

# **Math Vision: Fine-tuning Multi-modal LLMs for Geometry Problem Solving**

**Atul Shreewastav (THA077BCT013)**

**Bidhan Acharya (THA077BCT015)**

**Nischal Paudel (THA077BCT028)**

**Yugratna Humagain (THA077BCT047)**

**Under supervision of  
Er. Bibat Thokar and Er. Sushil Dyopala**

Department of Electronics and Computer Engineering  
Institute of Engineering, Thapathali Campus

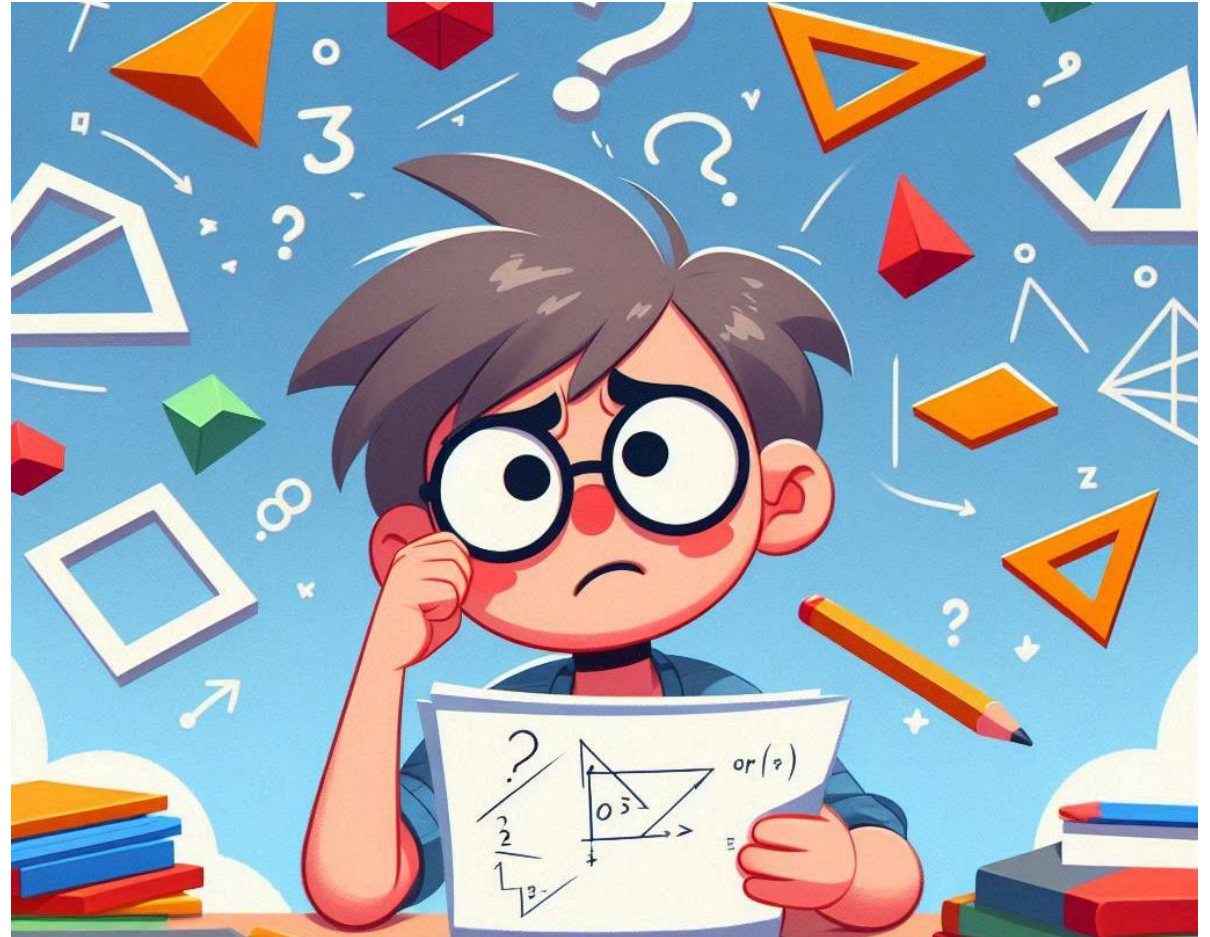
June 21, 2024

# Outlines

- Motivation
- Introduction
- Objectives
- Scope of Project
- Proposed Methodology
- Dataset Exploration
- Expected Results
- Project Application
- Gantt Chart / Project Budget
- References

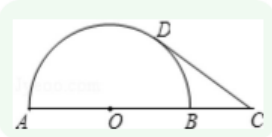
# Motivation

- Existing LLMs struggle with geometric shapes.
- High incidence of hallucinations and inaccuracies.
- Geometry problem solving is essential for education, engineering, and architecture



# Problem with existing LLMs

GPT4-V

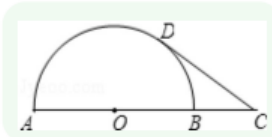


Describe the figure in detail.



The image shows a geometric figure consisting of a semicircle and a right-angled triangle. The semicircle has its diameter on the horizontal line segment AC, with its center marked as point O. Point D is at the topmost part of the semicircle, indicating the midpoint of the arc. The right-angled triangle is formed by the diameter AC of the semicircle and a vertical line segment BD dropping from point D to point B on the diameter AC. The triangle's hypotenuse is the radius of the semicircle, and the right angle is at point B.

