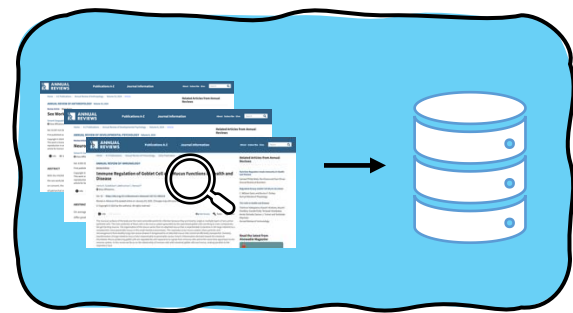


## Task Design

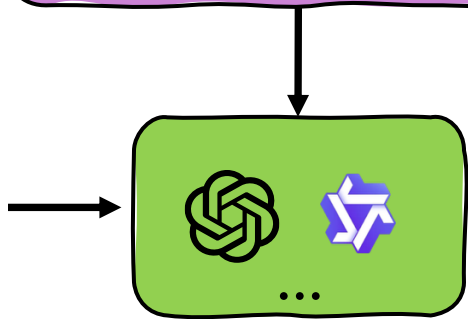
Task1: Reference Generation  
Task2: Abstract Writing  
Task3: Review Composition

## Metric Design

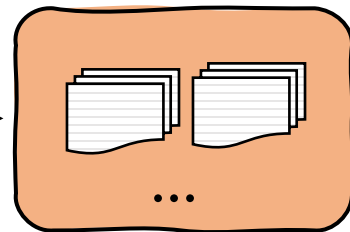
- Hallucination rate
- Factual consistency
- Semantic coverage



Data Collection



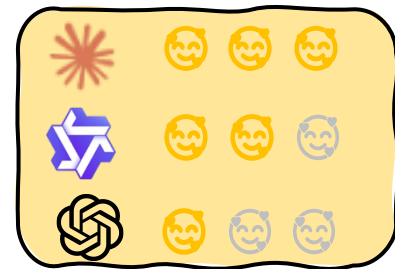
LLMs



Generated Text



Evaluator



Leaderboard