검색엔진을 활용한 챗봇

목차

사전조사 & 아이디어

사전조사 & 아이디어

1. 본질적으로는 **RAG** 시스템과 같다.
   a. **Embedding Vector DB**를 활용하는가, 검색엔진을 활용하는가-의 차이만 존재
   b. 검색 성능의 고도화 및 최적화는 전적으로 **Search Engine**과 **Query**에만 영향을 받을 것
2. 사용자 질의는 특정 도메인에 한정되지 않고 범용 질의에 모두 대응할 수 있어야 한다.
   a. 특정 공신력있는 웹페이지를 특정 도메인에 직접 연결시키는 것은 확장성을 제한하는 일이 될 수 있다. (위키피디아 등)
3. 상용 파운데이션 모델(**GPT**)의 웹 검색 기능을 벤치마크로 삼고 실질적인 효과와 특이점을 발견할 수 있는 방향으로 메커니즘을 구현해야한다.
4. **LLM** 챗봇이 검색 엔진을 활용할 때, 어떤 경우에 효용이 있는지를 파악해야한다.
   a. 최신 **or** 실시간 정보
   b. 지역(실시간 위치), 장소, 사람등의 보다 세부적인 정보
   c. 불확실하거나, 논란이 있는 정보
   d. **LLM**이 학습하지 않은 데이터에 대한 정보

프로젝트 목표

Project Pipeline Overview

01 User Query

02 Search Engine

03 Text Context Preprocessing

04 Summary & Generation

Search Engine

검색 중...

Text Preprocessing

Generation

오늘(5월 5일) 서울은 흐리고 선선하며, 밤부터 가끔 비가 내릴 가능성이 있습니다. 현재 기온은 약 16°C이며, 밤 최저기온은 9°C로 예상됨

Query

오늘 서울 날씨 어때?

2025년 5월 5일 17시 기준 내일 날씨 예보입니다.
내일 새벽부터 저녁 사이 남부지방과 제주도에는 가끔 비가 내리는 곳이 있겠습니다. 그 밖의 지역에서도 약하게 비가 오는 곳이 있겠으니 실시간 레이더 영상과 최신 기상정보를 꼭 확인해 주세요!

