



1. Generating questions is hard.

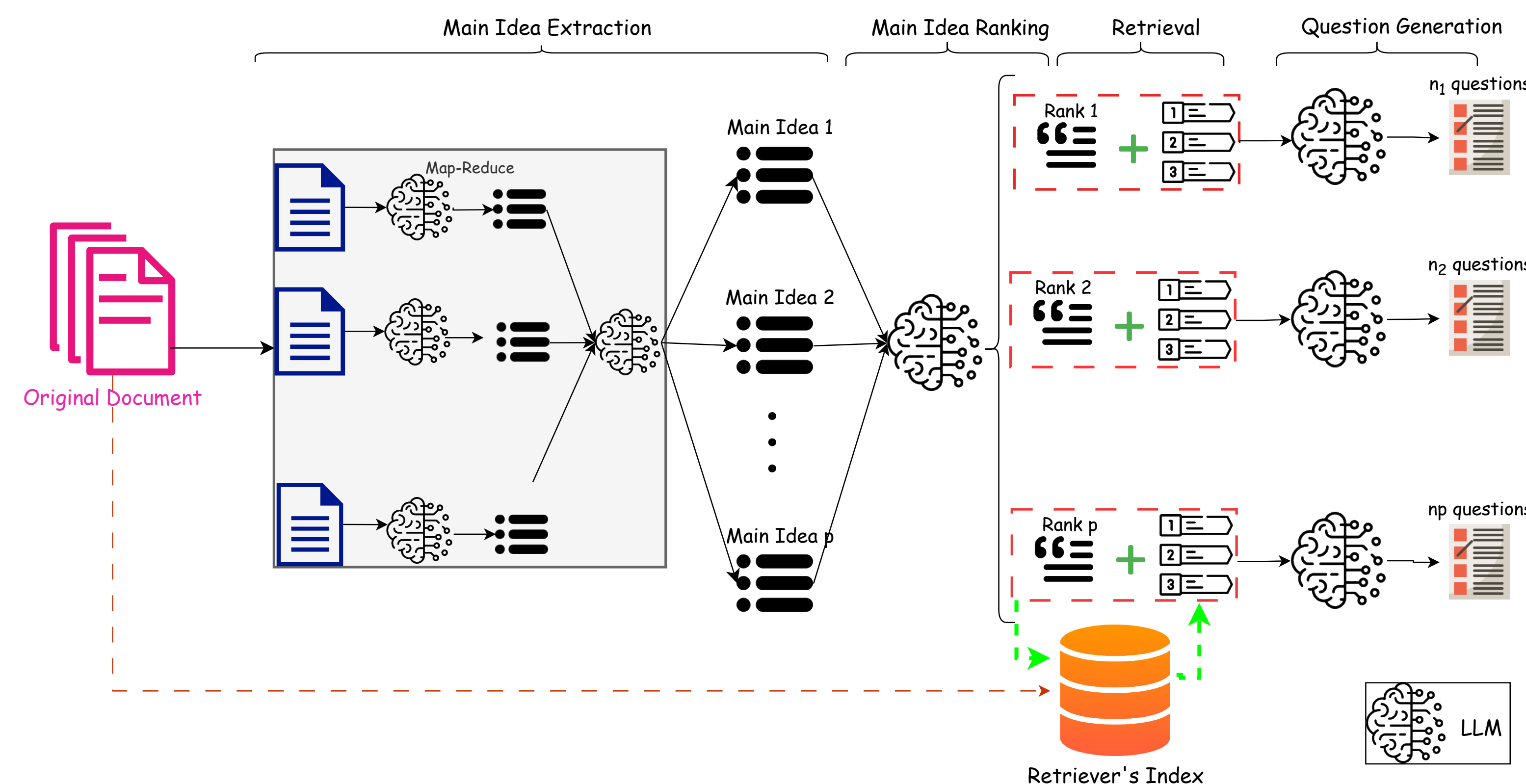


- Time-intensive
- High cognitive load
- Unscalable existing solutions

2. Goals

