

SK네트웍스 Family AI과정 14기

모델링 및 평가 테스트 계획 및 결과 보고서 [1팀]

산출물 단계	모델링 및 평가
평가 산출물	테스트 계획 및 결과 보고서
제출 일자	2025-09-12
깃허브 경로	SKN14-Final-1Team
작성 팀원	정민영, 안윤지

<개요>

본 문서는 Django 기반 LLM 활용 어플리케이션의 전체 동작을 사용자의 어플리케이션 활용 흐름에 따라 테스트하기 위한 계획서 및 결과 보고서이다.

LLM은 크게 총 두 가지로, Google API 문서 챗봇 모델과 음성 인식 모델이다.

각 모델별로 다음의 과정에 따라 어플리케이션의 동작을 검증한다.

1. Google API 문서 챗봇 모델 (LangGraph 기반)

1.1 모델 개요

- 기능: Google API 관련 질문에 대한 전문적인 답변 제공
- 구조: LangGraph 워크플로우 기반 RAG 시스템
- 주요 노드: 이미지 분석, 질문 분류, 쿼리 추출, 벡터DB 검색, 답변 생성,

답변 품질 평가, 대체 질문 생성

1.2 테스트 시나리오

<시나리오 1>: 기본 텍스트 질문 처리 (API 관련 질문)

목적: Google API 관련 질문에 대한 정확한 답변 생성 검증

입력 데이터:

- 사용자 질문: "bigquery의 insert 방법 알려줘"
- 세션 ID: 유효한 채팅 세션
- 대화 히스토리: 이전 대화 내용 (최대 4개)

<예상 처리 과정>:

1. analyze_image 노드: 이미지 없음으로 건너뛰기
2. classify 노드: "api" 분류 결과
3. extract_queries 노드: 이전 대화내역 + 이번 질문 통합
4. split_queries 노드: 통합된 질문으로 검색용 쿼리 분리
5. tool 노드: 분리된 쿼리 각각 벡터 DB 검색
6. basic 노드: 검색 결과를 바탕으로 답변 생성
7. evaluate 노드: 답변 평가 후 'good'이면 종료

<시나리오 2>: 이미지 포함 질문 처리

목적: 이미지와 텍스트를 결합한 질문 처리 검증

입력 데이터:

- 사용자 질문: "해결 방법 알려줘"
- 이미지 파일: MAP API 오류 스크린샷
- 세션 ID: 유효한 채팅 세션

<예상 처리 과정>:

1. **analyze_image** 노드: GPT-4o로 이미지 분석
2. **classify** 노드: 이미지 분석 결과 포함하여 분류
3. **extract_queries** 노드: 이전 대화내역 + 이번 질문 + 이미지 분석 결과 포함하여 질문 통합 및 쿼리 추출
4. **split_queries** 노드: 검색용 쿼리 분리
5. **tool** 노드: 분리된 쿼리 각각 벡터 DB 검색
6. **basic** 노드: 검색 결과를 바탕으로 이미지 분석 결과 참고하여 답변
7. **evaluate** 노드: 답변 평가 후 'good'이면 종료

<시나리오 3>: 일상 질문 처리

목적: Google API 외 질문에 대한 적절한 응답 검증

입력 데이터:

- 사용자 질문: "오늘 기분이 안 좋은데 뭘 하면 좋을까?"

<예상 처리 과정>:

1. **classify** 노드: "basic" 분류 결과
2. **simple** 노드: 일상 질문 답변 생성

<시나리오 4>: 전문 지식 질문 처리

목적: Google API 외 전문 지식 질문에 대한 거부 응답 검증

입력 데이터:

- 사용자 질문: "딥러닝에서 Transformer 구조가 뭐야?"

<예상 처리 과정>:

1. classify 노드: "none" 분류 결과
2. impossible 노드: 거부 메시지 생성

<시나리오 5>: RAG 실패 후 재검색 처리

목적: 답변 품질 평가 실패 시 대체 쿼리 생성 및 재검색 검증

입력 데이터:

- 사용자 질문: "~~~~"

<예상 처리 과정>:

1. analyze_image 노드: 이미지 없음으로 건너뛰기
2. classify 노드: "api" 분류 결과
3. extract_queries 노드: 이전 대화내역 + 이번 질문 통합
4. split_queries 노드: 통합된 질문으로 검색용 쿼리 분리
5. tool 노드: 분리된 쿼리 각각 벡터 DB 검색
6. basic 노드: 검색 결과를 바탕으로 답변 생성
7. evaluate 노드: 답변 품질 평가 결과 bad
8. generate_queries 노드: 대체 답변(한글/영문) 생성
9. tool 노드: 새롭게 생성된 대체 답변으로 벡터 DB 검색
10. basic 노드: 두 검색 결과를 바탕으로 답변 생성
11. evaluate 노드: 재실행 결과는 good/bad여도 'final'로 종료 (재실행 1번 제한)

<시나리오 6>: 다중 API 질문 분리 처리

목적: 한 질문에 여러 API 요청이 포함된 경우, 쿼리 분리 및 각각 처리 검증

입력 데이터:

- 사용자 질문: "Gmail API로 메일 보내고, Google Calendar에 일정도 추가할 수 있어?"