LLaVA1.5

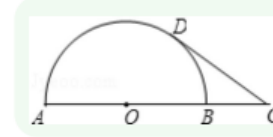


Describe the figure in detail.



The image features a white circle with a line drawn through its center. The line is labeled with the letter "d" and extends from the center of the circle to the edge. The circle is divided into four equal parts, with each part labeled with the letters "a," "b," "c," and "d." The circle is also labeled with the number "0" in the top left corner. The overall appearance of the image suggests a mathematical or geometric concept.

MiniGPT-4



Describe the figure in detail.



The image shows a right triangle with a hypotenuse of length 10 and legs of length 5 and 12. The triangle is labeled with the letters a, b, and c, which represent the vertices of the triangle. The hypotenuse is labeled with the number 10, and the legs are labeled with the numbers 5 and 12. The triangle is drawn in black ink on a white background.

# Objectives

- Fine-tune a Multi-modal Large Language Model (MLLM) to solve geometric problems.
- Enhance accuracy in geometric problem-solving.
- Reduce hallucinations and irrelevant results.
- Provide precise, step-by-step solutions.

# Scope of Project

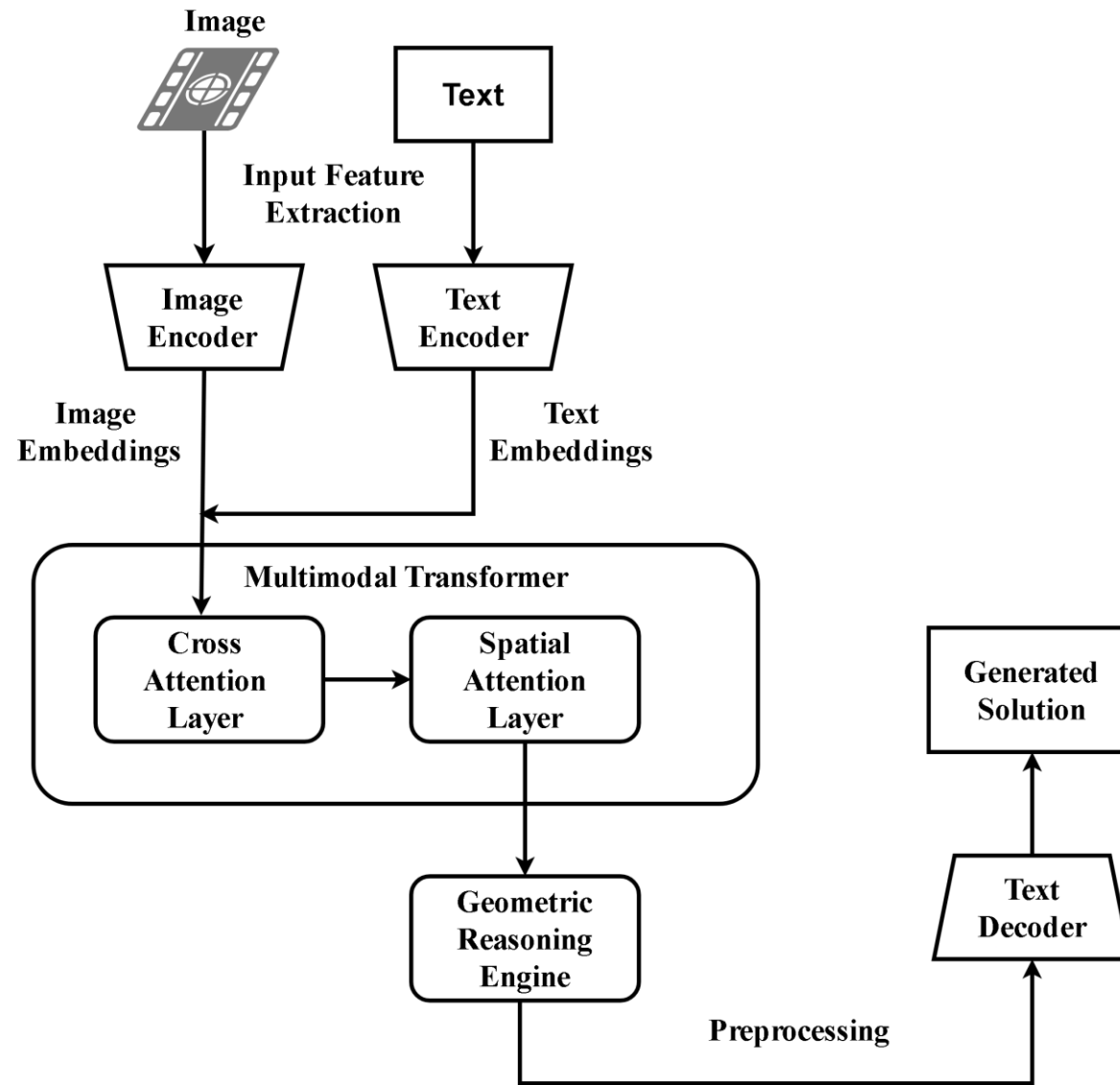
## **Project Capabilities:**

- Interpret and solve geometry problems from text and images.
- Providing detailed solutions in text and mathematical expressions.
- Minimizing model hallucinations for accurate problem interpretation.

