발간등록번호
11-1741050-100023-01

연구결과보고서

AI를 활용한 비공개기록물 공개재분류 및 비식별화 서비스 모델 연구

A Study on an AI-Enabled Service Model for Access Reclassification and Personal Data De-identification of Restricted Government Records for Public Release

주관연구기관 : ㈜가온아이

국가기록원

주 의

1. 이 보고서는 국가기록원에서 시행한 용역연구개발사업의 연구결과보고서입니다.

2. 이 보고서 내용을 발표할 때에는 반드시 국가기록원에서 시행한 용역연구개발사업의 연구결과임을 밝혀야 합니다.

3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 아니 됩니다.

편집순서 2 : 제출문

제 출 문

국가기록원장 귀 하

이 보고서를 “AI를 활용한 비공개기록물 공개재분류 및 비식별화 서비스 모델 연구((주)가온아이/손기영)” 과제의 연구결과보고서로 제출합니다.

2025. 11. 25.

주관연구기관명 : ㈜가온아이

주관연구책임자 : 손 기 영

편집순서 3 : 목차

목 차

편집순서 4 : 요약문

연구결과보고서 요약문

연구과제명	AI를 활용한 비공개기록물 공개재분류 및 비식별화 서비스 모델 연구		
중심단어	AI, 공개재분류, 개인정보 비식별화		
주관연구기관	㈜가온아이	**주관연구책임자**	손 기 영
연구기간	2025 . 4 . 29 - 2025 . 11. 25		

본 연구는 국가기록원의 공개재분류 및 비식별화 업무에 인공지능(AI)을 적용함으로써, 기록관리 업무의 효율성과 신뢰성을 동시에 향상시킬 수 있음을 입증하였다.

우선, 공개재분류 업무는 법령·정책·사회적 가치가 복합적으로 작용하는 영역으로, 기존 AI 분류기(Classifier) 모델의 고정된 패턴 학습 방식만으로는 정확한 판단이 어려움을 확인하였다. 이에 본 연구는 RAG(Retrieval Augmented Generation) 기반 AI 모델을 도입하여, 과거 재분류 사례 및 관련 판단기준을 지식저장소(Knowledge Base) 에서 검색·활용하는 구조로 구축하였다. 그 결과, 모델은 근거 중심의 판단(Explainable Decision)이 가능해졌으며, 기준 변경에도 재학습 없이 즉시 대응할 수 있는 유연성을 확보하였다.

또한, 공공기록물의 비식별화 과정에서는 정규식·규칙·문맥·LLM 검증을 결합한 하이브리드 검출 구조를 적용하였다. 숫자형 정보(주민등록번호·계좌번호 등)는 다단계 검증 파이프라인으로, 문자형 정보(이름·주소·기관명 등)는 LLM 프롬프트 기반 문맥 탐지 방식으로 처리하여 정확도를 향상시켰다. 이를 통해 검출된 비식별화 대상 정보는 공개재분류 과정과 연계되어, 이용자 정보는 제외하고 자동 마스킹하는 이용자 맞춤형 비식별화 기록물 제공 기능이 구현되었다.

이를 통해 AI 모델을 활용한 공개재분류 및 비식별화 대상 정보 검출이 가능하며, 그 결과를 열람·활용 단계에서 비식별화 처리된 기록물로 안전하게 제공할 수 있음을 확인하였다.

향후 발전 방향으로는 문서요약·판단·문맥이해·비식별화 등 업무 특성별로 최적화된 AI 모델을 병렬로 활용하는 멀티 LLM 오케스트레이션(Multi-LLM Orchestration) 구조를 제안하였다. 다양한 LLM의 특성을 분석·조합함으로써 공개재분류에 특화된 AI 에이전트 구성이 가능하며, 최신 LLM의 도입을 통해 성능과 적용 범위가 한층 확대될 것으로 기대된다.

편집순서 5 : 요약문(영문)

Summary

Title of Project	A Study on an AI-Enabled Service Model for Access Reclassification and Personal Data De-identification of Restricted Government Records for Public Release		
Key Words	AI, public reclassification, de-identification of personal information		
Institute	KAONi Co., Ltd.	**Project Leader**	Sohn Kiyoung
Project Period	2025 . 4 . 29 - 2025 . 11. 25		

This study demonstrated that applying Artificial Intelligence (AI) to the National Archives' declassification and de-identification processes can significantly enhance both the efficiency and reliability of records management operations.

First, it was confirmed that the declassification process involves a complex interplay of legal, policy, and social value judgments, making it difficult to achieve accurate results through conventional AI classifier models based solely on fixed pattern learning. To address this limitation, the study adopted a Retrieval Augmented Generation (RAG)-based AI model, which retrieves and utilizes past reclassification cases and decision criteria from a Knowledge Base. As a result, the model was able to provide evidence-based and explainable decisions and demonstrated structural flexibility that allows it to adapt immediately to changes in classification criteria without retraining.

In addition, for the de-identification of public records, a hybrid detection framework combining regular expressions, rule-based validation, contextual analysis, and LLM-based verification was applied. Numerical information (e.g., resident registration numbers, account numbers) was processed through a multi-stage validation pipeline, while textual information (e.g., names, addresses, organization names) was detected using LLM prompt-based contextual analysis, improving overall accuracy. The detected de-identification targets were linked with the declassification process, enabling the implementation of a user-customized de-identified records provision function that automatically masks personal information when accessed.

Through this approach, it was verified that AI models can effectively perform both declassification and de-identification, and that the resulting records can be safely provided in a de-identified form during access and utilization stages.

For future development, the study proposes adopting a Multi-LLM Orchestration structure that applies optimized AI models in parallel according to task characteristics such as document summarization, decision-making, contextual understanding, and de-identification. By analyzing and integrating the capabilities of various LLMs, it will be possible to configure specialized AI agents for declassification tasks. Furthermore, with the continuous advancement of high-performance LLMs, the system's performance and scope of application are expected to expand significantly.

편집순서 6 : 총괄연구과제의 연구결과

총괄연구개발과제 연구결과

제1장 총괄연구개발과제의 최종 연구개발 목표

1.1. 총괄연구개발

1.1.1. 연구개발의 배경 및 필요성

우리나라는 아시아에서 첫 번째로 정보공개를 법으로 제정한 국가이다. 1996년 12월에 제정된「공공기관의 정보공개에 관한 법률」은 1998년 1월부터 2025년 현재까지 27년째 시행되고 있다. '국민의 알권리를 보장하고 더 많은 정보를 바탕으로 국정운영에 대한 참여를 유도하기 위한' 이 법을 통해 정보공개 청구 제도가 정착되었다. 정보공개에 대한 수요는 최근 10년 동안 큰 폭으로 증가하였다. 2024년 기준, 정보공개포털을 통해 원문을 내려받은 수는 총 4,443,239건[1]이다. 정보공개포털이 개편되었던 2014년의 원문 다운로드 수는 253,310건이었다[2].

증가하는 정보공개 수요에 따라 각 공공기관에서는 「공공기록물 관리에 관한 법률」 제35조에 근거해 5년 주기로 기록물 공개재분류를 수행하고 있다. 국가기록원에서는 5년 주기 재분류 대상량이 매년 약 244만건에 달한다. 원내 소장량 전체 중 약 65.9%가 주기적 재분류 대상에 해당한다(국가기록원 서비스정책과, 2025, 27). 게다가 매년 신규 기록물이 등록되기 때문에 재분류 대상량은 매년 누적되는 업무 구조이다. 이 같은 규모와 주기성 때문에 기존 인력만으로는 누적되는 재분류 기록물을 적시에 처리하기 어렵고, 업무 프로세스의 지능화와 자동화가 필요하다.

기록물 열람 서비스를 제공하기 전에 꼭 확인해야 하는 것 중 하나는 개인정보의 유무이다. 정보공개는 국민의 알 권리를 보장하기 위한 제도이지만, 동시에 사생활 보호와 국가안보, 기업 영업비밀 보호 등의 가치를 함께 고려해야 한다. 정보공개제도에는 「개인정보 보호법」이 병행 적용된다. 개인정보를 가리기 위해 마스킹하는 과정을 거쳐야 청구인에게 열람 서비스를 제공할 수 있다. 하지만 단순 반복적인 마스킹 업무로 업무 적체가 발생하고 있어 신속한 정보제공에 어려움이 있다.

---

1) 중앙행정기관, 지방자치단체, 교육청, 공공기관의 원문다운로드 수를 합산하였다.
2) https://www.open.go.kr/infOthbc/numberInfoOthbc/orgDwld.do 정보공개포털 참조. (2025년 08월 29일 최종 접속).

이는 비단 국가기록원만의 고민은 아닐 것이다. 각급 기관에서도 공개재분류, 정보공개를 수행한다. 국가기록원에서 업무 프로세스를 개선하고 시범 모델을 제시하고, 각 기관이 해당 모델을 각자의 환경에 맞게 도입한다면, 국가 단위의 예산 절감 효과를 기대할 수 있다.

공개재분류 및 개인정보 비식별화 업무를 자동화하려는 국가기록원의 노력은 이전에도 있었다. 대표적으로 2020, 2021년의 국가기록원 연구용역이 있다. 2020년에는 “전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 적용방안 연구”, 2021년에는 “전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 적용”을 수행한 바 있다(국가기록원, 2020 ; 2021). 해당 연구를 통해 공개재분류 자동화에 AI 적용 가능성을 충분히 확인하였다. 2023년에는 한국지능정보사회진흥원(NIA)과 국가기록원에서 “국가기록물 대상 초거대 AI 학습을 위한 말뭉치 데이터 구축” 사업을 통해 말뭉치 데이터 세트 및 질의응답 데이터를 구축하였다(AI Hub, 2023). 해당 사업을 통해 국가기록물 관리에 AI를 적용할 토대를 마련하였다. 이러한 흐름의 연장선에서 이번 연구에서는 AI를 통해 실무에 도움이 되는지 기술적 검증을 하고자 한다.

1.1.2. 연구개발의 목표

본 연구개발의 목표는 아래와 같다.

- AI를 활용한 지능형 공개재분류 수행 모델 연구 개발
- 비공개대상정보의 자동식별 및 마스킹을 통한 대국민 열람 제공 모델 연구 개발

이에 따른 연구질문은 다음과 같다.

- 국가기록원의 공개재분류 업무에 가장 적합한 AI 모델은 무엇일까?
- AI 공개재분류 및 비공개대상정보 비식별화 모델을 실무에 적용하기 위해 어떤 작업이 필요할까?
- AI 공개재분류 및 비공개대상정보 비식별화 모델의 기계 학습을 위해서는 기록 정보 데이터를 어떻게 구성하여야 할까?
- 각급 기록물관리기관으로 확산 및 보급하기 위한 과제는 무엇일까?

1.2. 총괄연구개발과제의 목표 달성도

○ 일정

- 2025. 04. 29 ~ 2025. 11. 25(총 210일)

○ 진도

- 2025년 11월 25일 기준 진척율 100%

연 구 개 발 과 제	진척율	최종 진척율
○ 착수	100%	100%
○ 요구분석	100%	100%
○ “공개재분류 업무 지능화”를 위한 업무 분석 및 기술동향 조사	100%	100%
- 업무 프로세스 분석, AI 도입 분야 진단	100%	100%
- AI 기술 검토	100%	100%
- 국내외 유사 사례 조사	100%	100%
- 종합 분석 및 보고서 작성	100%	100%
○ “비공개대상정보의 검출·비식별화, 대국민 제공”을 위한 기술동향·사례 조사	100%	100%
- 비공개정보 자동 검출 기술동향 조사	100%	100%
- 국내외 시스템 구축 사례 조사	100%	100%
- 종합 분석 및 보고서 작성	100%	100%
○ 데이터 수집 및 정제 방안 마련	100%	100%
- AI 모델 학습을 위한 데이터 구성 및 정제 프로세스(안)마련	100%	100%
- 공개재분류/비공개정보 식별 관련 파인튜닝을 위한 학습데이터 수집, 정제, 구축	100%	100%
○ AI를 활용한 “공개재분류 지원 모델” 연구 개발	100%	100%
- 모델 설계	100%	100%
- 모델 학습, 모델 평가, 모델 개선 및 튜닝	100%	100%
- 모델 배포	100%	100%
○ “이용자 맞춤형 비식별화 기록물 제공 기능” 연구 및 개발	100%	100%
- 분석	100%	100%
- 설계	100%	100%
- 구현	100%	100%
- 시험	100%	100%
○ AI를 도입한 공개재분류부터 비공개대상정보를 식별·비식별화하여 대국민 서비스 제공까지의 단계를 구현한 PoC 개발 및 시범운영	100%	100%
- 환경 구성 및 배포	100%	100%
- 시범 운영	100%	100%
○ 기록관리시스템 기능 연계 제안	100%	100%
- 기록관리시스템(CAMS, RMS)에 기능을 구현할 수 있는 설계 제안	100%	100%
- 각급 기록물관리기관으로 확산 보급 방안 제안	100%	100%
- 협의체 구성, 운영(관련 분야 교수, 기술진)	100%	100%
전 체	100%	100%

(표 1 - The Progress Research Development)

1.3. 총괄연구개발과제의 추진 일정

연 구 개 발 과 제	사 업 기 간													
	5월		6월		7월		8월		9월		10월		11월	
○ 착수														
○ 요구분석														
○ “공개재분류 업무 자능화”를 위한 업무 분석 및 기술동향 조사														
- 업무 프로세스 분석, AI 도입 분야 진단														
- AI 기술 검토														
- 국내외 유사 사례 조사														
- 종합 분석 및 보고서 작성														
○ “비공개대상정보의 검출·비식별화, 대국민 제공”을 위한 기술동향·사례 조사														
- 비공개정보 자동 검출 기술동향 조사														
- 국내외 시스템 구축 사례 조사														
- 종합 분석 및 보고서 작성														
○ 데이터 수집 및 정제 방안 마련														
- AI 모델 학습을 위한 데이터 구성 및 정제 프로세스(안)마련														
- 공개재분류/비공개정보 식별 관련 파인튜닝을 위한 학습데이터 수집, 정제, 구축														
○ AI를 활용한 “공개재분류 지원 모델” 연구 개발														
- 모델 설계														
- 모델 학습, 모델 평가, 모델 개선 및 튜닝														
- 모델 배포														
○ “이용자 맞춤형 비식별화 기록물 제공 기능” 연구 및 개발														
- 분석														
- 설계														
- 구현														
- 시험														
○ AI를 도입한 공개재분류부터 비공개대상정보를 식별·비식별화하여 대국민서비스 제공까지의 단계를 구현한 PoC 개발 및 시범운영														
- 환경 구성 및 배포														
- 시범 운영														
○ 기록관리시스템 기능 연계 제안														
- 기록관리시스템(CAMS, RMS)에 기능을 구현할 수 있는 설계 제안														
- 각급 기록물관리기관으로 확산 보급 방안 제안														
- 협의체 구성, 운영(관련 분야 교수, 기술진)														