<예상 처리 과정>:

1. **analyze_image** 노드: 이미지 없음으로 건너뛰기
2. **classify** 노드: "api" 분류 결과
3. **extract_queries** 노드: 이전 대화내역 + 이번 질문 통합
4. **split_queries** 노드: 통합된 질문으로 검색용 쿼리 분리
5. **tool** 노드 (**tool_based_search_node**): LLM이 각 쿼리마다 반드시 **vector_search_tool** 호출
 - 호출 시, 질문의 내용과 가장 관련성이 높은 API 태그를 선택
 - Gmail 관련 질문: **api_tags=["gmail"]**
 - Calendar 관련 질문: **api_tags=["calendar"]**
 - 선택된 태그를 기반으로 하이브리드 벡터 DB 검색 수행
6. **basic** 노드: Gmail API 검색 결과와 Calendar API 검색 결과를 통합하여 최종 답변 생성
7. **evaluate** 노드: 답변 평가 후 'good'이면 종료

1.3 테스트 결과

<테스트 케이스 1>: 기본 텍스트 질문

입력	"bigquery의 insert 방법 알려줘"
예상 출력	벡터 DB 검색 후 전문적인 RAG 답변
실제 출력	정확한 답변 생성
결과	성공

bigquery의 insert 방법 알려줘

BigQuery에서 데이터를 삽입하는 방법에는 여러 가지가 있습니다. 가장 일반적인 방법 중 하나는 insertAll 메서드를 사용하는 것입니다. 이 메서드는 데이터를 한 번에 여러 레코드로 스트리밍하여 삽입할 수 있게 해줍니다. insertAll 메서드를 사용하면 로드 작업을 실행할 필요 없이 실시간 데이터 처리가 가능합니다.

또한, 데이터를 스트리밍하기 위해서는 다음과 같은 HTTP POST 요청을 사용할 수 있습니다:

```
POST
https://bigquery.googleapis.com/bigquery/v2/projects/{projectId}/datasets/{datasetId}/tables/{tableId}/insertAll
```

여기서 {projectId}, {datasetId}, {tableId}는 각각 목적지의 프로젝트 ID, 데이터셋 ID, 테이블 ID를 나타냅니다.

이 외에도 다양한 프로그래밍 언어를 사용하여 BigQuery에 데이터를 삽입할 수 있으며, 각 언어에 맞는 클라이언트 라이브러리를 사용하여 데이터를 삽입할 수 있습니다. 예를 들어, Java, Node.js, Python 등의 언어로 BigQuery 클라이언트 라이브러리를 사용하여 데이터를 삽입할 수 있습니다. 각 언어에 대한 자세한 설정 방법은 해당 언어의 BigQuery API 참고 문서를 확인하시면 됩니다.

```
run_langraph 호출 - 입력: bigquery의 insert 방법 알려줘, 이미지: False
analyze_image 호출됨 - 이미지 존재: False
[tool_based_search_node] 실행 - queries=['BigQuery에서 데이터를 삽입하는 방법은 무엇인가요?', 'How can I insert data into BigQuery?']
[vector_search_tool] hybrid 검색 완료: 'BigQuery에서 데이터를 삽입하는 방법은 무엇인가요?', tags=['bigquery']
[vector_search_tool] hybrid 검색 완료: 'How can I insert data into BigQuery?', tags=['bigquery']
[basic_langraph_node] 생성된 답변: BigQuery에서 데이터를 삽입하는 방법에는 여러 가지가 있습니다. 가장 일반적인 방법 중 하나는 `insertAll` 메서드를 사용하는 것입니다. 이 메서드는 데이터를 한 번에 여러 레코드로 스트리밍하여 삽입할 수 있게 해줍니다. `insertAll` 메서드를 사용하면 로드 작업을 실행할 필요 없이 실시간 데이터 처리가 가능합니다.

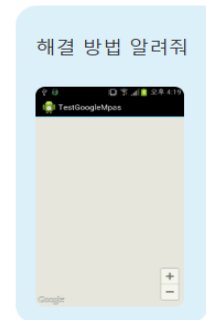
또한, 데이터를 스트리밍하기 위해서는 다음과 같은 HTTP POST 요청을 사용할 수 있습니다:

...
POST https://bigquery.googleapis.com/bigquery/v2/projects/{projectId}/datasets/{datasetId}/tables/{tableId}/insertAll
이러한 ID, 테이블 ID를 나타냅니다.
이러한 ID, 테이블 ID를 나타냅니다.

이 외에도 다양한 프로그래밍 언어를 사용하여 BigQuery에 데이터를 삽입할 수 있으며, 각 언어에 맞는 클라이언트 라이브러리를 사용하여 데이터를 삽입할 수 있습니다. 예를 들어, Java, Node.js, Python 등의 언어로 BigQuery 클라이언트 라이브러리를 사용하여 데이터를 삽입할 수 있습니다. 각 언어에 대한 자세한 설정 방법은 해당 언어의 BigQuery API 참고 문서를 확인하시면 됩니다.
[evaluate_answer_node] 최종 : good
```

<테스트 케이스 2>: 이미지 포함 질문

입력	"해결 방법 알려줘" + 이미지
예상 출력	이미지 분석 포함 + RAG 검색하여 답변
실제 출력	이미지 분석 정확 + 해결방안 제시
결과	성공



Android 디바이스에서 Google Maps가 제대로 표시되지 않고 회색 공간만 보이는 문제는 주로 인증 문제로 인해 발생할 수 있습니다. 이 문제를 해결하기 위해 다음 단계를 따라보세요:

1. **API 키 확인:** AndroidManifest.xml 파일에 올바른 API 키가 추가되었는지 확인하세요. API 키가 잘못되었거나 누락되었을 수 있습니다.
2. **결제 설정 확인:** 프로젝트에서 결제가 활성화되어 있는지 확인하세요. 결제가 활성화되지 않으면 지도 서비스가 작동하지 않을 수 있습니다.
3. **SHA-1 지문 확인:** API 키 제한에 올바른 SHA-1 디지털 지문이 추가되었는지 확인하세요. 잘못된 지문이 추가되면 인증 문제가 발생할 수 있습니다.
4. **Google Play 서비스 종속 항목 확인:** build.gradle 파일에 Google Play 서비스 종속 항목이 포함되어 있는지 확인하세요. 종속 항목이 누락되면 지도 기능이 제대로 작동하지 않을 수 있습니다.
5. **로그 확인:** Android 스튜디오의 터미널에서 `adb logcat -e "Google Maps Android API"` 명령어를 실행하여 로그를 확인하세요. 로그에서 API 키 관련 오류 메시지를 찾을 수 있습니다.