## **Project limitations:**

- Specialized in geometry, not addressing other math fields.
- Model effectiveness depends on quality and diversity of training datasets.
- Regular updates and maintenance needed to incorporate new geometric theories and techniques.

# Proposed Methodology-[1] (System Architecture)



# Proposed Methodology-[2] (System Architecture)

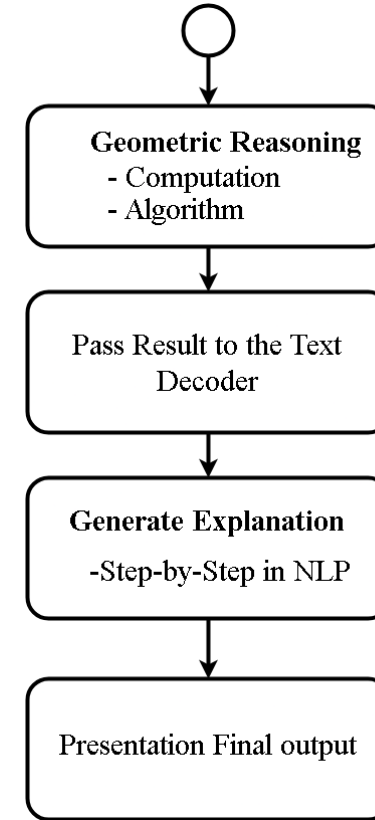
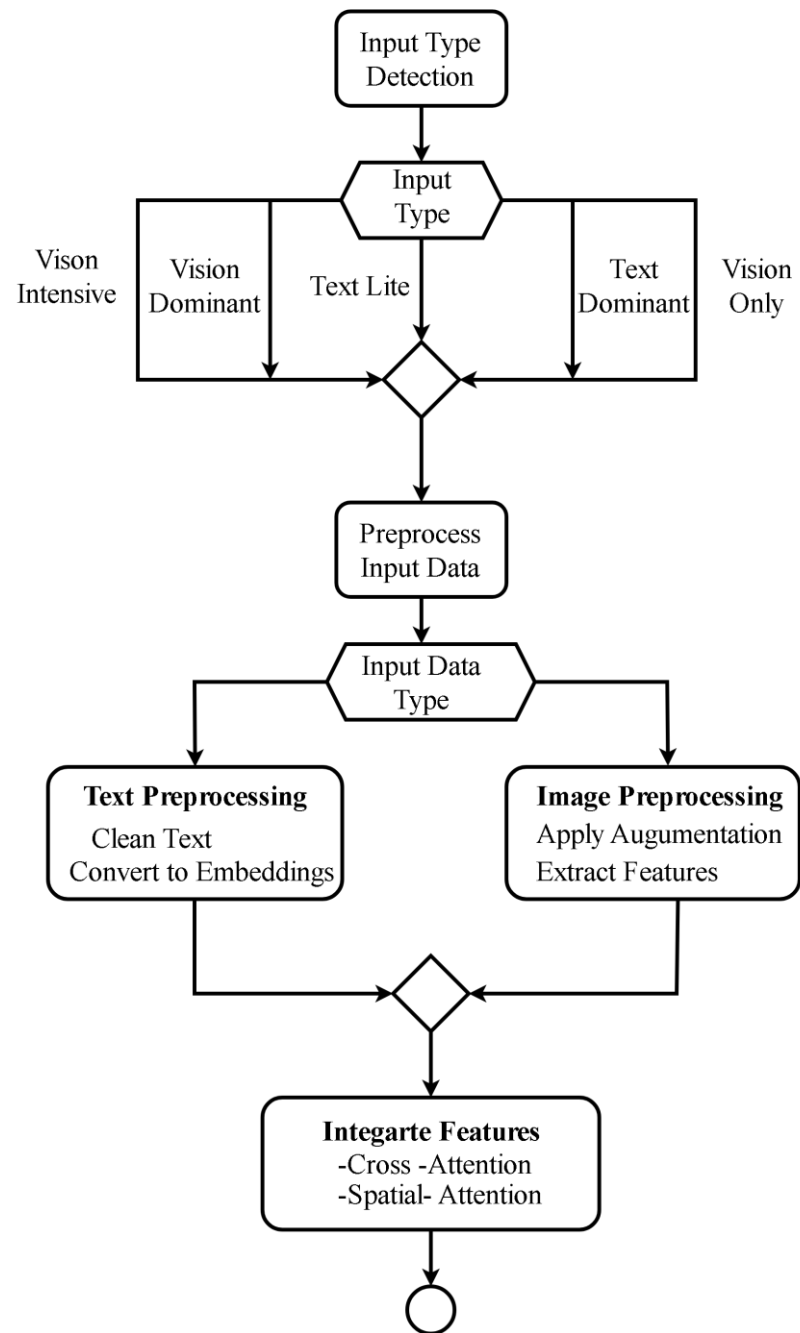
- Text Encoder:
  - Converts text into dense vector representation using models like BERT or GPT-3.
  - Captures semantic meaning of textual problem descriptions.
- Image Encoder:
  - Uses CNNs like ResNet to extract visual features from geometric diagrams.
  - Transforms visual data into compact representations.



