

### 1 Multi-Domain Coverage



Sciences



Medicine



Mathematics



Knowledge

#### Complex Reasoning



- Inductive
- Deductive
- Abductive
- Analogical

### 2 Multi-Modal Support

#### Datasets



Text-only



Multimodal

#### LLMs / LVLMs



Qwen



Intern



Llama



GPT

More Series and Various Sizes

### 3 Multi-Dimensional Evaluation

#### Collaborative Debate



Agent1 Response 1



Agent2 Response 1



Agent3 Response 1



Agent2 Response 1



Agent3 Response 1



Agent1 Response 2



Agent2 Response 2



Agent3 Response 1



Agent1 Response 1



Agent2 Response 2

⋮

X-Verify

Analysis →

#### Adversarial Debate



Affirmative side of debate



Negative side of debate



Neutral side of debate



Affirmative Response 1



Negative Response 1



Affirmative Side

Negative Side

Round1 - Final Answer

⋮

Accuracy

Cost

- Inference Time
- Tokens