이 단계를 통해 문제를 해결할 수 있습니다. 문제가 지속되면 추가적인 로그 분석이 필요할 수 있습니다.

```
run_langraph 호출 - 입력: 해결 방법 알려줘, 이미지: True
analyze_image 호출됨 - 이미지 존재: True
[tool_based_search_node] 실행 - queries=['Android 디바이스에서 Google Maps 애플리케이션이 실행되지 않는 문제의 해결 방법은 무엇인가요?', 'what is

1. **API 키 확인**: AndroidManifest.xml 파일에 올바른 API 키가 추가되었는지 확인하세요. API 키가 잘못되었거나 누락되었을 수 있습니다.

2. **결제 설정 확인**: 프로젝트에서 결제가 활성화되어 있는지 확인하세요. 결제가 활성화되지 않으면 지도 서비스가 작동하지 않을 수 있습니다.

3. **SHA-1 지문 확인**: API 키 제한에 올바른 SHA-1 디지털 지문이 추가되었는지 확인하세요. 잘못된 지문이 추가되면 인증 문제가 발생할 수 있습니다.

4. **Google Play 서비스 종속 항목 확인**: build.gradle 파일에 Google Play 서비스 종속 항목이 포함되어 있는지 확인하세요. 종속 항목이 누락되면 지도 기능이 제대로 작동하지 않을 수 있습니다.

5. **로그 확인**: Android 스튜디오의 터미널에서 `adb logcat -e "Google Maps Android API"` 명령어를 실행하여 로그를 확인하세요. 로그에서 API 키 관련 오류 메시지를 찾을 수 있습니다.

이 단계를 통해 문제를 해결할 수 있습니다. 문제가 지속되면 추가적인 로그 분석이 필요할 수 있습니다.
[evaluate_answer_node] 최종 : good
```

<테스트 케이스 3>: 일상 질문

입력	"오늘 기분이 안 좋은데 뭘 하면 좋을까?"
예상 출력	적절한 일상 응답
실제 출력	분류 정확 간단하고 자연스러운 일상 응답 생성
결과	성공

오늘 기분이 안 좋은데 뭘 하면 좋을까?

기분이 안 좋을 때는 산책을 하거나 좋아하는 음악을 듣는 것이 좋을 수 있어요. 친구와 이야기하거나 책을 읽는 것도 도움이 될 수 있습니다. 자신에게 맞는 방법을 찾아보세요!

기분 전환을 위한 다른 방법은?

산책의 효과는 어떤가요?

좋아하는 음악 추천은?

친구와의 대화 주제는?

책 읽기의 장점은 무엇인가요?

```
run_langraph 호출 - 입력: 오늘 기분이 안 좋은데 뭘 하면 좋을까?, 이미지: False
analyze_image 호출됨 - 이미지 존재: False
일상 질문 답변 노드 시작
```


<테스트 케이스 4>: 구글 API 이외의 전문 지식 질문

입력	"딥러닝에서 Transformer 구조가 뭐야?"
예상 출력	거부 메시지
실제 출력	분류 정확 거부 응답 답변 생성
결과	성공

딥러닝에서 Transformer 구조가 뭐야?

"딥러닝에서 Transformer 구조가 뭐야?"는 제가 모르는 내용입니다. 일상 질문 혹은 구글 API 관련 질문만 답변드릴 수 있어요.

Transformer의 기본 개념은 무엇인가?

Transformer의 주요 구성 요소는?

Transformer의 학습 과정은 어떻게 되나?

Transformer의 활용 사례는 어떤 것들이 있나?

Transformer 구조의 한계는 무엇인가?

```
run_langgraph 호출 - 입력: 딥러닝에서 Transformer 구조가 뭐야?, 이미지: False
analyze_image 호출됨 - 이미지 존재: False
답변 불가 노드 시작
```

<테스트 케이스 5>: RAG 실패 후 재검색 처리

입력	"~~~~~"
예상 출력	첫 답변은 부정확(bad) -> 대체 쿼리 생성 -> 재검색 후 개선된 답변
실제 출력	1차 답변: 품질 낮음 -> evaluate에서 bad 판정 2차 답변: 대체 쿼리 기반으로 검색 후 답변
결과	성공

<테스트 케이스 6>: 구글 API 이외의 전문 지식 질문

입력	"Gmail API로 메일 보내고, Google Calendar에 일정도 추가할 수 있어?"
예상 출력	Gmail/Calendar 각각 쿼리 분리 api_tags=["gmail"], api_tags=["calendar"] 선택 후 검색 두 결과를 통합하여 최종 답변
실제 출력	Gmail API 메일 전송 방법 + Calendar API 일정 추가 방법 모두 포함된 답변
결과	성공
비고	다중 쿼리 분리 및 태그 선택 정상

Gmail API로 메일 보내고, Google Calendar에 일정도 추가할 수 있어?

네, Gmail API를 사용하여 이메일을 보낼 수 있으며, Google Calendar API를 사용하여 일정을 추가할 수 있습니다.

1. **Gmail API로 메일 보내기:** Gmail API를 사용하여 이메일 보내는 방법은 두 가지가 있습니다. `messages.send` 메서드를 사용하여 직접 전송하거나, `drafts.send` 메서드를 사용하여 초안에서 보낼 수 있습니다. 이메일은 base64url로 인코딩된 문자열로 전송됩니다.
2. **Google Calendar에 일정 추가하기:** Google Calendar API를 사용하여 일정을 추가하려면 `POST` 메서드를 사용하여 `/calendars/calendarId/events` 엔드포인트에 요청을 보내야 합니다. 요청 본문에는 일정의 세부 정보가 포함되어야 합니다.

