

ESG 이중 중대성 평가 서비스 개발 가이드라인

[0단계: RAG 인프라 및 지식 기반 구축]

0.1. 벡터 데이터베이스(Vector DB) 설정

- 도구: Pinecone 사용.
- 구조: 표준별/언어별 데이터 격리를 위해 네임스페이스 Namespace)를 다음과 같이 설계.
 - **gri-en**: GRI 영문 원본 자료.
 - **sasb-kr**: SASB 산업별 국문 기준서.
 - **issb-kr**: ISSB 국문 공시 표준.
 - **kssb-kr**: KSSB 국문 공시 표준.
- 임베딩 모델: Gemini 또는 OpenAI의 최신 임베딩 모델 적용.

0.2. 데이터 전처리 및 적재 (Knowledge Base)

- GRI (영문): Sector Standards 영문 원본 PDF 파싱. 특히 Section 2. Likely material topics와 그 하위 Topic Standards disclosures의 계층 구조를 보존하여 저장.
- SASB/ISSB/KSSB (국문): 한국어 번역본 PDF 파싱. 표(Table)와 핵심 조항 텍스트를 구조화하여 저장.
- 청킹(Chunking) 전략: 조항 단위 또는 500자 내외로 분할하여 검색 정확도 확보.

[1단계: 이슈풀(Issue Pool) 구축 - 데이터 추천 및 수집]

1.1. 표준 기반 지능형 추천 (RAG 활용)

- 기능: 유저가 선택한 산업군 키워드를 기반으로 0단계에서 구축한 RAG DB를 검색.
- 로직: 유사도가 높은 GRI/SASB 이슈를 추출하여 유저에게 후보군으로 제안.

1.2. 미디어 분석 에이전트 (News Scraping)

- 데이터 수집: 유저에게 키워드 입력을 받아 Google Search API 를 연동하여 최근 1년치 관련 뉴스 수집.
- AI 라벨링 (Gemini 2.5 Flash): 수집된 뉴스 본문을 읽고 PRD 2.3항의 ‘이슈배치’에 첨부된 정규화 기준 표(노출 빈도 및 부정 맥락 비중)에 따라 1~5점 점수 부여.

1.3. AI 보고서 파서 (PDF Parser)

- 기능: 유저가 자사의 전년도 보고서나 경쟁사의 ESG 보고서(PDF) 내 ‘중요성평가(중대이슈)’ 페이지를 업로드하면 기존 중대 이슈 리스트를 추출.

[2단계: 설문조사 시스템 - 주관적 데이터 수집]

2.1. 이해관계자 관리 및 맞춤형 발송

- 그룹 분류: '내부(경영진/실무진)'와 '외부(고객/주주/NGO 등)' 필수 구분.
- 설문 UI: 비관여자도 이해하기 쉬운 직관적인 문구(PRD 2단계 참고)를 적용한 리커트 척도 폼 구축.

2.2. 실시간 데이터 적재

- 자체 구축: 응답 즉시 DB에 기록하여 외부 플랫폼 연동 없이 실시간 분석 연계.

[3단계: 매트릭스 분석 및 결과 도출 - 데이터 통합 및 증빙]

3.1. RAG 기반 자동 스코어링

- 객관적 점수 산출: 이슈별로 RAG DB를 재검색하여 해당 이슈가 표준 문서의 어떤 섹션(Core/Additional 등)에 위치하는지 판별 후 1~5점 자동 스코어링.
- 최종 산식 적용: 설문 점수(6: 가중치 평균)와 RAG/미디어 점수(4)를 합산하여 최종 X, Y 좌표 확정.

3.2. 이중 중대성 매트릭스 시각화 (Traceability)

- 차트 구현: Recharts 기반 Scatter Chart.
- 추적성(Traceability) 기능: 차트 점 클릭 시 **RAG가 참조한 표준 문서 원문 조항(영문/국문)**과 뉴스 근거를 하이라이트하여 팝업 노출.

[4단계: 기술 인프라 및 보안]

4.1. AI 모델 활용: 데이터 분류(Flash 모델), 최종 추론 및 리포팅(Pro 모델).

4.2. 저장소: 이슈풀 및 설문 데이터 보관을 위한 관계형 DB(PostgreSQL 등).

4.3. 인증: 기업별 데이터 격리를 위한 로그인 시스템.