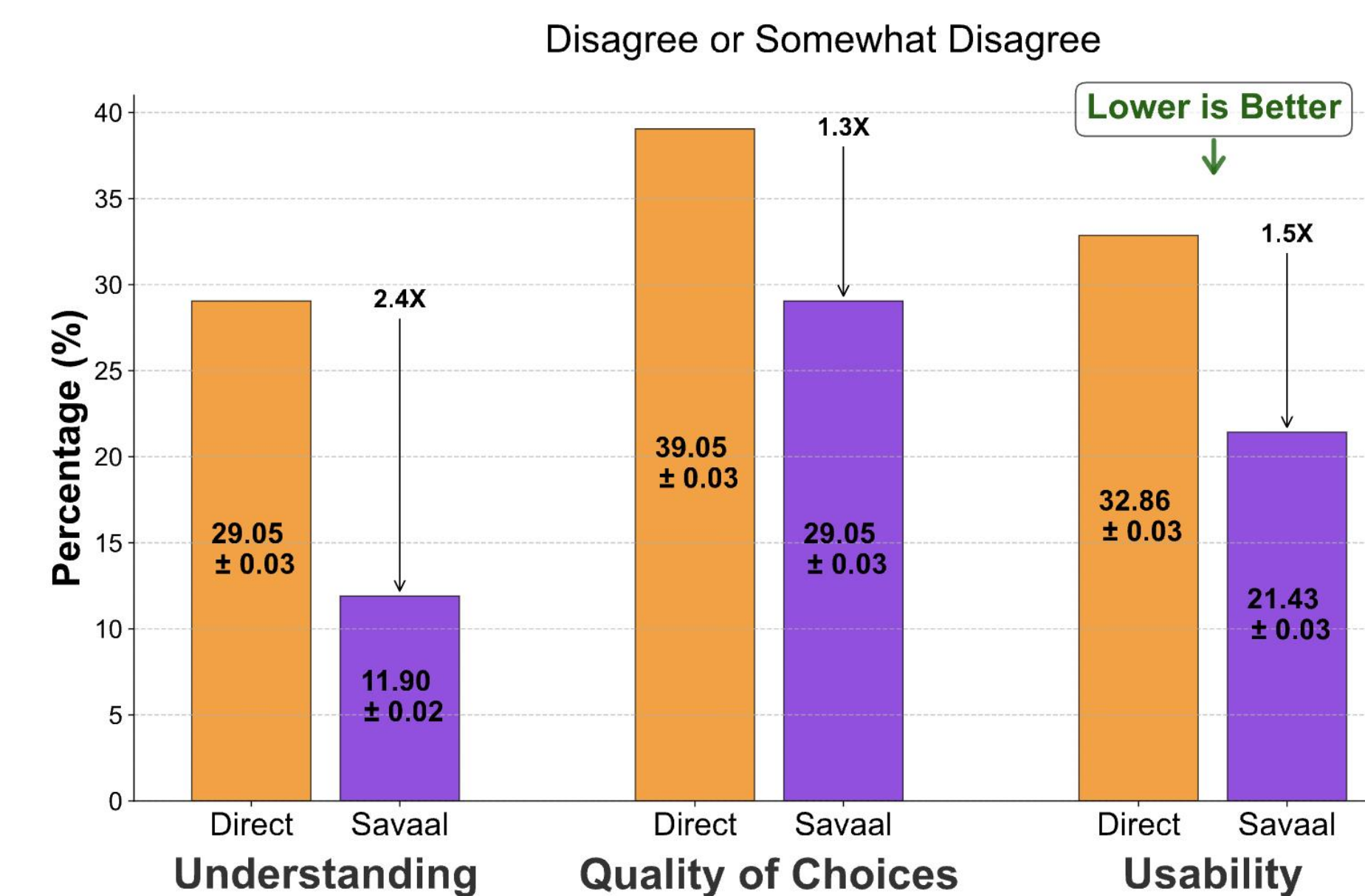
- Generates questions automatically
- Scales in context size, cost, and quality
- Probes deeper understanding
- Works across domains

