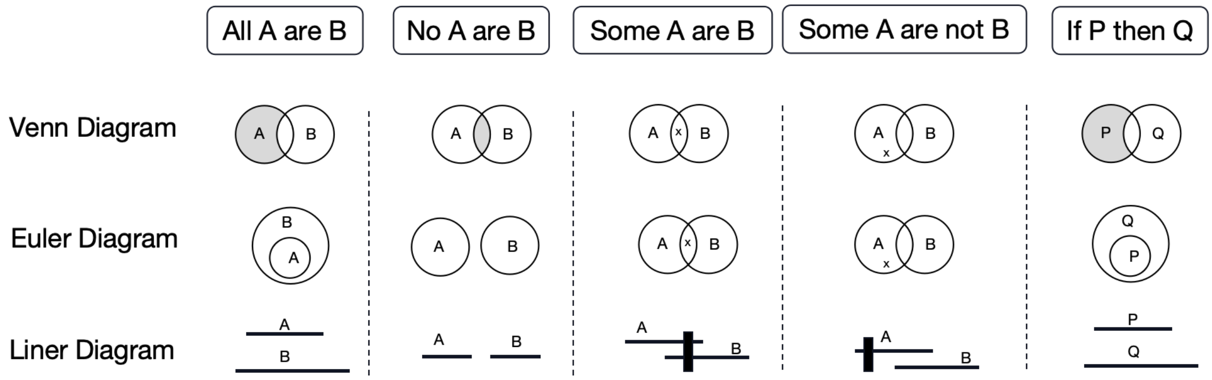

Figure 6: Venn diagram, Euler diagram, and Linear diagram of basic statements in syllogism and conditional reasoning. X marks and black rectangle shows the presence of at least one element.

Entity 1	Entity 2	Entity 3
librarians	happy people	healthy people
dragons	mythical creatures	things that exist
guns	weapons	dangerous things
politicians	dishonest people	people who lie
electronics	computers	expensive things
whales	mammals	big things
trees	plants	tall things
vegetables	foods	healthy things
flowers	animals	reptiles
famous actors	wealthy people	old people
diamonds	gems	transparent things
vehicles	things that move	buildings

Table 5: 12 entity tuples used to construct syllogism dataset.

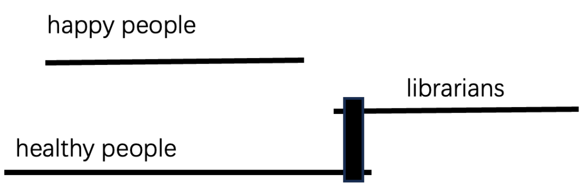

Figure 7: Another possible Linear diagram of "All happy people are healthy people, Some healthy people are librarians, Some librarians are happy people."

Form	Label	Template
1AAA	valid	All M are P, All S are M, All S are P.
1EAE	valid	No M are P, All S are M, No S are P.
1AII	valid	All M are P, Some S are M, Some S are P.
1AEE	invalid	All M are P, No S are M, No S are P.
1IAI	invalid	Some M are P, All S are M, Some S are P.
2AEE	valid	All P are M, No S are M, No S are P.
2EAE	valid	No P are M, All S are M, No S are P.
2AII	invalid	All P are M, Some S are M, Some S are P.
2AAI	invalid	All P are M, All S are M, Some S are P.
2IAI	invalid	Some P are M, All S are M, Some S are P.
3AAI	valid	All M are P, All M are S, Some S are P.
3AII	valid	All M are P, Some M are S, Some S are P.
3IAI	valid	Some M are P, All M are S, Some S are P.
3AAA	invalid	All M are P, All M are S, All S are P.
3EAE	invalid	No M are P, All M are S, No S are P
3AEE	invalid	All M are P, No M are S, No S are P.
4AEE	valid	All P are M, No M are S, No S are P.
4AAI	valid	All P are M, All M are S, Some S are P.
4IAI	valid	Some P are M, All M are S, Some S are P.
4AAA	invalid	All P are M, All M are S, All S are P.
4EAE	invalid	No P are M, All M are S, No S are P.
4AII	invalid	All P are M, Some M are S, Some S are P.

Table 6: 22 syllogism forms used to construct dataset.