(표 2 - The Research Schedule)

제2장 총괄연구개발과제의 최종 연구개발 내용 및 방법

2.1. 총괄연구개발과제의 연구범위와 방법

2.1.1. 연구범위

○ “공개재분류 업무 지능화”를 위한 업무 분석 및 기술동향 조사
- 업무 프로세스를 분석, AI 도입 분야 진단
- 분야 진단에 도입가능한 AI 기술 제시
- 국내외 유사 구축 사례 조사

○ “비공개대상정보의 검출·비식별화, 대국민 제공”을 위한 기술 동향·사례 조사
- 비공개정보 자동 검출 기술동향 조사
- 국내외 시스템 구축 사례 조사

○ 데이터 수집 및 정제 방안 마련
- AI 모델 학습을 위한 데이터 구성 및 정제 프로세스(안) 마련
- 공개재분류/비공개정보 식별 관련 파인튜닝을 위한 학습데이터 수집, 정제, 구축
- 국가기록원 선행 연구, 사업 분석 및 활용 방안 제시

○ AI를 활용한 “공개재분류 지원 모델” 연구 개발
- 기록물의 비공개대상정보 식별 및 비공개 호수 제시, 유형 자동분류 기능
- 전자문서 및 붙임자료(한글, 엑셀 등 파일유형)에 대한 비공개 대상정보 식별, 패턴학습
- 공개재분류 검토서 작성 지원 및 공개기준 관리 모델

○ “이용자 맞춤형 비식별화 기록물 제공 기능” 연구 및 개발
- 비공개대상정보의 자동식별 및 비식별화 기능
- 이용자 맞춤형 비식별화 기록물 제공 기능
- 자동식별 및 비식별화 처리 완료 데이터의 저장 방안 제시
- 비식별화된 ‘비공개대상정보’를 공개재분류 지원 모델의 파인튜닝 학습데이터로 활용할 방안 제시

○ AI를 도입한 공개재분류부터 비공개대상정보를 식별·비식별화하여 대국민 서비스 제공까지의 단계를 구현한 PoC 개발 및 시범운영

○ 기록관리시스템 기능 연계 제안
- 기록관리시스템(CAMS, RMS)에 기능을 구현할 수 있는 설계 제안
- 각급 기록물관리기관으로 확산 보급 방안 제안

2.1.2. 연구방향 및 방법

❍ 국가기록원 공개재분류 업무 프로세스 분석과 AI 기술 적용 방안 마련
- 공개재분류 관련 법령 및 지침 분석
- 업무담당자 면담을 통해 실무에서의 각 호수 별(1-8호) 판단 과정 분석
- AI 도입 가능한 분야 진단 및 적용 방안 도출

❍ AI 기술 동향 및 국내외 사례 조사·분석을 통한 시사점 도출
- AI 기술 동향 분석, 공개재분류 및 비공개대상정보 비식별화 업무에 적용 방안 도출
- 기록물 공개재분류 관련 자동화·지능화 연구 동향 조사·분석
- 비공개대상정보 비식별화 자동화·지능화 연구 동향 조사·분석

❍ AI를 활용한 “공개재분류 지원 모델” 개발
- 국가기록원 업무 환경에 적절한 AI 모델 선정
- 선정 모델에 국가기록원 공개재분류 업무 프로세스 학습 및 파인튜닝
- 파인튜닝을 위한 학습데이터 수집, 정제, 구축 방안 도출
- 자동식별 및 비식별화 처리 완료 데이터의 저장 방안 제시

❍ 국가기록원 및 각급 기록관리기관의 특성을 반영한 AI “공개재분류 지원 모델”의 적용 로드맵
- 기록관리시스템 환경 분석 및 AI “공개재분류 지원 모델” 연계 방향 설정
- 각급 기록물관리기관으로 확산 보급 방안 및 향후 과제 제안

2.2. 공개재분류 업무 지능화

2.2.1. 연구의 방법 및 절차

○ “공개재분류 업무 지능화” 관련 주요 선행연구 조사
- 문헌조사를 통해 기록물의 공개재분류와 관련된 국내외 선행연구 조사
- 공개재분류 기술의 현황 파악

○ “공개재분류 업무 지능화”를 위한 업무 분석
- 법령, 매뉴얼, 지침, 조사를 통해 업무 프로세스 분석
- 현재 업무 프로세스의 한계점 분석
- 인공지능 적용 가능성 분석

○ “공개재분류 업무 지능화”를 위한 기술동향 조사
- 인공지능 기술 관련 개념 정의
- 인공지능 기술 현황 파악
- 공개재분류 업무 지능화에 적용할 수 있는 기술 분석

○ “공개재분류 업무 지능화”를 위한 사례 조사
- 국내외 사례 조사 및 시사점 분석

2.2.2. 선행연구

공개재분류 업무 지능화에 대한 수요와 가능성은 이전부터 제기되어 왔다. 관련 주요 선행연구를 정리하면 아래와 같다.

○ 국가기록원(2020~2021) : 전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 적용방안 연구
- 비공개정보 기준 도출 및 비공개정보를 패턴화하여 정규식으로 도출
- 일부 호수(1, 5호)의 경우에 키워드 및 패턴 추출 어려움
- AI는 인명 패턴 인식 및 기계학습 영역에 한정하여 적용
- 패턴분석과 기계학습을 결합하면 정확도가 높아지는 것을 확인
- KoBERT 기본모델은 인명 검색 속도가 느리고 정확성에 한계(구조 개선시 속도향상)
- TinyBERT 기본모델을 통해 KoBERT모델의 성능 향상을 시도하였으나 코드충돌의 문제 발생
- 공개재분류서버와 마스킹서버의 개발을 위한 설계 제안서 작성

❍ 국가기록원(2021) : 기록관리 AI 기술 적용을 위한 공통 학습데이터 세트 구축 연구
- 향후 기록관리 업무에 AI 기술을 적용하기 위한 공통 기초 학습데이터 파일럿 구축
- 기계가독 텍스트 추출 후 메타데이터에 태그를 정의함으로써 공통 태그셋 도출
- “그룹별 직제령 비교표”와 “계열별 기능어 분류표”를 기록물 공개재분류에 활용

❍ 한국지능정보사회진흥원(2023) : 국가기록물 대상 초거대 AI 학습용 말뭉치데이터 구축 사업
- Llama2를 이용하여 국가기록물 및 정부간행물을 활용한 초거대 AI 학습용 말뭉치 데이터 세트 및 질의응답 데이터 구축
- 데이터는 초거대 언어 모델(LLM, Large Language Model)을 기반으로 한 국가기록물 및 기타 문서 대상 질의응답 인공지능 모델 개발에 활용
- 초거대 언어 모델은 대규모 말뭉치를 이용하여 입력 문장의 다음 토큰을 예측하는 Auto-regressive Model 방식(사전 학습하여 언어 자체에 대한 이해도)
- 한국어 사용자의 의도를 잘 파악하기 위한 지시문 학습을 진행하여 대화형 인공지능 모델을 개발
- 사용자의 의도를 파악하여 지문에서 질문에 해당하는 답변을 생성하는 대화형 인공지능 QA 태스트 등의 차이점에 대한 고려 필요
- (질문에 대해 지문에서 스팬형의 정답을 찾는 KorQuAD 데이터 세트과 같은 스팬형 기계독해와의 성능비교는 부적절) NarrativeQA 학습데이터 및 해당 리더보드를 통해 간접 비교
- BERT를 적용하여 유해 질의 분류 작업을 진행

❍ 김해찬솔, 안대진, 임진희 외(2017) : 기계학습을 이용한 기록 텍스트 자동분류 사례 연구. 정보관리학회지, 2017(4), 321-344
- 지도학습 방식으로 서울시의 결재문서를 엑소브레인(ETRI)을 통해 자동분류 테스트
  ◦ 엑소브레인은 플랫폼 선정 당시 기준, 한글 처리가 가능한 유일한 인공지능 플랫폼
  ◦ 엑소브레인의 기술 중 형태소 분석, 개체명 인식, 분류 등 3종의 기술을 활용
  ◦ 엑소브레인의 형태소 분석 기술은 45개의 세종 태그셋을 기반
- 기록물 원본을 분류하는 것이 아니라 기록물을 정제한 텍스트를 분류하는 기능
- 향후 인공지능 기술을 위한 데이터 추출 및 XML 기반의 보존포맷 필요성을 강조
- 기능분류에 관한 연구로서 공개재분류를 위한 연구는 아니지만 공개재분류 연구개발에 상당 부분 적용가능
- 연구 당시에는 한국어를 처리할 모델이 부족했지만 현 시점에서는 다양한 선택지가 있기 때문에 자동 분류의 높은 성능 기대
- XML 기반의 문서보존포맷이 적합하다는 제언은 2022년에 「NAK 37:2022 전자기록물 보존포맷 선정기준(v1.0)」이 제정되면서 일부 실현

❍ 대통령기록관(2024) : 대통령기록관리 AI 도입 방안 정책 연구 결과보고서

- 자연어 처리 및 텍스트 유사도 분석으로 비공개 유형 자동분류 가능성을 제시
- 한국어의 특성상 토큰화 선행 작업이 필요
- NER(Named Entity Recognition, 개체명 인식)로 개인정보 유무 식별 가능성을 제시
- 생산시점부터 AI가 활용할 수 있는 메타데이터 확보를 제언

❍ Giovanni Colavizza, Tobias Blanke, Charles Jeurgens 외(2021) : Archives and AI : An overview of Current Debates and Future Perspectives

- AI를 통해 일부 기록관리 업무를 자동화하기 위한 사전 작업으로 워크플로우 개발과 데이터 준비를 언급
- 이를 위한 역량으로는 데이터, 알고리즘 등 전문지식을 언급

❍ David Canning, Lise Jaillant(2025) : AI to review government records: new work to unlock historically significant digital records

- 영국 내각부(Cabinet Office) 기록을 대상으로 가치 판단 작업의 AI 자동화를 시도
- 대규모 데이터 일괄 분석(distant viewing)과 선별 자료 심층 검토(close reading)를 병행 적용
- 데이터 클렌징 솔루션 기반으로 AI 시스템 구축 파일럿을 2022년에 진행하였고 2023년부터 운영

2.3. 비공개대상정보의 검출 및 비식별화

2.3.1. 연구의 방법 및 절차

○ “비식별화 업무 지능화” 관련 주요 선행연구조사
- 비공개 대상 정보 검출 기술의 발전
- 연구기술 및 적용 사례

○ “비식별화 업무 지능화”를 위한 업무분석 및 기술동향 조사
- 법령, 지침, 조사를 통해 업무 분석
- 지능형 비식별화 자동화 기술 및 적용 방향
- 개인정보 비식별화
- 국내 개인정보 비식별화 솔루션

○ “비식별화 업무 지능화”를 위한 사례조사
- 국내외 사례 조사 및 시사점 분석

2.3.2. 선행연구

디지털 전환과 데이터 기반 행정·산업 환경의 확산은 기록물 관리 및 개인정보 활용의 새로운 국면을 열고 있다. 공공·민간 부문 모두에서 데이터 활용성 확대가 국가 경쟁력과 직결되는 상황에서, 개인정보의 안전한 보호는 필수적 과제로 인식되고 있다(개인정보보호위원회, 2024). 그러나 기존의 비식별화 기법은 기술적· 제도적 한계로 인해 활용성과 보호성의 균형을 달성하는 데 어려움을 겪어왔다. 디지털 전환과 데이터 기반 행정·산업 환경의 확산은 기록물관리 및 개인정보 활용의 새로운 국면을 열고 있다. 공공·민간 부문 모두에서 데이터 활용성 확대가 국가 경쟁력과 직결되는 상황에서, 개인정보의 안전한 보호는 필수적 과제로 인식되고 있다. 기록물의 개인정보 검출 및 비식별화 연구는 공공기관의 기록물 관리와 개인정보 보호를 위한 핵심분야로 특히 행정 기록물, 법률문서, 의료기록 등에서 개인정보를 안전하게 처리하고 활용하기 위한 다양한 기술적 접근이 이루어지고 있다(이재용, 2022). 선행연구는 기술적으로 진화된 내용을 2015년부터 2025년까지 기록물의 개인정보 검출 및 비식별화 연구는 크게 4가지 단계로 구분할 수 있다.

