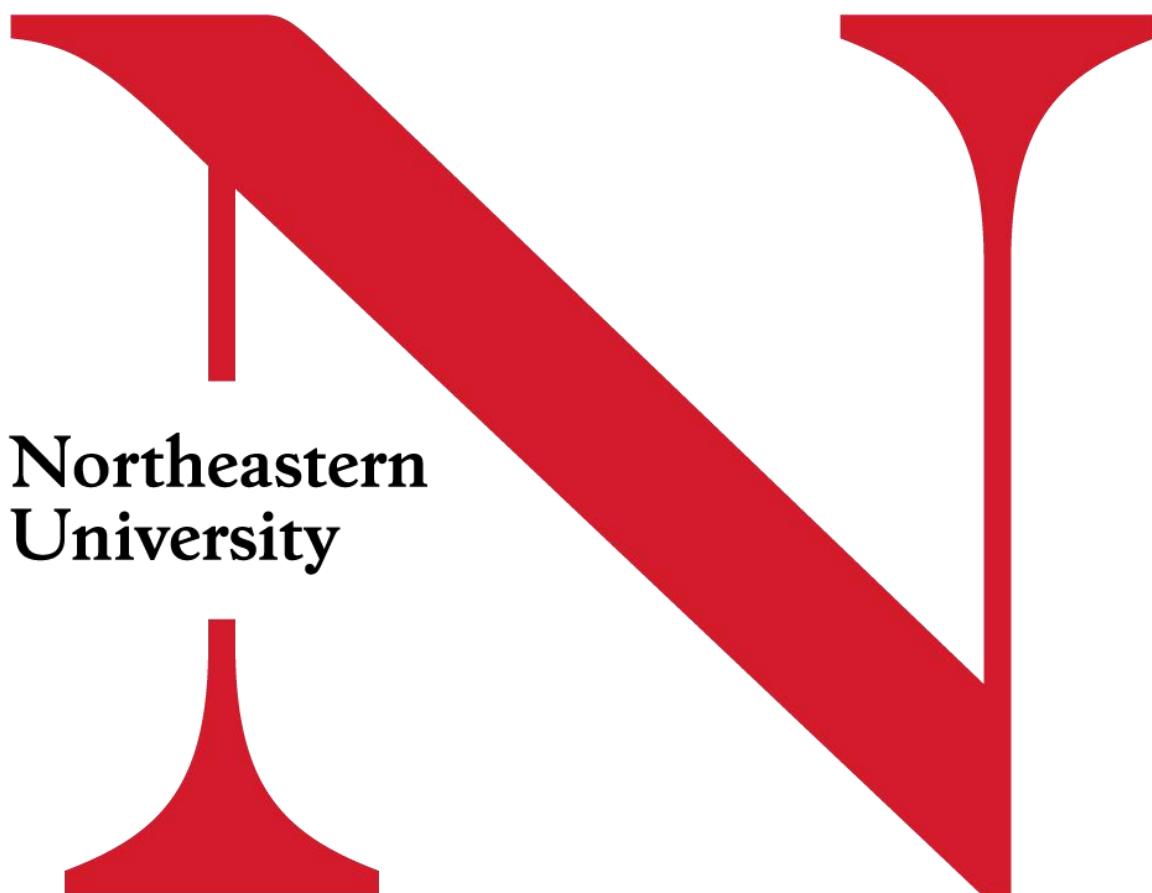


MaAs On Benchmark

yang lu

A large, semi-transparent watermark of the Northeastern University logo is positioned in the lower right quadrant. The logo consists of a red 'N' shape with a white center, set against a white background with red borders.