# Proposed Methodology-[3] (System Architecture)

- Multimodal Transformer:
  - Cross-Attention Layer: Aligns and integrates textual data and visual information.
  - Spatial Attention Layer: Focuses on specific image regions relevant to the problem.
- Geometric Reasoning Engine:
  - Applies geometric algorithms using libraries like SymPy for calculation.
- Text decoder:
  - Generates coherent, step-by-step solutions in natural languages.

# Proposed Methodology-[4] (Workflow)



# Proposed Methodology-[5] (Input Types)

- Vision Intensive:
  - Primarily visual data with minimal text.
- Vision Dominant:
  - Significant visual data with moderate text.
- Text Lite:
  - Primarily text with some or no visual data.
- Text Dominant:
  - Mainly textual data with minimal visual input.
- Vision Only:
  - Solely visual data without text.

# Proposed Methodology-[6] (Workflow)

- Data Preprocessing:
  - Text Data: Tokenize, clean, and convert to embeddings.
  - Image Data: Apply augmentation and extract features.
- Feature Extraction:
  - Text Encoder: Generates semantic embeddings.
  - Image Encoder: Extracts visual features.
- Multimodal Integration:
  - Cross-Attention: Aligns text and images features.
  - Spatial Attention: Focuses on critical image regions.

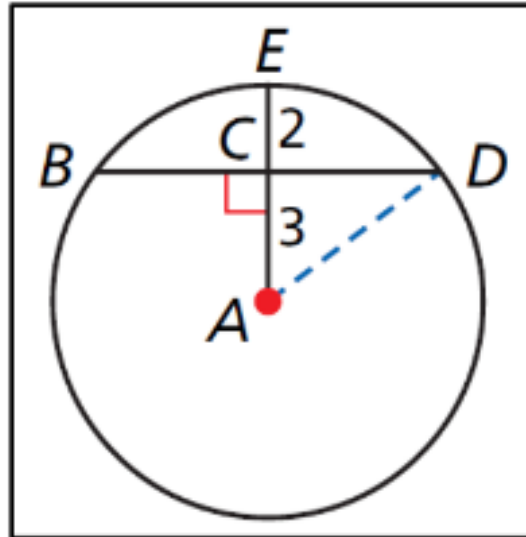
# Proposed Methodology-[7] (Workflow)

- Geometric Reasoning:
  - Utilizes symbolic mathematics for calculations.
- Solution Generation:
  - Text decoder generates detailed, step-by-step solutions.
- Output:
  - Comprehensive explanation of the solution in a user-friendly format.

# Dataset Exploration-[1]

Attribute	Details
Dataset	CASIA-PGPS9K
Quantity	9,000 plane geometry problems
Structure	<ul style="list-style-type: none"><li>• <b>Diagram:</b> An image depicting the geometric problem.</li><li>• <b>Structural Clauses:</b> List of structural lines and relationships in the diagram.</li><li>• <b>Semantic Clauses:</b> List of numerical or relational data relevant to the problem (e.g., lengths, angles).</li><li>• <b>Textual Problem:</b> The problem statement or question based on the diagram and clauses</li></ul>
Origin	Compiled by the Chinese Academy of Sciences Institute of Automation (CASIA)

# Dataset Exploration-[2]



## Structural Clauses

- line B C D
- line A D
- line E C A
- $\odot A$  lies on E B D

## Semantic Clauses

- $BD \perp EA$  on C
- $EC = 2(N0)$
- $CA = 3(N1)$

**Textual Problem:** Find BD.

**Answer:** 8.00

**Solution Program:**

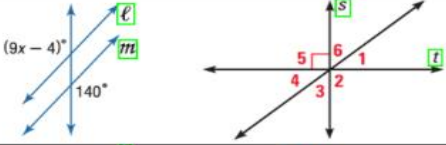
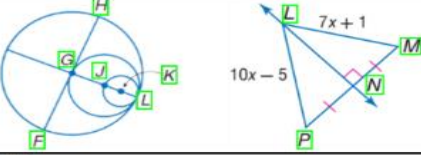
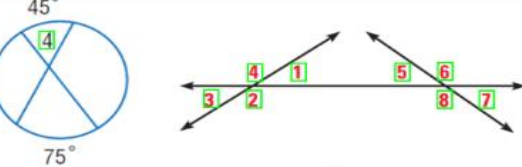
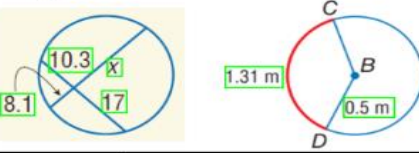
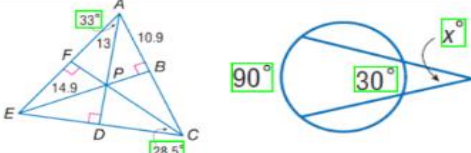
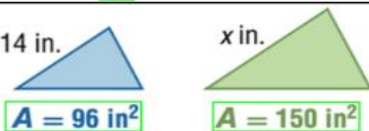
Sum N0 N1 V0 Gougu N1 V1 V0 Multiple V1 C2 V2 Get V2