○ 규칙 기반 기법(2015년 ~ 2017년) : 정규 표현식(Regex, Regular Expression) 및 키워드 기반 탐지
- 행정 기록물의 개인정보 검출 : 주민등록번호, 전화번호 등
- 법률 문서의 비식별화 : 법률 문서 내 개인정보를 식별하고 제거

○ 기계학습 기반 접근(2018년 ~ 2020년) : NER(Named Entity Recognition) 이용
- 행정 기록물의 비식별화 : 인물, 기관명, 날짜 등

- 의료 기록의 비식별화 : 전자건강기록(EHR[3])에서 환자 이름, 진단명, 병원명 등
- 공공 기록물의 비식별화 : 국가기록원 등에서 기록물 공개 전 개인정보를 자동 검출 / 비식별화

○ AI 기반 비식별화(2021년 ~ 2023년) : BERT[4], GPT 등 Transformer 기반 모델
- 기록물 문서에서 문맥을 고려하여 개인정보를 정확하게 식별하는 연구 진행
- 공공기관에서 대량의 기록물을 효율적 처리위한 자동 비식별화 시스템 도입
- 법률 문서에서 개인정보를 식별 및 비식별화하는 기술 발전으로 법적 요구사항 충족

○ 통합 기록물관리 시스템(2024년 ~ 2025년) : 제도적 비식별화 절차의 표준화 논의
- 공공기관의 기록물의 수집, 분류, 비식별화, 공개까지의 전 과정을 통합 관리하는 시스템 개발
- 기록물의 비식별화 절차와 기준 표준화, 관련 법령 및 정책을 수립하여 개인정보 보호 강화
- AI를 활용한 기록물 자동 분류, 검색, 비식별화, 공개를 수행하는 시스템의 공공기관 도입

○ 비식별화(De-identification) 기술의 진화
- 비식별화 기술은 데이터 활용성과 개인정보 보호 간의 균형을 유지한 방향으로 발전
- 비식별화 기법은 데이터 마스킹(Data Masking), 가명처리(Pseudonymization), k-익명성(k-Anonymity), l-다양성(l-Diversity), t-근접성(t-Closeness), 차등 프라이버시(Differential Privacy) 등
- 텍스트, 이미지, 음성 등 비정형 데이터로의 확장 연구 진행중
- 의료 분야에서는 EHR(Electronic Health Record) 내 환자 식별자(환자명, 병원명, 날짜 등)의 자동 탐지 및 비식별화 연구 진행중
- 행정기록 분야에서는 공공문서 내 주민정보, 민원인 정보, 위치정보 등의 자동 검출 및 제거를 위한 AI 기반 기술 도입중

구분	기술단계	세부내용
1단계	• Rule → Model → Policy Intelligence로 진화	• 초기에는 단순 규칙(Rule-based) 중심이었으나, → AI 기반 문맥 인식(Model-based) → 정책·법규 자동 반영형(Policy-aware Intelligence) 으로 발전 중.
2단계	• 정형 → 비정형 → 멀티모달 데이터 확장	• 텍스트 중심에서 이미지, 영상, 음성까지 확장되며, 복합형 비식별화 기술이 요구됨.
3단계	• 보호 → 활용 균형 중심 패러다임 전환	• 단순 보호 목적에서, 데이터 분석·AI 학습을 위한 “활용 가능한 비식별화(utility-preserving privacy)”로 진화.
4단계	• 지능형 거버넌스·검증체계 구축	• AI 모델의 처리과정 기록, 재식별 위험 지표화, 메타데이터 기반 이력관리(Audit trail) 체계가 중요 요소로 부상.

(표 3 - De-identification RoadMap)

3) Electronic Health Record, 전자건강기록(환자의 건강 및 진료 정보를 디지털 형태로 기록·저장·관리하는 시스템)
4) BERT(Bidirectional Encoder Representations from Transformers) 2018년 구글이 발표한 자연어 처리(NLP) 모델

❍ 해외에서는 의료·행정 영역을 중심으로 발전

- 미국의 DEFT(De-identification Evaluation Framework for Text) 프로젝트 : 의료 텍스트의 비식별화 성능을 평가하는 표준 프레임워크 제시
- 영국의 BoB (Better of Both Worlds) 프로젝트 : 공공 데이터 비식별화를 위한 AI·규칙 기반 하이브리드 모델 개발
- 캐나다, 프랑스 : DICOM(의료영상 데이터) 및 대규모 임상 텍스트(Clinical Notes)에 대한 비식별화 연구(Campbell et al., 2021)

2.4. 데이터 수집 및 정제 방안

초기 단계에서 파인튜닝(Fine-tuning)을 통한 분류기 기반의 AI 모델 적용으로 시작하였으나, 연구의 진행 과정에서, 최신기술 검토 및 외부전문가 자문을 통해, RAG(Retrieval Augmented Generation) 기반의 AI 모델 구축으로 발전하여 데이터 수집 및 정제 방안 또한, 파인튜닝 데이터 정제방안에서 RAG 구현을 위한 벡터DB 구축으로 수정되었다.

RAG 구현을 위한 벡터DB를 구축하기 위해 국가기록원 고유업무 비공개기록물 공개재분류 세부유형기준서와 공통업무 비공개기록물 공개재분류의 세부유형기준서를 참조하여 벡터DB를 구축하였다.

2.4.1. 데이터 수집

고유업무 비공개기록물 공개재분류 세부유형기준서와 공통업무 비공개기록물 공개재분류의 세부유형기준서를 분석하여 카테고리 별 상위유형, 세부유형, 세세부유형, 기능 및 추가 설명, 기록물 철명, 주요내용, 공개검토 기간구분, 공개유형, 사유 및 근거의 정보를 구축하고, 기준서상 검토항목을 분석하여 중요내용에 대한 질의를 생성하는 것이다. 고유업무에 대해서는 조직정보 항목을 추가로 추출한다.

2.4.2. 추출 데이터

- 카테고리: 원형숫자형식과 숫자를 조합하여 대분류 업무로 구분
- 상위유형: 최상위 대범위로 분류된 업무유형(예: 감사, 인사)
- 세부유형: 상위유형 다음 단계로 분류된 중간단계의 업무유형
- 세세부유형: 세부유형 다음 단계로 분류된 최하위의 업무유형
- 기능 및 추가 설명: 수행되는 업무의 처리 기능을 설명
- 기록물 철명: 재분류 대상인 기록물의 철명을 나열함
- 주요내용: 업무내용을 기술한 항목
- 공개검토 기간구분: 30년경과 미경과에 따른 검토 항목
- 공개유형: 공개/부분공개,비공개에 해당
- 사유 및 근거: 관련된 공개유형의 호수
- 검토항목: 공개유형을 결정하기 위해 기준이 되는 추가적으로 검토하는 항목

2.4.3. 추출 데이터의 주요유형

○ 고유업무 세부기준 : 총 38개 기관, 214개 기준서 요약 사례

번호	기 관 명	세부 기준수	유형 현황			주요 유형
			상위	세부	세세부	
1	감사원	3	2	3	-	(감사) 감사결과처분요구서, 결산검사보고서 (소청 · 소원 · 소송) 심사 및 재심의 청구
2	고용노동부	4	3	3	3	(인 · 허가)면허 · 자격관리, (수사)노동사건 수사기록, 판정 · 결정 및 조정, (기타) 사업장정보집
3	공정거래 위원회	3	2	2	3	(인 · 허가)소비자단체등록 (기타)공정거래위원회 심의 · 의결(심의안건, 의결조서)
4	과학기술 정보통신부	2	1	2	-	(인 · 허가)무선국허가 및 주파수 할당, 재단법인 설립
5	관세청	3	2	3	-	(소청 · 소원 · 소송)관세심사 및 심판 (수사)관세범칙(사건조사, 사건대장)
6	교육부	2	2	2	-	(인 · 허가)사립학교법인설립, (학적관리)학위수여자 등록부, 신입생명부, 졸업대장
7	국가인권 위원회	3	1	1	3	(기타)인권침해 및 차별행위에 대한 사건조사 및 구제, 구금 · 보호시설 방문조사, 위원회 심의 · 의결 · 조정
8	국무조정실	2	2	2	-	(소청 · 소원 · 소송)조세심판, (기타)국무회의
9	국민권익 위원회	3	1	1	3	(소청 · 소원 · 소송)행정심판에 대한 청구처리 및 통지, 위원회 심의 · 의결, 의결서
10	국세청	12	4	11	3	(수사)조세범칙, (인 · 허가)법인설립신고, 사업자등록, 법인세적, 주류제도 및 판매면허 등, (기타) 세무조사
11	대학	1	1	1	-	(학적관리)국 · 공립 대학 학위수여자 등록 신청, 학위등록 및 신입생(편입생) 명부, 학력조회, 졸업대장
12	원자력안전 위원회	2	1	2	-	(인 · 허가)방사선동위원소 사용허가, 면허관리
13	지역교육청	1	1	1	-	(학적관리)학적부, 생활기록부, 제적부, 졸업대장
14	경찰청	12	6	12	-	(수사) 각종 사건부, 명부, 수사기록 등 (정보·보안) 공안사범 등, (인·허가) 총포소지허가 등
15	대검찰청	9	4	9	-	(수사) 각종 사건부, 명부·색인부, 수사기록 (법무·행형) 검찰기록관리, 판결문·결정문·약식명령
16	법무부	14	3	9	6	(법무·행형) 수용자기록, 사면·복권, 소년범죄명부 (국적관리) 국적관리, (기타) 외국인 기록 등
17	보건복지부	3	1	3	-	(사회·복지) 의료장비구입, 보건장학금, 해외입양
18	식품의약품 안전처	5	1	5	-	(인·허가) 마약취급/의료기기/의약외품/외약품/화장품 제조 및 수입 등 허가
19	인사혁신처	9	1	9	-	(인사·징계) 인사기록, 채용 및 임용, 교육훈련 등
20	여성가족부	5	1	3	4	(기타) 여성정책, 청소년유해매체심사 등

번호	기 관 명	세부 기준수	유형 현황			주요 유형
			상위	세부	세세부	
21	해양경찰청	**9**	3	9	1	(수사) 각종 사건부, 명부, 수사기록 (시설) 함정건조 및 관리, 함정도면, 함정대장
22	행정안전부	**6**	4	6	-	(감사) 감사, (인사·징계) 상훈관리 (기타) 국무회의록 및 안건철, 감정서
23	금융위원회	**5**	1	5	-	(인·허가) 금융기관설립·취소·합병 인가, 보험업허가 등
24	기획재정부	**2**	1	2	2	(기타) 경제개발계획, 조세재정관련 위원회
25	농림축산식품부	**3**	1	3	-	(인·허가) 농지관련 인허가, 농지개발, 농지전용확대
26	농촌진흥청	**5**	2	4	3	(인·허가) 농약관련 허가 (기타) 유전자원종자관리, 조사야장 등
27	문화체육관광부	**9**	3	8	2	(인·허가) 정기간행물 관련 허가, 표준영정 등 (문화재관리) 문화재조사, 유물관리 등 (기타) 국제경기대회 및 문화행사 등
28	문화재청	**12**	3	12	-	(인·허가) 문화재수리업자 등록 (건설·토목) 전수관·전수회관 건립 (문화재관리) 발견·발굴매장문화재 처리, 문화재 현상변경, 문화재위원회 등
29	병무청	**1**	1	1	-	(병무) 병적카드
30	산림청	**12**	4	6	8	(국유재산) 국유림 매수 및 처분, 경계측량 및 표주설치 등 (인·허가) 해외산림자원개발허가, 자연휴양림 지정 (건설·토목) 휴양림 운영·관리 (기타) 산림입지자원조사, 산림기술개발 등
31	산업통상자원부	**5**	2	5	2	(소청·소원·소송) 광업조정위원회 (인·허가) 광업권등록, KS표시허가 등 (기타) 외국인투자
32	중소벤처기업부	**2**	2	2	1	(인·허가) 시험분석감정신청, (기타) 기술지도
33	특허청	**1**	1	1	-	(인·허가) 변리사등록부
34	해양수산부	**9**	3	8	2	(인·허가) 선박건조허가 및 어선건조공사, 외항화물운송사업등록 등 (건설·토목) 해운항만관련 건축시설공사 등 (기타) 해양오염 등
35	환경부	**10**	2	9	2	(인·허가) 폐기물처리업허가, 야생동식물수출입허가 등 (기타) 환경영향평가 등
36	국토교통부 ('21년 신규)	**23**	4	22	2	(인·허가) 도시계획(변경)협의, 택지개발사업택지공급변경승인, 도로건설사업승인 설계방침 승인 등 (건설토목) 경인운하건설, 공항건설, 도로건설 등 (토지) 지도제작, 토지수용재결, (기타) 국제협력 등
37	지방자치단체 ('21년 신규)	**1**	1	1	-	(기타) 국가장
38	질병관리청	**1**	1	1	-	(감염병 관리) 질병관리 ※ 보건복지부 17-1-2에서 이동

(표 4 - Table of Detailed Criteria for Inherent Duties)

2.5. AI를 활용한 “공개재분류 지원 모델” 연구 개발

2.5.1. 연구개요

본 연구는 인공지능(AI)을 활용하여 비공개 기록물의 공개여부를 자동 판정·지원하는 「공개 재분류 지원 모델」을 개발하는 것을 목표로 한다.

특히, 「공공기록물 관리에 관한 법률」 제35조 및 「공공기관의 정보공개에 관한 법률」 제9조에 근거한 비공개 대상 정보의 세부 기준을 AI가 이해하고 적용할 수 있도록 하는 체계를 구축함으로써, 기록물 공개 판단의 일관성과 업무 효율성 향상을 주요 과제로 삼았다.

이를 위해 업무분석, AI 적용 방향성에 대해서 검토하고, 향후 알고리즘 및 프로세스 설계 과정을 거쳐, 모델 검증(Proof of Concept: PoC) 등의 과정을 단계적으로 수행하였다.

