

# 테스트 계획 및 결과 보고서

## 1. 프로젝트 목표 및 개요

- 본 프로젝트는 ESG 경영 관련 정보를 정확하게 제공하는 챗봇을 개발하는 것을 목표로 함.
- GPT-3.5-mini 모델을 ESG 모범규준으로 파인튜닝하고, ESG 우수기업의 경영 보고서를 Retrieval-Augmented Generation(RAG) 방식으로 결합하여 보다 신뢰성 있는 응답을 생성하는 챗봇을 개발함.

## 2. 테스트 계획

- **목적:** 다양한 프롬프트 전략에 따른 응답 품질 비교를 통해 최적의 방식 도출
- **대상:** ESG 챗봇
- **방법:** 동일한 질문을 세 가지 유형의 프롬프트에 적용하여 각각 3회씩 테스트를 수행함. 응답에 대해 정확성, 관련성, 구조화 수준을 기준으로 평가하였으며, 각 응답에서 환각(Hallucination) 발생 여부와 정보의 일관성 또한 함께 분석함.

## 3. 테스트 케이스

번호	프롬프트	내용
1	프롬프트 1	답변의 구조만 제공
2	프롬프트 2	답변의 구조 + 평가 기준에 대한 설명 포함
3	프롬프트 3	답변 구조 + 평가 기준 + few-shot 예시 포함

ESG 주제 특성에 맞게 구조화된 응답과 사례 기반 정보가 도출되도록 프롬프트를 설계함.

## 4. 테스트 결과 및 분석

- 프롬프트 1: 형식은 존재하지만 내용이 피상적이며, 실제 ESG 사례나 평가 지표와 연관된 정보는 부족함.

<pre>system_prompt = f"""너는 ESG 전문가야.  아래 제공된 문서 내용을 기반으로 질문에 답변해줘. 다음 규칙을 반드시 따라야 해:  1. 답변은 항상 구체적이고 상세해야 함 2. 가능한 한 많은 관련 정보를 포함해야 함 3. 출처와 맥락을 유지하면서 답변해야 함 4. 숫자, 통계, 구체적인 정책이나 프로그램이 있다면 반드시 포함 5. 답변은 다음 구조를 따라야 함:    - 핵심 답변 요약    - 구체적인 내용 설명    - 관련 정책이나 프로그램    - 성과나 목표 (있는 경우)  제공된 문서의 내용을 벗어나는 답변은 하지 말고, 모르는 경우 모른다고 해야 해.</pre>	<p>질문을 입력하세요 (종료하려면 'quit' 입력): 탄소배출량을 줄이기 위한 실천 방안은 어떤 것들이 있나요?</p> <p>제공된 문서 필드: {"section": {"\$eq": "Environment"}}</p> <p>답변:</p> <p>기업은 탄소배출량을 줄이기 위해, 생산 프로세스 개선 및 친환경 제품 개발을 통해 탄소 중립 목표에 수렴하고 이를 실천하고 있습니다. 실제로, 상무사면리는 유망한 배출 목표값을 2023년 4.3%에서 2023년 4.32로 낮춤 계획이며, 친환경 제품인 거품발 사면투를 활용해 이산화탄소 배출을 줄이고 있습니다.</p>
---	---

- 프롬프트 2: 평가 기준에 대한 설명을 제공하면서 ESG 보고서와의 연결성이 향상됨. 구체적인 기업 사례는 부족할 수 있으나 응답 신뢰도는 높아짐.

<pre>system_prompt = f"""너는 ESG 전문가야.  아래 제공된 문서 내용을 기반으로 질문에 답변해줘. 다음 규칙을 반드시 따라야 해:  1. 답변은 항상 구체적이고 상세해야 함 2. 가능한 한 많은 관련 정보를 포함해야 함 3. 출처와 맥락을 유지하면서 답변해야 함 4. 숫자, 통계, 구체적인 정책이나 프로그램이 있다면 반드시 포함 5. 답변은 다음 구조를 따라야 함:    - 핵심 답변 요약    - 구체적인 내용 설명    - 관련 정책이나 프로그램    - 성과나 목표 (있는 경우)  제공된 문서의 내용을 벗어나는 답변은 하지 말고, 모르는 경우 모른다고 해야 해.</pre>	<p>질문을 입력하세요 (종료하려면 'quit' 입력): 탄소배출량을 줄이기 위한 실천 방안은 어떤 것들이 있나요?</p> <p>제공된 문서 필드: {"section": {"\$eq": "Environment"}}</p> <p>답변:</p> <p>***핵심 답변 요약*** 탄소배출량 감축을 위해 다양한 실천 정책 강화, 재생에너지 전환 등 다양한 방안을 적용하고 있습니다.</p> <p>***구체적인 내용 설명*** 기업은 다양한 실천 정책을 추진하고, 재생에너지 전환을 구체화하거나 전속구매계약 체결에 적용에너지 비동일 확대하고 있습니다. 또한, 전가자 전환과 기후를 같이 하는 기후 변화 대응을 위한 ESG경영을 추진하고 있습니다.</p> <p>***관련 정책이나 프로그램*** KTH는 ESG 이니셔티브를 강화하여, ESG경영을 위한 ESG경영을 100% 달성할 목표로 하고 있습니다. 또한, ESG경영을 위한 ESG경영을 강화하고 있습니다.</p>
---	---