Form	Label	Example
1AAA	valid	All expensive things are electronics, All computers are expensive things, All computers are electronics.
1AEE	invalid	All big things are mammals, No whales are big things, No whales are mammals.
2AEE	valid	All plants are tall things, No trees are tall things, No trees are plants.
2AII	invalid	All foods are healthy things, Some vegetables are healthy things, Some vegetables are foods.
3AAI	valid	All old people are wealthy people, All old people are famous actors, Some famous actors are wealthy people.
3AAA	invalid	All transparent things are gems, All transparent things are diamonds, All diamonds are gems.
4AEE	valid	All buildings are things that move, No things that move are vehicles, No vehicles are buildings.
4AAI	valid	All happy people are healthy people, All healthy people are librarians, Some librarians are happy people.
4EAE	invalid	No healthy people are happy people, All happy people are librarians, No librarians are healthy people.
4AII	invalid	All weapons are dangerous things, Some dangerous things are guns, Some guns are weapons.

Table 7: Syllogism examples.

Form	Label	Example
P Q	valid	If a car experiences a flat tyre, it is necessary to have it towed. The car has a flat tyre. The car is towed.
P not Q	invalid	If an individual possesses a car, they have the capability to embark on an extended expedition by road. Judy has a car, Judy does not embark on an extended expedition by road.
Not P Q	invalid	If someone is operating a car, they are obligated to adhere to the regulations governing traffic. Shelley is not operating a car, Shelley adheres to the regulations governing traffic.
Not P not Q	invalid	If one wishes to actively participate in an online event or activity, it is imperative for them to have a stable connection to the internet. Mark is not participating in an online event, Mark does not have a stable connection to the internet.
Q P	invalid	If an individual decides to have a pet dog, it becomes their responsibility to provide proper care for their beloved companion. Samantha provides proper care for her beloved companion, Samantha has a pet dog
Q not P	invalid	If a person commits a bank robbery, they will face legal charges for the act. Tom faces legal charges for the act, Tom does not commit a bank robbery
Not Q Not P	valid	If someone decides to go to the store, it means they have the intention to make a purchase. Linda does not have the intention to make a purchase, Linda does not go to the store
Not Q P	invalid	If an individual puts in dedicated effort and devotes ample time to studying, they will excel in their exams. Sara does not excel in her exams, Sara puts in dedicated effort and devotes ample time to studying

Table 8: Conditional reasoning examples.

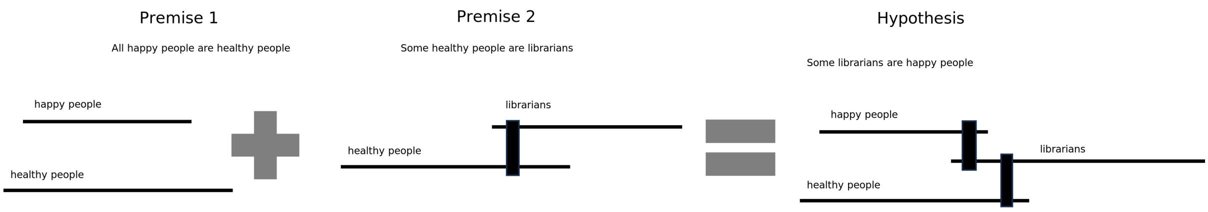

Figure 8: Linear diagram of "All happy people are healthy people, Some healthy people are librarians, Some librarians are happy people."

	AII	AAI	IAI	EAE	AEE	AAA
Llava-1.5-7b	52.86	53.57	46.43	51.85	39.81	50.00
+Venn Diagram	62.50	54.46	44.64	**59.26**	50.00	51.56
+Euler Diagram	59.82	61.61	47.32	54.63	46.30	57.81
+Linear Diagram	56.25	60.71	52.68	54.63	40.74	51.56
Qwen2-VL-7B-Instruct	74.11	60.71	53.57	**59.26**	53.70	70.31
+Venn Diagram	65.18	69.64	56.52	45.37	**60.19**	75.00
+Euler Diagram	**81.25**	**78.57**	**72.32**	49.07	55.56	67.19
+Linear Diagram	77.68	66.07	69.64	50.93	50.93	73.44
LLama3.2-vision-11B	70.54	54.46	46.43	53.70	51.85	73.44
+Venn Diagram	60.71	49.11	52.68	43.52	52.78	62.50
+Euler Diagram	62.50	40.18	49.11	48.15	42.59	73.44
+Linear Diagram	66.07	50.89	57.14	51.85	53.70	**76.56**

Table 9: Accuracy for each form in syllogism reasoning.