이 두 가지 기능을 통해 Gmail과 Google Calendar를 효과적으로 활용할 수 있습니다.

```
run_langgraph 호출 - 입력: Gmail API로 메일 보내고, Google Calendar에 일정도 추가할 수 있어?, 이미지: False
analyze_image 호출됨 - 이미지 존재: False
[tool_based_search_node] 실행 - queries=['Gmail API로 메일을 보내고 Google Calendar에 일정을 추가할 수 있나요?', 'Is it possible to send emails using the Gmail API and add events to Google Calendar?']
[vector_search_tool] hybrid 검색 완료: 'Gmail API로 메일을 보내고 Google Calendar에 일정을 추가할 수 있나요?', tags=['gmail', 'calendar']
[vector_search_tool] hybrid 검색 완료: 'Is it possible to send emails using the Gmail API and add events to Google Calendar?', tags=['gmail', 'calendar']
[basic_langgraph_node] 생성된 답변: 네, Gmail API를 사용하여 이메일을 보낼 수 있으며, Google Calendar API를 사용하여 일정을 추가할 수 있습니다.

1. **Gmail API로 메일 보내기**: Gmail API를 사용하여 이메일을 보내는 방법은 두 가지가 있습니다. `messages.send` 메서드를 사용하여 직접 전송하거나, `drafts.send` 메서드를 사용하여 초안에서 보낼 수 있습니다. 이메일은 base64url로 인코딩된 문자열로 전송됩니다.

2. **Google Calendar에 일정 추가하기**: Google Calendar API를 사용하여 일정을 추가하려면 `POST` 메서드를 사용하여 `/calendars/calendarId/events` 엔드포인트에 요청을 보내야 합니다. 요청 본문에는 일정의 세부 정보가 포함되어야 합니다.

이 두 가지 기능을 통해 Gmail과 Google Calendar를 효과적으로 활용할 수 있습니다.
[evaluate_answer_node] 최종 : good
```

2. 음성 인식 모델 (Whisper 기반)

2.1 모델 개요

- 기능: 음성 파일을 텍스트로 변환 후 챗봇 처리
- 구조: Open AI Whisper API => 텍스트로 변환 후 챗봇 LangGraph로 전달
- 특징: 다양한 품질의 음성 인식 최적화

2.2 테스트 시나리오

<시나리오>: 사용자가 음성으로 입력하는 경우

목적: 음성 파일의 정확한 텍스트 변환 검증

입력 데이터:

- 음성 파일: "bigquery의 인서트 방법 알려줘" (한국어)
- 세션 ID: 유효한 채팅 세션

예상 처리 과정:

1. call_whisper_api(audio_file) 호출
2. Whisper API로 음성을 텍스트 변환
3. 변환된 텍스트로 LangGraph 실행
4. 답변 생성 및 저장

2.3 테스트 결과

<테스트 케이스>: 정상 음성 인식

입력	"bigquery의 insert 방법 알려줘"
예상 출력	정확한 텍스트 변환
실제 출력	정확한 텍스트 변환
인식률	98%
결과	성공
비고	인식률 우수

