

MaAs On Benchmark

yang lu



**Northeastern
University**

Test Overview

- Test Dataset: theoremqa
- Number of Test Samples: 50 questions
- model: MaAS
- Success Rate: 100%
- Test Environment: DeepSeek–reasoner model

Search Failure Possibilities

- Although all current tests were successful, search may fail in the following situations:
- Complex problems requiring dynamic workflow switching
- Real-time computation problems requiring external tool calls
- Understanding problems with multimodal inputs

Analysis of Most Complex Successful Cases

TheoremQAParser → TheoremQAKnowledgeRetriever → TheoremQAReasoner

↓
Identified as
calculation
problem

↓
Retrieved complex
analysis concepts
(empty)

↓
Applied residue
theorem

↓
Laurent series
expansion

↓
Essential singularity
handling

↓
Answer: $-\frac{\pi}{3}$

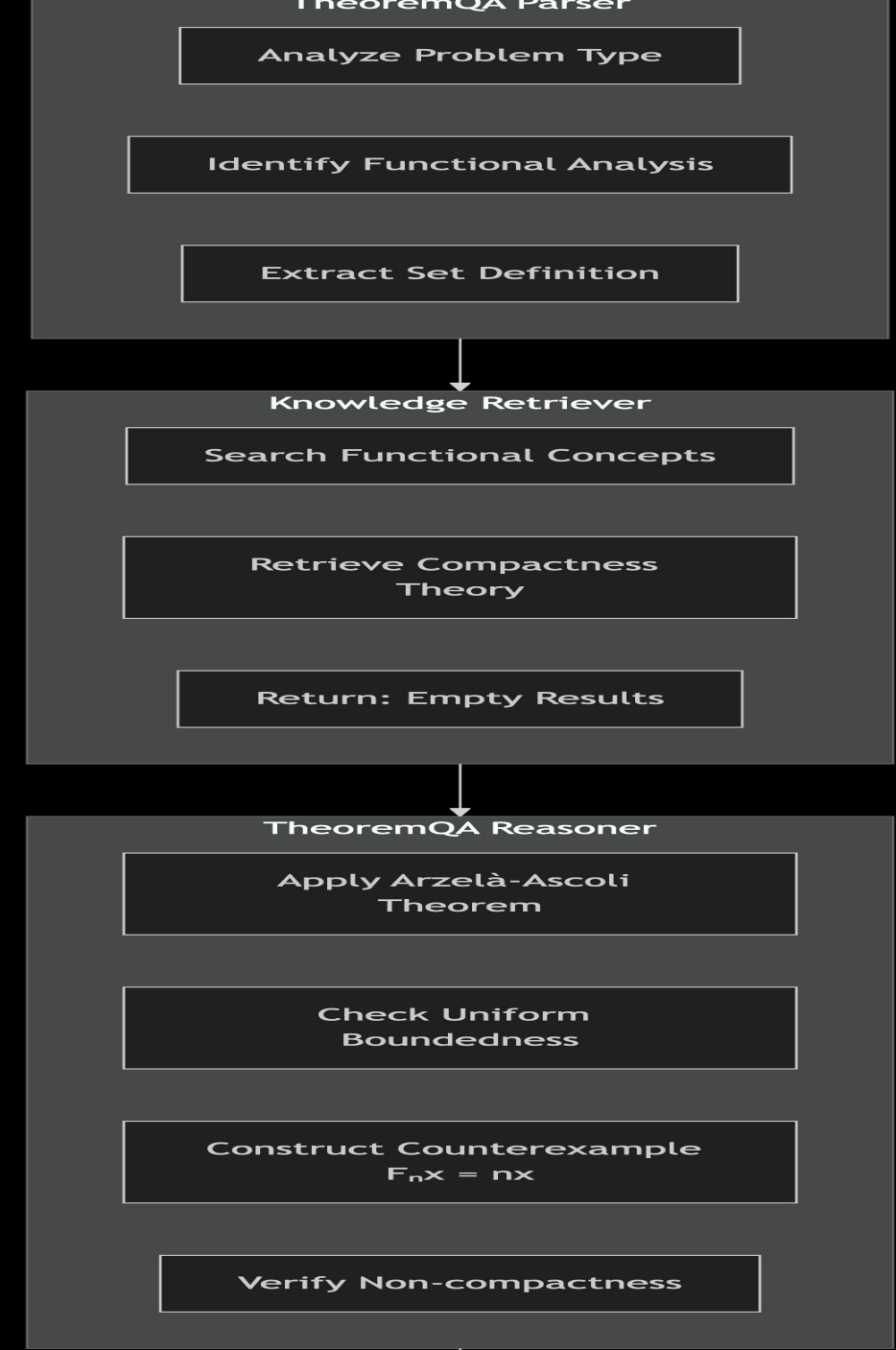
- Problem:

Compute $\int_{|z|=1} z^2 \sin(1/z) dz$.

The answer is Ai
with i denoting the
imaginary unit, what
is A ?

Analysis of Most Complex Successful Cases

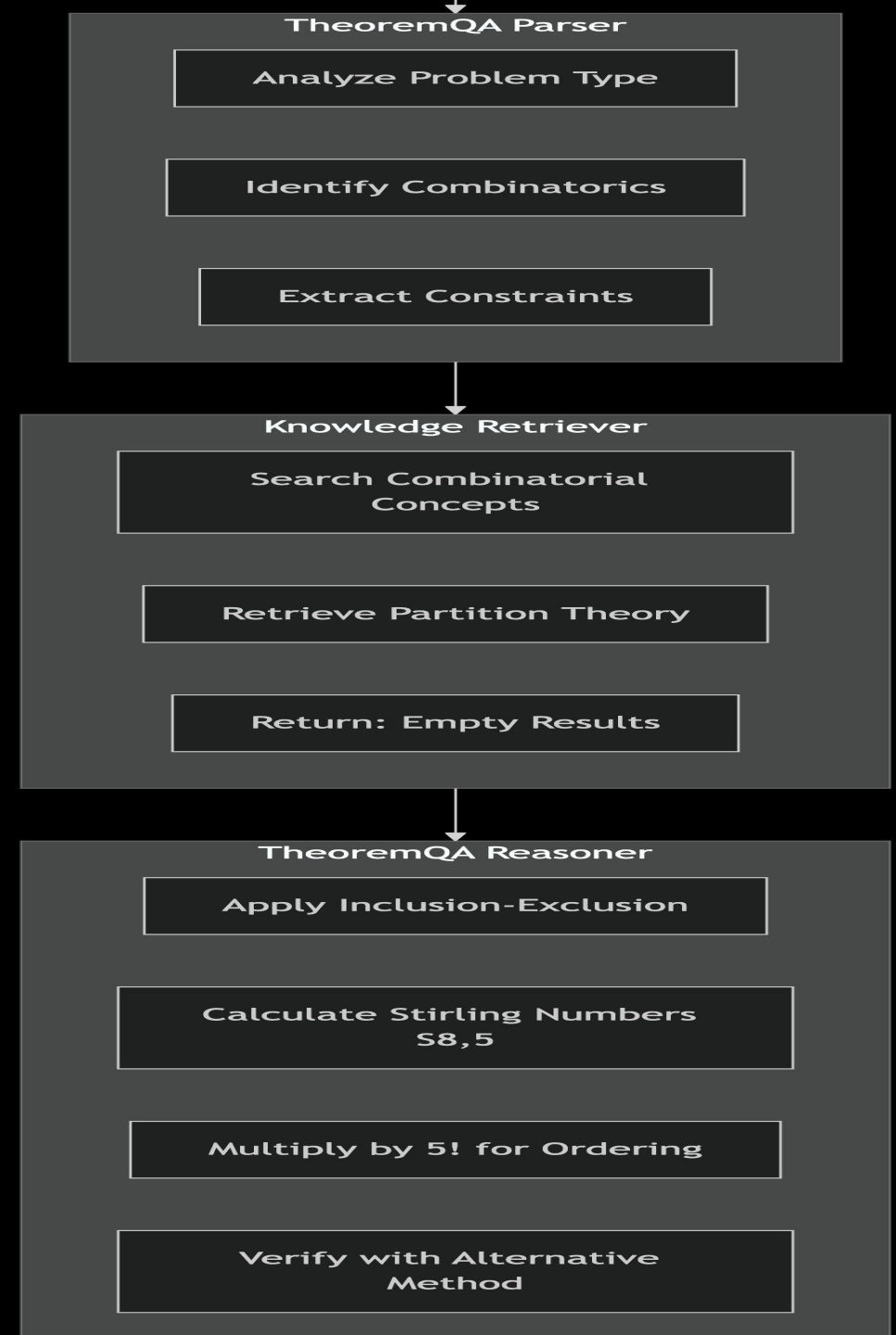
- Problem: Determine if the set $\{F(x) = \int_0^x f(t) dt \mid f \in M\}$ is sequentially compact