3. Savaal's Question Generation Pipeline



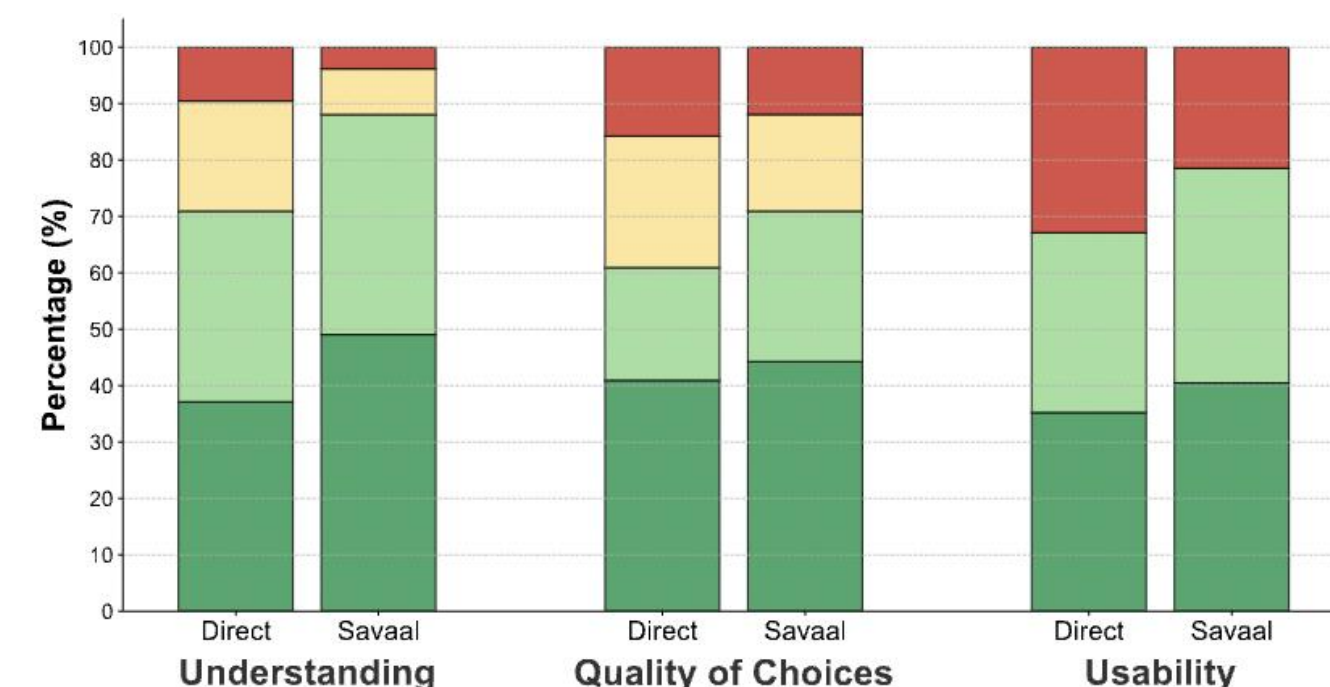
4. Human Expert Evaluation

- 76 experts across 1600 questions
- Likert Scale: Agree, Somewhat Agree, Somewhat Disagree, Disagree

