



CLIKCA

RAG 기반 업무 보조 AI 비서

SK Networks Family AI Camp 13기

1팀: 남궁건우, 이명인, 우지훈, 홍채우, 김승호

프로젝트 개요

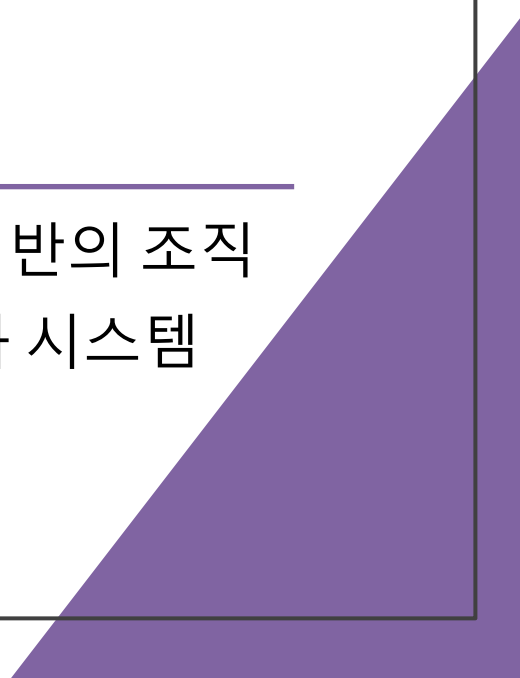
- 단순 검색을 넘어, 문서·회의·일정·메일까지 통합 관리하는 올인원 AI 비서
- 서비스 슬로건: '한 번의 클릭, 모든 업무 완결'
- 핵심 기능: 문서요약, 일정관리, 이메일, To-Do, 초안작성, QnA

추진 배경

신입/부서이동자의 50.2%가
업무 적응에 최소 3개월 소요

문제점: 정보 단편화, 검색 한
계, 반복 작업 증가

필요성: RAG+LLM 기반의 조직
내 정보 통합·자동화 시스템



서비스 개요

서비스명: CLIKCA (Click + Assistant)