Analysis of Most Complex Successful Cases

- Problem:

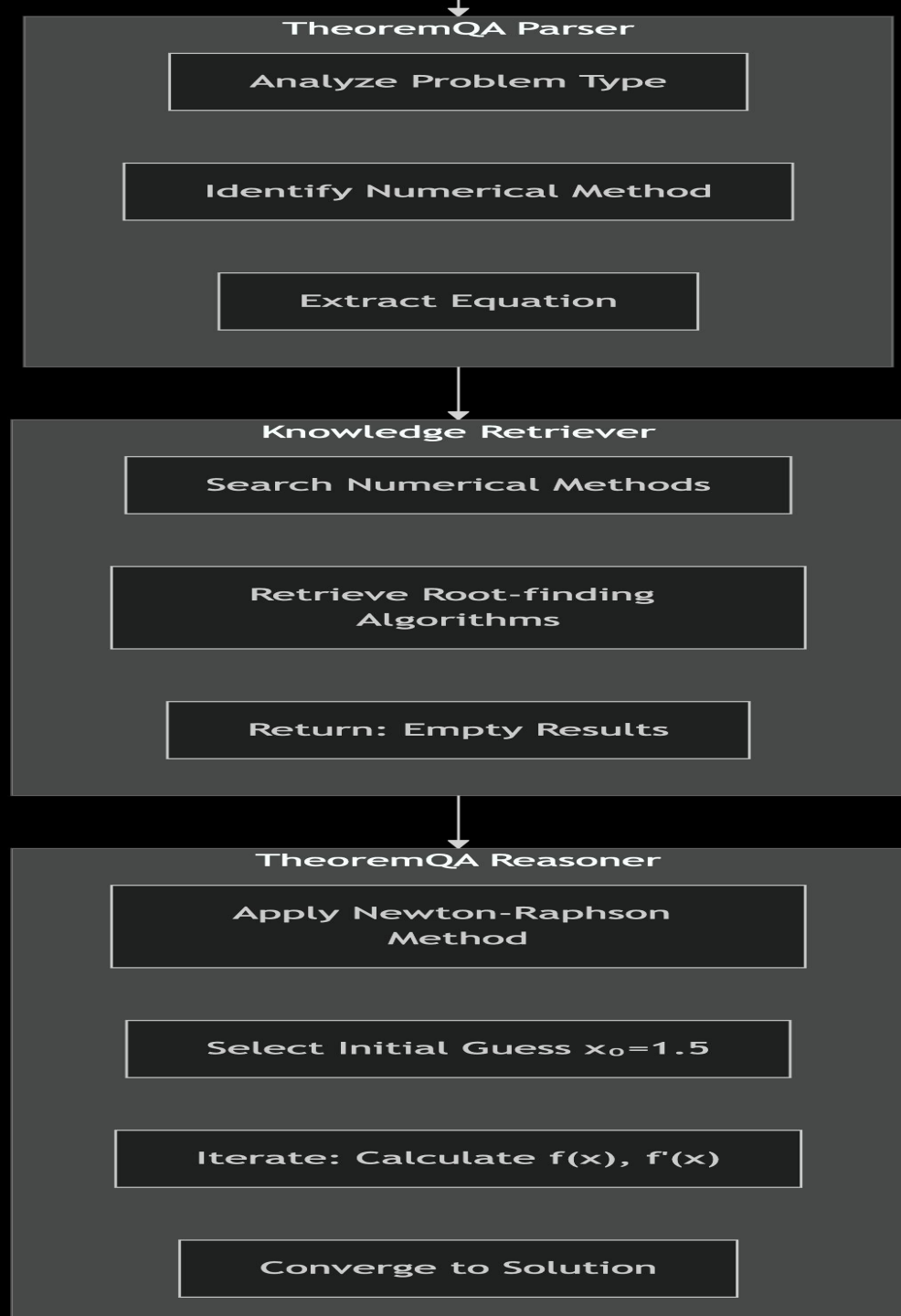
How many ways to divide 8 elements into 5 non-empty ordered subsets?