프로젝트 목표

과업 세부 목표

AI 파이프라인 구축

LangChain 프레임워크를 핵심으로 활용하여 사용자 쿼리를 분석하고 가공

적절한 검색 엔진을 통해 정보 검색을 실행

최종 검색 결과를 전달받아 요약하고 검증

사용자 인터페이스 구현

Steamlit 라이브러리를 활용하여 실시간으로 질문을 입력하고 챗봇이 답변하는 구조

이전 대화 기록을 확인할 수 있는 웹 기반 채팅 인터페이스 구현

웹 데이터 전처리 & context 화

원본 Text(HTML)에서 광고, 메뉴 등 사용자 쿼리와 관련없는 Text는 제거하고 핵심 본문 내용은 명확하게 유지하여 Context 구성

LangChain

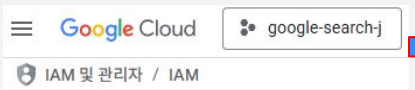
활용 요소

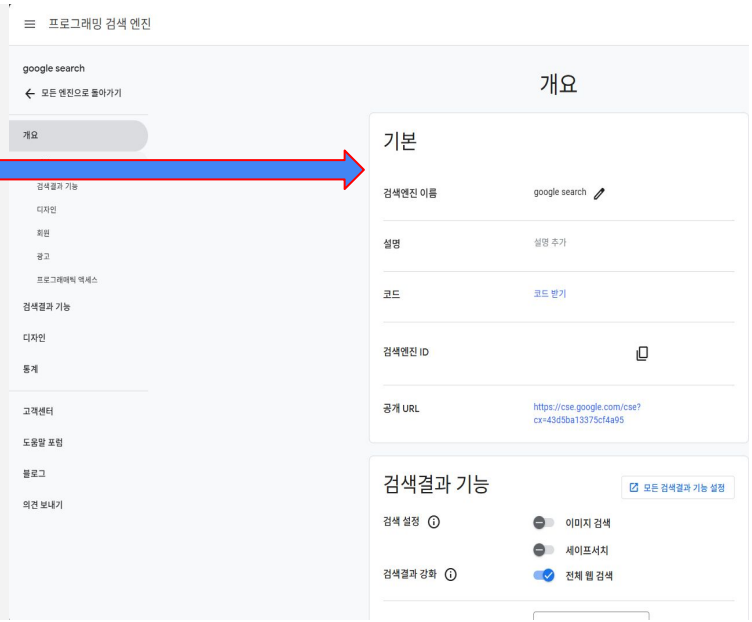
- **LLM**
  - **gpt-4o-mini**
- **Prompt Engineering**
  - **Sequential Chain**
    - 의존성 있는 **Chain**을 연속적으로 처리
  - **MultiPrompt Chain**
    - **Router** 등을 활용한 동적 흐름 제어
  - **Router Chain**
- **Agent**
  - **ReAct**
    - **Thought → Action → Observation**

Search Engine

CSE(Custom Search Engine)

- 프로그래밍 검색
  - **credentials.json**
    - 서버 통신 전용 구글 계정
    - **CSE 호출 시 구글 API 접근 가능**
  - **구글 검색 기반 & 전 세계 웹페이지 연동**
  - **특정 도메인 제한 가능 (Site Filtering)**
  - **신뢰도 높고, 형식화된 출력 결과물**
  - **100건/일 무료**
  - **Tiltle, Link(URL) 반환**