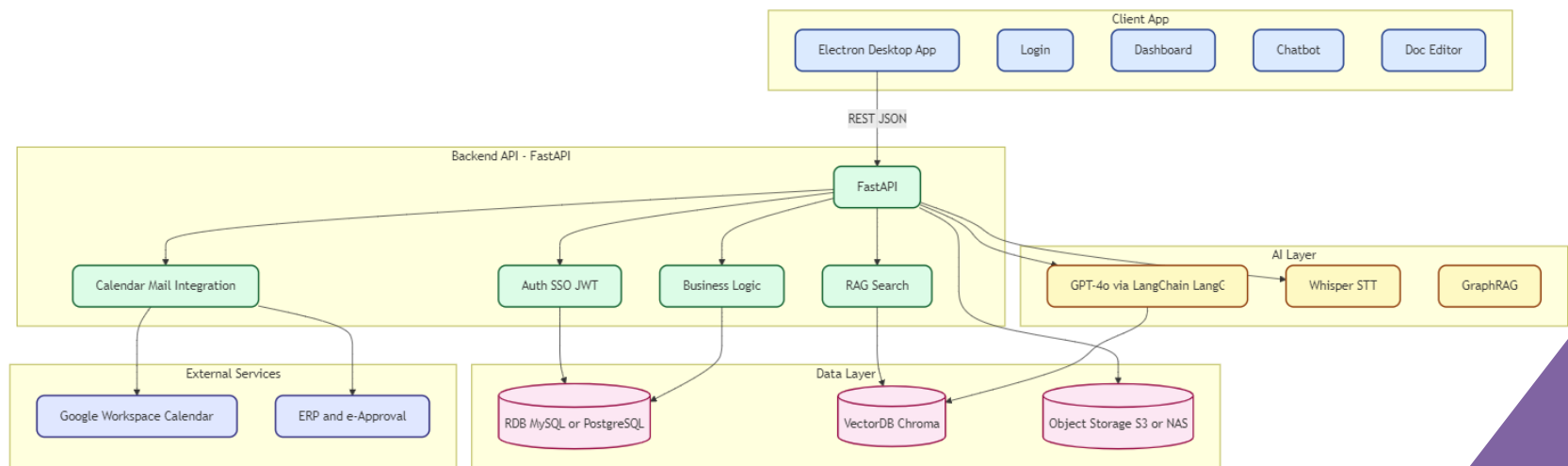
차별성: 의미검색+요약,
API 양방향 연동

활용 시나리오: 규정 질의
, 회의록→일정, 초안 작성

주요 기능

- **회의록/문서 요약:** STT → 요약.액션아이템
- **일정관리:** 자동 추출.등록.충돌감지
- **이메일 생성:** 템플릿+자동발송
- **To-Do 브리핑:** 태스크 추출.리포트
- **문서 초안 작성:** 유사 사례+템플릿 기반
- **업무 QnA:** RAG 기반 검색.원문 링크
- **대시보드:** 일정.메일.문서 통합

시스템 아키텍처



로고

Sign-in

관리자

사용자

2045781352



☒ 아이디 저장

로그인

[아이디 찾기](#) | [비밀번호 찾기](#)

로고

어서오세요 홍길동님!

구글 계정을 연동하세요



로그인 화면

- 사용자와 관리자 계정 분리
- ID 및 PW 찾기 기능
- 최초 Login시 구글계정 연동
- 로그인 시 ID를 하드디스크에 저장하여 재부팅 시에도 마지막 접속 ID 유지

아이디 찾기

이메일을 입력하세요

코드 요청

인증 코드를 입력하세요

코드 확인

바꿀 아이디를 입력하세요

취소

저장

비밀번호 찾기

이메일을 입력하세요

코드 요청

인증 코드를 입력하세요

코드 확인

새 비밀번호

새 비밀번호 확인

취소

저장

ID 및 PW 찾기

- 연동된 구글 계정으로
인증메일 발송
- 인증코드 입력 후
계정정보 변경



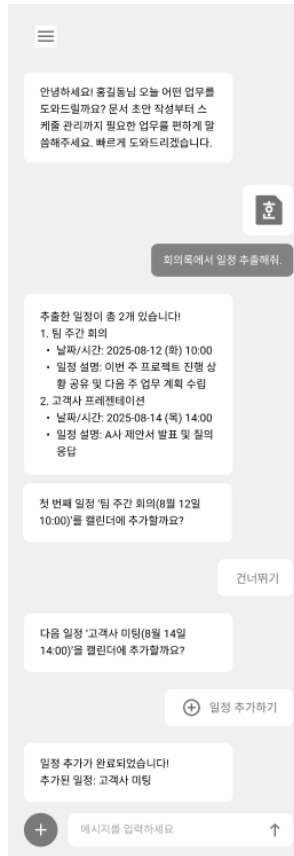
무엇이든 물어보세요.