# Dataset Exploration-[3]

Attribute	Details
Dataset	CASIA-PGDP5K
Quantity	5,000 complex plane geometry problems
Structure	<ul style="list-style-type: none"><li>• <b>Diagram:</b> Diagram: An image depicting the geometric problem.</li><li>• <b>Structural Clauses:</b> List of structural lines and relationships in the diagram.</li><li>• <b>Semantic Clauses:</b> List of numerical or relational data relevant to the problem (e.g., lengths, angles)</li><li>• <b>Textual Problem:</b> The problem statement or question based on the diagram and clauses</li></ul>
Origin	Compiled by the Chinese Academy of Sciences Institute of Automation (CASIA)



# Dataset Exploration-[4]

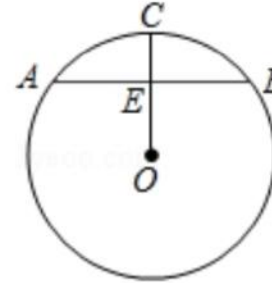
line	
point	
angle	
len	
degree	
area	

# Dataset Exploration-[5]

Attribute	Details
Dataset	Geo QA
Quantity	1,000 geometry problems
Structure	<ul style="list-style-type: none"><li>• <b>Image:</b> An image depicting the geometric problem.</li><li>• <b>Problem Type:</b> The type of problem (e.g., Length Calculation)</li><li>• <b>Knowledge Points:</b> Key geometric concepts and theorems used to solve the problem (e.g., Vertical Diameter, Pythagorean Theorem).</li><li>• <b>Problem Solving Explanations:</b> Detailed step-by-step solution to the problem.</li></ul>
Origin	Available on Papers with Code, compiled by Luckyjhg

# Dataset Exploration-[6]

As shown in the figure, in  $\odot O$ ,  $AB$  is the chord,  $OC \perp AB$ , if the radius of  $\odot O$  is 5 (N0) and  $CE=2$  (N1), then the length of  $AB$  is ()



A. 2   B. 4   C. 6   D. 8

**Answer:** D. 8

**Problem Type:** Length Calculation

**Knowledge Points:** Vertical Diameter, Pythagorean Theorem

**Problem Solving Explanations:**

$OE = OC - CE = 5 - 2 = 3$ . According to the Pythagorean Theorem,

$AE = \sqrt{OA^2 - OE^2} = \sqrt{5^2 - 3^2} = 4$ . Thus,  $AB = 2AE = 8$ .

**Annotated Programs:**

Minus | N0 | N1 | PythagoreanMinus | N0 | V0 | Double | V1

Step1: Minus(N0, N1) =  $5 - 2 = 3$  (V0)

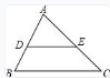
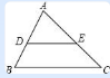
Step2: PythagoreanMinus(N0, V0) =  $\sqrt{5^2 - 3^2} = 4$  (V1)

Step3: Double(V1) =  $2 \times 4 = 8$  (V2)

# Dataset Exploration-[7]

Attribute	Details
Dataset	MathVerse
Quantity	1,000 geometry problems
Structure	<ul style="list-style-type: none"><li>• <b>Sample Index:</b> Unique identifier for each problem instance.</li><li>• <b>Problem Index:</b> Index of the problem within the dataset.</li><li>• <b>Problem Version:</b> Specifies the version of the problem, e.g., Text Dominant.</li><li>• <b>Problem Solving Explanations:</b> Detailed step-by-step solution to the problem.</li><li>• <b>Question:</b> Text description of the geometry problem.</li><li>• <b>Image:</b> Associated image illustrating the geometry problem.</li><li>• <b>Answer:</b> Correct answer to the problem.</li><li>• <b>Question Type:</b> Type of question, e.g., multi-choice.</li><li>• <b>Metadata:</b> Additional metadata including source, split, subfield, and subject.</li><li>• <b>Query WO:</b> Query without certain elements.</li><li>• <b>Query COT:</b> Query with chain of thought reasoning.</li><li>• <b>Question Form:</b> Format of the question.</li></ul>
Origin	Available on Hugging Face, contributed by AI4Math