2.5.2. 업무분석

○ 목적 및 필요성

- “비공개 기록물의 30년 경과 공개원칙” 적용 및 비공개 기록물의 주기적 공개여부 검토(5년)를 통해 비공개 기록물의 신속한 공개 전환 유도
- 국민의 알권리 충족을 위한 기록물의 적극적 공개 및 기록정보 서비스 기반 마련

○ 관련 법령

- 『공공기록물 관리에 관한 법률』 제35조(기록물의 공개여부 분류)
- 『공공기관의 정보공개에 관한 법률』 제9조(비공개 대상정보)

○ 대상 기록물

- 생산 후 30년 경과 비공개/미분류 기록물
- 이관 또는 재분류 이후 5년 경과 비공개/미분류 기록물
- 전시, 편찬 등 대국민 서비스를 위한 기획성 기록물

○ 공개 유형

구 분	내 용
공 개	○ 기록물철에 포함된 개별 건 전체가 비공개대상이 없는 경우 - '정보공개법 제9조 제1항의 각 호'에 해당하는 정보가 없는 경우
부분공개	○ 기록물철에 비공개 건과 공개 건이 혼합되어 있는 경우 - '정보공개법 제9조 제1항 각 호'에 해당하는 부분과 공개 가능한 부분이 혼합되어 있는 경우
비 공 개	○ 기록물철에 포함된 개별 건 전체가 비공개인 경우 - 기록물 내 정보 전체가 '정보공개법 제9조 제1항 각 호'에 해당하는 경우

(표 5 - Type of Disclosure)

○ 『공공기관의 정보공개에 관한 법률』 제9조 에 따른 비공개 대상 정보

구분	규 정	예 시
1호	**다른 법률 또는 법률이 위임한 명령**(국회규칙·대법원규칙·헌법재판소규칙·중앙선거관리위원회규칙·대통령령 및 조례에 한한다)**에 의하여 비밀 또는 비공개 사항으로 규정된 정보**	o 공직자의 재산등록사항(공직자윤리법) o 납세자의 과세정보(국세기본법) o 노동위원회 회의록(노동위원회법) o 국가정보원의 조직·소재지(국가정보원법)
2호	**국가안전보장·국방·통일·외교관계 등에 관한 사항**으로서 공개될 경우 국가의 중대한 이익을 현저히 해할 우려가 있다고 인정되는 정보	o 대북한 관련 정보수집·분석자료 o 국방투자사업 관련 문서 o 외국 및 국제기구 간 교섭에 관한 문서
3호	공개될 경우 **국민의 생명·신체 및 재산의 보호에 현저한 지장을 초래할 우려가 있다고 인정되는 정보**	o 범죄의 피의자, 참고인 등의 명단 o 생화학테러 대비 기술개발사업 o 국가중요시설의 경비에 관한 사항
4호	**진행중인 재판에 관련된 정보와 범죄의 예방, 수사, 공소의 제기 및 유지, 형의 집행, 교정, 보안처분에 관한 사항**으로서 공개될 경우 그 직무수행을 현저히 곤란하게 하거나 형사피고인의 공정한 재판을 받을 권리를 침해한다고 인정할 만한 상당한 이유가 있는 정보	o 재판과 관련된 소장, 답변서 등 o 피의자 신문조서 o 수형자의 신분기록 o 수사기록, 의견서, 내사자료 등
5호	**감사·감독·검사·시험·규제·입찰계약·기술개발·인사관리·의사결정과정 또는 내부검토과정에 있는 사항** 등으로서 공개될 경우 업무의 공정한 수행이나 연구·개발에 현저한 지장을 초래한다고 인정할 만한 상당한 이유가 있는 정보	o 불시 직무감찰 계획 o 공무원 채용시험 출제 관련 정보 o 입찰예정가격을 예측할 수 있는 단가 o 연구·개발 중에 있는 과제에 관한 정보
6호	해당 정보에 포함되어 있는 성명·주민등록번호 등 「**개인정보 보호법**」 **제2조제1호에 따른 개인정보**로서 공개될 경우 개인의 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보	o 민원인의 인적사항 o 인허가 신청인의 주민등록번호, 주소 등 o 수험생의 성적, 학력, 주소 등
7호	**법인·단체 또는 개인**(이하 "법인등"이라 한다)**의 경영·영업상 비밀에 관한 사항**으로서 공개될 경우 법인 등의 정당한 이익을 현저히 해할 우려가 있다고 인정되는 정보	o 단체의 자금, 인사 등에 관한 정보 o 업체의 신공법 등에 관한 정보 o 법인·단체의 경영상태에 관한 정보
8호	공개될 경우 **부동산 투기·매점매석 등으로 특정인에게 이익 또는 불이익을 줄 우려가 있다고 인정되는 정보**	o 보호구역 지정고시 전의 정보 o 물품가격 변동 등에 관한 정보 o 지역개발 등에 대한 공고, 고시 전 정보

(표 6 - Information Subject to Non-Disclosure under the Law)

○ 본 연구 개발로 AI 가 적용되는 업무 프로세스

- 국가기록원의 기록관리 프로세스 중 공개재분류, 열람제공의 프로세스에 본 연구개발 결과가 적용될 수 있으며, 이중 본 연구에서 개발된 인공지능(AI) 기술은 “공개여부 검토” 및 “공개여부 확인 후 기록물 제공” 단계에 적용된다.

- “공개여부 검토” 단계에서는 LLM(Large Language Model)이 기록물의 성격을 분석하고, 과거 재분류 사례와 비공개 사유 데이터를 근거로, 비공개 사유 해당 여부를 자동 예측하거나 근거를 제시하도록 하였다. 이는 심의자에게 판단 근거를 제공함으로써 업무 효율성을 높이고, 판단 일관성을 확보하는데 기여한다.

- 또한 “공개여부 확인 후 기록물 제공” 단계에서는, “공개여부 검토” 과정에서 AI가 식별한 비공개 정보를 비식별화 하며, 이용자 맞춤형 기록물 제공이 가능하도록 하였다.

(그림 1 - Records Management Process)

2.5.3. 공개재분류 업무의 AI 적용 방향

국가기록원의 영구기록물에 대한 공개재분류 업무는 단순한 문서 분류나 내용 분석을 넘어, 시의성·정책성·사회적 파급력을 종합적으로 고려해야 하는 복합적 판단 영역이다. 따라서 일반적으로 활용되는 AI 기반 분류기(Classifier) 모델의 직접적인 적용은 적정하지 않다. 분류기 모델은 일정한 기준에 따라 고정된 데이터를 분류하는 데에는 효과적이나, 재분류 과정처럼 비공개 사유의 유효성이 시점과 환경에 따라 변동되는 업무에는 한계가 존재한다.

이에 본 연구는 국가기록원이 축적해온 기록물 분류 기준, 사례, 행정적 경험을 기반으로 인공지능이 이를 학습·활용할 수 있는 방안을 모색하였다. 특히, 최근 발전한 대규모 언어모델(LLM)의 문맥 이해 및 추론 능력을 접목하여, 과거의 분류기준과 사례 데이터를 종합적으로 반영한 지능형 재분류 지원 모델을 설계하였다. 이를 통해 모델이 단순 분류를 넘어, 기록물의 내용·시점·공개 문맥을 종합적으로 고려한 판단을 제시할 수 있음을 확인하였다.

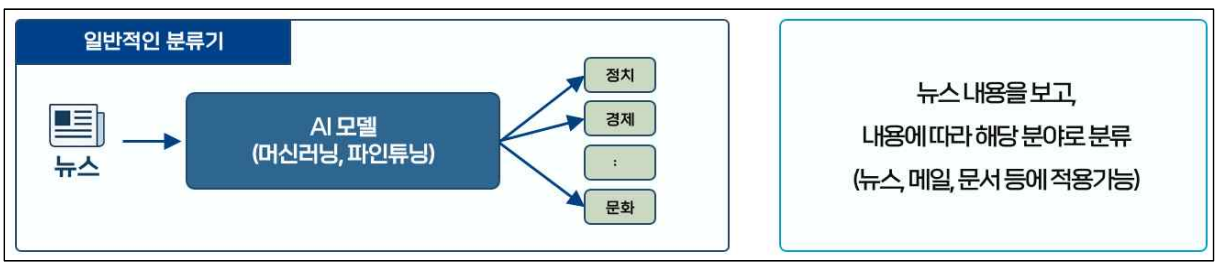
나아가, 이러한 접근의 타당성을 검증하기 위해 개념검증(Proof of Concept, PoC) 단계를 수행하였다. PoC 결과, LLM 기반 모델은 기존의 기계학습형 분류기 대비 의미있는 분류 성과가 있어, 업무 특성에 부합하는 재분류 판단 지원이 가능하다고 판단된다.

○ AI 기반 분류기 모델의 정의 및 개요

- AI 기반 분류기 모델은 입력된 데이터를 사전에 정의된 여러 범주 중 하나로 자동 분류하는 인공지능 알고리즘으로, 대규모 학습데이터를 통해 패턴과 특징(feature)을 추출하고 이를 기반으로 새로운 데이터의 속성을 예측한다. 일반적으로 지도학습(Supervised Learning) 방식이 적용되며, 입력 데이터와 정답(label) 간의 관계를 학습하여 모델의 분류 정확도를 점진적으로 향상시킨다. 이러한 모델은 텍스트, 이미지, 음성 등 다양한 비정형 데이터(Unstructured Data)를 처리할 수 있으며, 특히 자연어처리(NLP)와 컴퓨터비전(CV) 분야에서 광범위하게 활용되고 있다.

- 대표적인 응용 사례로는 뉴스 기사 자동 분류, 스팸메일 탐지, 소셜미디어 게시물 분석 등이 있다. 예를 들어 이메일이나 SNS, 커뮤니티 플랫폼에서 AI 분류기는 텍스트를 분석하여 스팸, 광고, 허위 정보, 일반 게시물 등으로 구분하거나, 특정 주제·감정·의도에 따라 자동으로 분류할 수 있다. 이러한 방식은 대량의 데이터 처리와 실시간 의사결정을 지원함으로써 업무 효율성을 높이고, 정보의 신뢰성 및 접근성을 개선하는 데 기여한다.

- 요약하면, AI 기반 분류기 모델은 데이터의 자동 분류와 정보의 체계적 관리를 가능하게 하는 핵심 기술로, 인공지능 응용체계의 기초를 구성하는 주요 알고리즘 중 하나라 할 수 있다.



(그림 2 - AI-based Classification Model)

○ 공개재분류 업무의 특성과 AI 기반 분류기 모델 적용 한계

- 본 연구에서는 국가기록원의 공개재분류 업무 특성을 분석하고, 이에 따라 인공지능(AI) 기반

분류기 모델 적용의 타당성을 검토하였다. 공개재분류란 기록물이 생산된 이후 일정 기간이 경과한 시점에서, 기존의 비공개 결정 사유를 재검토하고 변경된 사회·행정 환경에 맞추어 공개 여부를 다시 판단하는 절차를 의미한다. 이러한 과정은 행정정보의 투명성 제고와 공공기록의 접근성 향상을 위한 핵심 절차로, 법령과 정책 변화, 사회적 파급효과 등을 종합적으로 고려하여 수행된다.

- 공공기록물의 공개·비공개 기준은 「공공기관의 정보공개에 관한 법률」 제9조를 근거로 하며, 예를 들어, 비공개 사유 8호는 “공개될 경우 부동산 투기, 매점매석 등으로 특정인에게 이익 또는 불이익을 줄 우려가 있는 정보”를 의미한다. 이 기준은 기록물이 생산될 당시의 사회·경제적 상황을 반영하지만, 시간이 경과함에 따라 그 적용 범위가 달라질 수 있다. 수서 - 평택 간 고속철도 건설계획 문서는 계획 당시에는 비공개(8호) 대상이었으나, 해당 사업이 언론을 통해 이미 공개되고 국가기록원으로 이관된 시점에서는 더 이상 동일한 비공개 사유가 되지 않을 수 있다. 이러한 사례는 재분류 시점의 사회적·정책적 맥락을 반영한 판단이 필요함을 보여준다.

- 이와 같은 특성은 인공지능 모델, 특히 AI 기반 분류기(Classifier) 모델의 적용에 본질적 제약을 발생시킨다. 분류기 모델은 일반적으로 뉴스, 이메일, 스팸 등과 같이 시간이 지나도 내용의 분류 기준이 고정적인 데이터에 적합하다. 즉, 한 번 학습된 규칙이나 패턴이 지속적으로 유효한 환경에서는 높은 정확도를 보인다. 그러나 공개재분류 업무는 시간적 변화, 정책적 상황, 사회적 가치 판단이 복합적으로 작용하는 영역으로, 단순한 데이터 패턴 학습만으로는 정확한 판단을 내리기 어렵다. 특히 비공개 사유의 타당성은 단순 문서 내용뿐만 아니라 당시의 행정 결정 맥락, 외부 공개 여부, 이해관계자 영향도 등을 함께 고려해야 하므로, 정형화된 분류기 학습모델보다는 RAG(Retrieval Augmented Generation) 기반의 문맥적 추론과 전문가적 판단을 결합한 접근이 요구된다. 이는 인공지능이 단순히 문서의 내용을 분류하는 수준을 넘어, 정책적 판단의 근거를 제시하고 설명 가능한 판단 과정(Explainable AI)을 구현하는 방향으로의 발전이 필요함을 시사한다.

- 결론적으로, 공개재분류 업무는 시의성, 맥락성, 사회적 파급력을 동반한 복합적 판단 영역으로, 고정된 기준에 따라 학습된 AI 분류기 모델만으로는 적절히 대응하기 어렵다. 따라서 본 과제에서는 재분류 경험과 도메인 지식을 통합한 지식 기반 판단형 AI 모델 구축이 보다 현실적이며, 공공기록 관리의 특수성을 반영한 인공지능의 활용으로 방향을 설정하였다.

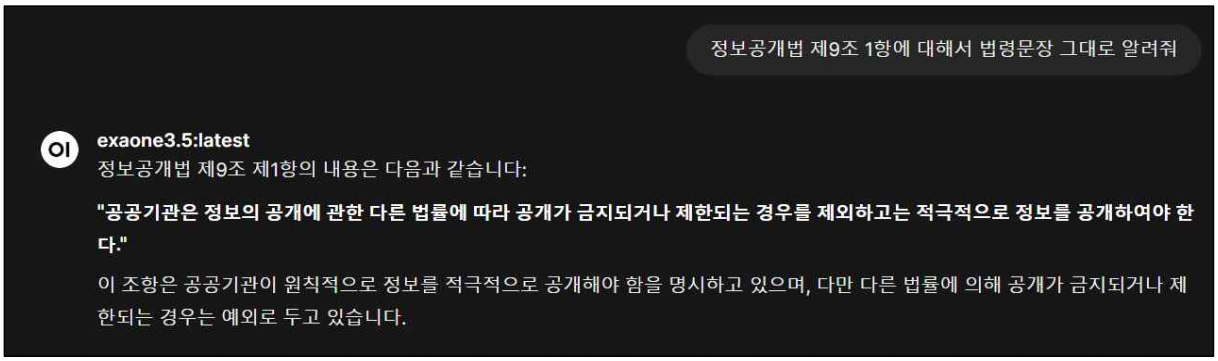
○ 공개재분류 유형 판단 기준의 적용 한계

- 공개재분류 업무에서 핵심이 되는 판단 기준은 「공공기관의 정보공개에 관한 법률」 제9조 제1항 각 호(1호~8호)에 근거한다. 본 연구에서는 이러한 법령 기준을 인공지능의 언어 이해 능력을 활용하여 대규모 언어모델(LLM, Large Language Model)을 통해 자동 판단하는 가능성을 검토하였다. 그러나 실제 적용 과정에서 다음과 같은 한계가 확인되었다.

- 우선, 제9조 제1항의 각 호는 표현이 광범위하고 포괄적이며, 법령 해석상 맥락 의존성이 매우 높다. 예를 들어 1호부터 8호까지의 비공개 사유는 행정·경제·사회적 상황에 따라 달리 해석될 수 있으며, 동일한 문서라도 시점과 사안의 특성에 따라 적용 기준이 변동된다. 이와 같은 특성으로 인해 LLM이 단순한 언어적 유사도 기반으로 판단할 경우, 정확도가 저하되고 부정확한 분류 결과가 도출되는 사례가 다수 확인되었다.

