

Stage I: Dataset Construction

Real-World Scenario



Tasks

Enumeration

- ◆ Max Degree
- ◆ Triangle
- ◆ Maximum Clique

Exploration

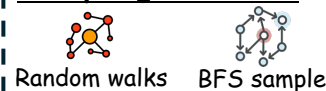
- ◆ Distance-K
- ◆ PathSum
- ◆ Diameter

Intuition

- ◆ MST
- ◆ Maximum K-core
- ◆ DistanceThreshold

Graphs

Sampling Methods



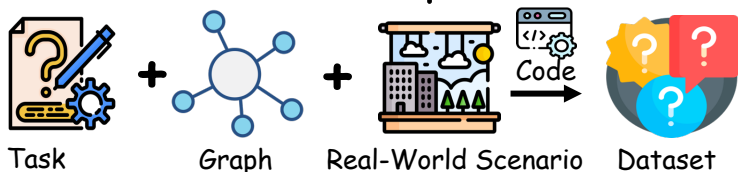
Graph Type

Undirected, Weighted, Tree

Node Range:

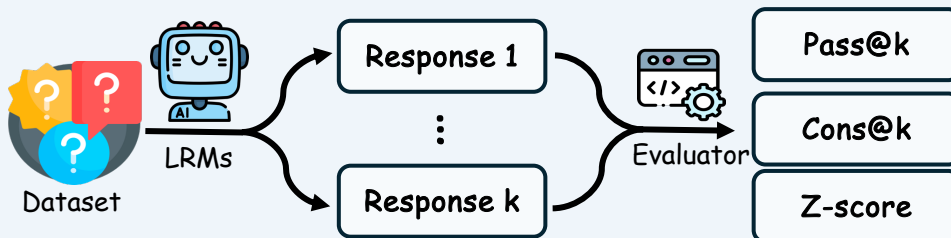
Level 1: 8-15,
Level 2: 16-30
Level 3: 31-50,
Level 4: 51-80,
Level 5: 81-120
Level 6: 121-240

Dataset Construction Pipeline



Stage II: Evaluation

RQ1: Exploring Reasoning Taxonomies LLMs Excel At



RQ2: Error Analysis

Algorithm Selection

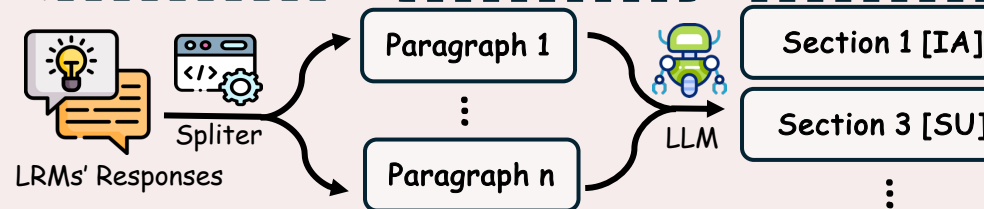
- Incorrect Algorithm
- Suboptimal Algorithm

Algorithm Execution

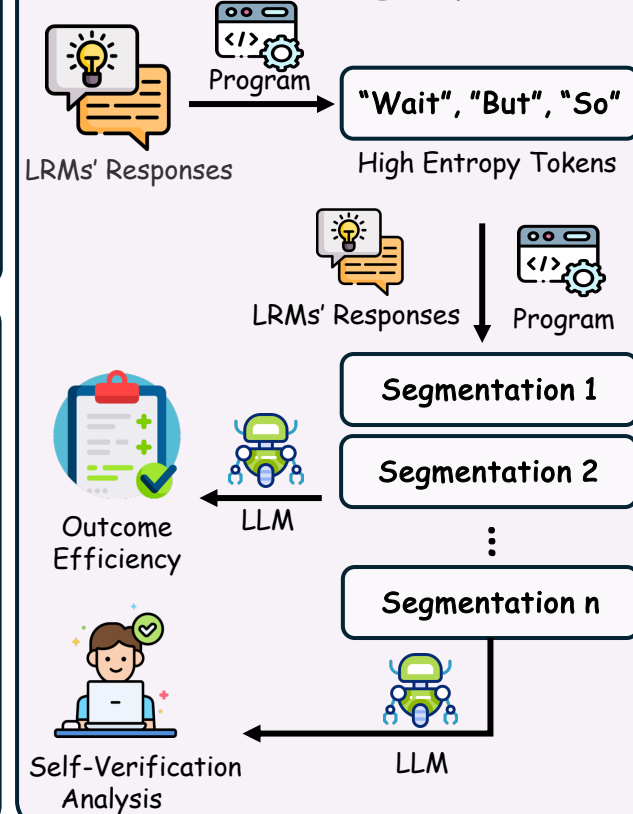
- State Update
- Missing Element
- Condition Misjudgement

Output Quality

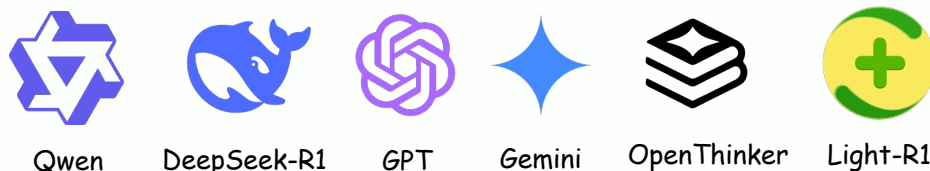
- Redundancy
- Information
- Graph Memory



RQ3: Overthinking Exploration



Reasoning Models



Non-Reasoning Models