	P Q	P not Q	not P Q	not P not Q	Q P	Q not P	not Q P	not Q not P
Llava-1.5-7b	**100.00**	90.00	**100.00**	90.00	0.00	**100.00**	**100.00**	10.00
+Venn Diagram	90.00	**100.00**	**100.00**	**100.00**	20.00	**100.00**	95.00	5.00
+Euler Diagram	90.00	**100.00**	**100.00**	**100.00**	0.00	**100.00**	95.00	15.00
+Linear Diagram	90.00	**100.00**	**100.00**	95.00	0.00	90.00	90.00	5.00
Qwen2-VL-7B-Instruct	85.00	**100.00**	**100.00**	90.00	30.00	**100.00**	**100.00**	10.00
+Venn Diagram	90.00	95.00	95.00	25.00	15.00	95.00	95.00	70.00
+Euler Diagram	**100.00**	**100.00**	**100.00**	10.00	**35.00**	**100.00**	95.00	**85.00**
+Linear Diagram	85.00	**100.00**	95.00	20.00	30.00	95.00	95.00	70.00
LLama3.2-vision-11B	**100.00**	95.00	**100.00**	15.00	15.00	**100.00**	**100.00**	55.00
+Venn Diagram	95.00	95.00	**100.00**	30.00	15.00	**100.00**	95.00	70.00
+Euler Diagram	80.00	95.00	**100.00**	20.00	15.00	95.00	95.00	55.00
+Linear Diagram	90.00	**100.00**	95.00	25.00	25.00	**100.00**	**100.00**	65.00

Table 10: Accuracy for each form in conditonal reasoning.

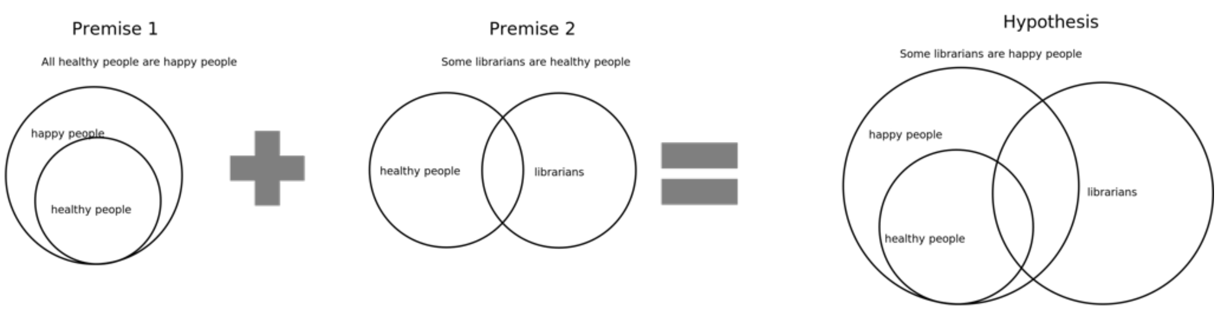

Figure 9: Euler diagram of "All healthy people are happy people, Some librarians are healthy people, Some librarians are happy people"

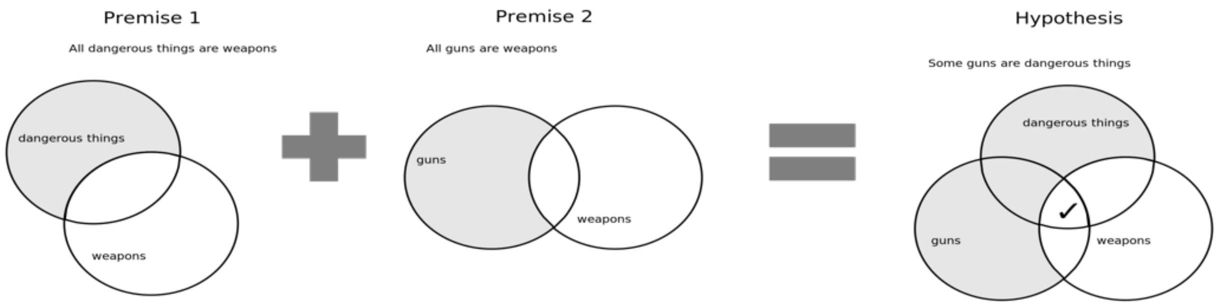

Figure 10: Venn diagram of "All dangerous things are weapons, All guns are weapons, Some guns are dangerous things"