- 또한, LLM이 학습한 정보공개법 관련 데이터는 대부분 과거 시점의 법령 텍스트나 일반적 해석 사례에 기반하고 있어, 국가기록원에서 실제로 적용하는 최신 법령 및 세부 판단기준(1호~8호)과 불일치하는 경우가 존재한다. 따라서 LLM의 판단을 직접적으로 활용하기보다는, 모델이 참고하는 법령 데이터의 정확성과 최신성 검증이 선행되어야 한다.

- 아래 그림은 LLM exaon3.5 에 “정보공개법 제9조에 대해서 법령문장 그대로 알려줘” 라는 질의에 답하는 결과로, 현행 법령을 다르게 알고 있음을 보여준다. 이는 모델이 사전에 학습된 내용을 미리 검증하고 질의에 어떻게 답할지를 예측해 볼 필요가 있다.



(그림 3 - LLM Chat Screen)

- LLM은 지속적으로 최신모델이 발표되고 있으며, 최신모델을 검증하고, 선정함에 따라서 이러한 문제는 해결될 수도 있다.

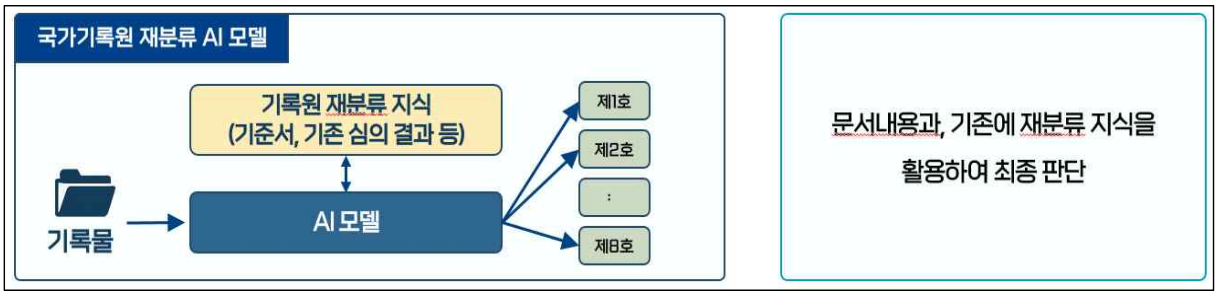
○ AI 모델의 판단 근거로 활용될 지식 기반 데이터 구조 설계 방향

- 본 연구에서는 인공지능 모델이 재분류 판단을 수행할 때, 신뢰성 있는 지식 근거를 활용할 수 있도록, 지식 기반 데이터 구조(Knowledge-based Data Structure) 로 설계 방향을 수립하였다. 기존 방식인 학습용 데이터 세트 구축 후 라벨링(Labeling)을 통한 파인튜닝(Fine-tuning) 접근은, 공개재분류 업무의 복잡한 맥락과 법적 기준을 일관성 있게 반영하기 어렵다는 한계가 있었다. 또한, LLM에 직접적인 판단을 요청할 경우, 사전학습 데이터의 불완전성 및 최신 법령 반영 미비로 인해 정확한 결과를 보장하기 어렵다는 점이 확인되었다.

- 이에 따라 본 연구는 기존의 공개재분류 지식과 사례를 LLM이 직접 참조·활용할 수 있는 구조적 접근, 즉 RAG(Retrieval Augmented Generation) 기반 모델 구성을 채택하였다. 모델은 추론 시 외부 지식저장소(Knowledge Base, RAG)를 실시간으로 검색하고, 그 결과를 판단 근거로 반영하도록 설계하였다. 이를 위해 재분류 사례를 포함한 RAG용 문서 코퍼스(Document Corpus, AI 학습용 텍스트 데이터 집합)를 구축하여 모델의 판단 정확성과 설명 가능성을 강화하였다.

- 또한, 국가기록원의 재분류 이력 데이터, 세부기준, 기관별 사례자료를 메타데이터 기반으로 구조화함으로써, LLM이 판단해야 할 정보의 범위를 명확히 제한하였다. 이를 통해 LLM이 불필요한 일반 지식을 참조하지 않고, 도메인 특화된 근거 중심 판단(knowledge-driven reasoning)을 수행하도록 하여 정확도와 신뢰성을 동시에 향상시키는 방향으로 설계하였다.

- 아래 그림은 국가기록원의 RAG 기반 AI 모델의 구성도이다.



(그림 4 - RAG Diagram)

○ 사전학습 언어모델과 RAG 기반 AI 모델의 비교 및 특성

- 본 연구에서는 국가기록원 공개재분류 업무에 적합한 인공지능 모델 구조를 검토하기 위해, 사전학습 언어모델(KoBERT 등)과 RAG(Retrieval Augmented Generation) 기반 AI 모델의 특성을 비교·분석하였다.

- 먼저, 사전학습 언어모델은 명확한 규칙과 고정된 기준을 기반으로 작동하며, 모델 학습을 위해 대규모의 학습데이터와 파인튜닝(Fine-tuning)이 필요하다. 이러한 방식은 데이터 기준이 일정할 경우 효율적이지만, 기준이 변경되거나 새로운 데이터가 추가될 경우 전체 모델을 재학습해야 하는 한계가 있다. 또한 추론 단계에서는 주어진 입력에 대한 단순 분류나 예측에 집중하기 때문에, 결과에 대한 이유나 근거를 제시하는 논리적 판단 기능은 제한적이다.

- 반면, RAG 기반 AI 모델은 사전학습된 지식 외에도 지식데이터(Knowledge Data)와 벡터 DB(Vector Database)를 참조하여 외부 정보를 동적으로 활용한다. 이 방식은 불명확한 분류나 새로운 기준이 등장하더라도, 지식데이터를 추가·갱신하는 것만으로 모델의 성능을 유지할 수 있

다. 또한 프롬프트 엔지니어링(Prompt Engineering)이나 Few-shot Learning 기법을 활용하여 별도의 대규모 재학습 없이도 적응이 가능하다. 다만, GPU를 활용한 검색·추론 과정이 필요하기 때문에 상대적으로 속도와 비용 측면에서는 기존 모델보다 높은 자원이 요구된다.

- 특히, RAG 모델은 단순 분류 결과뿐 아니라 문맥과 논리적 근거를 기반으로 추론(reasoning)이 가능하며, 결과에 대한 설명이나 판단 사유를 함께 제시할 수 있다는 점에서 공개재분류와 같이 근거 제시가 중요한 행정 분야에 적합하다. 이러한 비교 결과는 본 과제가 추구하는 지식기반 판단형 AI 모델의 방향성을 구체화하는 핵심 근거로 활용되었다.

- 아래 그림은 사전학습모델과, RAG 기반 AI모델을 비교한 결과이다.

항목	사전학습언어모델(KoBERT 등)	RAG 기반 AI 모델
학습 필요 여부	필요 (파인튜닝)	지식데이터 ⇒ 벡터DB 구성 필요 프롬프트 사용 또는 Few Shot Learning
데이터 관리	명확한 규칙, 기준 중요, 기준이 고정적인 경우 기준이 변경될 경우 전체 재학습 필요	불명확한 분류도 가능, 지식데이터 추가 변경이 쉽게 구성 가능
새로운 모델로 변경시	전체 재학습(다시 만들어야 함)	모델만 변경 가능
속도/비용	빠르고 가벼움	상대적으로 느림, GPU 필요
추론 능력	없음	있음 (문맥, 논리 판단 가능), 이유 사유 생성 가능

(그림 5 - Comparison Table)

○ AI 모델 구성의 변화 및 고도화 과정

- 본 연구과제는 초기 단계에서 분류기(Classifier) 중심의 AI 모델을 기반으로 시작하였으나, 연구의 진행 과정과 외부 전문가 자문을 통해 보다 고도화된 RAG(Retrieval Augmented Generation) 기반의 지능형 모델로 발전하였다.

- 이러한 변화는 연구 데이터의 성격, 업무 특성, 그리고 최신 인공지능 기술 동향을 종합적으로 고려한 결과이다.

- 초기 제안요청서 및 제안서 단계에서는 파인튜닝(Fine-tuning)을 통한 분류기 학습 모델을 중심으로 개발이 계획되었으나, 이는 제한된 범위 내에서 특정 카테고리 분류나 의사결정 지원에 효과적인 방식으로, 명확한 정답(label)이 존재하는 데이터를 학습시켜 성능을 검증하는 접근이었다.

- 이후 착수보고회 단계에서 고성능 LLM이 발표되고 활용됨을 고려할 필요가 있어, 이에 따라 분류기 학습 모델과 대규모 언어모델(LLM)을 결합한 하이브리드(Hybrid) 방식으로 계획하였다. 이

방식은 분류기의 정형적 판단력과 LLM의 문맥 이해 능력을 결합하려는 계획이었다.

- 다음 단계에서는 외부 전문가 자문회의를 통해 최신 AI 기술 동향과 업무분석을 통한 분류기 모델의 한계점 등이 검토되었으며, 그 결과 RAG 기반 AI 모델의 도입이 권장되었다.

- RAG(Retrieval Augmented Generation)는 검색(Retrieval)과 생성(Generation)을 결합한 구조로, 단순히 학습된 정보만을 활용하는 것이 아니라 외부 지식저장소(Knowledge Base)에서 관련 문서를 검색하고 이를 근거로 판단하는 방식을 말한다.

- 최종적으로 본 연구는 RAG 기반 AI 모델(exaone, GPT-oss 등)을 중심으로 개발을 진행하였다. 이를 위해 과거의 재분류 경험과 축적된 지식을 기반으로 벡터 DB(Vector Database)를 구축하였으며, LLM이 문서 내용을 모두 읽고 해당 벡터 DB를 참조하여 최종 판단을 수행하도록 시스템을 설계하였다. 이 과정은 기존의 단순 문서 분류형 모델을 넘어, 지식 검색과 추론을 결합한 AI 모델임을 의미한다.

2.6. “이용자 맞춤형 비식별화 기록물 제공 기능” 연구 및 개발

2.6.1. 연구개요

공공기록물의 공개 과정에서 개인정보 및 법인정보의 노출은 법적·기술적 리스크를 초래할 수 있는 중대한 문제이다. 특히 「공공기관의 정보공개에 관한 법률」 제9조 제1항 제6호, 제7호에 따르면, ‘개인의 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보’, ‘법인·단체의 경영·영업상 비밀에 관한 사항으로서 공개될 경우 법인 등의 정당한 이익을 현저히 해할 우려가 있다고 인정되는 정보’ 는 비공개 대상으로 분류된다.

따라서 국가기록원과 같은 기관에서는 기록물을 공개할 경우 비공개 대상정보를 정확히 검출하고, 이를 비식별화(De-identification)하여 안전하게 공개할 수 있는 체계가 필수적이다.

본 연구에서는 AI 기반 비공개정보 검출 및 비식별화 시스템을 설계하였으며, 이 시스템은 다양한 형태의 비정형 문서 비공개 대상 정보를 탐지하고, 의미적 문맥과 규칙 검증을 결합하여 검출한다.

검출 대상은 주민등록번호, 계좌번호, 신용카드번호 등 15종의 숫자와 이름, 주소, 법인(단체)명 등 문자형이 포함되어 검출하도록 하였다.

2.6.2. 연구의 필요성

○ 일반적인 검출 방식의 한계

- 기존의 개인정보 검출 시스템은 주로 정규표현식(Regular Expression) 기반의 패턴 매칭 기술을 중심으로 구성된다. 이 방식은 특정 형식을 가진 데이터(예: 주민등록번호: 13자리, 사업자등록번호: 3-2-5 형식 등)를 빠르게 탐지할 수 있으나, 다음과 같은 한계가 존재한다.

첫째, 문맥 인식의 부재이다.

- 정규식 기반 탐지는 숫자열이나 문자 패턴이 실제 개인정보인지 단순 코드인지 구분할 수 없으므로, “계좌번호”, “신청번호” 등 유사한 형식을 가진 일반 정보도 개인정보로 오탐(誤探)하는 경우가 발생한다.

둘째, 표기 변형에 취약하다.

- 공백, 하이픈, 전각문자, 줄바꿈 등 다양한 표기 변형이 포함된 문서에서는 정규식이 완전 일치하지 않아 검출 누락이 발생할 수 있다.

셋째, 검증 절차의 부재이다.

- 주민등록번호나 사업자등록번호의 경우, 형식은 일치하더라도 실제 존재하지 않는 번호일 가능성이 있어, 단순 정규식 매칭만으로는 신뢰성 있는 비공개정보 검출이 어렵다.

2.6.3. 연구의 방법 및 절차

본 연구에서는 다음과 같은 다단계 접근 방법을 통해 비공개정보 검출 및 비식별화 시스템을 설계하여 기존 방식의 한계를 극복하고자 하였다.

○ 본 연구에 적용한 하이브리드 검출 방법

- 이러한 기존 방식의 한계를 극복하기 위해 본 연구에서는 숫자형과 문자형을 구분하고, 숫자형의 경우 4단계 하이브리드 검출 파이프라인을 구성하고, 문자형의 경우, LLM 프롬프팅방식으로 총 5단계를 거치도록 하였다

- 본 파이프라인은 숫자형의 경우 “정규식 기반 후보 탐색 → 규칙 기반 검증 → 문맥 키워드 검증 → LLM 문맥 보조 검증”의 순차 구조로 이루어져 있으며, 각 단계가 상호 보완적으로 작동하여 정확도 향상과 오탐 최소화를 동시에 달성한다.

- 문자형 정보는 LLM에 프롬프팅을 통한 검출방식으로 문서의 문맥을 기반으로 검출하여 검출된 숫자형 정보와 통합한다.

○ 단계별 주요 검출방안

- 1단계: 정규식 기반 후보 추출

  비식별화 유형별로 정의된 정규표현식을 활용하여 텍스트를 스캔한다.

  전처리 단계에서 문자의 형태를 통일(전각→반각, 하이픈 정규화, 공백 제거)하여 다양한 표기 변형을 허용한다.

  예를 들어, 주민등록번호(YYMMDD-XXXXXXX), 사업자등록번호(XXX-XX-XXXXX), 신용카드번호(15~17자리) 등 규칙적 패턴을 탐색한다.

- 2단계: 규칙 기반 검증

  정규식으로 검출된 후보값에 대해 추가적인 형식 검증 및 의미 검증 규칙을 적용한다.