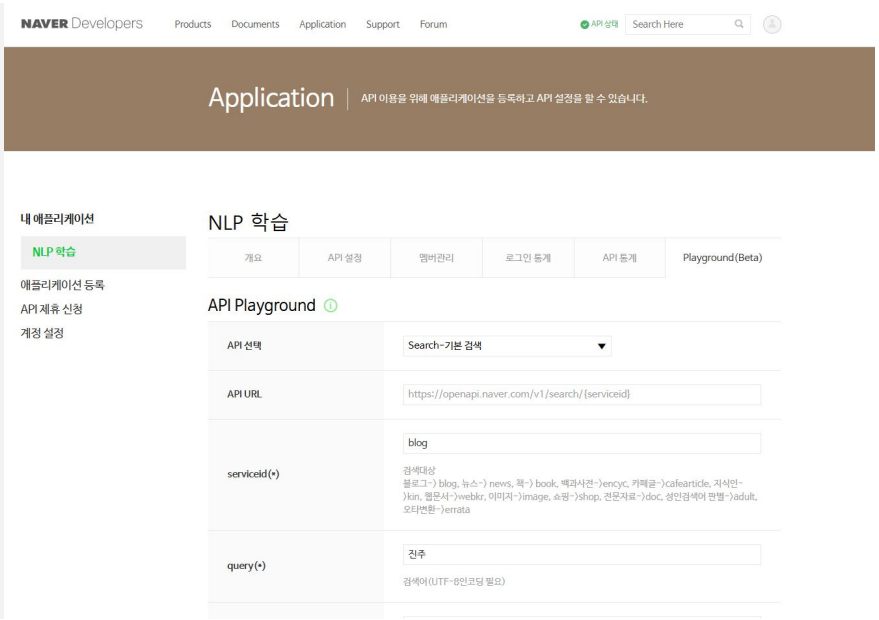




Search Engine

Naver Developers

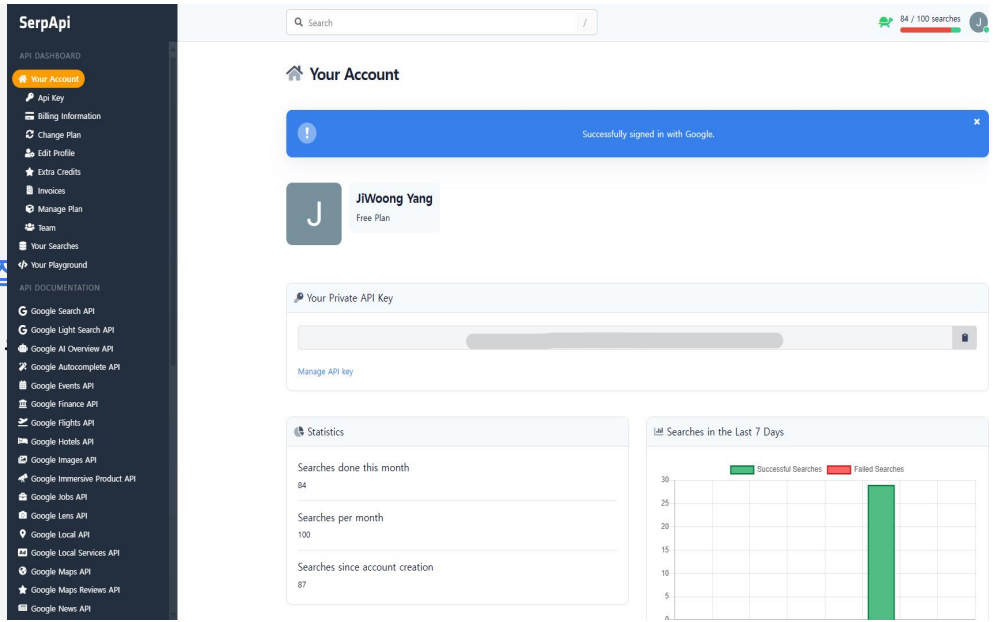
- 애플리케이션 - NAVER Developers
  - **NAVER Client ID 발급 후, 사용**
  - **한국어 콘텐츠 특화 (뉴스, 블로그 등)**
  - **Query에 대한 검색 결과 100건 제한**
  - **25000건/일 무료**
  - **Tiltle, Link(URL) 반환**



Search Engine

SerpAPI

- Dashboard - SerpApi
  - 실제 사용자 웹 UI 기반으로 수집
    - 실시간 정보에 강력
  - 다양한 포맷을 **JSON** 구조로 지
  - **100건/월 무료**
  - **Tiltle, Link(URL) 반환**



