MaAs On Benchmark

yang lu

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

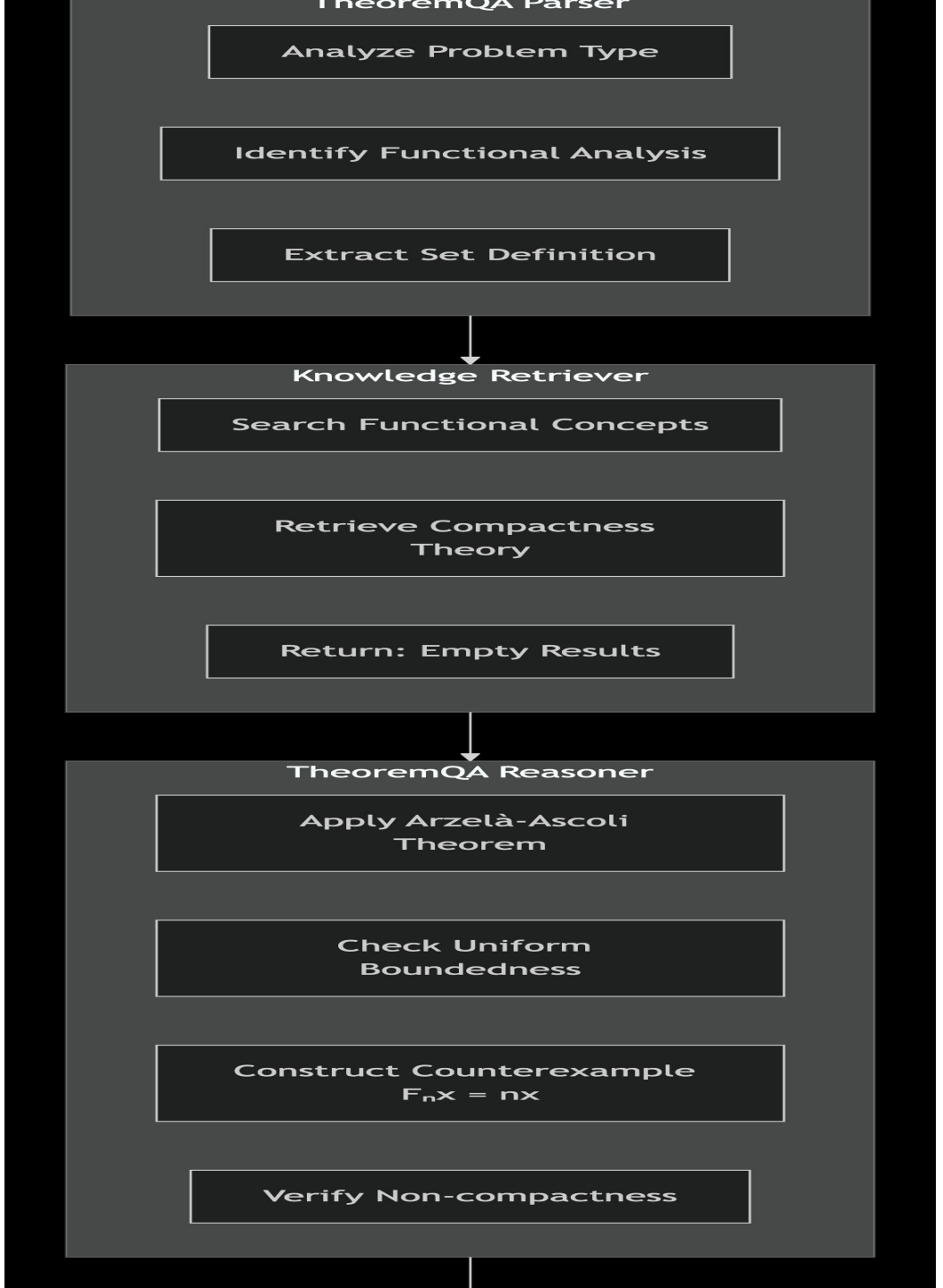
↓ Answer: $-\frac{\pi}{3}$

- 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