  주민등록번호: 생년월일 유효성, 성별 및 내·외국인 코드, 마지막 자릿수 검증

  신용카드번호: Luhn 알고리즘 적용

  사업자등록번호: 국세청 체크섬 알고리즘 적용

- 3단계: 문맥 키워드 검증

  계좌번호, 법인등록번호 등은 형식이 유사한 숫자열이 많기 때문에, 주변 60자 내에서 “계좌”, “입금”, “은행”, “송금” 등의 문맥 키워드가 함께 등장하는 경우에만 유효한 정보로 판정한다.

- 4단계: LLM 기반 문맥 보조 검증

  규칙 기반으로 판별이 어려운 케이스는 대규모 언어모델(LLM: Large Language Model)에 전달하여 문맥을 기반으로 최종 판정 및 신뢰도 점수를 산출한다. 이 단계에서 LLM은 주변 문맥의 의미적 패턴을 분석하여, 해당 숫자열이 실제 비식별화 대상 정보인지 여부를 판단하게 하였다.

- 5단계: LLM 기반 검출(문자형)

  규칙 기반으로 판별이 어려운 문자형의 경우 LLM이 문맥을 기반으로 검출하고, 비공개 정보에 추가한다.

- 고신뢰도 유형 즉시 확정처리

  이메일, 전화번호, 차량번호 등은 정규식 검출만으로도 충분히 신뢰도가 높아, 별도의 추가 검증 없이 검출 확정 처리한다.

2.6.4. 시스템의 기술적 특징

○ 본 연구에서 개발한 비공개정보 검출 시스템은 다음과 같은 기술적 특징을 갖는다.

① 오탐 최소화:

정규식 → 규칙 → LLM 순차 검증 구조를 통해 단순 숫자열이나 코드값의 오탐 가능성을 크게 줄인다.

② 문맥 인식 기반 판단:

단순 문자열 매칭이 아니라 주변 키워드와 의미적 관계를 함께 분석하여, 문맥적으로 비공개 대상 정보에 해당하는지를 판단한다.

③ 형식 유연성 확보:

공백, 하이픈, 줄바꿈 등 다양한 표기 변형을 허용하는 정규화 처리로, 실제 문서 환경에 대응할 수 있도록 하였다.

④ 중복 제거 및 효율적 관리:

동일한 비식별화 정보가 여러 번 검출되는 경우, 중복 제거 알고리즘을 통해 단일 결과로 통합한다.

2.7. PoC 개발 및 시범운영

2.7.1. 연구 개요

본 연구의 최종 목표는 AI 기반 공개재분류부터 비공개대상정보 식별·비식별화, 그리고 대국민 서비스 제공까지의 전 과정을 통합한 시범체계(PoC: Proof of Concept)를 구축하는 것이다.
이를 통해 실제 업무 환경에서의 적용 가능성을 검증하고, 기술적 완성도 및 행정 실무 적용성을 동시에 확보하고자 하였다.
PoC는 사업 종료 전 1개월의 기간 동안 시범운영을 할 예정이며, 공개재분류 업무의 자동화와 비공개정보 검출·비식별화 기능을 하나의 서비스 흐름 내에서 통합하여 볼 수 있도록 구현하였다.

2.7.2. 개발구성 및 환경

PoC 환경은 다음과 같은 2단계 구조로 설계되었다.

○ AI 모델 개발 환경 (외부 클라우드 기반)

- 개발 단계에서는 GPU 자원이 충분히 확보된 외부 클라우드 환경을 활용하여 모델을 테스트·튜닝하였다.
- 개발환경에서는 기록물의 보안을 고려하여, 인터넷 및 이미 알려진 문서를 사용하거나 직접 문서를 작성하여 테스트하였다.
- 최신 LLM을 적용하고 테스트하기 위해, 다수의 모델을 설치하고 향후 PoC 환경(워크스테이션급)에 적용할 것을 감안하여 모델의 크기 등도 고려하였다.

항목	내용	비고
AI 모델	Gemma 3:4.3B, gpt-oss:20.9B, Exaone 3.5:7.8B, Exaone-Deep:7.8B, DeepSeek R1:70.6B, LLaMA 3.3:70.6B, LLaMA 4:108.6B, gpt-oss:120B, chatgpt, LLaVA:7B, Qwen2.5VL:8.3B	13종을 설치하여 비교 분석
임베딩 모델	OllamaEmbeddings (bge-m3)	벡터 DB 구성 시 텍스트 벡터화
벡터DB	Qdrant	문서 임베딩 결과 저장 및 검색 기능
프레임워크	LangChain	LLM, 벡터DB, 프롬프트 체인을 연결하는 핵심 프레임워크
H/W	RTX 4090 × 2장	

(표 7 - SW/HW Configuration List)

○ PoC 운영 환경 (내부 행정망 기반)

- 실증운영은 기관 내부 행정망에 설치된 워크스테이션급 서버에서 수행되었다.
- SW 환경은 동일하며, RTX 4070 × 2장으로 구성하여, 별도의 클라우드 의존 없이 독립적으로 서비스가 동작하도록 구성하였다.
- 내부 직원들이 직접 접속하여 테스트 및 결과 검증을 수행할 수 있도록 환경을 구성하였다.

2.7.3. 시스템 구성 및 구현방법

PoC는 실제 업무 흐름을 모사하여 다음과 같은 순서로 작동되도록 구현되었다.

○ 대상 문서 입력 및 업로드

- PoC의 입력 데이터는 텍스트 기반의 PDF 문서이며, CAMS 시스템의 문서 조회 기능을 엑셀 업로드 프로세스로 대체하여 기존 CAMS와는 별개로 작동하되 화면 UI는 CAMS와 동일하게 구성하였다.

○ AI 재분류 기능 수행

- 사용자가 CAMS 내 통합 화면에서 'AI 재분류' 버튼을 클릭하면, 내부에 구축된 LLM 모델이 해당 문서를 분석하여 비공개 유형, 공개여부 판단, 비공개 대상 정보 검출을 수행하여, 그 결과를 메타데이터 형태로 저장하도록 구성하였다.
- 검출된 비공개대상정보는 JSON 형식으로 저장되어 추후 열람신청 단계에서 활용된다.

○ 열람·활용단계에서 비식별화 처리

- 재분류 과정 중 LLM은 문서 내 비공개대상정보(개인·법인정보)를 자동으로 검출하며, 열람신청 시 비식별화 모듈은 해당 내용을 삭제·대체 또는 마스킹 처리한다.

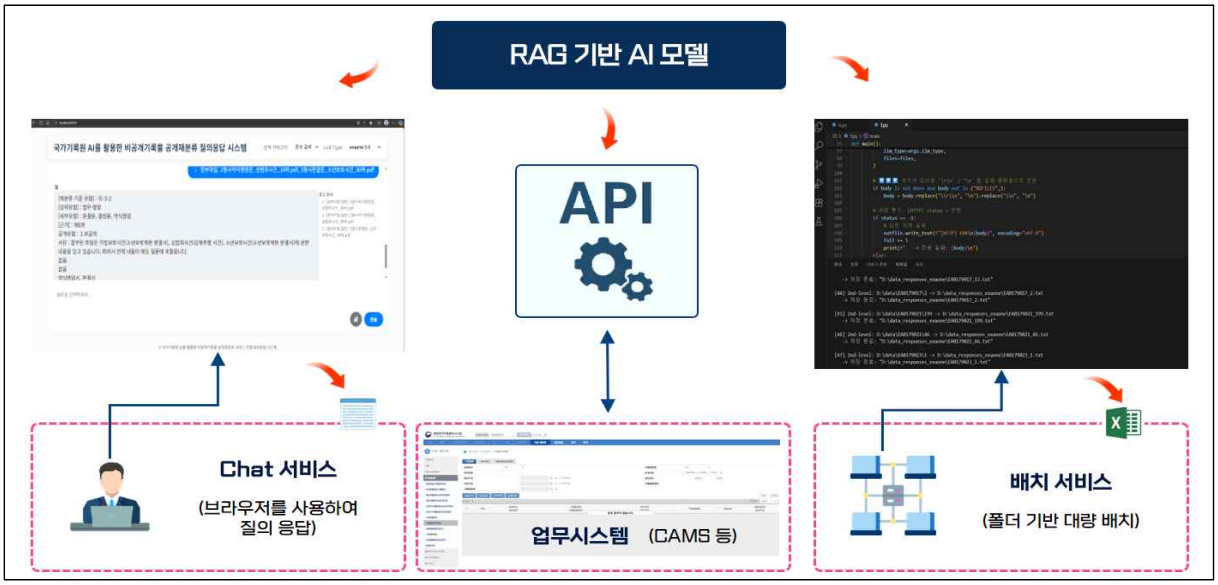
2.7.4. 사용자 인터페이스

○ 개발환경의 사용자 인터페이스

- PoC의 테스트 인터페이스는 Chat 서비스 형태로 구성되어, 사용자가 질문 또는 명령을 입력하면 LLM이 즉시 재분류 결과와 비식별화 상태를 응답하는 구조로 설계되었다.

- 이러한 대화형 구조를 통해 결과 확인까지의 시간을 단축하고, 다양한 문서 유형에 대한 반응성을 실시간으로 검증할 수 있도록 하였다. 또한 다양한 프롬프팅을 적용하여 효과적인 프롬프트를 완성하는데 효과적으로 활용하였다.

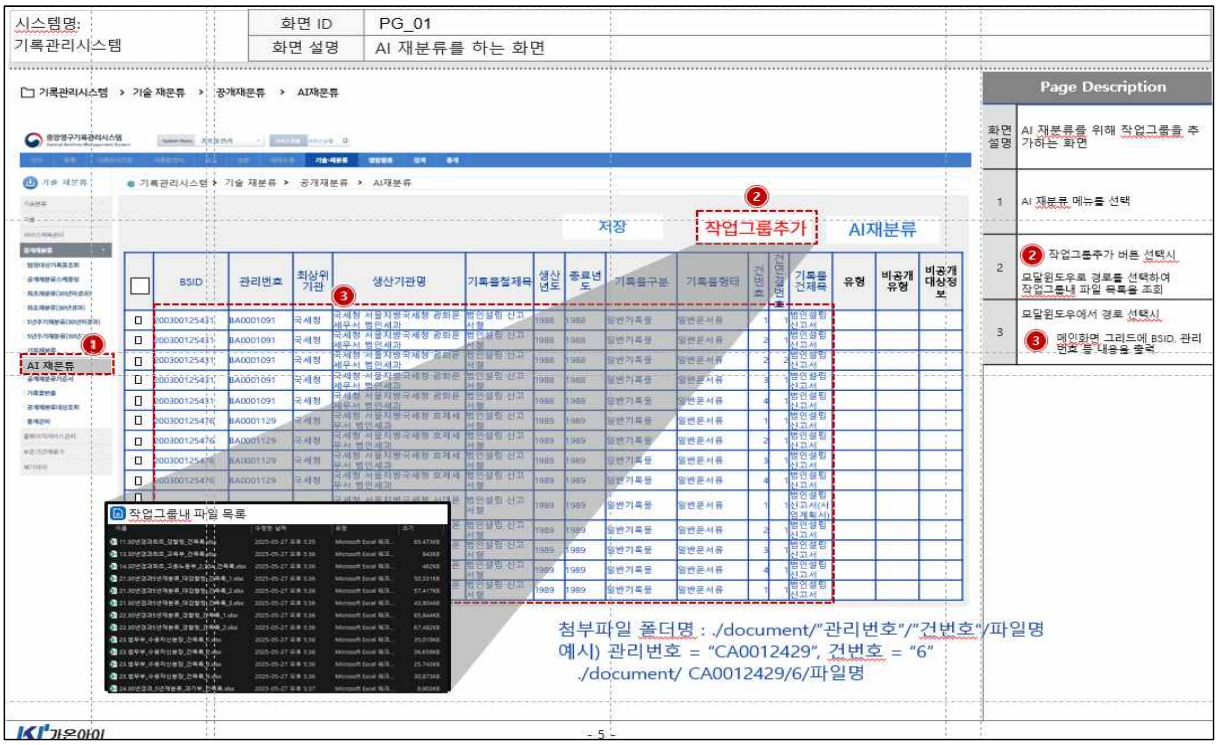
- 또한, 다양한 테스트를 위해서 Char서비스, API 서비스, 대용량 파일 처리로 나누어 외부 인터페이스를 개발하였으며, 테스트 후 PoC 에서는 주로 API서비스를 활용하여 CAMS와 유사한 UI로 테스트가 가능하게 하였다.



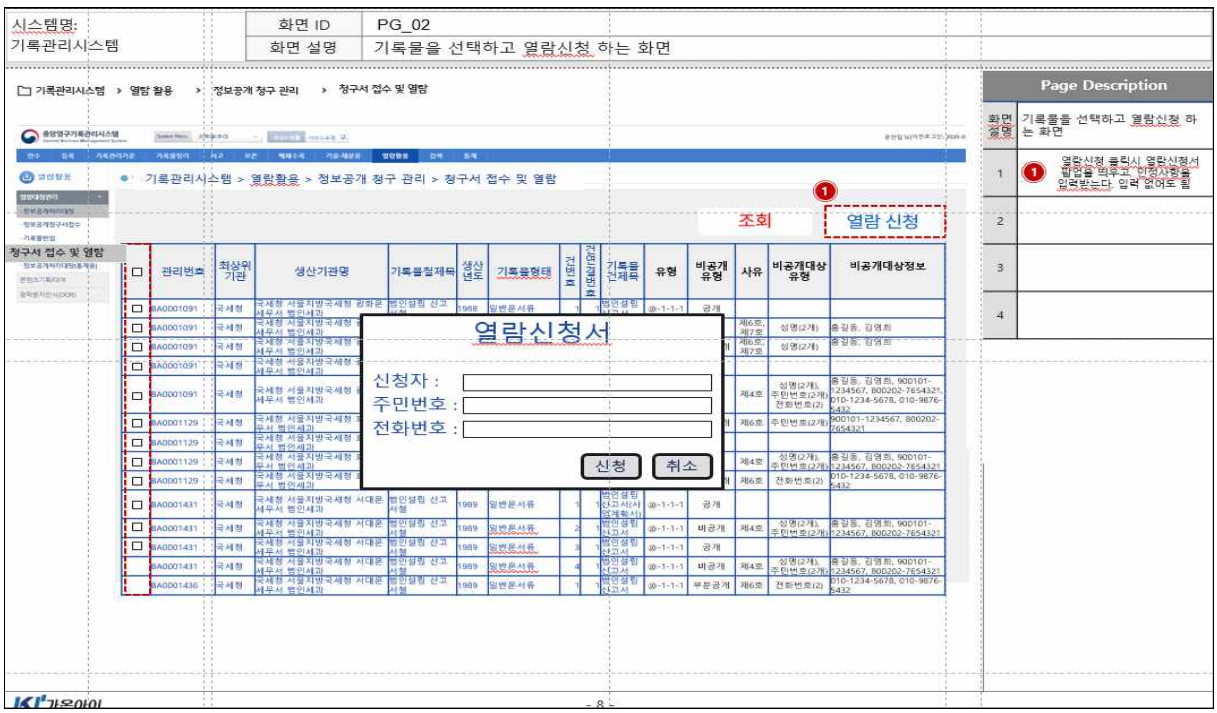
(그림 6 - Development Module)

(그림 7 - Chat Service)

- 시스템 화면은 기존 CAMS 시스템 UI와 동일하게 구성하여 사용자 경험(UX)을 유지하도록 하였으며, 초기화면 설계서는 아래와 같이 작성되었다.