HTML 전처리

Search Engine

CSE

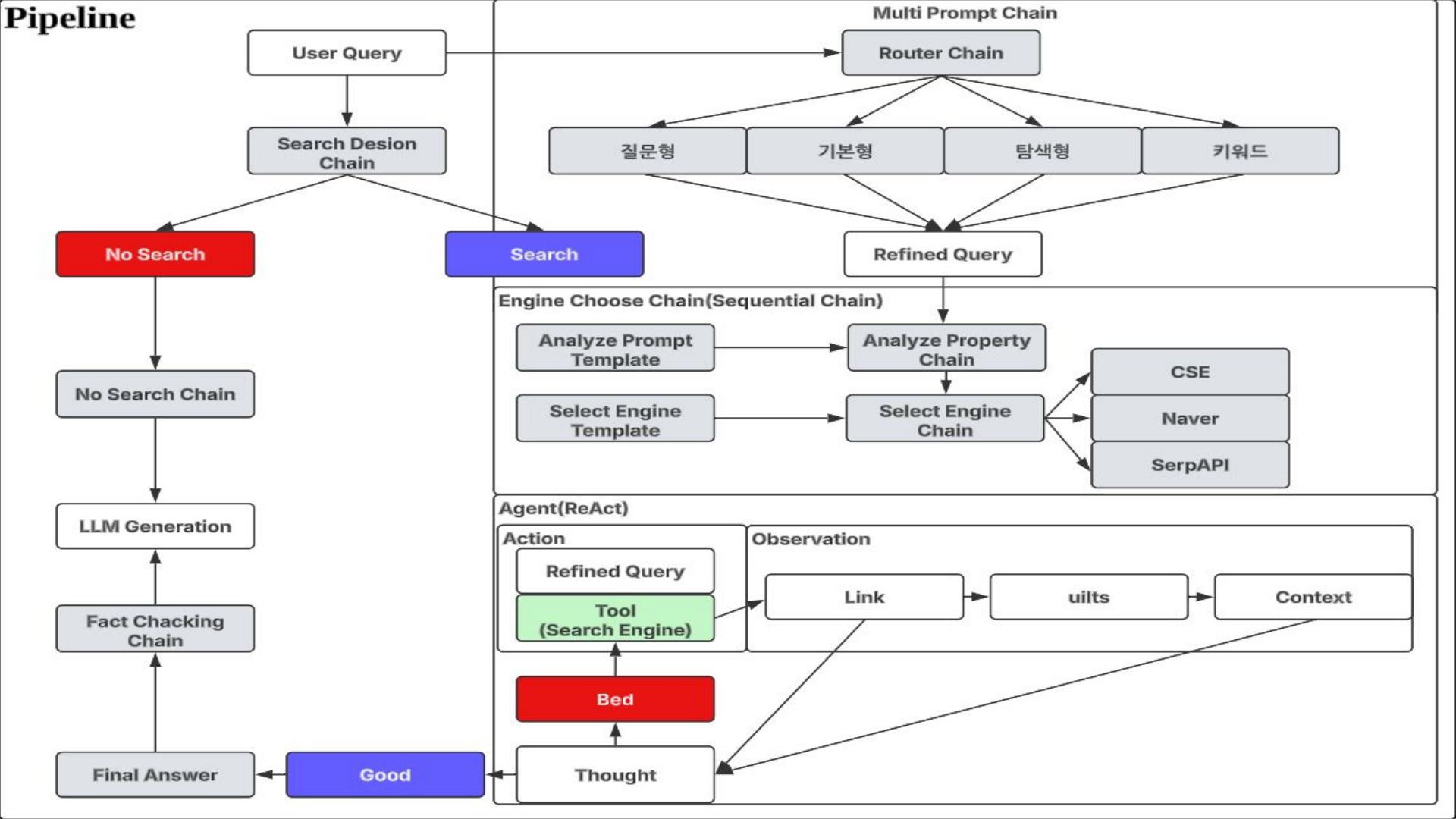
1. readability 라이브러리 활용
2. fallback : “style”, “noscript”, “form” 등 불필요한 테그 빼고 전부 불러오기

NAVER

1. 동일 형식 사이트에서 특정 CSS 리스트 활용 (news, blog 등)
2. readability 라이브러리 활용
3. fallback : “style”, “noscript”, “form” 등 불필요한 테그 빼고 전부 불러오기

SerpAPI

1. readability 라이브러리 활용
2. fallback : “style”, “noscript”, “form” 등 불필요한 테그 빼고 전부 불러오기
3. 특정 분야(금융 등)에선 UI Answer box 기반 정보 추출


Pipeline
Multi Prompt Chain
User Query
Router Chain
Search Desion Chain
질문형
기본형
탐색형
키워드
No Search
Search
Refined Query
Engine Choose Chain(Sequential Chain)
Analyze Prompt Template
Analyze Property Chain
CSE
Select Engine Template
Select Engine Chain
Naver
SerpAPI
No Search Chain
Agent(ReAct)
LLM Generation
Action
Observation
Refined Query
Link
uilts
Context
Fact Chacking Chain
Tool (Search Engine)
Bed
Final Answer
Good
Thought