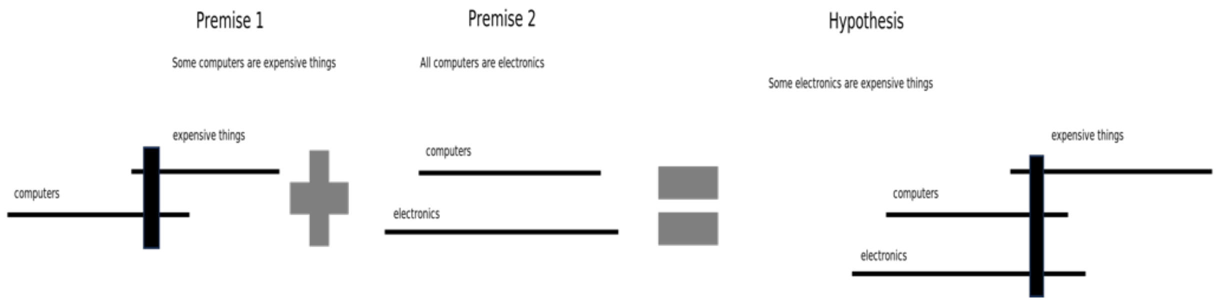

Figure 11: Linear diagram of "Some computers are expensive things, All computers are electronics, Some electronics are expensive things"

Figure 12: Euler diagram of "If one is employed in an office setting, it is customary to dress formally. David is employed in an office setting, David dresses formally."

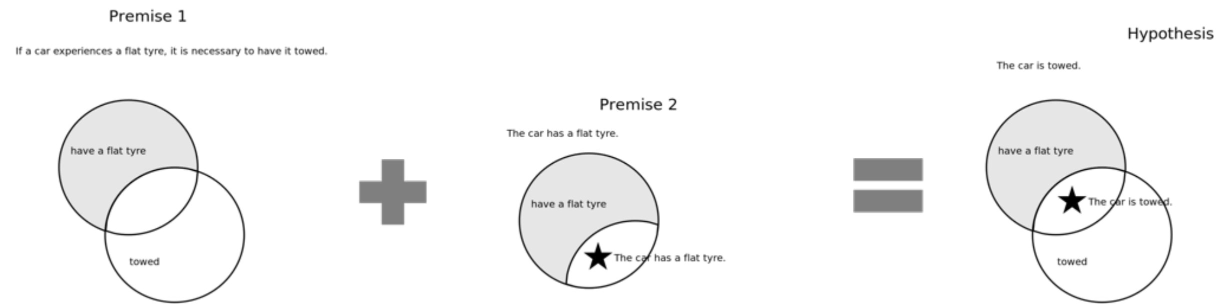

Figure 13: Venn diagram of "If a car experiences a flat tyre, it is necessary to have it towed. The car has a flat tyre, The car is towed."

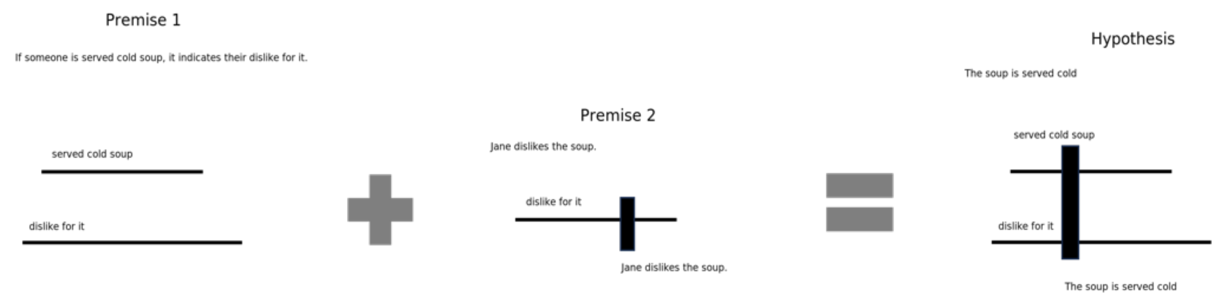

Figure 14: Linear diagram of "If someone is served cold soup, it indicates their dislike for it. Jane dislikes the soup, The soup is served cold."

