

## RAG 평가요소

- ① 검색 평가
  - ② LLM 답변 평가
- RAG 평가는 이 2개의 요소가 합쳐진 형태

• 검색 평가 ; 질문을 했을 때 문서가 잘 검색이 되는 가? (Retrieval 평가)  
→ 질문에 대한 답변의 근거가 되는 문서를 잘 검색했는 가?

• LLM 답변 평가 ; LLM이 검색된 문서를 기반으로 잘 답변을 하는 가? (Generation 평가)

⇒ RAG 평가는

- 1) 'Retrieval' 평가 진행 후,
- 2) 최종적인 'LLM 답변 평가'를 합쳐서 이루어짐.

즉, RAG 평가 중에 LLM 평가가 포함되어 있다고 볼 수도 있음.

### ※ Task 별 LLM 평가의 요소

- 1) QA Task
- 2) 요약 Task
- 3) 번역 Task
- 4) 대화 Task

Task 별로 다소 방법이 차이는 있겠으나,  
공통점은, '답변 품질 자체를 평가' 한다는 점.

### ※ RAG와 LLM의 추가적인 평가요소들

- 1) Answer Quality → 기본
- 2) Response Time
- 3) Cheap Cost
- 4) AI Safety

→ Use Case와 도메인에 따라  
중요점으로 고려해서 평가해야 함.

### ※ Agent 평가

Agent → LLM에게 역할 (페르소나)을 부여한 것.

→ 즉, LLM으로 수행하는 여러 Task가 Agent와 유사

⇒ LLM의 평가 요소를 숙지하면, 이를 Agent 평가에 활용한다음