Model Evaluation Overview

Test Data set

고려사항

질문 다양성
광범위한 질의에 대응 가능해야함

정답 검증 가능성
출처가 존재해야함

실시간성
답안이 실시간 정보에서 비롯되어야함

신뢰성
출처가 신뢰할 수 있어야함

성능 평가 FLOW

예시 질문
30문항

Final Model
Benchmark Model

QA set 구축

LLM Evaluation

결과 분석

모델 평가 개요

- 비교 모델

Model	LLM Version
Final Model	gpt-4o-mini
Chat GPT	gpt-4o

- 평가 내용

항목	설명
LLM Evaluation	- 사용자의 경험적인 측면을 반영한 정성적 평가 및 점수화 (사용자 쿼리 이해도, 대화 자연스러움 등)

- 활용 데이터셋 예시

- ooo 주식 얼마야?
- 오늘 서울 날씨 어때?
- 강남역 맛집 추천해줘

LLM Evaluation Results

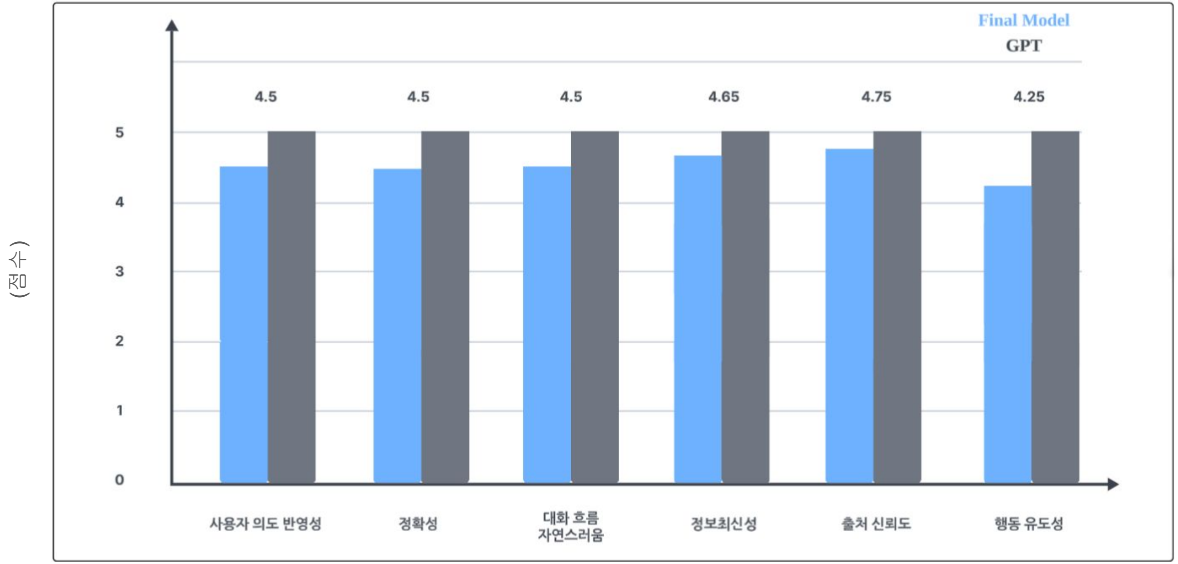
평가 방법

- 사용 데이터 컬럼

Question	예시 질문 30문항
Answer	모델 생성 응답

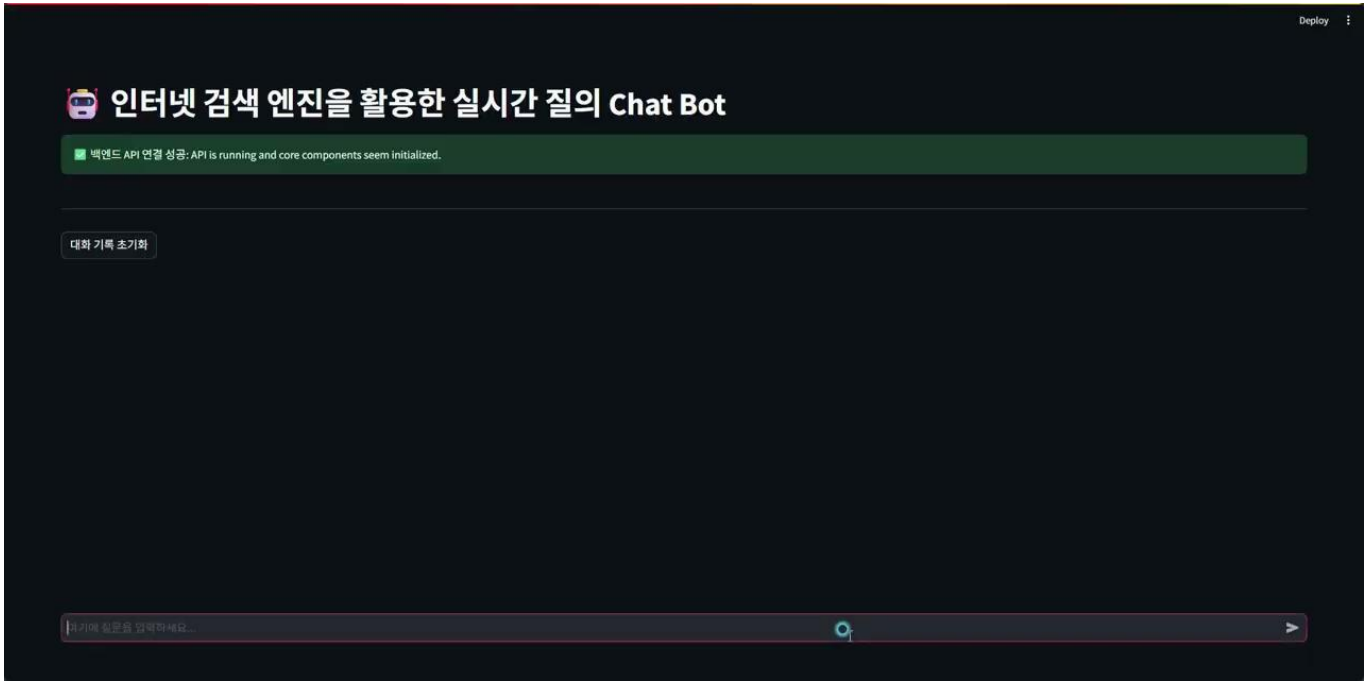
- 평가 내용 및 진행 방법
  - **사용자 의도 반영성**
    생성된 응답이 사용자의 의도를 잘 반영하여 생성되었는가?