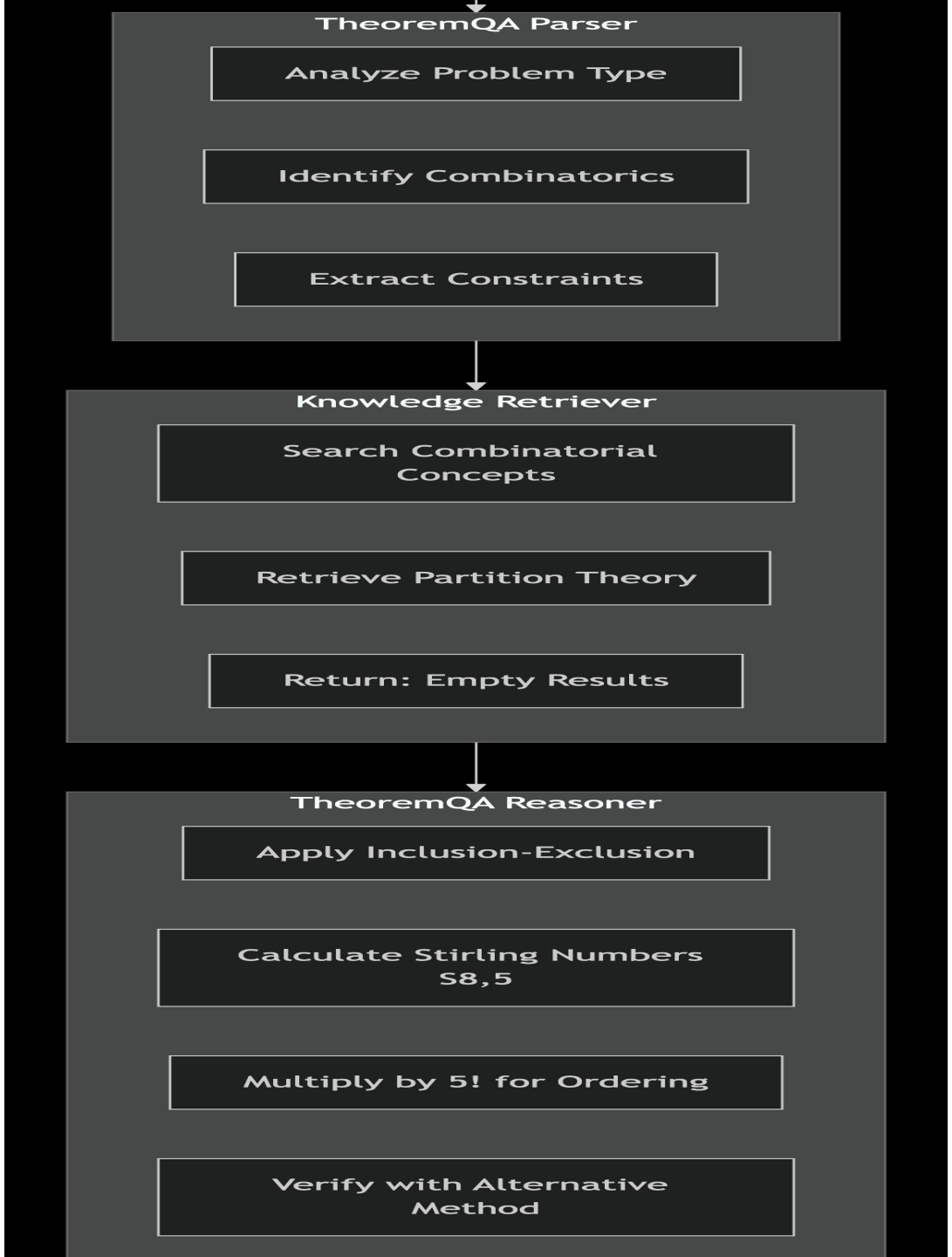
- Problem:Determine if the set ${F(x)=\int_0^x f(t) dt | f \in M }$ is sequentially compact



Analysis of Most Complex Successful Cases

- Problem:

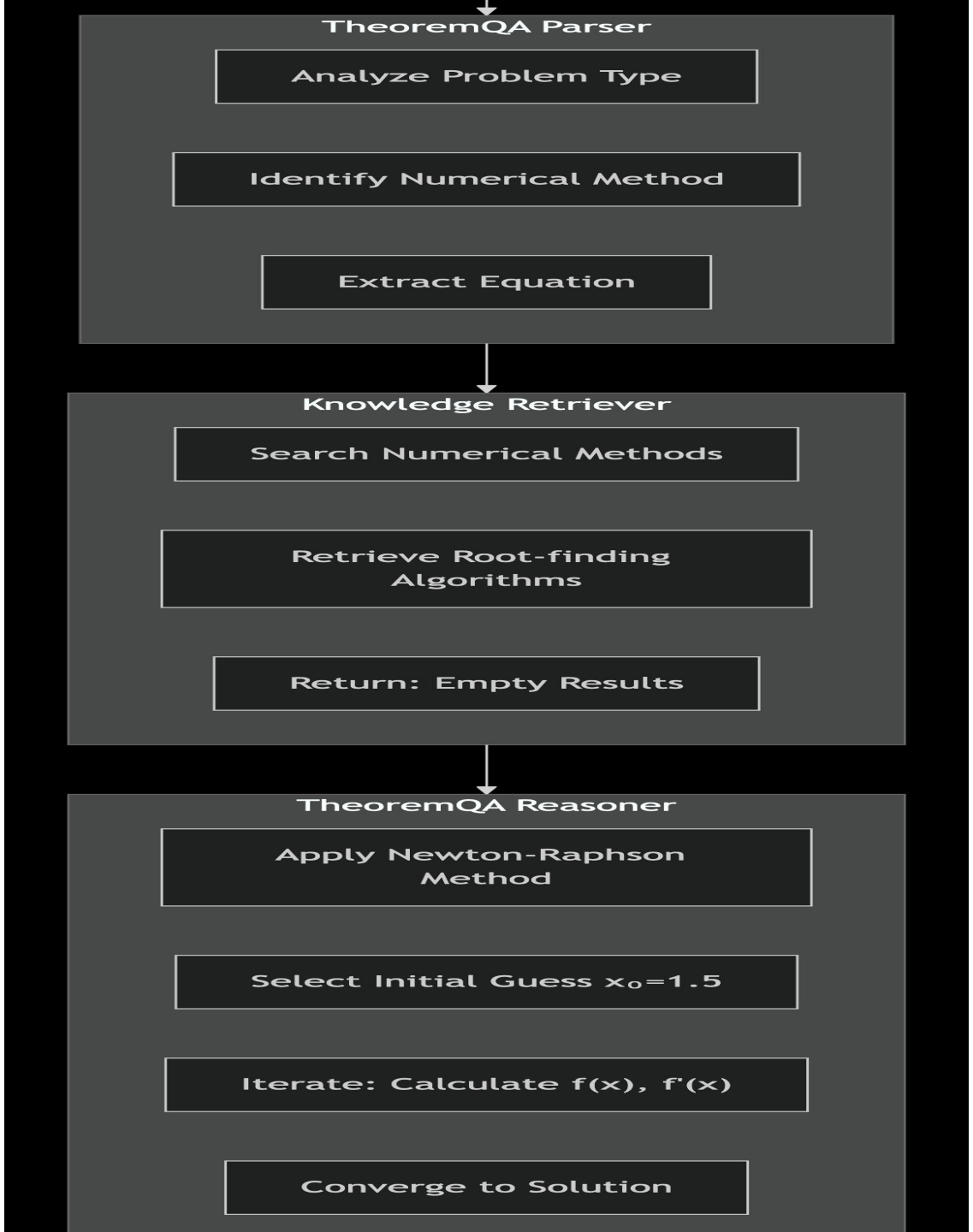
How many ways to divide 8 elements into 5 non-empty ordered subsets?