+ 메시지를 입력하세요



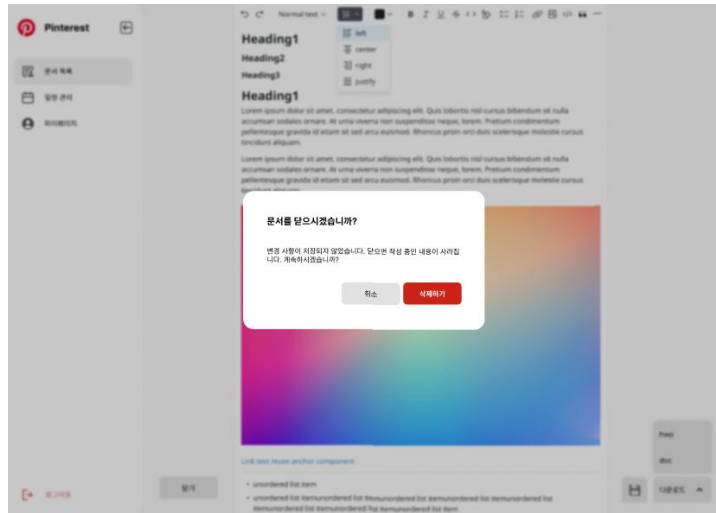
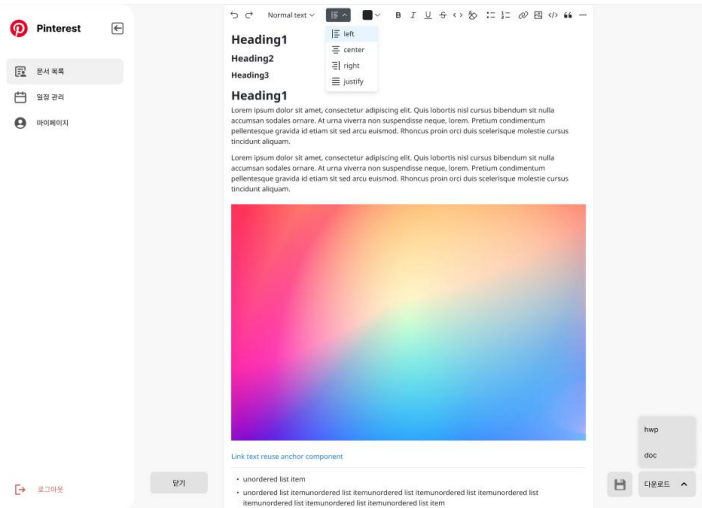
ChatBot