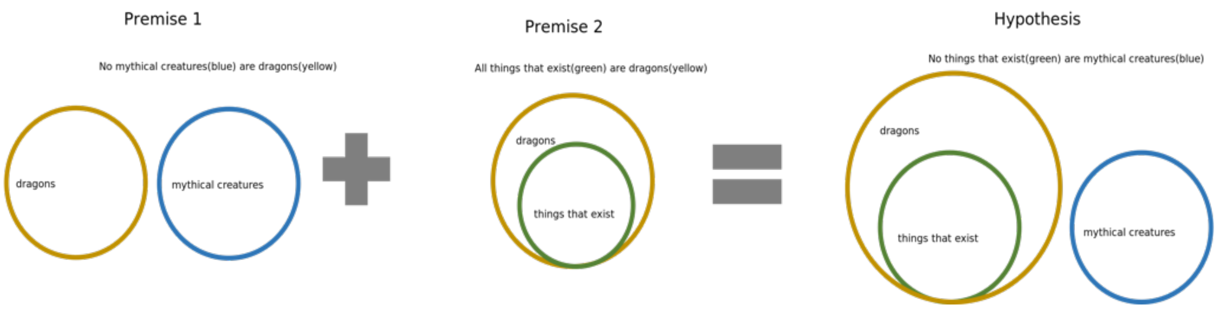

Figure 15: Euler diagram with color alignment of "No mythical creatures are dragons, All things that exist are dragons, No things that exist are mythical creatures"

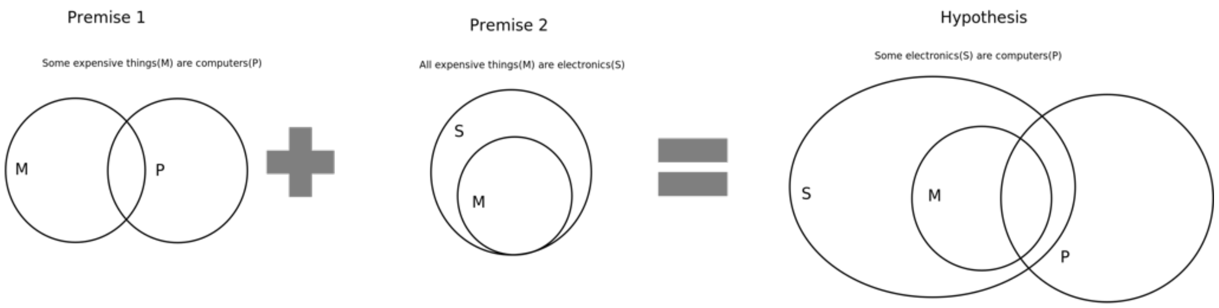

Figure 16: uler diagram with symbol alignment of "Some expensive things are computers, All expensive things are electronics, Some electronics are computers"

```
Determine the correct logical relationship between the given premises and the
hypothesis.
- Answer "valid" if the hypothesis follows logically from the premises.
- Answer "invalid" if the hypothesis does not follow logically from the premises.
Premise1: {}
Premise2: {}
Hypothesis: {}
Given a detailed explanation of your answer.
```

Figure 17: Prompt of only text input.

```
Determine the correct logical relationship between the given premises and the
hypothesis.
- Answer "valid" if the hypothesis follows logically from the premises.
- Answer "invalid" if the hypothesis does not follow logically from the premises.

Please analyze whether the relationships described in the hypothesis diagram are certain
to occur (valid) or are possible occurrences (invalid).
Premise1: {}
Premise2: {}
Hypothesis: {}
Given a detailed explanation of your answer.
```

