



# Pi-GPS: Enhancing Geometry Problem Solving by Unleashing the Power of Diagrammatic Information

Junbo Zhao, Ting Zhang, Jiayu Sun, Mi Tian, Hua Huang



## Motivation & Background

Geometry Problem Solving (GPS) is critical for intelligent education, requiring the integration of visual diagrams and textual descriptions.

- **Key Challenge:** Textual ambiguities often hinder problem comprehension, while diagrams can effectively resolve these ambiguities.
- **Existing Approaches:** Symbolic and neural methods struggle with precise alignment between text and diagrams, leading to errors in reasoning.
- **Our Perspective:** Text ambiguity is a major bottleneck in GPS, yet largely overlooked in previous research.

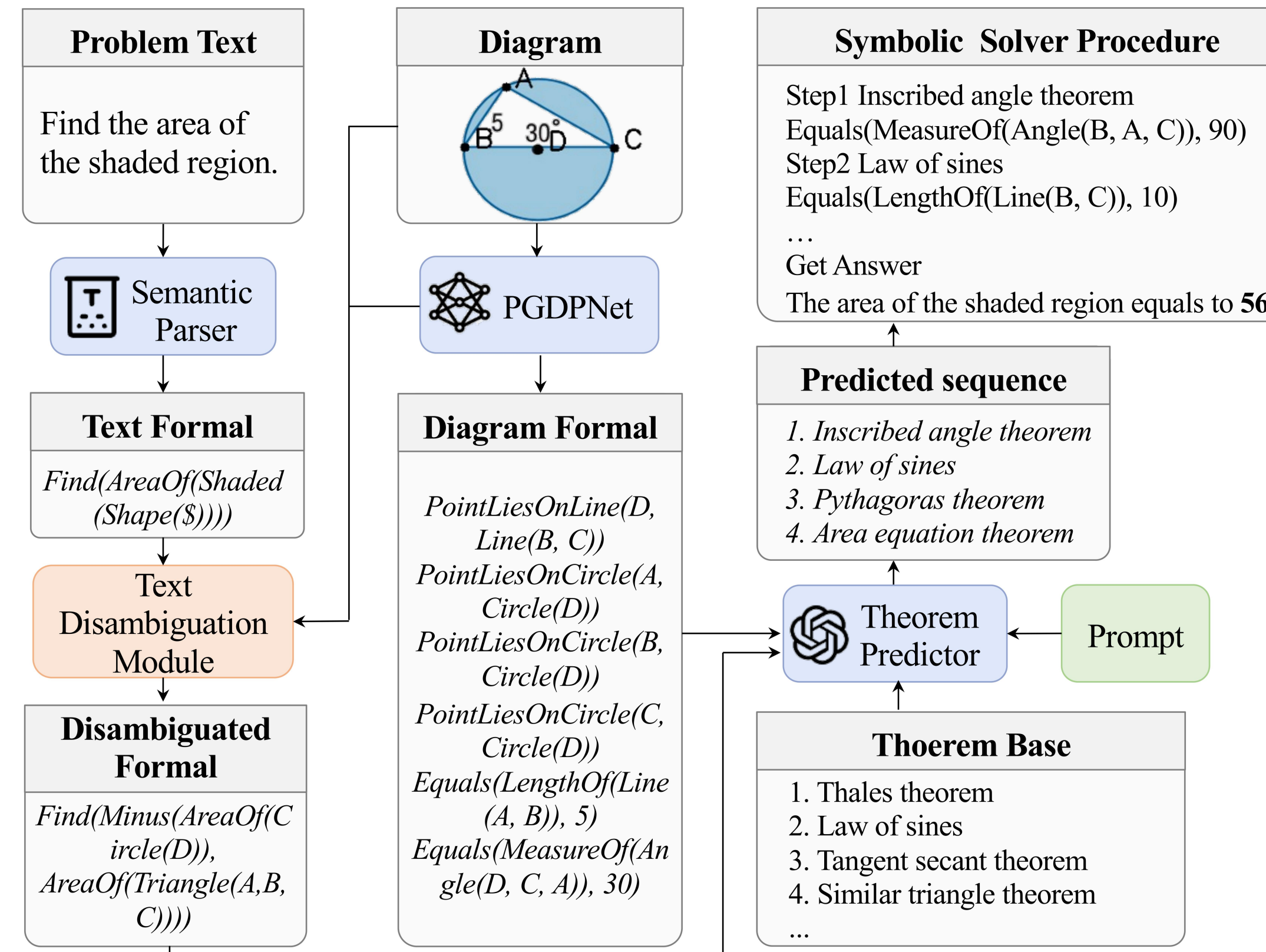
Problem Text	Parsed Text Formal Language	Diagram
The <b>rectangle</b> is inscribed into the <b>circle</b> . Find the exact circumference of the <b>circle</b> .	$\text{InscribedIn}(\text{Rectangle}(\$), \text{Circle}(\$))$ $\text{Find}(\text{CircumferenceOf}(\text{Circle}(\$)))$	
The two <b>polygons</b> are similar. Find UT.	$\text{Similar}(\text{Polygon}(\$1), \text{Polygon}(\$2))$ $\text{Find}(\text{LengthOf}(\text{Line}(U, T)))$	
Find the area of the <b>shaded region</b> . Round to the nearest tenth.	$\text{Find}(\text{AreaOf}(\text{Shaded}(\text{Shape}(\$))))$	