- GPT 모델과 연동하여 질문가능
- 문서 초안 작성 요청
- 회의록 업로드 요청
- 일정 등록 요청



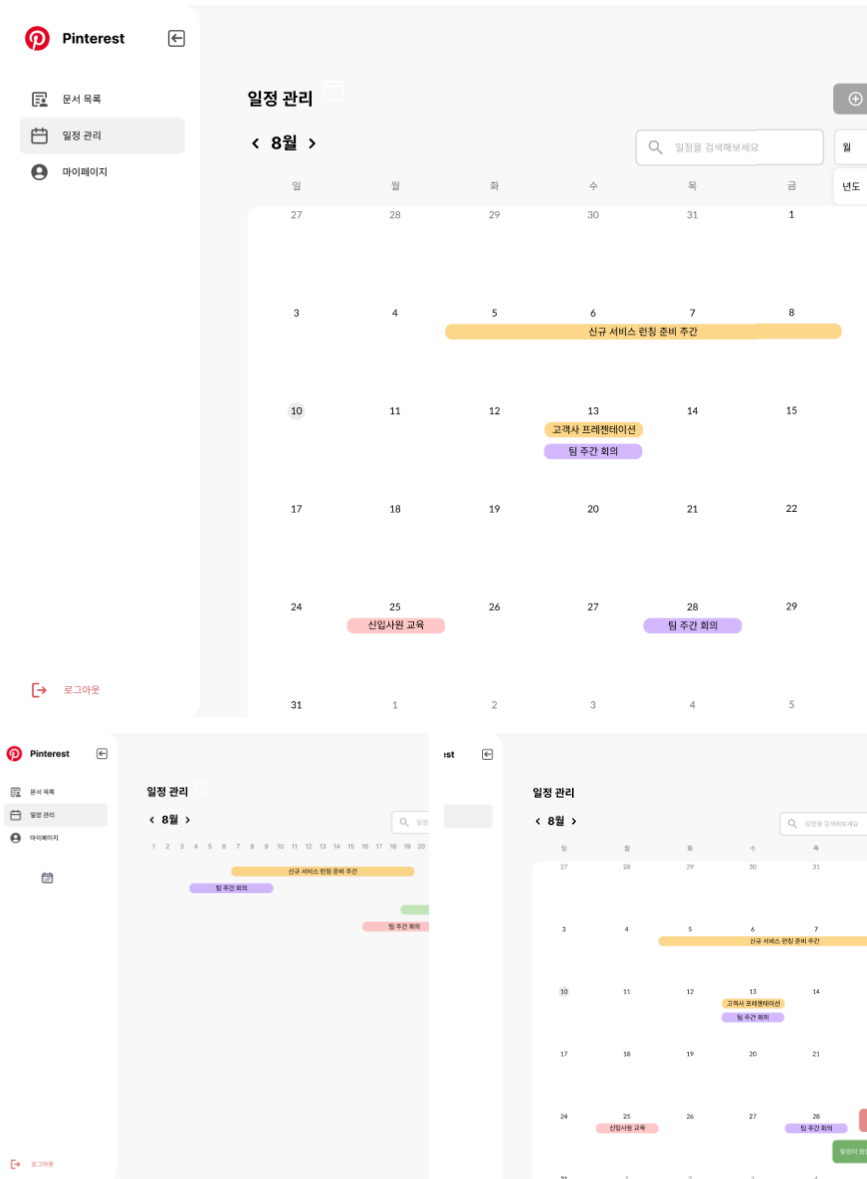
ChatBot

- 문서초안 작성 요청
- 회의록 업로드 요청
- 일정 등록 요청



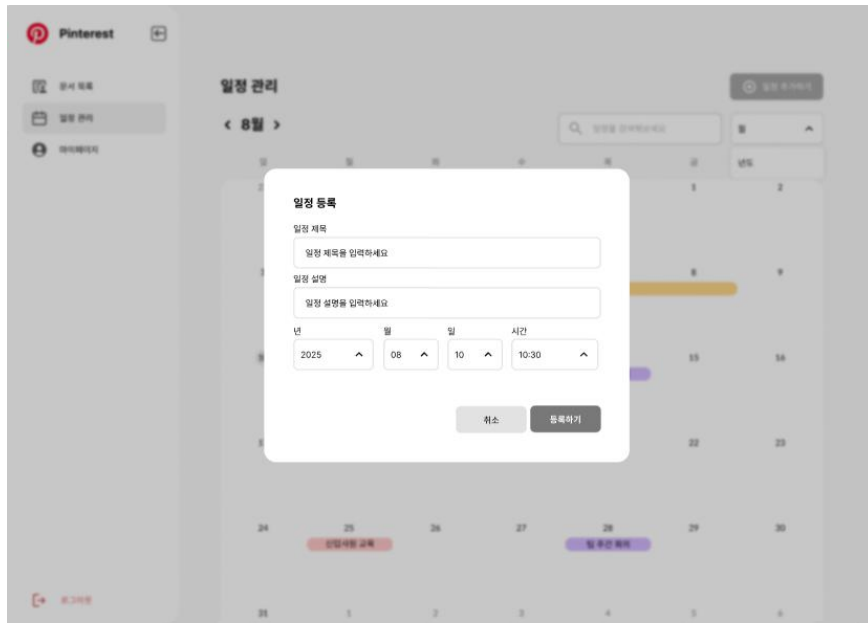
문서 편집

- 한글 파일 편집
- ChatBot을 통해양식에 맞는 초안 작성
- 사용자가 실시간 수정 가능



캘린더

- 구글 캘린더와 연동
- 개인 일정 등록 및
사내 일정 등록
- 간트 차트 형태 및
달력 형태 가능

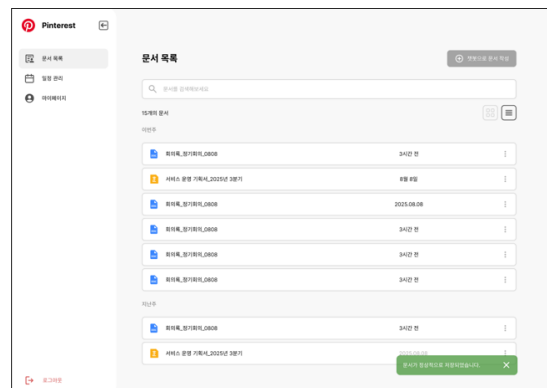
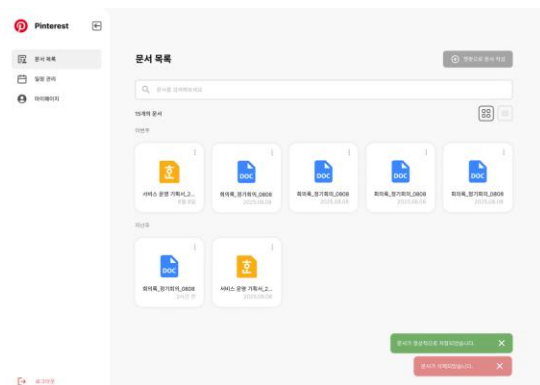
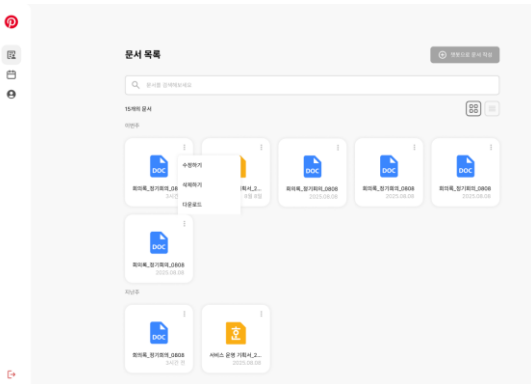