  - **정확성**
    질문에 대한 정확한 답변을 제시하는가?
  - **대화 흐름 자연스러움**
    대화 흐름이 자연스럽고 매끄러운가?
  - **정보최신성**
    응답 내용이 현재 데이터를 반영하는가?
  - **출처 신뢰도**
    제공된 정보 출처가 공신력 있는 곳인가?
  - **행동 유도성**
    사용자가 행동을 할만큼 유용한가?
  - **평가 LLM:** GPT-4o
  - **다중 척도:** GPT의 답변을 만점으로 잡고 각 지표별 1점 ~ 5점 기준 제공 후, 점수 평가 및 이유 생성 지시

LLM 평가 결과 (1~5점 척도)



API 호출 제약(쿼리 당 1~3건) 및 text clipping(2000자)을 고려할 때, 전 부문에서 GPT 웹 검색과 크게 떨어지지 않는 점수 차를 보임.

Demo



https://drive.google.com/file/d/1UEvnHUDzTlagTll1h2bJDw0MzdOtNJKC/view?usp=sharing

Future Work

Future Work

1. Agent에게 Query 재작성 및 요약까지 Tool 로 관리하게 하는 방안
2. 검색 엔진의 호출 시간을 줄이기 위해 검색 엔진을 병렬로 연결할 필요가 있음
3. 평가 데이터셋 구축과 비교 평가를 보완할 필요성

감사합니다