Problem Text	Formal / Disambiguated Formal	Diagram
The <b>rectangle</b> is inscribed into the <b>circle</b> . Find the exact circumference of the <b>circle</b> .	$\text{InscribedIn}(\text{Rectangle}(\$), \text{Circle}(\$))$ $\text{Find}(\text{CircumferenceOf}(\text{Circle}(\$)))$	
<b>Q</b> is the centroid and $BE = 9$ . Find $BQ$ .	$\text{IsCentroidOf}(\text{Point}(Q), \text{Shape}(\$))$ $\text{IsCentroidOf}(\text{Point}(Q), \text{Triangle}(A, C, B))$	
Find the area of the <b>shaded region</b> . Assume that the <b>polygon</b> is regular unless otherwise stated.	$\text{Regular}(\text{Polygon}(\$))$ $\text{Find}(\text{AreaOf}(\text{Shaded}(\text{Shape}(\$))))$ $\text{Regular}(\text{Triangle}(A, E, G))$ $\text{Find}(\text{Minus}(\text{AreaOf}(\text{Triangle}(A, E, G)), \text{AreaOf}(\text{Circle}(D))))$	

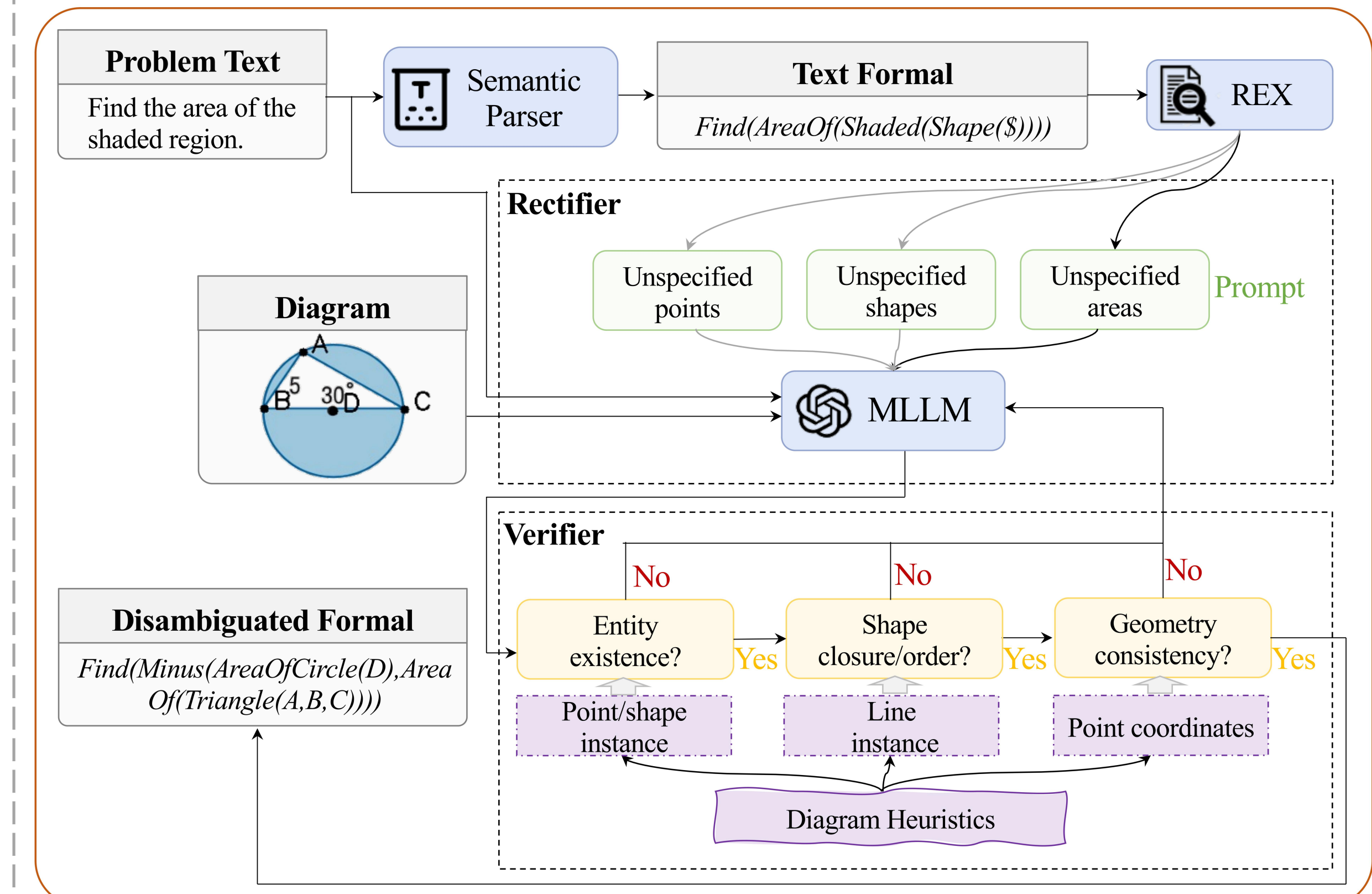
## Method: The Pi-GPS Framework

- **Parser:** text (rule-based); diagram (PGDPNet).
- **Disambiguator:** rectifier (MLLM; three ambiguities); verifier (diagram heuristics match entities/structure; curb hallucinations).
- **Reasoner:** theorem predictor (LLM/o3-mini); symbolic solver (executes sequence; interpretable).

### Pi-GPS framework



### Text disambiguation module



## Experiments & Results

### Key Findings:

- Pi-GPS achieves a nearly 10% improvement over previous state-of-the-art methods.
- The text disambiguation module is critical, delivering consistent performance gains.
- The verifier ensures reliable alignment and prevents MLLM hallucination.

Method	Completion	Choice
Ours w/o Text disam.	63.2	72.3
+ Rectifier (general prompt)	62.4	71.9
+ Rectifier (specific prompt)	64.2	73.3
+ Verifier	<b>70.6</b>	<b>77.8</b>

Task	Models	Completion	Choice
Direct solv. (MLLM)	GPT-4o	34.8	58.6
	Gemini 2	38.9	60.7
Direct solv. (LLM)	GPT-4o	36.5	59.7
	DeepSeek-R1	63.9	72.2
Theorem pred. (LLM)	o3-mini	66.4	75.5
	o3-mini w/o Text disam.	61.4	70.4
	o3-mini (ours)	<b>70.6</b>	<b>77.8</b>

Table 4. Illustrating the roles of the rectifier and verifier in the text disambiguation module on Geometry3K.