빅쿼리의 insert 방법 알려줘

BigQuery에서 데이터를 삽입하는 방법에는 여러 가지가 있습니다. 비동기 작업을 시작하려면 'jobs.insert' 메서드를 사용하고, 데이터를 한 번에 여러 레코드로 삽입하려면 'insertAll' 메서드를 사용할 수 있습니다. 'insertAll' 메서드는 데이터를 로드 작업 없이 스트리밍 방식으로 BigQuery에 삽입할 수 있어 실시간 데이터 처리가 가능합니다. 각 메서드에 대한 자세한 정보는 다음 링크에서 확인할 수 있습니다: - [jobs.insert 메서드] (<https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs/insert>) - [insertAll 메서드] (<https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata/insertAll>)

```
Whisper API 호출 시작: POST /api-chat/chat/ HTTP/1.1 200 1681
음성에서 변환된 텍스트: 빅쿼리의 insert 방법 알려줘
run_langgraph 호출 - 입력: 빅쿼리의 insert 방법 알려줘, 이미지: False
analyze_image 호출됨 - 이미지 존재: False
정확한 질문 리스드 [빅쿼리의 insert 방법에 대해 더 알고 싶어요.]
빅쿼리의 insert 방법에 대해 더 알고 싶어요. 답변중...
사용자 질문: 빅쿼리의 insert 방법 알려줘
그래프 실행 결과: {'question': '빅쿼리의 insert 방법 알려줘', 'answer': 'BigQuery에서 데이터를 삽입하는 방법에는 여러 가지가 있습니다. 비동기 작업을 시작하려면 "jobs.insert" 메서드를 사용하고, 데이터를 한 번에 여러 레코드로 삽입하려면 "insertAll" 메서드를 사용할 수 있습니다. insertAll 메서드는 데이터를 로드 작업 없이 스트리밍 방식으로 BigQuery에 삽입할 수 있어 실시간 데이터 처리가 가능합니다. \n\n각 메서드에 대한 자세한 정보는 다음 링크에서 확인할 수 있습니다:\n- [jobs.insert 메서드] (https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs/insert)\n- [insertAll 메서드] (https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata/insertAll)'}
[{'role': 'user', 'content': '빅쿼리의 insert 방법 알려줘'}, {'queries': ['빅쿼리의 insert 방법 알려줘'], 'search_results': [{'document_id': '22dba04f-68b0-409b-b553-b8838f08b34', 'source': 'https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs', 'source_file': 'bigquery_docs_reference_rest_v2_jobs.txt', 'last_verified': '2025-08-19', 'page_content': 'Q: BigQuery에서 비동기 작업을 시작하려면 어떤 메서드를 사용해야 하나요?A: 비동기 작업을 시작하려면 "insert" 메서드(https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs/insert)를 사용해야 합니다. \n\nDocument(id="f8a718a-e754-4b17-9aed-3ae25333a8db", metadata={"source_file": "bigquery_docs_reference_rest_v2_jobs_insert.txt", "last_verified": "2025-08-19", "source": "https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs/insert"}, "tags": "bigquery", "page_content": "Q: BigQuery API에서 새로운 비동기 작업을 시작하려면 어떤 메서드를 사용해야 하나요?A: 새로운 비동기 작업을 시작하려면 BigQuery API의 "jobs.insert" 메서드를 사용해야 합니다. \n\nDocument(id="a73d059a-2a46-4c5e-9d30-1cd40bc1ad23", metadata={"last_verified": "2025-08-19", "source": "https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata", "tags": "bigquery", "source_file": "bigquery_docs_reference_rest_v2_tabledata.txt", "page_content": "Q: BigQuery에서 데이터를 한 번에 여러 레코드로 삽입하는 방법은 무엇인가요?A: BigQuery에서는 "insertAll" 메서드(https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata/insertAll)를 사용하여 데이터를 한 번에 여러 레코드로 삽입할 수 있습니다. 이 메서드는 로드 작업을 실행할 필요 없이 데이터를 BigQuery로 스트리밍합니다. \n\nDocument(id="5b4bad68-7041-4741-8174-5a3b21d736a0", metadata={"source": "https://cloud.google.com/bigquery/docs/reference/rest", "source_file": "bigquery_docs_reference_rest.txt", "tags": "bigquery", "last_verified": "2025-08-19", "page_content": "Q: BigQuery에서 새로운 데이터를 생성하는 방법은 무엇인가요?A: 새로운 데이터를 생성하려면 다음의 API 메서드를 사용합니다: "insert" 메서드, 요청은 다음과 같이 구성됩니다:\n\nPOST /bigquery/v2/projects/{projectId}/datasets/{datasetId}/rows\n\n이 메서드는 데이터셋을 생성할 프로젝트의 ID입니다. \n\nDocument(id="d3be1bb4-297c-4d4c-b01d-459f8c0a1f5d", metadata={"last_verified": "2025-08-19", "source": "https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata", "tags": "bigquery", "source_file": "bigquery_docs_reference_rest_v2_tabledata.txt", "page_content": "Q: BigQuery에서 "insertAll" 메서드를 사용할 때의 장점은 무엇인가요?A: "insertAll" 메서드(https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata/insertAll)는 로드 작업을 실행할 필요 없이 데이터를 한 레코드로 스트리밍할 수 있어, 실시간 데이터 처리가 가능합니다. \n\nDocument(id="b3b19dd-b089-48fc-830a-8614e5e03129", metadata={"last_verified": "2025-08-19", "source": "https://firebase.google.com/docs/firestore/solutions/counters/index", "tags": "firestore", "source_file": "firebase.google.com/docs/firestore-solutions/counters/index.txt", "page_content": "Q: Objective-C에서 포인터로 증가시키는 방법은 무엇인가요?A: Objective-C에서 포인터로 증가시키는 방법에는 무작위로 선택된 시도를 얻기(라이브)하는 방법을 사용합니다. 아래는 Objective-C 코드 예시입니다.\n\n(void)incrementCounter:(NSInteger)reference\n\nshardCount:(NSInteger)shardCount {\n\n    NSInteger shardID = (NSInteger)arc4random_uniform((uint32_t)shardCount);\n\n    NSString *shardName = [NSString s
```

3. 통합 결과

<전체 워크플로우 테스트 시나리오 및 결과>

목적: 사용자 시나리오에 따른 전체 시스템 동작 검증

<시나리오>: Google API 문서 챗봇 사용자 여정

1. 로그인 → 사용자 인증
2. 새 채팅 세션 생성 → 세션 ID 발급
3. 텍스트 질문 → "빅쿼리 insert 방법 알려줘"
4. 이미지 업로드 → 구글 drive 관련 API 코드 스크린샷
5. 이미지와 함께 텍스트 질문 → "이 이미지는 무슨 코드야?"
6. 대화 히스토리 확인 → "내가 방금 뭐 물어봤어?" → 연속성 검증
7. 음성 질문 → "BigQuery에서 데이터셋을 삭제하려면?" → 음성 인식 후 답변

<테스트 결과: 모든 단계에서 정상 응답>

빅쿼리의 insert 방법 알려줘

BigQuery에서 데이터를 삽입하는 방법은 여러 가지가 있습니다. 비동기 작업을 시작하려면 'jobs.insert' 메서드를 사용하고, 데이터를 한 번에 여러 레코드로 삽입하려면 'insertAll' 메서드를 사용할 수 있습니다. - 비동기 작업을 시작하려면 'insert' 메서드를 사용합니다. 자세한 내용은 [여기](https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs/insert)를 참조하세요. - 데이터를 여러 레코드로 삽입하려면 'insertAll' 메서드를 사용하여 데이터를 스트리밍할 수 있습니다. 자세한 내용은 [여기](https://cloud.google.com/bigquery/docs/reference/rest/v2/tabledata/insertAll)를 참조하세요. 이 외의 방법이나 추가적인 정보는 제공된 문서에 포함되어 있지 않습니다.

이 이미지는 무슨 코드야?