캘린더

- 일정 등록
- 수정 및 삭제

문서 관리

- 열람한 문서 및 생성한 문서 목록
- 수정, 저장, 삭제 가능
- 리스트 형식 및 카드 형식 선택 가능



사원 목록

15명의 사원

부서 선택

직급 선택

검색할 사원 이름을 입력하세요

사원 계정 추가

#	이름	부서	직급	이메일	
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	우벳베	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮
1	홍길동	인사부	팀장	rfehd0101@gmail.com	⋮

사원 정보가 정상적으로 수정되었습니다.

사원 정보가 복사되었습니다.

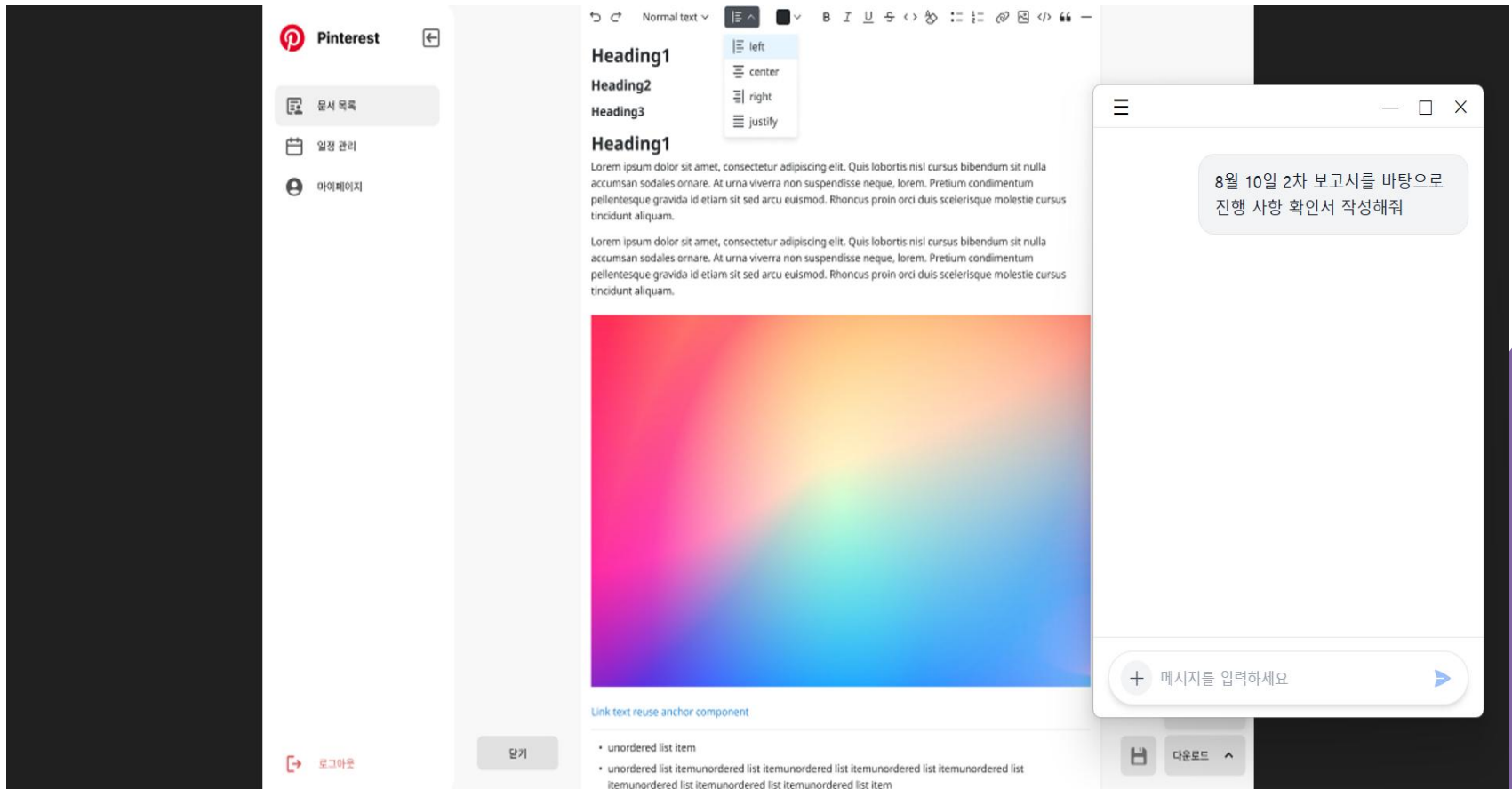
사원 정보가 정상적으로 등록되었습니다.