(그림 8 - UI Design Specification 1)



(그림 9 - UI Design Specification 2)

2.8. 기록관리시스템 기능 연계 제안

2.8.1. 기록관리시스템 기능연계 방안

○ 연계방안 개요

- CAMS 및 RMS는 Spring Framework를 기반으로 한 MVC(Model-View-Controller) 아키텍처에 Java 및 JSP를 주요 개발 언어로 사용하는 Java EE(Web) 아키텍처이며, 공개재분류 AI 모델과의 연계방안을 제시하고자 한다.
- 또한 향후 기록물 생산관리통합시스템으로 전환을 위한 통합구조를 고려하여야 한다.

○ 연계 내용

- 전자정부프레임워크 기반의 CAMS 시스템과 공개재분류 AI 모델간의 서비스 간 효율적 연계 구조를 설계한다.
- 공개재분류 및 비식별화 대상 정보의 데이터 처리·저장 체계를 확립한다.
- 향후 기록물 생산관리통합시스템 구축 시의 확장성 및 상호운용성 확보가 필요하다.

○ 연계 구조

- 본 연구과제에서 AI 모델은 API Call 방식으로 설계되었으며, CAMS 시스템과 AI 공개재분류 모델 간의 데이터 교환 또한 업무의 표준화 측면에서 API Call 기반의 연계 방식으로 설계가 필요하다.
- 요청(Request): CAMS에서 API 호출을 통해 분석 대상 기록물의 메타데이터 및 본문 정보를 전달하며, 전달하는 예시는 아래와 같다.

```
curl -X POST http://192.168.0.1:8000/chat \
  -F 'question=대검찰청, 30년 경과' -F 'llm_type=gpt_oss' -F 'files=@sample.pdf'
```

- 응답(Response): AI 분석 서비스는 JSON 포맷으로 분석 결과(공개·부분공개·비공개 등)를 반환하며, 응답결과는 아래와 같이 표현할 수 있다.

```
{
  "re_type": "CM②-4",
  "up_type": "소원·소청·소송",
  "dw_type": "행정심판",
  "re_reason": ", (제4호, 제5호) 제6호",
  "doc_summary": "**문서 유형:** 행정심판 청구서,..... (청구 기간)",
  "disclosure_type": "비공개",
  "re_reason2": "판단 대상 문서는 행정심판 청구서이며, .... 않았습니다.",
  "pii_counts": {
    "전화번호": 2,
    "name": 2,
    "address": 1
  },
```

```
  "pii_value": {
    "extractions": [
      {
        "type": "전화번호",
        "value": "010-1234-5678",
        "validated": true,
        "validators": [
          "regex",
          "rules"
        ],
        "confidence": 0.9
      },
      {
        "type": "name",
        "value": "洪吉童",
        "validators": "Only_LLM",
        "confidence": 0.9,
        "validated": true
      },
      {
        "type": "address",
        "value": "서울특별시 강남구 테헤란로 123 (역삼동, 000빌딩 10층)",
        "validators": "Only_LLM",
        "confidence": 0.9,
        "validated": true
      }
    ]
  },
  "chunk_index": "Z02-0038"
}
```


- 전송 프로토콜: HTTPS 기반 RESTful API 구조로 구성하며, 전송 데이터는 표준 스키마(JSON Schema)를 따른다.

○ 데이터 흐름 및 관리 방식

- 입력 단계 : CAMS에서 문서 메타데이터 및 원문 텍스트 본문을 AI 분석서비스로 API 송신한다.
- 분석 단계 : AI 모델이 공개재분류 기준에 따라 문서 내용을 자동 분류 및 비식별화 항목 식별한다.
- 출력 단계 : 결과값(JSON)을 CAMS로 반환하여, 분류 결과를 기록물 속성으로 저장한다.
- 저장 단계 : 분석 결과 및 생성 데이터는 기록물 생산관리통합시스템 구축시 '보존·활용 꾸러미'에 포함되하여, 장기 저장 및 활용 기반으로 관리한다.

2.8.2. 기록관리기관으로 확산·보급방안

○ 대상기관 범위 및 고려 사항

- 본 사업의 확산 대상은 전국 지자체 및 공공기관 중 기록물관리시스템(RMS, Records Management System)을 구축·운영 중인 기관을 중심으로 설정한다
- 중앙행정기관 및 산하기관의 RMS와의 연계성 검토를 통해, 향후 통합 서비스 확장이 가능한 구조를 고려해야 한다.
- 특히, 기관별 시스템 환경(플랫폼 구조, 데이터 표준, API 구성 등)을 사전 분석하여 기술적 호환성 확보가 필요하다.

○ 기술적 측면 검토

- 각 기관이 보유한 공개재분류 지식데이터를 수집·분석하고, 이를 기반으로 RAG(Retrieval Augmented Generation) 구조를 설계하여야 한다.
- RAG 설계 시 다음 사항을 고려해야 한다.
  ① 문서 검색 모듈, 임베딩 DB, 응답 생성기 등 핵심 구성요소의 표준화
  ② 기관별 데이터 포맷 차이를 최소화하기 위한 데이터 표준 스키마 정의
  ③ 모델 학습용 지식데이터 세트의 전처리 작업 및 비식별화 처리 기준 수립
- AI 모델은 공개·비공개 판단 및 비식별화 항목 자동 추출 등 기록관리 업무 자동화 서비스를 전제로 구축한다.
- 기술 검증을 위해 시범기관 기반의 파일럿 RAG 모델을 운영하여, 확산 전 단계별 안정성을 확보한다.

○ 운영적 측면 검토

- 기관별 시스템 운영 여건을 고려하여 이중 운영모델(독립형·공동형) 로 구분한 추진 전략이 필요하다.
  ① 독립형 모델 : 개별 기관이 자체 RMS 환경에 맞추어 AI 서비스를 구축·운영
  ② 공동형 모델 : 인접 지자체 또는 동일 업무 특성을 가진 기관 간 공동 AI 서비스 구축 및 데이터 공유 체계 운영
- 공동형 모델의 경우, 데이터 보안 및 접근권한 관리, 책임 범위 명확화, 모델 학습 데이터의 공동 관리체계 마련이 필요하다.
- 운영 표준화를 위해 중앙기록관리기관이 기술 사양 및 인터페이스 표준을 관리하고, 지자체는 이를 준용하여 자체 서비스로 확산하는 체계를 갖추어야 한다.

○ 행정·재정적 측면 검토

- 사업 추진을 위한 예산 확보 체계와 행정지원 절차를 사전에 수립해야 한다.
  ① 기관 예산 구조 내 정보화사업 항목 또는 RMS 고도화 사업 항목을 활용하여 예산 반영 가능성을 검토
  ② 국가기록원 등 중앙기관과 협력하여 공동 예산 편성 및 국비지원 연계 방안 마련
- 예산 및 사업 추진 절차는 다음과 같은 단계적 로드맵을 전제로 한다.
  ① 1단계(파일럿 구축) - 시범기관 중심의 AI 서비스 적용 및 성능 검증
  ② 2단계(확산 단계) - 성과를 기반으로 광역 및 기초지자체로 점진적 확대
  ③ 3단계(통합 운영 단계) - 전국 단위의 기록물관리 통합플랫폼 내 서비스 모듈로 편입
- 사업 전반의 행정적 효율성을 위해 중앙·지자체 간 협력 거버넌스 구축 및 사업관리지침 표준화가 필요하다

○ 확산·보급 방안은 기술·운영·행정 세 영역을 고려하여야 함
- 기관별 RMS 환경의 이질성, 데이터 표준화 수준, AI 서비스 관리 주체의 명확화가 사전 해결되어야 한다. 따라서 확산 추진 이전에 다음과 같은 선행 과제가 필요하다.
  ① RAG 기술표준 시범 검증 및 공통 API 정의
  ② 기관별 데이터 진단 및 표준화 수준 평가
  ③ 공동 구축형 모델의 법·제도적 근거 마련
  ④ 예산지원 및 운영비 배분체계 설계

제3장 총괄연구개발과제의 최종 연구개발 결과

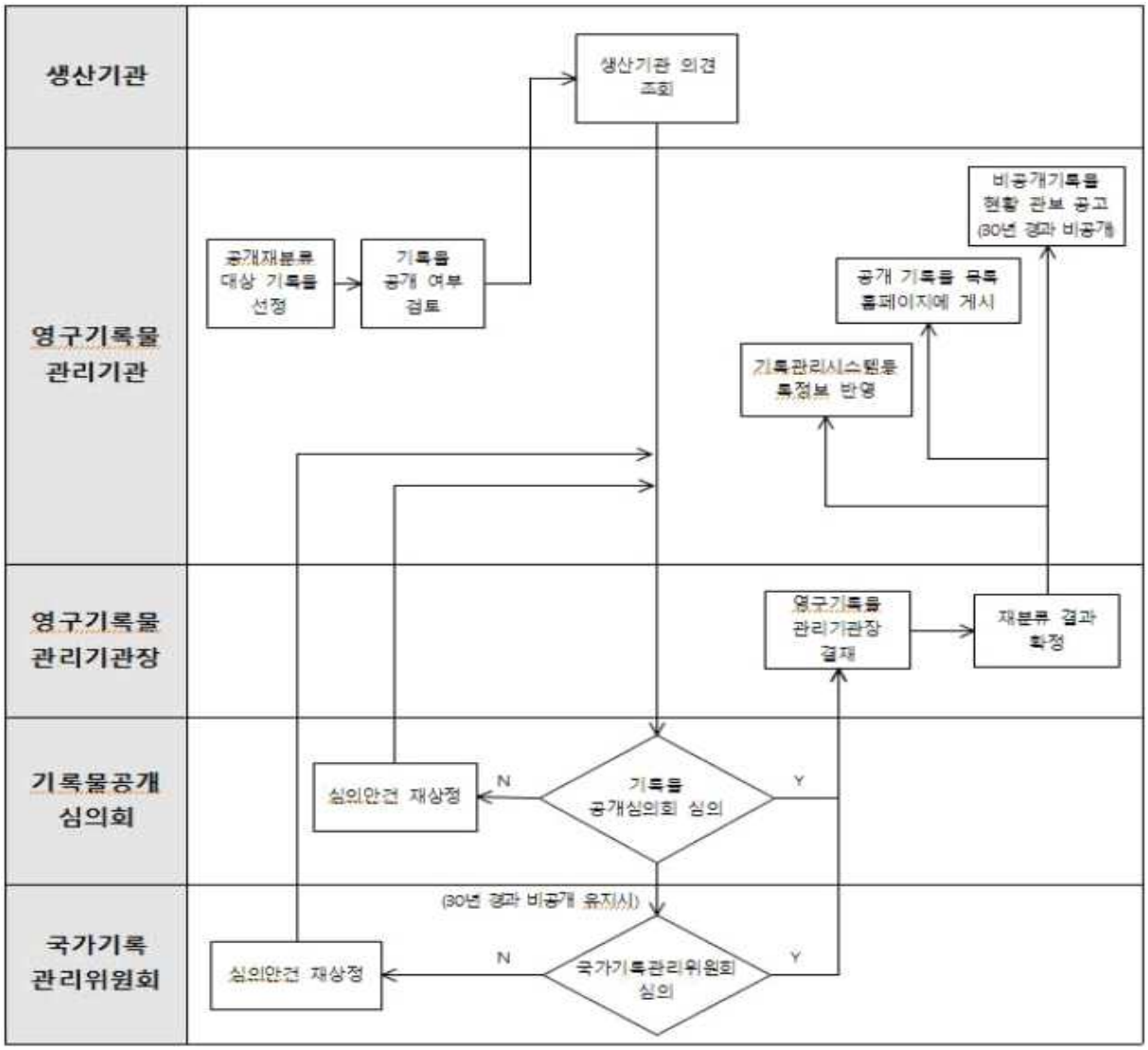
3.1. 공개재분류 업무 지능화

3.1.1. 업무 분석

공개재분류란 “기록물의 비공개 혹은 부분공개 여부를 검토하여 재분류하는 것”이다(국가기록원, 2024). 공개재분류 업무의 법적 근거는 「공공기록물 관리에 관한 법률」 제35조(기록물의 공개 여부 분류)이다. ①항에 따라 국가기록원으로 이관되기 전에 공개 여부를 재분류한다. ②항에 따라 재분류 이후 5년 주기로 공개재분류를 실시한다. 이 작업은 ③항에 따라 생산연도 종료 후 30년까지 반복된다. 동법 시행령 제72조(기록물의 공개여부 분류)에서는 비공개 시에는 비공개 사유를 제출할 의무, 공개 여부를 구분하여 관리할 의무 등을 추가로 명시하고 있다.

기록물의 공개 여부를 결정하는 근거법은 「공공기관의 정보공개에 관한 법률」 제9조(비공개 대상 정보)이다. ①항 1-8호에 비공개 대상 정보를 정의하고 있다. 영구기록물관리기관이 각급 기관으로부터 이관받아 소장하고 있는 기록물은 공개를 원칙으로 한다. 각 호의 범위에 해당하지 않는 기록물은 공개하여 비공개 대상 정보를 최소화하여야 한다. 정보공개 여부를 심의할 조직은 동법 제12조(정보공개심의회)에 따라 설치된다.

기록관리 공공표준(국가기록원, 2024b), 지침(국가기록원, 2024a ; 2025), 매뉴얼(국가기록원, 2010)은 보다 구체적인 절차를 안내하며, 공개재분류 업무의 프로세스는 다음과 같다.