Category	Method	Geometry3K		PGPS9K	
		Completion	Choice	Completion	Choice
MLLMs	Qwen-VL [7]	22.1	26.7	20.1	23.2
	GPT-4o [1]	34.8	58.6	33.3	51.0
	Claude 3.5 Sonnet [6]	32.0	56.4	27.6	45.9
	Gemini 2 [15]	38.9	60.7	38.2	56.8
	NGS [8]	35.3	58.8	34.1	46.1
Neural Methods	Geoformer [10]	36.8	59.3	35.6	47.3
	SCA-GPS [23]	-	76.7	-	-
	GOLD* [40]	-	62.7	-	60.6
	PGPSNet-v2-S* [43]	65.2	76.4	60.3	69.2
	LANS (Diagram GT)* [18]	72.1	82.3	66.7	74.0
	Inter-GPS [21]	43.4	57.5	-	-
Neural-symbolic Methods	GeoDRL [25]	57.9	68.4	55.6	66.7
	E-GPS [36]	-	67.9	-	-
	Pi-GPS (ours)	<b>70.6</b>	<b>77.8</b>	<b>61.4</b>	<b>69.8</b>

Table 2. Comparison on Geometry3K and PGPS9K. Our method achieves the best performance (highlighted in bold) compared to the neural-symbolic methods. Note that all baselines except LANS use parsed results, while LANS uses textual clauses and point positions from diagram annotations. \* indicates that GOLD, PGPSNet and LANS are trained on the larger dataset, PGPS9K.

## Conclusion & Limitation

- **Contribution:** Pi-GPS adds a rectifier-verifier that leverages diagram context to disambiguate text, boosting geometric problem-solving.
- **Impact:** Underscores the importance of ambiguity resolution in multimodal mathematical reasoning.
- **Limitation:** Still constrained by brittle text-to-formal parsing, weak detection of subtle diagram relations, and an incomplete theorem base.

Problem Text:	Diagram:
A plane travels from Des Moines to Phoenix, on to Atlanta and back to Des Moines. Find distance in from Phoenix to Atlanta if trip was 3482.	
⚠ Text parser cannot resolve the question text into the correct formal language: $\text{Find}(\text{LengthOf}(B, C))$ .	
Problem Text:	Diagram:
Find the area of the shaded region.	
⚠ Diagram parser cannot parse the tangent relationship between circles: $\text{Tangent}(\text{Circle}(D), \text{Circle}(E))$ .	
Problem Text:	Diagram:
Find the area of the regular polygon.	
⚠ The absence of established theorems on regular hexagons precludes the deductive reasoning.	