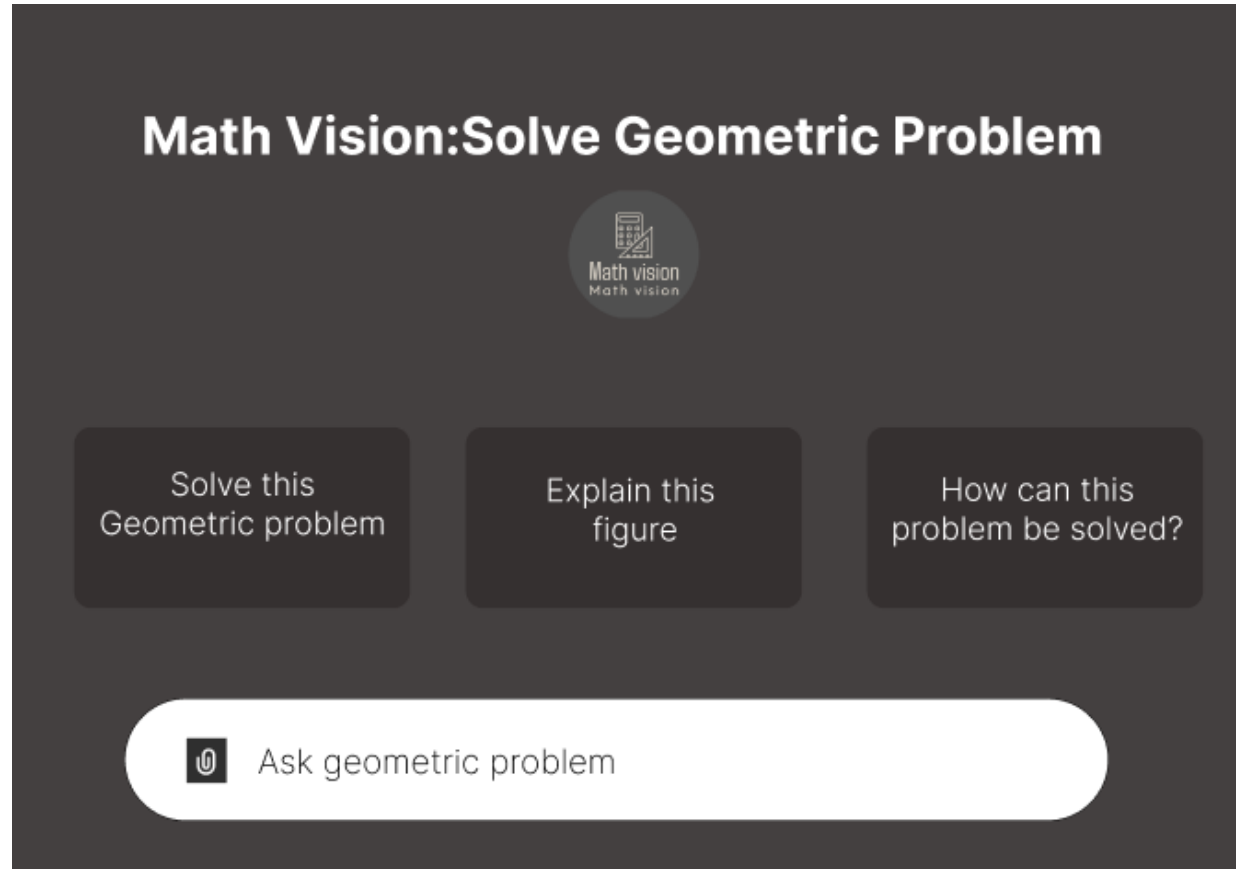
# Dataset Exploration-[8]

problem_index string · classes	problem_version string · classes	question string · lengths	image image · width (px)	answer string · classes	question_type string · classes	metadata dict	query_wo string · lengths	query_cot string · lengths	question_for_eval string · lengths
1 0.1%	Text Lite 20%	132*264 28%	63*741 76.8%	D 8.8%	multi-choi... 55.3%		233*365 27.4%	270*402 27.4%	138*269 29.2%
1	Text Dominant	As shown in the figure, in...		D	multi-choice	{ "source": "GeoQA", ...	Please directly answer the...	Please first conduct reasoning...	As shown in the figure, in trianl...
1	Text Lite	As shown in the figure, it is known that angle A = 80.0, angle B = 60.0, DE parallel BC, then the size of angle CED is () Choices: A:40° B:60° C:120° D:140°		D	multi-choice	{ "source": "GeoQA", "split": "testmini", "subfield": "Angle", "subject": "Plane Geometry" }	Please directly answer the question and provide the correct option letter, e.g., A, B, C, D. Question: As shown in the figure, it is known that angle A = 80.0, angle B = 60.0, DE parallel BC, then the size of angle CED is () Choices: A:40° B:60° C:120° D:140°	Please first conduct reasoning, and then answer the question and provide the correct option letter, e.g., A, B, C, D, at the end. Question: As shown in the figure, it is known that angle A = 80.0, angle B = 60.0, DE parallel BC, then the size of angle CED is () Choices: A:40° B:60° C:120° D:140°	As shown in the figure, it is known that angle A = 80.0, angle B = 60.0, DE parallel BC, then the size of angle CED is () Choices: A:40° B:60° C:120° D:140°

# Instrument Requirements

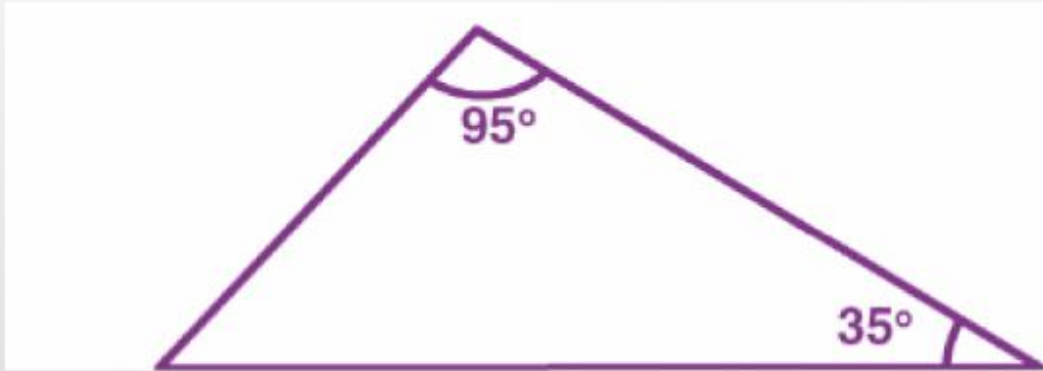
- Software Requirements
  - Operating System: 64 bit Linux-based (Ubuntu 20.04 LTS)
  - Programming language: Python
  - Libraries: Deep Learning Frameworks, Geometric Computation, NLP tools, Image Processing
  - IDE: Jupyter Notebook and VS Code
  - Data and Model management: Git, HDF5, TorchServe
- Hardware Requirements
  - Multicore CPU(Intel Xeon or AMD Ryzen)
  - High Performance GPU(NVIDIA's Tesla V100 or A100)

# Expected Results-[1] (Landing Page)



# Expected Results-[2] (Question-Answer)

Case 1st : only image is provided



The image shows a triangle with the following angles labeled:

- One angle is  $95^\circ$ .
- Another angle is  $35^\circ$ .