Analysis of Most Complex Successful Cases

- Problem:

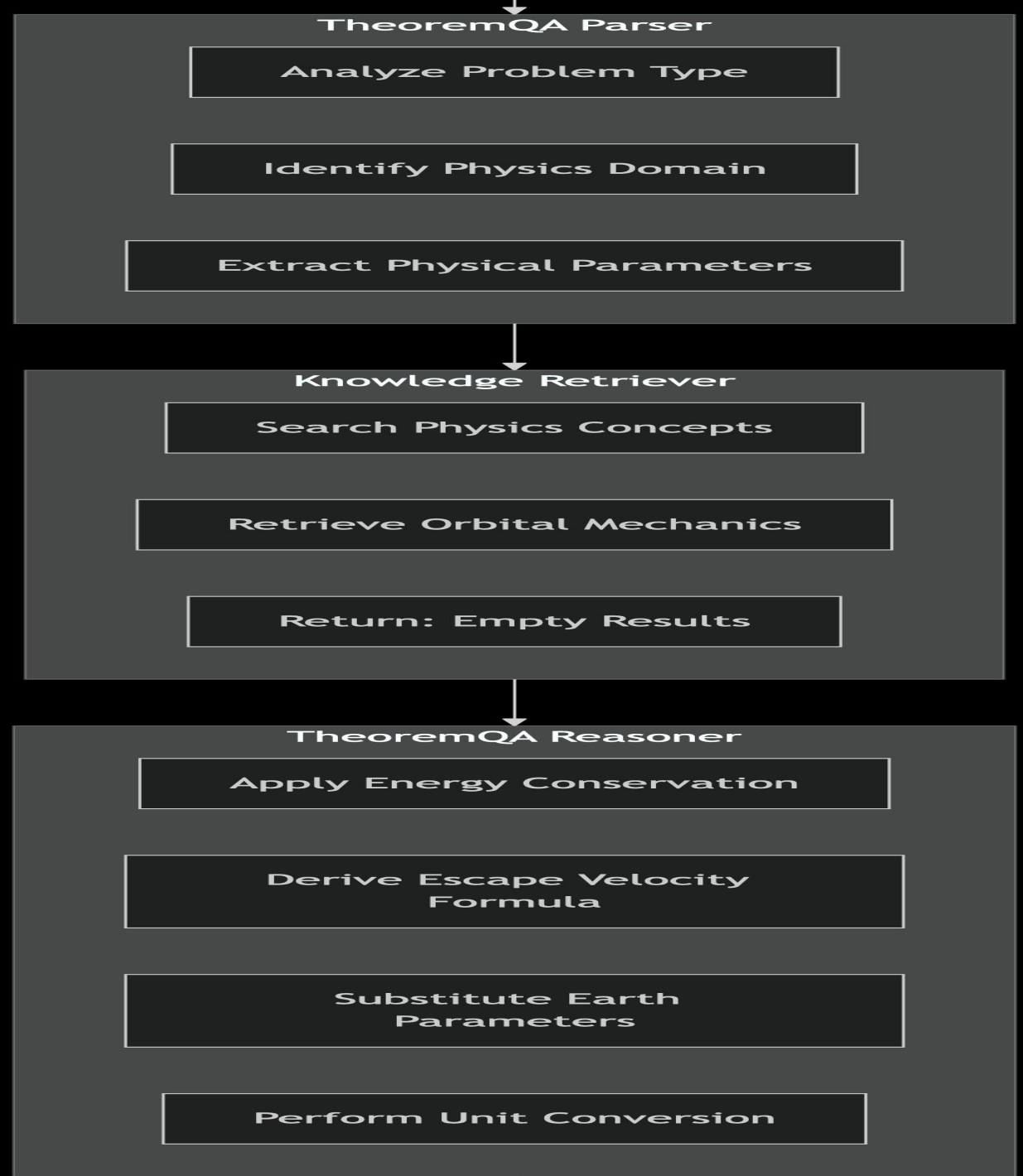
Solve $2x^3 + e^x = 10$
using Newton-Raphson
method