- 프롬프트 3: few-shot 예시 제공으로 인해 응답의 구조, 표현 방식, 사례의 질이 현저히 향상되었으며 환각 가능성이 낮아졌음. 정보의 깊이와 일관성에서 가장 우수한 성과를 보임.

<pre>system_prompt = f"""너는 ESG 전문가야.  아래 제공된 문서 내용을 기반으로 질문에 답변해줘. 다음 규칙을 반드시 따라야 해:  1. 답변은 항상 구체적이고 상세해야 함 2. 가능한 한 많은 관련 정보를 포함해야 함 3. 출처와 맥락을 유지하면서 답변해야 함 4. 숫자, 통계, 구체적인 정책이나 프로그램이 있다면 반드시 포함 5. 답변은 다음 구조를 따라야 함:    - 핵심 답변 요약    - 구체적인 내용 설명    - 관련 정책이나 프로그램    - 성과나 목표 (있는 경우)  제공된 문서의 내용을 벗어나는 답변은 하지 말고, 모르는 경우 모른다고 해야 해.</pre>	<p>질문을 입력하세요 (종료하려면 'quit' 입력): 탄소배출량을 줄이기 위한 실천 방안은 어떤 것들이 있나요?</p> <p>제공된 문서 필드: {"section": {"\$eq": "Environment"}}</p> <p>답변:</p> <p>***핵심 답변 요약*** 탄소배출량 감축을 위해 에너지 효율화와 재생에너지 확대 등 다양한 방안을 적용하고 있습니다.</p> <p>***ESG 핵심 지표 (KPI)***</p> <ul style="list-style-type: none"><li>- 온실가스 감축: Scope 1+2 배출량 2030년까지 50% 감축 목표</li><li>- 재생에너지 사용률: 2023년 45%, 2030년 목표 80%</li><li>(출처: KTH ESG 보고서, p.1-17)</li></ul> <p>***단계별 실행 로드맵***</p> <ul style="list-style-type: none"><li>- (0~6개월): 유틸리티 설비 고효율 전환, 폐열 회수 시스템 구축</li><li>- (7~12개월): 전기차로 영업 차량 전환, 직원 전력거래계약 확대</li></ul> <p>***요약*** 탄소배출 감축을 위해 에너지 효율화와 재생에너지 확대 등 다양한 실천 방안을 필요로 합니다.</p>
---	---

## 5. 성능 개선 시도 정리

- **프롬프트 엔지니어링 고도화:** 평가 기준과 사례를 포함한 few-shot 예시를 통해 보다 정제된 답변 유도
- **벡터 검색 정확도 개선:** ESG 보고서 데이터를 적절한 청크 단위로 분할하고 임베딩 모델 변경 실험 수행
- **환각 방지 강화:** few-shot 방법과 정답지 기반 응답 및 답변 내 근거 명시 유도하여 환각을 방지
- 

## 6. 향후 개선 방향

- **산업군 확대:** 다양한 산업군의 ESG 보고서 추가 수집 및 반영
- **실시간 정보 반영:** ESG 뉴스·규제 등 최신 정보 통합
- **맞춤형 추천 강화:** 기업 규모/산업에 따른 전략 자동 추천 기능 강화
- **max\_token:** 조금 더 자세한 답변을 만들기 위한 토큰 수 증가
- **메모리 기능 도입:** 이전 대화 내용을 기억하여 문맥 기반의 연속적 대화가 가능하도록 개선

## 7. 결론

본 테스트를 통해 ESG 챗봇의 응답 품질을 결정짓는 주요 요소로서 프롬프트 설계의 중요성을 확인함. 특히 평가 기준과 few-shot 예시를 포함한 구조화된 프롬프트는 환각을 줄이고 ESG 보고서를 기반으로 한 구체적이고 신뢰성 높은 응답을 유도하는 데 탁월한 효과를 보임. 향후 개선 방향으로 사용자 대화 맥락을 기억할 수 있는 메모리 기능 도입, 다양한 산업군 확장, 실시간 ESG 정보 통합, 기업 맞춤형 전략 추천 기능 강화, 응답의 깊이를 보장하기 위한 토큰 수 조정 등이 있으며, 이러한 고도화를 통해 실무 적용성과 사용자 만족도를 더욱 높일 수 있을 것으로 기대됨.