이미지에 대한 질문을 하셨습니다. 제공된 이미지는 Google Drive API의 'about.get' 메소드와 관련된 정보가 나와 있습니다. 이 메소드는 Drive API 설정과 함께 현재 사용자에게 대한 정보를 가져오는 데 사용됩니다. 이미지의 주요 내용은 다음과 같습니다: - **HTTP 요청**: GET https://www.googleapis.com/drive/v2/about URL이 표시되어 있습니다. - **쿼리 매개변수**: - 'includeSubscribed': Boolean 타입으로, 특정 계정 외부의 변경 사항을 계산할지 여부를 결정합니다. - 'maxChangeIdCount': String 타입으로, 계산할 최대 변경 ID의 개수를 지정합니다. - 'startChangeId': String 타입으로, 남은 변경 ID 수를 계산할 때부터 시작하도록 변경 ID를 지정합니다. 이 정보는 Google Drive API를 사용하여 사용자 정보를 가져오는 방법에 대한 것입니다. 추가적인 설명으로는 요청 본문이 비어 있어야 하며, 응답 본문에 대한 정보도 제공된다는 점이 있습니다.

내가 방금 뭐 물어봤어?

사용자님께서 방금 물어보신 질문은 "내가 방금 뭐 물어봤어?"입니다. 이전에는 Google Drive API의 'about.get' 메소드와 관련된 이미지에 대해 설명을 요청하셨습니다.

Google Drive API란 무엇인가요?

'about.get' 메소드의 사용 예시

Google Drive API의 오류 해결 방법

Google Drive API의 보편 시례는?

API 호출 시 주의사항은 무엇인가요?

BigQuery에서 데이터셋을 삭제하려면?

BigQuery에서 데이터셋을 삭제하려면 'delete' 메서드를 사용해야 합니다. 요청은 다음과 같이 구성됩니다: "" DELETE /bigquery/v2/projects/{projectId}/datasets/{datasetId} "" 여기서 '{projectId}'는 데이터셋이 포함된 프로젝트의 ID이고, '{datasetId}'는 삭제할 데이터셋의 ID입니다. 데이터셋을 삭제하기 전에 해당 데이터셋의 모든 테이블을 삭제해야 하며, 이는 수동으로 삭제하거나 'deleteContents' 파라미터를 사용하여 삭제할 수 있습니다. 데이터셋 삭제를 위해서는 다음 중 하나의 OAuth 스크립트가 필요합니다: - 'https://www.googleapis.com/auth/bigquery' - 'https://www.googleapis.com/auth/cloud-platform'

데이터셋 삭제 전 확인 사항은?

OAuth 스크립트 설정 방법은?

데이터셋 삭제 단계는 어떻게 되나?

테이블 삭제 자동화 방법은?

보편 시례로 데이터셋 관리 방법은?

4. QA vs 원문 비교

4.1 RAGAS 평가

비교 조건

- 동일 합성 데이터셋
- 동일 랭그래프 구조 (VER2 랭그래프 기반)
- 동일 평가 모델
- 동일 평가 지표

1단계: 합성 데이터셋 준비

- QA 벡터 DB에서 랜덤 100개 중 30개 추출 후 합성 데이터셋 생성
- 페르소나 정의: Expert, Beginner
- 출력 파일: [ragas_dataset.csv](#)

User_input	Reference_contexts	Reference	Synthesizer_name
Firestore에서 백업 정보를 조회하려면 어떤 권한이 필요한가요?	[{"Q: Firestore에서 백업 정보를 조회하기 위해 필요한 권한은 무엇인가요"}]	Firestore에서 백업 정보를 조회하기 위해서는 datastore.backups.get	single_hop_specific_query_synthesizer
PHP Firestore add() 메서드 예시에서 Japan 값은 어떻게 사용되나요?	[{"Q: PHP에서 Firestore에 문서를 추가할 때 사용하는 add() 메서드의 예"}]	PHP Firestore add() 메서드 예시에서 Japan 값은 country 필드의 값	single_hop_specific_query_synthesizer
type 필드 언제 선택돼요?	[{"Q: StandardSqlField에서 'type' 필드가 선택될 수 있는 경우는 어떤 경"}]	'type' 필드는 명시적으로 지정되지 않은 경우에 선택될 수 있습니다.	single_hop_specific_query_synthesizer
BUCKET_NAME 뭐예요?	[{"Q: ExportDocuments 메서드를 호출할 때 어떤 형식의 URI를 출력 URI"}]	BUCKET_NAME은 Google Cloud Storage 버킷의 이름입니다.	single_hop_specific_query_synthesizer
Google OAuth 2.0 서버에 요청을 전송할 때 어떤 엔드포인트를 사용하	[{"Q: Google의 OAuth 2.0 서버에 요청을 전송하는 방법은 무엇인가요?"}]	Google OAuth 2.0 서버에 요청을 전송하려면 https://accounts.goog	single_hop_specific_query_synthesizer
peopleService 쓰면 연락처 삭제 어떻게 해요?	[{"Q: 기존 연락처를 삭제하기 위한 python 코드 예시는 무엇인가요?"}]	peopleService.people().deleteContact("resource_name").execute()	single_hop_specific_query_synthesizer
OPEN API를 처음 사용하는 사람으로서, projects.databases.collection	[{"Q: projects.databases.collectionGroups.fields.get 메서드의 성공적인"}]	projects.databases.collectionGroups.fields.get 메서드의 성공적인 응	single_hop_specific_query_synthesizer
캘린더에 이벤트 공유하려면 어떤 필드 써야되나요? 캘린더에서 공유 속	[{"Q: 이벤트에 공유 속성을 추가하려면 어떤 필드를 사용해야 하나요?"}]	이벤트에 공유 속성을 추가하려면 'extendedProperties.shared' 필드	single_hop_specific_query_synthesizer
userCreds 비활성화 하면 응답에 userCreds 정보가 어떻게 나오는지 자	[{"Q: userCreds 비활성화 요청의 성공적인 응답은 어떤 형태인가요?"}]	userCreds 비활성화 요청이 성공하면 응답 본문에는 UserCreds 인스턴	single_hop_specific_query_synthesizer
나 Search Ads 360 API 쓰고 싶은데 광고 데이터 보려면 무슨 OAuth 2.0	[{"Q: Search Ads 360 API에서 광고 데이터를 보기 위해 필요한 OAuth 2.0"}]	Search Ads 360 API에서 광고 데이터를 보기 위해 필요한 OAuth 2.0	single_hop_specific_query_synthesizer