Analysis of Most Complex Successful Cases

Problem:

Calculate Earth's escape velocity



Terminal Results

EXPLORER

MAAS

maas

conrags

data

ext

prompts

provider

rag

strategy

theoremqa

__pycache__

__init__.py

theoremqa_coordinator.py

theoremqa_knowledge_agent.py

theoremqa_knowledge.py

theoremqa_parser.py

theoremqa_reasoner.py

tools

utils

__init__.py

_compat.py

config2.py

const.py

context_mixin.py

context.py

document.py

llm.py

logs.py

repo_parser.py

schema.py

test_basic.py

theoremqa_controller.py

my_maas_env

workspace

quick_test.py

README.md

requirements_clean.txt

requirements.txt

run_theoremqa.py

setup.py

test_basic.py

OUTLINE

TIMELINE

陌生单词

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

PORTS

Answer: No

Reasoning: To determine if W is a subspace of \mathbb{R}^2 , we must verify the three subspace criteria:

1. The zero vector of \mathbb{R}^2 , $[0, 0]$, must be in W . Since $x = 0 \geq 0$ and $y = 0 \geq 0$, $[0, 0] \in W$.

2. W must be closed under vector addition. For any $u = [x_1, y_1]$ and $v = [x_2, y_2]$ in W , where $x_1, y_1, x_2, y_2 \geq 0$, their sum $u + v = [x_1 + x_2, y_1 + y_2]$ has $x_1 + x_2 \geq 0$ and $y_1 + y_2 \geq 0$, so $u + v \in W$.

3. W must be closed under scalar multiplication. For any $u = [x, y]$ in W (with $x, y \geq 0$) and any scalar $c \in \mathbb{R}$, $c * u = [c * x, c * y]$. However, if c is negative (e.g., $c = -1$), then $c * u = [-x, -y]$, where $-x \leq 0$ and $-y \leq 0$, so $c * u \notin W$. For example, $u = [1, 1] \in W$, but $c * u = [-1, -1] \notin W$.

Since W fails closure under scalar multiplication, it is not a subspace of \mathbb{R}^2 .

2025-11-30 17:25:14.753 | WARNING | maas.utils.cost_manager:update_cost:49 - Model deepseek-reasoner not found in TOKEN_COSTS.

Reasoning completed: No (method: llm_based)

Success - Answer: No

14. Question: compute the line integral of $\int_K xy \, dx$, $\int_L xy \, dx$, where K is a straight line from $(0,0)$ to $(1,1)$ and L is the Parabola $y=x^2$ from $(0,0)$ to $(1,1)$. return the answer as a list

Starting processing: compute the line integral of $\int_K xy \, dx$, $\int_L xy \, dx$

Executing problem parsing...