Analysis of Most Complex Successful Cases

- Problem:

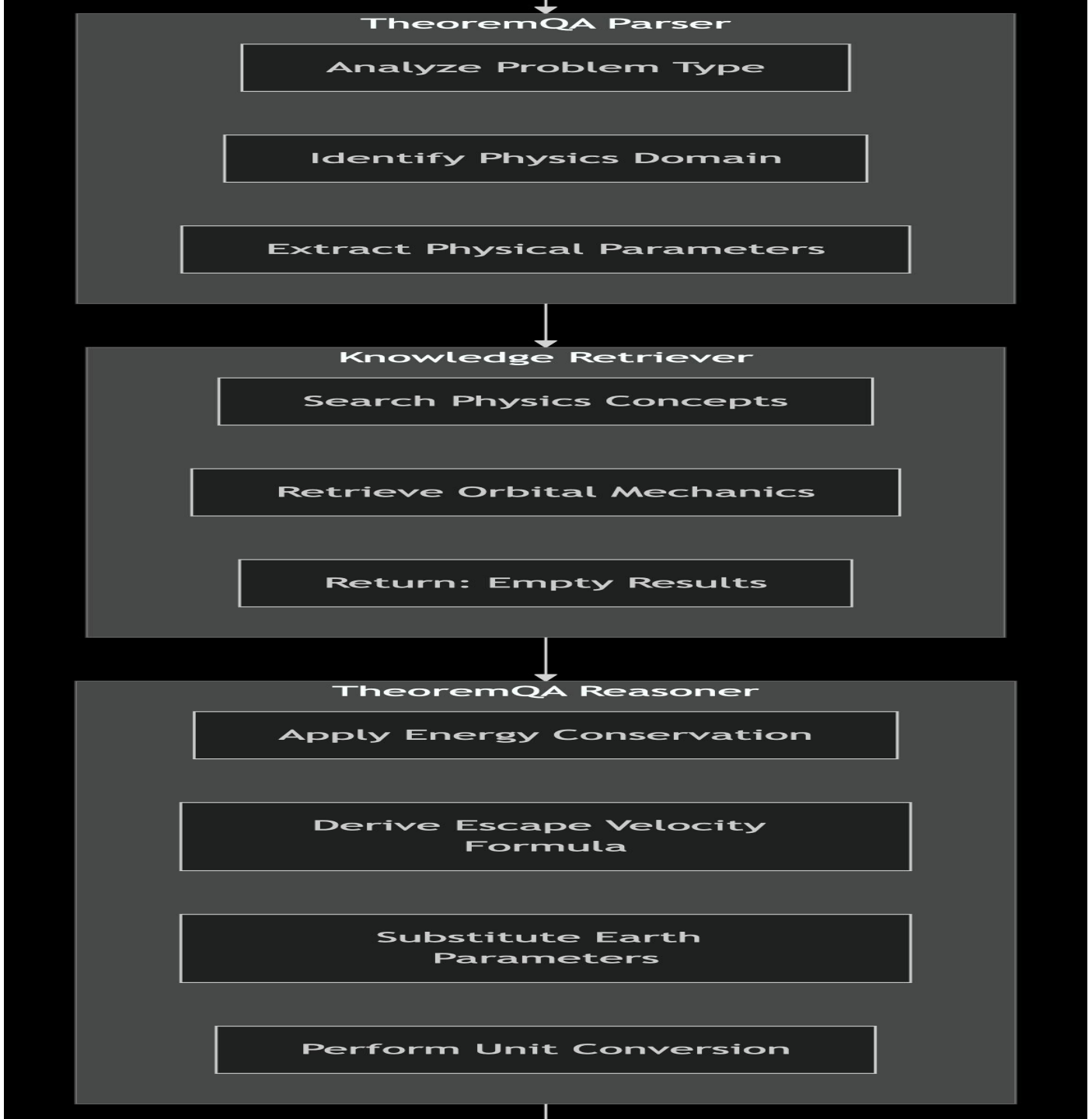
Solve $2x^3 + e^x = 10$ using Newton–Raphson method



Analysis of Most Complex Successful Cases

Problem:

Calculate Earth's escape velocity



Terminal Results