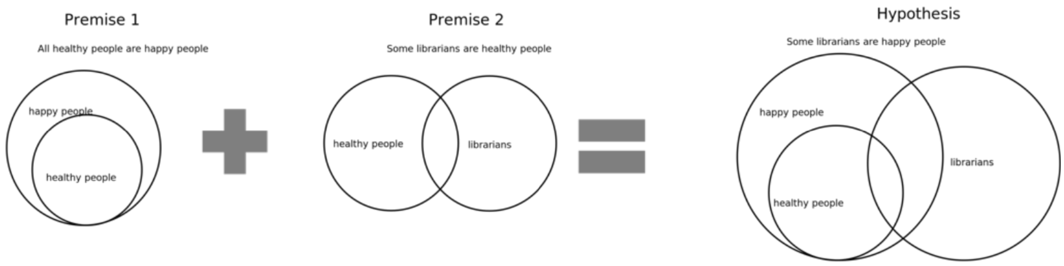

Figure 18: Prompt of text + Euler diagram input.

```
Determine the correct logical relationship between the given premises and the
hypothesis.
- Answer "valid" if the hypothesis follows logically from the premises.
- Answer "invalid" if the hypothesis does not follow logically from the premises.

Please analyze whether the relationships described in the hypothesis diagram are certain
to occur (valid) or are possible occurrences (invalid). Each circle in the diagram
represents a set, and the overlap of circles indicates an intersection between sets.
Shading represents an empty region. Check marks(✓) show the presence of at least one
element.

You SHOULD analyze whether the relationships described in the hypothesis diagram are
certain to occur (valid) or are possible occurrences (invalid).

Premise1: {}
Premise2: {}
Hypothesis: {}
```

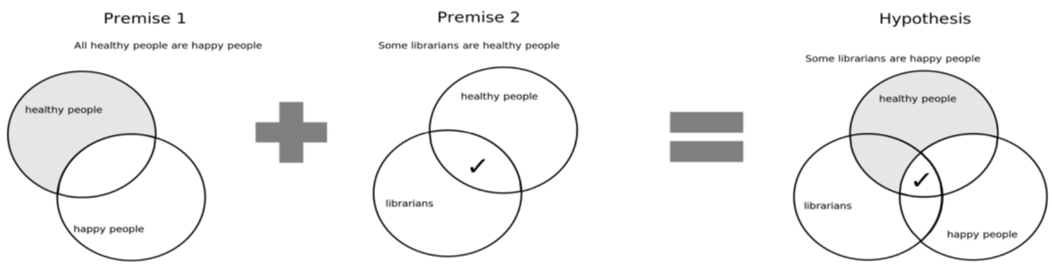

Figure 19: Prompt of text + Venn diagram input.

```
Determine the correct logical relationship between the given premises and the
hypothesis.
- Answer "valid" if the hypothesis follows logically from the premises.
- Answer "invalid" if the hypothesis does not follow logically from the premises.

Please analyze whether the relationships described in the hypothesis diagram are certain to
occur (valid) or are possible occurrences (invalid). Each line segment in the diagram
represents a set, and the overlap of line segments indicates an intersection between sets.

Premise1: {}
Premise2: {}
Hypothesis: {}
```

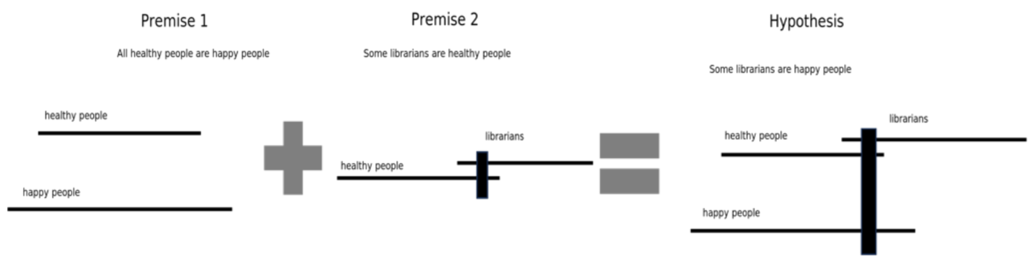

Figure 20: Prompt of text + Linear diagram input.