[TheoremQAParser] 解析问题: compute the line integral of $\int_K xy \, dx$, $\int_L xy \, dx$, where K is a straight line from $(0,0)$ to $(1,1)$ and L is the Parabola $y=x^2$ from $(0,0)$ to $(1,1)$. return the answer as a list

识别为: calculation, 概念: []

Selected workflow: ['parser', 'knowledge_retriever', 'reasoner']

Executing knowledge retrieval...

[TheoremQAKnowledgeRetriever] Retrieving question: compute the line integral of $\int_K xy \, dx$, $\int_L xy \, dx$

Retrieved knowledge for 0 concepts

Executing logical reasoning...

[TheoremQARreasoner] Reasoning question: compute the line integral of $\int_K xy \, dx$, $\int_L xy \, dx$

Answer: $[1/3, 1/4]$

Reasoning:

To compute the line integrals $\int_K xy \, dx$ and $\int_L xy \, dx$, where K is the straight line from $(0,0)$ to $(1,1)$ and L is the parabola $y = x^2$ from $(0,0)$ to $(1,1)$, we parameterize each path and evaluate the integral.

For path K (straight line):

- Parameterize with $x = t$, $y = t$, where t ranges from 0 to 1 .

- Then, $dx = dt$, and $xy = t * t = t^2$.

- The integral becomes $\int_0^1 t^2 \, dt$.

- Compute $\int_0^1 t^2 \, dt$ from 0 to 1 : $[t^3/3]_0^1 = (1/3) - 0 = 1/3$.

For path L (parabola $y = x^2$):

- Parameterize with $x = t$, $y = t^2$, where t ranges from 0 to 1 .

- Then, $dx = dt$, and $xy = t * t^2 = t^3$.

- The integral becomes $\int_0^1 t^3 \, dt$.

- Compute $\int_0^1 t^3 \, dt$ from 0 to 1 : $[t^4/4]_0^1 = (1/4) - 0 = 1/4$.

Thus, the line integrals are $1/3$ for K and $1/4$ for L , resulting in the list $[1/3, 1/4]$.

2025-11-30 17:26:04.438 | WARNING | maas.utils.cost_manager:update_cost:49 - Model deepseek-reasoner not found in TOKEN_COSTS.

Reasoning completed: $[1/3, 1/4]$ (method: llm_based)

Success - Answer: $[1/3, 1/4]$

15. Question: True of false: one can draw a simple connected planar graph with 200 vertices and 400 faces

Starting processing: True of false: one can draw a simple connected pla...

Executing problem parsing...

[TheoremQAParser] 解析问题: True of false: one can draw a simple connected planar graph with 200 vertices and 400 faces

识别为: boolean_verification, 概念: []

Selected workflow: ['parser', 'reasoner']

Executing logical reasoning...

[TheoremQARreasoner] Reasoning question: True of false: one can draw a simple connected pla...

Answer: false

Reasoning:

1. The problem involves determining whether a simple connected planar graph with 200 vertices and 400 faces can exist. A simple graph has no multiple edges or loops, and a planar graph can be drawn on a plane without edge crossings.

2. Euler's formula for connected planar graphs states that $V - E + F = 2$, where V is the number of vertices, E is the number of edges, and F is the number of faces (including the outer face).

3. Substituting the given values ($V = 200$, $F = 400$) into Euler's formula: $200 - E + 400 = 2 \rightarrow 600 - E = 2 \rightarrow E = 598$.

4. For a simple connected planar graph with $V \geq 3$, there is an inequality that must hold: $E \leq 3V - 6$. This inequality ensures that the graph is not too dense to be planar.

5. Calculate $3V - 6$ for $V = 200$: $3 * 200 - 6 = 600 - 6 = 594$.

6. Compare $E = 598$ with 594 : $598 > 594$, which violates the inequality $E \leq 3V - 6$.

7. Since the inequality is violated, no such simple connected planar graph can exist. Therefore, the statement is false.

2025-11-30 17:26:45.555 | WARNING | maas.utils.cost_manager:update_cost:49 - Model deepseek-reasoner not found in TOKEN_COSTS.