To find the third angle of the triangle, we can use the fact that the sum of the angles in any triangle is  $180^\circ$ .

Let's calculate the third angle.

$$\text{Third angle} = 180^\circ - 95^\circ - 35^\circ$$

$$\text{Third angle} = 180^\circ - 130^\circ$$

$$\text{Third angle} = 50^\circ$$

So, the third angle of the triangle is  $50^\circ$ .



Ask geometric problem



# Expected Results-[3]

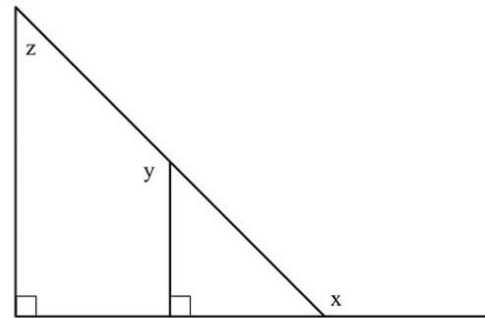
## (Question-Answer)

### Case 2nd: Image with description

#### Problem

In the figure, if  $z = 50^\circ$ , then  $x + y =$

- (A)  $230^\circ$
- (B)  $250^\circ$
- (C)  $260^\circ$
- (D)  $270^\circ$
- (E)  $290^\circ$



The angle given is  $z = 50^\circ$

Considering the big triangle,  $z + 90^\circ + \text{other angle} = 180^\circ$

other angle  $= 40^\circ$

The angle of the straight line is  $180^\circ$

Lets calculate  $x$ ,  $x + 40^\circ = 180^\circ$  we get  $x = 140^\circ$

Considering the small triangle,

$40^\circ + 90^\circ + \text{remaining angle} = 180$

remaining angle  $= 50^\circ$

Calculating the value of  $y = 180^\circ - 50^\circ = 130^\circ$ . since the angle of straight line is  $180^\circ$

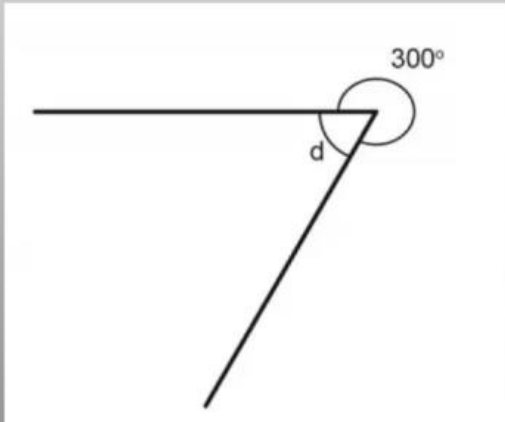
The required sum of  $x + y = 140^\circ + 130^\circ = 270^\circ$



Ask geometric problem

# Expected Results-[4] (Question-Answer)

Case 3rd: Image and description given  
Prompt given: Find the value of d



In the given figure, we need to find the value of angle d.

The figure shows an angle of  $300^\circ$ . This angle is measured in the exterior direction from one line segment to another.

The sum of the angles around a point is  $360^\circ$ . Since  $300^\circ$  is the exterior angle, the interior angle d is:

$$d = 360^\circ - 300^\circ$$

Calculating this:

$$d = 60^\circ$$

Therefore, the value of angle d is  $60^\circ$ .



Ask geometric problem

# Expected Results-[5] (Question-Answer)

## Case 4th : Prompt only

Prompt given: If one side of the triangle is 40 degree and the other one is 90 , then what is value of remaining angle?

In any triangle, the sum of the interior angles is always  $180^\circ$ .

Given:

- One angle is  $40^\circ$ .
- Another angle is  $90^\circ$ .

We need to find the remaining angle .

Using the sum of angles property:

$$40^\circ + 90^\circ + \text{remaining angle} = 180^\circ.$$

Solving for remaining angle:

$$130^\circ + \text{remaining angle} = 180^\circ.$$

$$\text{remaining angle} = 50^\circ$$

Therefore, the value of the remaining angle is  $50^\circ$ .