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 A_i
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

TheoremQA Parser

Analyze Problem Type

Identify Functional Analysis

Extract Set Definition

Knowledge Retriever

Search Functional Concepts

Retrieve Compactness Theory

Return: Empty Results

TheoremQA Reasoner

Apply Arzelà-Ascoli Theorem

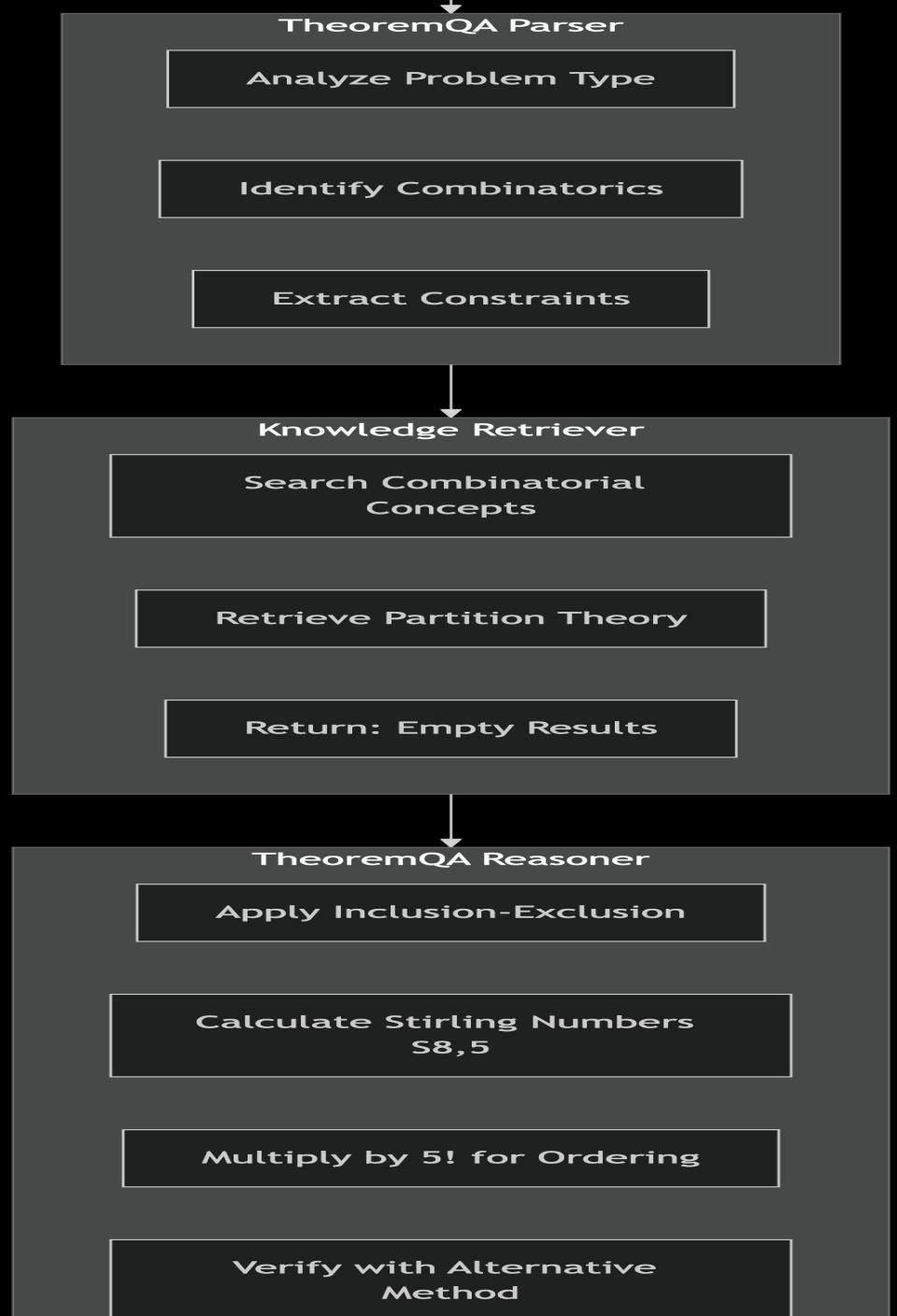
Check Uniform Boundedness

Construct Counterexample $F_n x = nx$

Verify Non-compactness

Analysis of Most Complex Successful Cases

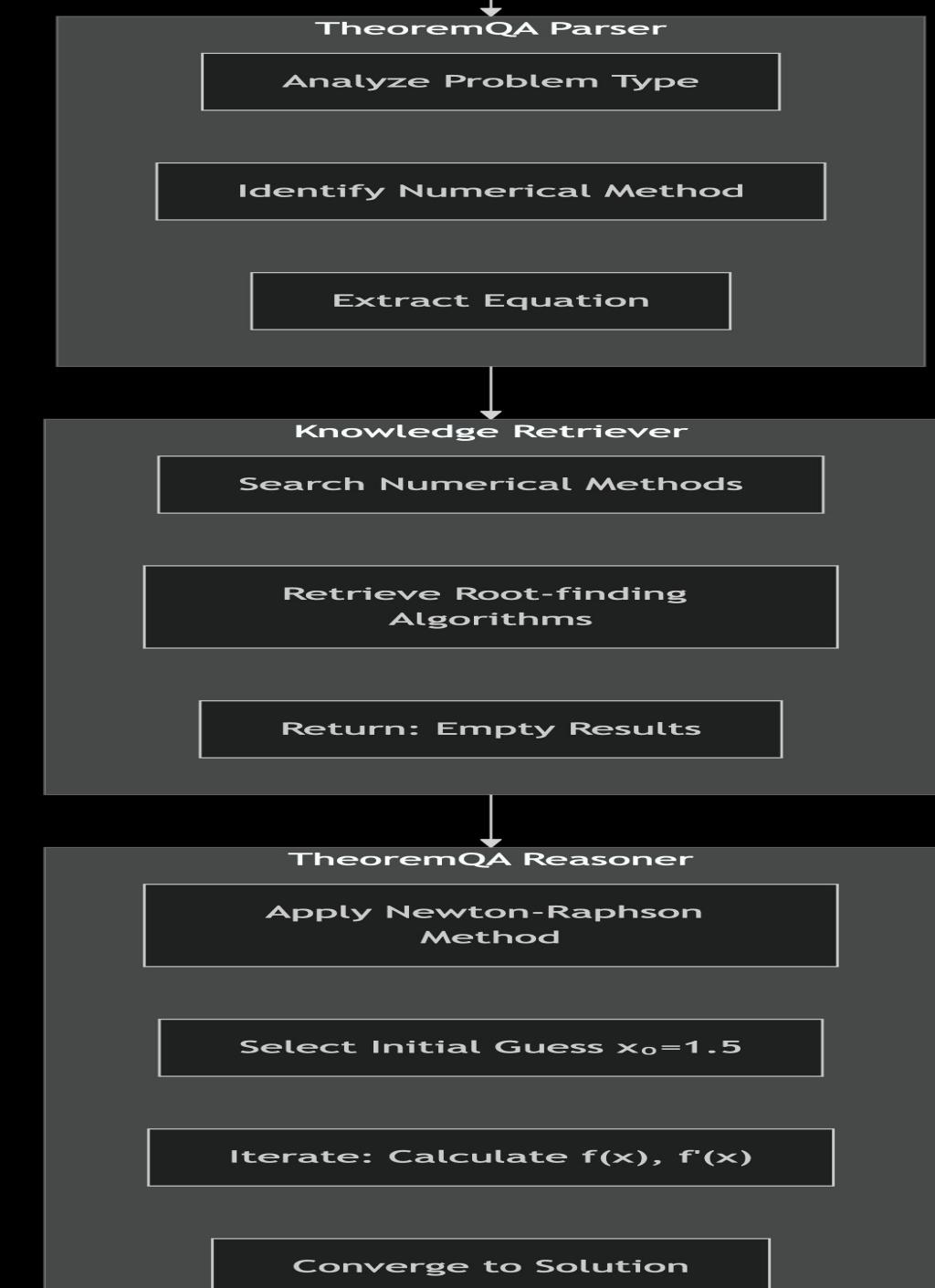
- Problem:
How many ways to divide 8 elements into 5 non-empty ordered subsets?