사원 정보가 삭제되었습니다.

관리자 페이지

- 사원 정보 수정, 삭제

사용 예시



진행 현황

- 분석·설계 단계: 100% 완료
- 데이터 수집·전처리: 80% 완료
(전처리 보고서 성과 포함)
- 프론트엔드: 화면 설계 완료,
일부 구현 중
- 백엔드: DB설계·RAG구현 착수
- 남은 과제: 백엔드 API 완성,
통합 테스트

PDF 문서의 Vector DB 구축

- 전처리 목표

PDF 문서의 텍스트, 표,
이미지 정보를 손실 없이
추출해 순수 텍스트로
변환하고 모든 비텍스트
데이터를 AI로 실제 텍스트
정보로 치환하는
파이프라인을 구축

1단계: 데이터 추출 및 구조화

PyMuPDF, camelot과 같은 라이브러리를 사용하여 PDF 문서를 구조적으로 분해 및 핵심 데이터 추출



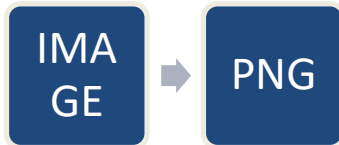
텍스트(Text) -> 마크다운 (.md):

문서의 기본 텍스트를 머리글/바닥글과 같은 불필요한 내용을 제외한 순수한 본문 내용만 페이지별로 추출



표(Table) -> CSV :

페이지 내 표를 탐지하여 그 구조를 그대로 보존하는 CSV 파일로 추출



이미지 -> PNG :

차트, 다이어그램, 사진 등 의미 있는 시각 정보들을 PNG파일로 저장

추출이 완료되면, 마크다운 본문에는 [표: page_1_table_1.csv]나 [이미지: page_1_image_1.png]와 같이 추출된 자산 파일의 경로가 명시적으로 연결되어, 다음 단계에서 AI가 처리할 대상과 그 위치를 명확히 합니다.

2단계: AI 기반 자동 서술 및 통합

이 단계에서는 1단계에서 추출한 개별 데이터들을 GPT-4o의 멀티모달(Multi-modal) 능력을 이용해 완전한 텍스트로 변환합니다.