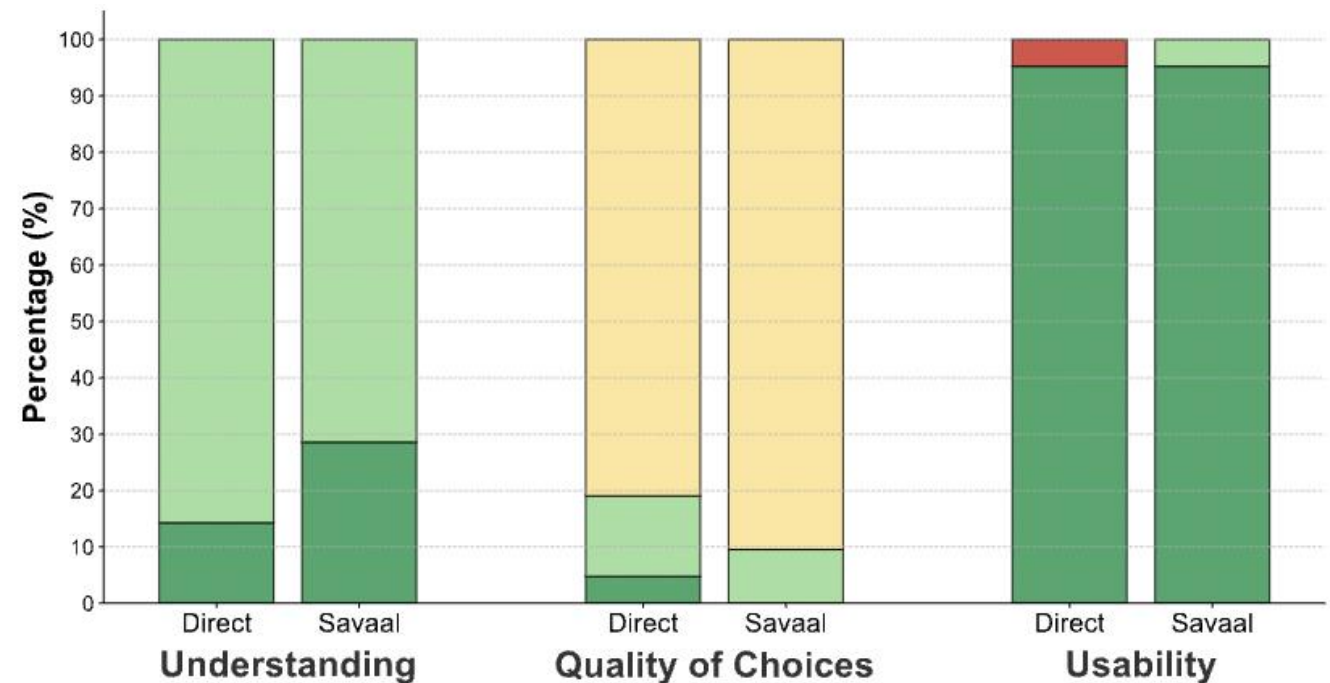


5. LLMs struggle judging questions.

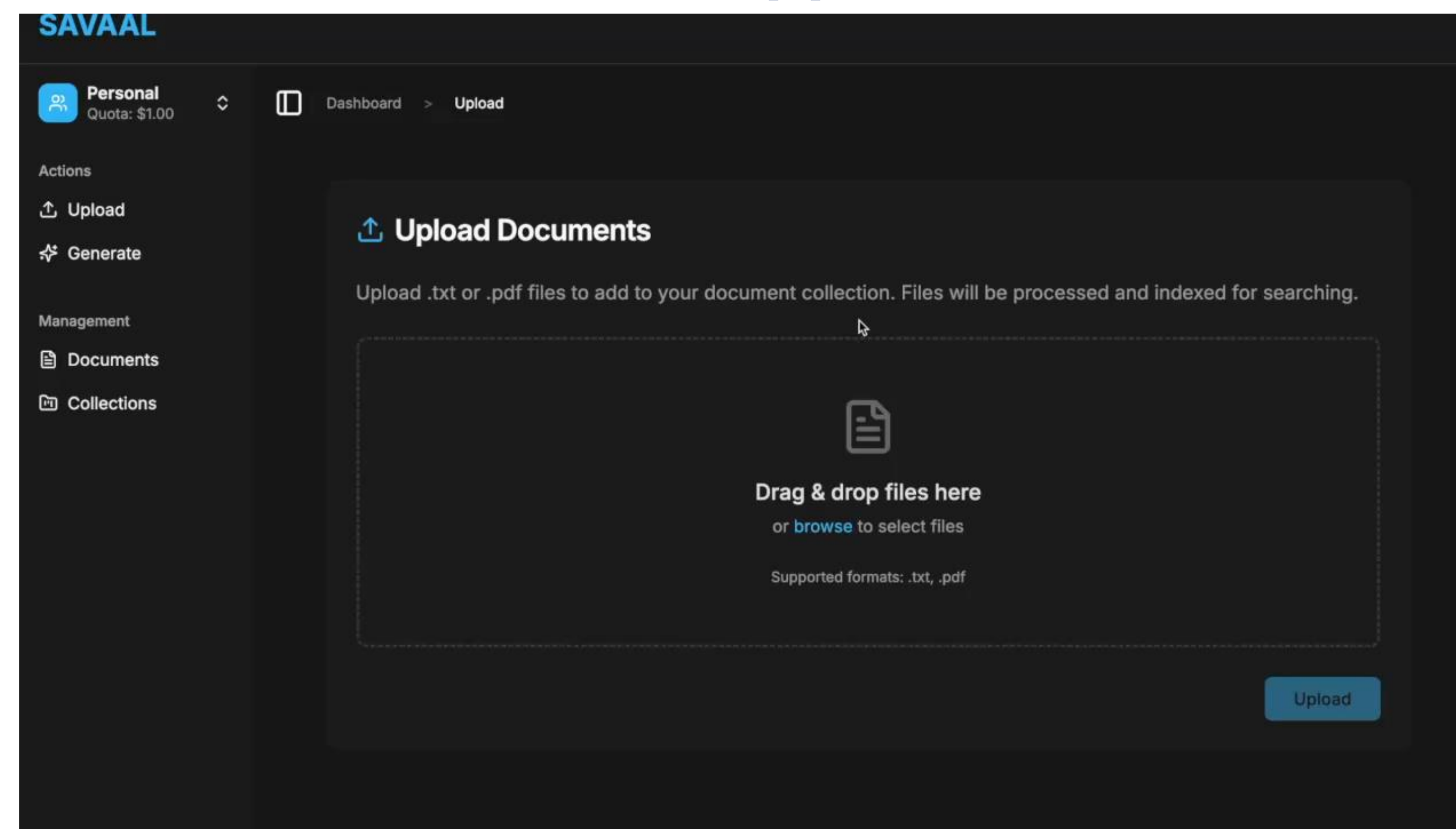
Human Expert Scores



GPT-4o Scores



6. Savaal's Application



7. Savaal has scalable cost.

Direct: direct prompting LLM