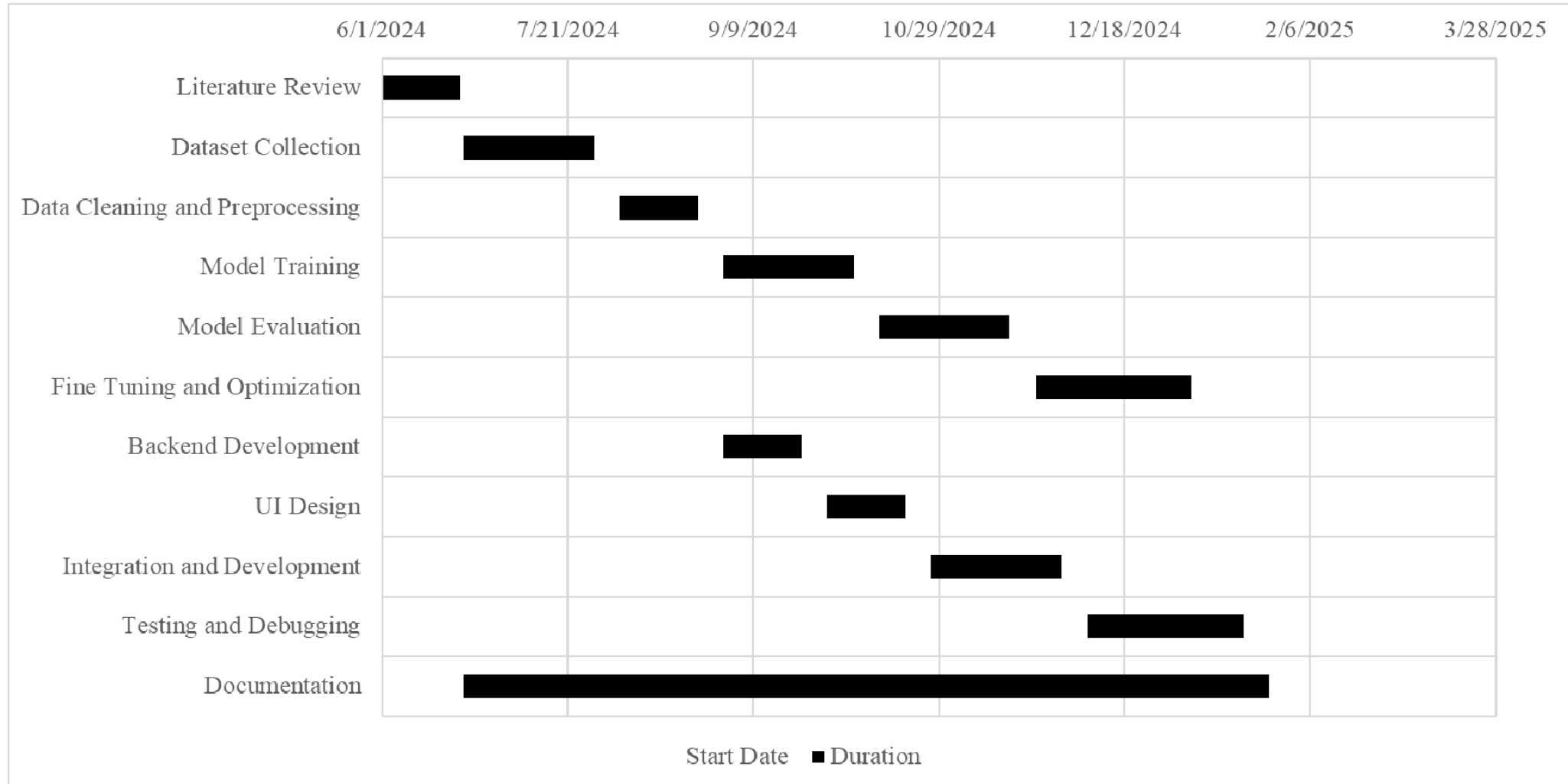


Ask geometric problem

# Project Application

- Education:
  - Intelligent tutoring systems, personalized learning.
- Engineering and Architecture:
  - Automatic geometric modeling, error detection
- Scientific Research:
  - Automatic geometric data analysis
- Robotics and Computer Vision:
  - Enhanced path planning, object recognition.

# Expected Project Schedule



# Expected Project Budget

<b>TASK</b>	<b>EXPECTED PRICE(NRS)</b>
Cloud Services	7000.00
Computing Resources - Google Colab Premium (170 hrs. approx.)	2040.00
Printing	2500.00
Miscellaneous	1500.00
<b>Total</b>	<b>13040.00</b>

# References-[1]

[1] J. Chen, J. Tang, J. Qin, et al., “Geoqa: A geometric question answering benchmark towards multimodal numerical reasoning,” in Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1- 6, 2021, C. Zong, F. Xia, W. Li, and R. Navigli, Eds., ser. Findings of ACL, ACL/IJCNLP 2021, Association for Computational Linguistics, 2021, 513–523. DOI: 10.18653/v1/2021.findings-acl.46.

[2] Y. Z. H. L. Z. G. P. Q. A. Z. P. L. K.-W. C. P. G.H. L. Renrui Zhang, Dongzhi Jiang, “Mathverse: Does your multi-modal llm truly see the diagrams in visual math problems?” in arXiv, 2024

# References-[2]

- [3] M.-L. Zhang, F. Yin, Y.-H. Hao, and C.-L. Liu, “Plane geometry diagram parsing,” arXiv preprint arXiv:2205.09363, 2022.
- [4] Y. Hao, M. Zhang, F. Yin, and L.-L. Huang, “Pgdp5k: A diagram parsing dataset for plane geometry problems,” in 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022, pp. 1763–1769.
- [5] M.-L. Zhang, F. Yin, and C.-L. Liu, “A multi-modal neural geometric solver with textual clauses parsed from diagram,” arXiv preprint arXiv:2302.11097, 2023.



**THANK YOU**