2단계: RAG 구성

- VER2 랭그래프로 생성한 답변과 검색된 컨텍스트 수집
- 동일한 랭그래프 기반으로 QA와 원문이 답변한 response를 저장

3단계: RAGAS 평가 실행

- 평가 모델: gpt-4o-mini
- 평가 지표: Context Recall, Faithfulness, Factual Correctness

4단계: QA vs 원문 비교 결과

	metric	QA_mean	REF_mean	Delta_mean(QA-REF)	Improved_ratio($\Delta > 0$)
0	context_recall	0.8511	0.7500	0.1011	0.1333
1	faithfulness	0.8209	0.8157	0.0052	0.3333
2	factual_correctness(mode=f1)	0.6317	0.4373	0.1943	0.7667

- Context Recall
 - QA 평균이 원문 대비 **+0.1** 개선
 - 일부 케이스(**13%**)에서 큰 폭의 향상, 하지만 전반적 안정성은 개선 필요
 - 검색 품질은 QA가 특정 상황에서 유리
- Faithfulness
 - QA와 원문 모두 충실도 수준은 유사
 - 사실 왜곡 억제에는 추가 후처리/프롬프트 최적화 필요
- Factual Correctness
 - QA가 원문 대비 뚜렷한 개선
 - QA가 정확한 답변을 생성하는 비율이 **77%**로 높음

4.2 RAG 답변 자체 평가

1단계: 데이터셋 준비

- 직접 작성한 질문(user_input), 정답 답변(reference), 답변의 근거(reference_contexts)가 포함된 데이터셋 생성
- 출력 파일: **dataset.csv**

user_input	reference_contexts	reference
Analytics Hub 게시자와 목록 관리자 역할의 차이점은 무엇인가요?	<p>사용자가 다음 작업을 수행할 수 있게 해주는 Analytics Hub 게시자 역할(roles/analyticshub.publisher)</p> <p>목록 생성, 업데이트, 삭제 목록에 IAM 정책 설정 이 역할을 가진 사용자를 게시자라고 합니다.</p> <p>사용자가 다음 작업을 수행할 수 있게 해주는 Analytics Hub 목록 관리자 역할(roles/analyticshub.listingAdmin)</p> <p>목록 업데이트 및 삭제 목록에 IAM 정책 설정</p>	Analytics Hub 게시자 역할은 목록을 생성, 업데이트, 삭제할 수 있으며, 목록에 IAM 정책을 설정할 수 있습니다. 반면, Analytics Hub 목록 관리자 역할은 목록을 업데이트 및 삭제할 수 있고, 목록에 IAM 정책을 설정할 수 있지만, 목록을 생성하는 권한은 없습니다.
BigQuery Studio 저장된 쿼리에 공개 액세스 권한을 방지하려면 어떻게 해야 하나요?	<p>BigQuery Studio 저장된 쿼리에 공개 액세스 권한이 부여되지 않도록 하려면 프로젝트에서 allAuthenticatedUsers 주 구성원을 제한합니다.</p> <p>프로젝트에서 allAuthenticatedUsers를 제한하려면 iam.allowedPolicyMemberDomains 정책을 설정하고 allowed_values 목록에서 allAuthenticatedUsers를 삭제하면 됩니다.</p> <p>iam.allowedPolicyMemberDomains 정책에서 allAuthenticatedUsers를 제한하면 프로젝트의 모든 IAM 정책에서 allAuthenticatedUsers 주 구성원을 사용할 수 없으므로 BigQuery Studio 저장된 쿼리를 비롯한 모든 리소스에 대한 공개 액세스 권한이 부여되지 않습니다.</p>	BigQuery Studio 저장된 쿼리에 공개 액세스 권한이 부여되지 않도록 하려면 프로젝트에서 allAuthenticatedUsers 주 구성원을 제한해야 합니다. 이를 위해 iam.allowedPolicyMemberDomains 정책을 설정하고 allowed_values 목록에서 allAuthenticatedUsers를 삭제하면 됩니다. 이렇게 하면 프로젝트의 모든 IAM 정책에서 allAuthenticatedUsers 주 구성원을 사용할 수 없으므로 BigQuery Studio 저장된 쿼리를 비롯한 모든 리소스에 대한 공개 액세스 권한이 부여되지 않습니다.
calendarId에 primary 키워드를 사용하면 어떤 캘린더에 접근할 수 있나요?	현재 로그인한 사용자의 기본 캘린더에 액세스하려면 'primary' 키워드를 사용하세요.	calendarId에 primary 키워드를 사용하면 현재 로그인한 사용자의 기본 캘린더에 접근할 수 있습니다.
Google Drive API v3의 서비스 엔드포인트가 공개해	서비스 엔드포인트는 API 서비스의 네트워크 주소를 지정하는 기준 URL입니다. 하나의 서비스는 여러 서비스 엔드포인트를 포함할 수 있습니다. 이 서비스에는 다음 서비스 엔드포인트가 포함되고, 그 아래의 모든 URI가 이 서비스 엔드포인트와 관련됩니다. https://www.googleapis.com	Google Drive API v3의 서비스 엔드포인트는 https://www.googleapis.com 입니다.