Analysis of Most Complex Successful Cases

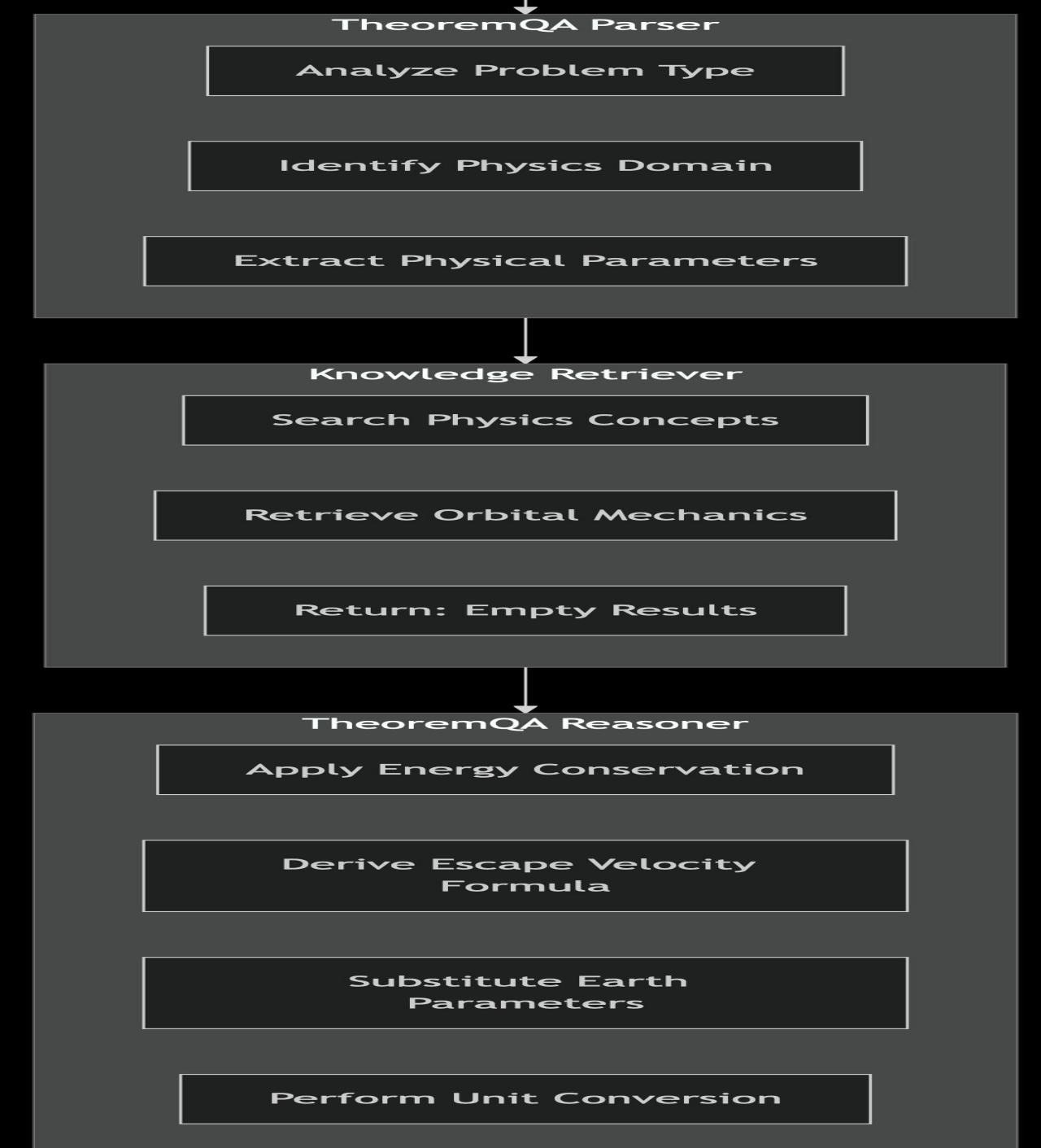
- Problem:

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



Analysis of Most Complex Successful Cases

Problem:
Calculate Earth's escape velocity



Terminal Results

The screenshot shows a VS Code interface with the following details:

- EXPLORER** sidebar on the left containing project files like `MAAS`, `maas`, `utils`, `tools`, etc.
- TERMINAL** tab selected at the top.
- Terminal Content:**
 - 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
- Bottom Status Bar:** Shows the date and time as 2025-11-30 17:26:45.555.
- Bottom Right:** A small icon representing a workspace or project.