2.1 텍스트 교정 및 복원

PDF에서 추출된 텍스트는 종종 단어가 깨지거나 오타자가 포함됩니다. AI는 이러한 오류를 수정하여 문장의 완성도를 높입니다.

•처리 방식: 깨진 텍스트 덩어리를 AI에 전달하며, "문맥에 맞게 자연스러운 문장으로 복원하되, 원본의 의미를 절대 변경하지 말라"고 지시합니다.

2.2 표 데이터 서술

단순히 표 데이터를 나열하는 것이 아닌, AI가 표의 시각적 맥락까지 이해하여 자연스러운 설명으로 변환합니다.

•처리 방식: camelot으로 추출한 CSV 데이터와 함께, 표가 포함된 페이지 전체의 이미지를 GPT-4o에 동시에 제공합니다. AI는 이미지를 보고 표의 제목과 단위를 파악한 후, CSV 데이터를 기반으로 내용을 서술합니다.

•핵심 프롬프트 지시사항:

- 1.페이지 전체 이미지에서 표의 제목과 단위를 찾아 명시할 것.
- 2.각 행의 데이터를 기계적으로 나열하지 말고, 완전한 문장으로 설명할 것.
- 3.원본에 없는 내용을 절대 분석, 요약, 추론하지 말 것.

•변환 예시:

[참고] 2024년 경영지침 체크리스트

연번	구분	개수
1	공공기관의 운영에 관한 법률 및 시행령	3
2	공기업 준정부기관의 경영에 관한 지침	60
3	공공기관의 혁신에 관한 지침	27
4	공기업 준정부기관 임원보수지침	4
5	공기업 준정부기관 예산운용지침	51
6	공공기관의 안전관리에 관한 지침	13
7	공기업 준정부기관 계약사무규칙	7
8	공공기관의 통합공시에 관한 기준	6
9	국민권익위원회 권고사항	4
10	대민업무 불공정관행 및 규제사항	6
계		181

표 내용: **제목: [참고1] 2024년 경영지침 체크리스트**

표의 첫 번째 항목은 '공공기관의 운영에 관한 법률 및 시행령'으로, 이에 해당하는 개수는 3입니다. 두 번째로는 '공기업 준정부기관의 경영에 관한 지침'이 있으며, 이 지침의 개수는 60입니다. 세 번째 항목은 '공공기관의 혁신에 관한 지침'으로, 해당 개수는 27입니다.

네 번째 항목은 '공기업 준정부기관 임원보수지침'이며, 이에 해당하는 개수는 4입니다. '공기업 준정부기관 예산운용지침'은 다섯 번째에 위치하며, 개수는 51입니다. 여섯 번째로는 '공공기관의 안전관리에 관한 지침'이 있으며, 그 개수는 13입니다.

일곱 번째 항목은 '공기업 준정부기관 계약사무규칙'으로 되어 있으며, 개수는 7입니다. 표의 여덟 번째 항목은 '공공기관의 통합공시에 관한 기준'으로 개수는 6입니다. 아홉 번째로 명시된 '국민권익위원회 권고사항'의 개수는 4입니다. 마지막으로 열 번째 항목은 '대민업무 불공정관행 및 규제사항'이며, 그 개수는 6입니다.

전체 항목의 개수 합계는 181입니다.

2.3 이미지 정보 서술

이미지에 담긴 시각 정보를 객관적인 텍스트로 변환합니다.

•처리 방식: 이미지 파일을 Base64로 인코딩하여 GPT-4o에 전달합니다. 이때 AI는 이미지가 정보 전달 목적(차트, 다이어그램 등)인지, 단순 장식용(로고, 배경 등)인지 먼저 판단합니다.

•핵심 프롬프트 지시사항:

- 1.의미 없는 장식용 이미지는 '배경 이미지'라고만 응답할 것.
- 2.차트의 경우 제목, 축, 단위, 범례, 데이터 포인트를 포함하여 설명할 것.
- 3.사진이나 다이어그램은 보이는 그대로의 객체, 텍스트, 상황을 묘사할 것.
- 4.절대 주관적인 해석이나 결론을 내리지 말 것.