2단계: RAG 답변 생성

- VER2 랭그래프로 생성한 답변과 검색된 컨텍스트 수집
- 동일한 랭그래프 기반으로 QA와 원문이 답변한 response를 저장

3단계: 개별 답변 평가

- 평가 모델: gpt-4o-mini
- 평가 지표: 다음 4가지 평가 기준으로 1-5점 척도로 점수 평가
 - 응답 정확성 (Answer Correctness) : 답변이 정답과 얼마나 일치하는가?
 - 컨텍스트 관련성 (Context Relevance) : 답변이 검색된 문서에만 기반했는가?
 - 컨텍스트 충실성 (Context Faithfulness) : 검색된 문서가 질문에 얼마나 관련 있는가?
 - 컨텍스트 충분성 (Context Recall) : 검색된 문서가 답변에 필요한 정보를 모두 포함했는가?
 - total_score : 총점
 - analysis: 종합적인 분석 의견

- 4단계: QA vs 원문 비교 결과

* QA	* 원문
===== 평가 요약 =====	===== 평가 요약 =====
평균 총점: 15.55	평균 총점: 17.85
응답 정확성 평균: 3.40	응답 정확성 평균: 4.30
컨텍스트 관련성 평균: 3.85	컨텍스트 관련성 평균: 4.45
컨텍스트 충실성 평균: 4.65	컨텍스트 충실성 평균: 4.65
컨텍스트 충분성 평균: 3.65	컨텍스트 충분성 평균: 4.45

- QA: 근거 문서에 충실하게 답변하지만, 정확성과 충분성에서 낮은 점수를 받아 실제 사용자 질문에 완전하게 답변하지 못하는 한계가 존재한다.
- 원문: 전반적으로 균형 잡힌 답변을 제공하며, 정확성, 관련성, 충분성에서 모두 우수한 점수를 획득하였다.

4.3 결론

- 평가 결과, QA 기반 RAG가 잘 답변하는 유형과 원문 기반 RAG가 잘 답변하는 질문 유형이 다른 경우가 있었다.
- 따라서 질문마다 QA 벡터 DB와 원문 벡터 DB를 모두 검색하는 하이브리드 방식을 활용하기로 최종적으로 결정하였다.

4.2 랭그래프 버전별 평가

비교 조건

- 동일 합성 데이터셋
- 동일 평가 모델 : gpt-4.1-mini
- 동일 평가 지표

평가 대상

- VER1. 쿼리 분리 없이, 질문만 받아서 벡터 db 검색한 후 답변하는 langgraph
- VER2. 질문 받으면 질문 히스토리+질문을 합친 후 쿼리 정제 및 분리한 후, 정제된 쿼리로 벡터 db 검색 후 답변하는 langgrpah
- VER3. 버전 2에서 retriever만 메타 필터링 가능한 셀프 쿼리 리트리버로 변경 (그래프 구조는 VER2와 동일)

평가 결과

버전	Context Recall	Faithfulness	Factual Correctness
VER 1	0.9020	0.8936	0.6021
VER 2	0.9737	0.8887	0.6342
VER 3	0.8353	0.7927	0.5379

- VER1 → VER2
 - Context Recall 크게 개선 (0.9020 → 0.9737)
 - Factual Correctness도 개선 (0.6021 → 0.6342)
 - Faithfulness는 유사 수준 유지

- 쿼리 정제 및 분리 과정이 검색 품질 및 답변 정확성 개선에 크게 기여
- VER2 → VER3
 - Self-Query Retriever 도입 결과 오히려 성능 하락
 - Context Recall과 Faithfulness, Factual Correctness 모두 낮아짐
 - 쿼리 정제 + 분리 방식(VER2)이 현재 기준 최적 구조

4.3 RAGAS 평가 한계 및 개선 방향

문제점

1. 토큰 제한 문제 (VER3)
 - 평가 과정에서 토큰 한계로 인해 일부 데이터가 정상적으로 평가되지 못한 경우가 있음
 - result_df에 NaN 값으로 저장된 케이스가 다수 존재
 - 따라서 평균값이나 통계 수치가 실제보다 왜곡될 가능성 있음
2. 합성 데이터셋 품질 문제
 - RAGAS가 생성하는 합성 질문 중 일부가 비현실적이거나 불명확함
 - 예: “범위 많으면 동의 어렵나요?”와 같이 메타 필터링 적용이 불가능한 질문
 - 실제 검색 및 답변 검증에 적합하지 않은 데이터가 포함될 위험

개선 방향

- 재평가 필요
 - 토큰 제한이 없는 환경에서 다시 평가 수행
 - 데이터 손실 없는 상태에서 정확한 평균/분포 재산출 필요

- 합성 데이터셋 대신 사람이 직접 설계한 평가 데이터셋 생성 예정

5. 결론 및 개선사항

5.1 주요 성과

1. Google API 문서 챗봇

- 질문 분류(api/basic/none)와 답변 생성 정확도 우수
- 다중 API 질문 분리 및 처리, 재검색(fallback)까지 정상 동작

2. 음성 인식(Whisper)

- 음성입력에 대한 우수한 인식률

3. 이미지 처리(GPT-4o Vision)

- 사용자가 입력한 이미지에 대한 정확한 분석 결과 생성
- 연속 질문에서도 이미지 분석 결과가 반영됨

4. 오류 처리 및 예외 대응

- RAG 품질 평가(evaluate) 및 대체 쿼리 생성(generate_queries)을 통한 fallback 정상 동작

5.2 향후 계획

1. 사용자 경험 개선 : 답변 스타일·길이·속도 최적화
2. RAG 평가 및 성능 최적화 지속 (프롬프트, 검색방식, Langgraph 구조 최적화)
3. 배포 환경에서의 답변 속도 및 품질 최적화