•변환 예시:



< 홈페이지 이해관계자 탭 화면 >

< 홈페이지 이해관계자 탭 화면 >

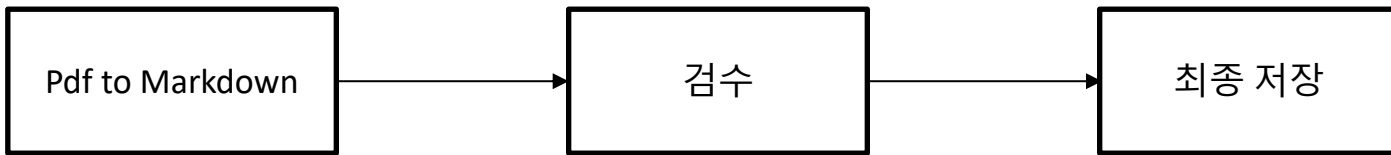
이미지 내용: 이 이미지는 이해관계자를 주제로 한 정보 다이어그램입니다. 상단에 "이해관계자"라는 제목이 있으며, 하단에는 "kobaco"가 추구하는 다양한 가치와 가치유형별 이해관계자들의 현황을 수록하였습니다."라는 문장이 있습니다.

이미지는 크게 두 부분으로 나눌 수 있습니다. 첫 번째 부분에서 가치유형별 핵심 키워드를 설명하고 있으며, 네 가지 카테고리로 나누어져 있습니다. 각각의 카테고리에는 아이콘과 함께 관련 키워드가 나열되어 있습니다.

1. 정책적 가치: 저울 모양의 아이콘이 있으며, 하단에 "투명, 책임, 공공성, 신뢰, 협력"이라는 키워드가 나열되어 있습니다.
2. 경제적 가치: 손바닥과 더하기 기호가 있는 아이콘이 있으며, "효율, 협력, 책임, 고객중심, 신뢰"라는 키워드가 있습니다.
3. 사회적 가치: 지구와 사람 모양의 아이콘이 있으며, "공정, 투명, 공익, 공감, 신뢰"라는 키워드가 포함되어 있습니다.
4. 기술적 가치: 원자 모양의 아이콘이 있으며, "혁신, 참여, 신뢰, 개방성, 협력"이라고 나와 있습니다.

두 번째 부분에서는 "가치유형별 이해관계자"라는 제목 아래에 네 가지 가치유형에 따른 이해관계자가 나열되어 있습니다.

- **결론 및 향후 계획**
- 이러한 '추출 → 분석 → 자동 서술'의 2단계 파이프라인을 통해, 저희는 원본 PDF의 정보 손실을 최소화하면서도, 기계가 이해하고 처리할 수 있는 **순수 텍스트 데이터**를 확보하는 기반을 마련했습니다.
- 다만, 아무리 정교한 AI 모델이라도 복잡한 문서에서는 일부 내용이 누락되거나 부정확하게 변환될 가능성이 여전히 존재합니다. 따라서 데이터의 신뢰도를 100%로 끌어올리기 위해, 저희는 다음과 같은 '**Human-in-the-loop(인간 참여형)**' 방식으로 최종 프로세스를 확정하고자 합니다.



1차 변환 : 현재 구축된 파이프라인을 이용해 PDF를 마크다운 파일로 신속하게 자동 변환합니다.

2차 검수 : 변환된 마크다운 파일을 직접 검수하여, AI가 놓쳤을 수 있는 미묘한 맥락이나 오류를 수정하고 내용의 완전성을 보장합니다.

최종 저장 : 사람의 검수를 거쳐 최종 확정된 고품질의 텍스트 데이터만을 벡터DB에 저장합니다. 이러한 접근은 AI의 속도와 효율성, 그리고 인간의 정확성과 깊이 있는 이해를 결합하여 가장 신뢰도 높은 데이터베이스를 구축하기 위한 최적의 전략입니다.

향후 계획

- 8월 중순: 백엔드·프론트 연동
- 8월 말: 내부 테스트 및 오류 수정
- 9월 초: 최종 발표·시연