(그림 10 - The Disclosure Status Reclassification Process (국가기록원, 2024b))

국가기록원의 경우, 공개재분류 기준서 및 검토서 작성에 몇 가지 불편사항이 있다(권미현, 2019).

- ❍ 공개재분류 기준서 작성 시 반복 문구 기재 작업의 피로함
- ❍ 기준서를 토대로 건 입력 시 공개 유무 오입력
- ❍ 업무 담당자에 따라 동일한 기관도 업무 세분류 값이 상이함
- ❍ 국가기록원 원내 규정에서 공개재분류 대상 기관 전체에 생산기관 의견조회를 하도록 정하고 있어 업무 부담[5)]
- ❍ 기 심의 이력이 있는 유형도 재분류 절차를 반복 준수
- ❍ 기준서 서식의 복잡성으로 인한 업무 과부하

공개재분류 업무에 AI를 적용하고, 현재 업무 부담을 겪고 있는 공개재분류 기준서 및 검토서 작성 보조

5) 「공공기록물 관리에 관한 법률」에서는 제35조 ⑤항을 통해 통일·외교·안보·수사·정보 분야의 기록물을 공개할 때에 생산기관 의견조회를 하도록 정하고 있다.

에 집중한다면, 전반적인 업무 효율을 바로 높일 수 있을 것이다. 해당 업무는 다음과 같이 세분화할 수 있다(권미현, 2019).

❍ 기관명 재분류
- CAMS(Central Archives Management System, 중앙영구기록물관리시스템)는 생산 당시 기관명을 저장
- 생산기관 의견조회는 현재 기관명을 기준으로 수행
- 직제규정, 해당기관 조직도 등을 참고하여 현재 기관명으로 재분류

❍ 기록물 유형 분류
- 23개 기록물 유형으로 분류 (6개 공통 업무, 17개 고유 업무로 구성)
- 공통업무 : ①인사·징계, ②법령·예규, ③감사, ④국유재산, ⑤소청·소송, ⑥인·허가
- 고유업무 : ①외교, ②수사, ③정보·보안, ④외사, ⑤법무·행형, ⑥재산, ⑦시설, ⑧문화재관리, ⑨개인, ⑩병무, ⑪건설·토목, ⑫토지, ⑬지방행정, ⑭사회·복지, ⑮국적관리, ⑯학적관리, ⑰기타

❍ 기록물 정보 기재
- 기록물 생산기관 및 처리부서 기재
- 기록물철명 기재
- 생산연도, 수량 기재
- 기록물 개요 및 상세 내용 기재
- 기록물 기존 공개 이력 기재

❍ 기록물 재분류 의견 및 사유 기재
- 「공공기관의 정보공개에 관한 법률」 제9조(비공개 대상 정보) ①항 1-8호에 따른 비공개 대상 정보 유무 파악
- 재분류 의견 기재
- 재분류 사유 및 근거 기재
- 비공개 시 해당 비공개 대상정보 기재

이 중 기관명 재분류 작업은 AI 없이도 기관 전거데이터를 통해 자동화가 가능할 것으로 보인다. 기록물 정보 기재 작업의 자동화도 AI가 필수 요구사항은 아니다. 다만 정형 텍스트를 대량으로 추출·구조화하고, 누락·불일치를 검증하는 데에는 AI가 강점을 발휘할 수 있다.

기록물 유형 분류 작업과 기록물 재분류 의견 및 사유 기재 작업은 단순 전거데이터나 기존 메타데이터 처리만으로는 한계가 있다. 기록물의 맥락과 본문 내용을 분석해야 하므로, 텍스트 분류·자연어 처리와 같은 AI 기술을 적용할 필요가 있다. 이러한 시도를 통해, 단순·반복적인 업무를 줄이고 업무 담당자는 수준 높은 의사결정에 집중할 수 있을 것이다.

3.1.2. 기술동향 조사

○ 인공지능(AI, Artificial Intelligence)

- 1956년 미국의 존 매카시(John McCarthy)가 Dartmouth Summer Research Project on Artificial Intelligence에서 처음 사용한 용어로(최수진, 2023) '(인간) 학습의 모든 측면이나 지능의 특징들이 충분히 정밀하게 기술된다면 기계가 모방할 수 있다(J. McCarthy, M. L. Minsky, N. Rochester, 1955)'는 가정에 기반하여 출발한 기술 개념이다.
- 존 매카시는 인공지능을 '지능적인 기계, 특히 지능적인 컴퓨터 프로그램을 만드는 과학이자 공학'이라고 정의하였다.[6)]
- 2025년 제정되고 2026년 시행 예정인 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」에서는 인공지능을 '학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것'으로 정의한다.
- 개념적으로 강한 인공지능과 약한 인공지능으로 구분된다. 강한 인공지능은 범용인공지능(AGI, Artificial General Intelligence)라고도 불리며, 일반적인 업무, 여러 업무를 수행한다. 약한 인공지능은 특정 분야에 특화된 인공지능이며, 대부분의 인공지능이 여기에 속한다(원동규, 이상필, 2016).

○ 머신러닝(Machine Learning)

- 머신러닝은 인공지능의 한 분야로, 패턴인식(Pattern Recognition)이라고도 불린다. 뇌의 학습을 모방하여 명시적인 프로그래밍 없이 컴퓨터가 학습하는 것을 의미한다. 머신러닝은 알고리즘을 활용하여 입력된 데이터를 모델이나 새로운 알고리즘으로 개발하고, 이를 새로운 데이터에 적용하는 것을 포함하는 개념이다.[7)]
- 머신러닝에서 인간이 수행해야 하는 사전 작업은 다음과 같다(Giles Hooker, Cliff Hooker, 2017)
  - 예측 작업 식별
  - 알고리즘 선택
  - 입출력 지정
  - 적절한 데이터 세트 확보

○ 딥러닝(Deep Learning)

- 2006년 캐나다의 제프리 힌튼(Geoffrey E. Hinton)이 고안하였으며 '컴퓨터가 스스로 외부 데이터를 조합하고 분석·학습하는 것'이다(최수진, 2023).

---

6) http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html Professor John McCarthy 참조. (2025년 09월 18일 최종 접속).

7) A State Archives and Records initiative for the NSW Government. (2017). Machine Learning and Records Management. 출처 : https://futureproof.records.nsw.gov.au/machine-learning-and-records-management/ (2025년 09월 18일 최종 접속).

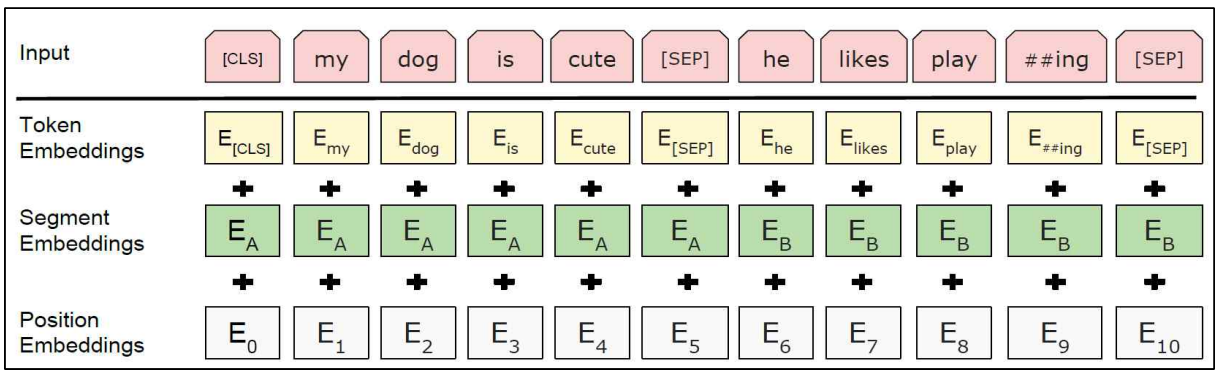
- 머신러닝을 다층으로 조합한 것이다. 딥러닝은 여러 층의 머신러닝을 거쳐 적절한 특징값(입력값)을 스스로 생성한다. 따라서 사람이 직접 특징량을 추출할 필요가 없는 것이 최대 장점이다. 딥러닝을 통해 인간이 포착하지 못한 특징값까지도 스스로 생성하는데, 이런 학습은 인간의 인식 과정과 유사하다(원동규, 이상필, 2016).

❍ 언어 모델(LM, Language Model)(박찬준, 이원성, 김윤기 외, 2023)

- 인간의 언어(자연어)를 컴퓨터로 처리하기 위해 컴퓨터가 이해할 수 있는 형태로 변환해주는 모델이다.
- 전통적인 언어 모델은 숫자체계를 기반으로 발전했다. 대표적인 예시로 원-핫 인코딩(one-hot encoding) 방법이 있다. 이 방법은 단어 간의 의미 연관성을 고려하지 못하는 한계를 가진다
- 의미기반 언어 모델은 단어를 밀집 벡터(dense vector)공간에 표현하여 유사한 단어를 가까운 공간에 학습시킨다. Word2Vec 모델이 가장 대표적이다. 같은 문자를 같은 벡터로 표현하기 때문에 문맥을 읽고 이해해야 하는 동음이의어 처리에 약하다.
- 문맥기반 언어 모델은 문맥 정보를 반영하여 언어를 이해한다. 초기 모델은 단방향 문맥 정보만 활용했지만, 이후 양방향 문맥을 전부 활용하는 모델이 개발되었다. 긴 길이의 텍스트에 약하다는 한계가 있다. Google의 BERT 모델도 문맥기반 언어 모델에 해당한다. 문맥기반 언어 모델 연구에서 학습 데이터의 크기와 모델의 성능 간의 상관관계가 주목받았고, 대규모 언어 모델의 개발로 이어졌다.

❍ BERT(Bidirectional Encoder Representations from Transformers)(Jacob Devlin, Ming-Wei Chang, Kenton Lee 외, 2018)

- 2018년에 구글에서 공개된 문맥기반 언어 모델이다.
- 트랜스포머 아키텍처 기반의 양방향 학습 모델이다.
- BERT 입력 표현은 해당 토큰 임베딩(Token Embeddings), 세그먼트 임베딩(Segment Embeddings), 위치 임베딩(Position Embeddings)의 조합으로 구성된다.



(그림 11 - BERT input representation)

- 기존의 단방향적인 문맥기반 언어 모델의 한계를 극복하기 위해 제안되었다.
- 아래의 두 가지 사전학습이 도입되어, 문장에서 가려진 단어나 구문을 양방향 문맥을 기준으로

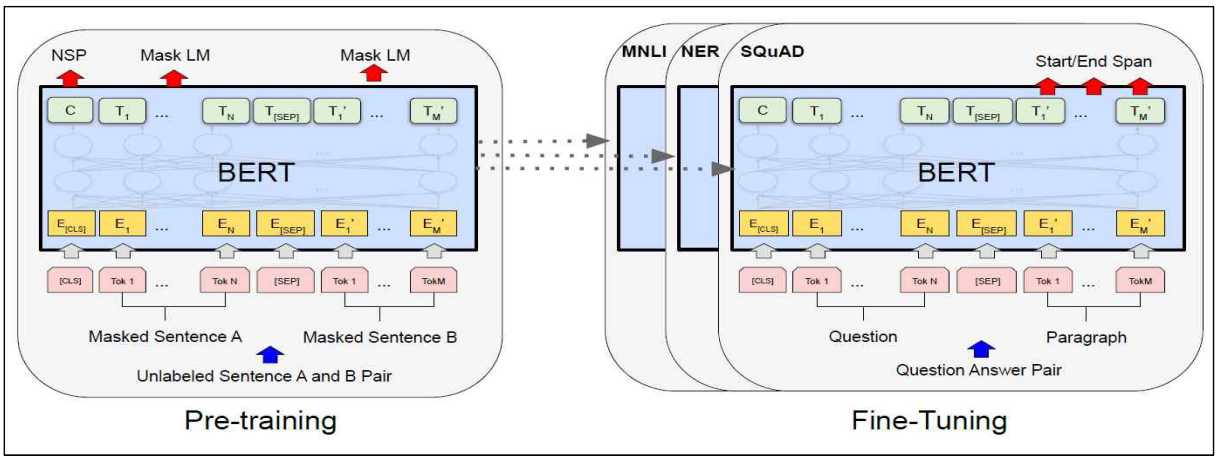
예측하는데 특화된 모델이다.

◦ MLM(Masked Language Model)

◦ NSP(Next Sentence Prediction)

- 11개의 자연어처리 태스크에서 SOTA(State-of-the-Art)를 달성했다.
- 입력 표현 구조가 통합적이고, 문장 간 관계 학습에 강하다는 특징이 있다.

사전학습과 파인튜닝에서 동일한 아키텍처 구조를 사용하여 별도의 조정 없이 사용할 수 있다.



(그림 12 - The Same Architectures between pre-training and fine-tuning for BERT)

❍ 대규모 언어 모델(LLM, Large Language Model)

- 방대한 파라미터를 갖고 데이터를 학습하는 언어 모델이다.
- 대규모 언어 모델은 다음과 같은 장점과 한계점을 가진다(장나은, 2024).
  - ◦ 장점
    - ▪ 방대한 지식 기반을 통한 심층적 응답
    - ▪ 자연스러운 언어 처리와 생성 능력
    - ▪ 다양한 태스크 수행 능력과 높은 범용성
    - ▪ 사용자 친화적인 인터페이스 제공
    - ▪ 다양한 산업 분야에서의 응용 가능성
  - ◦ 한계점
    - ▪ 학습데이터의 편향성을 그대로 학습하여 사회적 편향과 선입견을 반영한 답변
    - ▪ 사실이 아닌 것을 사실인 것처럼 답변하는 할루시네이션(Hallucination)
    - ▪ 짧은 텍스트에는 강하지만 긴 텍스트나 고맥락 언어에 약함
    - ▪ 일관성 부족과 확률적 응답의 문제로 높은 신뢰성을 요구하는 태스크에 부적절
    - ▪ 윤리적 문제와 악용 가능성
- 대규모 언어 모델을 구성하는 과정은 백본(Backbone) 생성과 파인튜닝으로 나뉜다(임철홍, 2023).
  - ◦ 백본(Backbone)
    - ▪ 사전학습으로 완성된 LLM을 말함

▪ 백본 생성 절차는 아래와 같음

· 원시 데이터 셋 수집 및 정제

· LLM 학습 모델 선정 및 파라미터 설정

· 사전 학습 수행

▪ 사전학습을 통해서 언어 능력 및 기본 지식을 모델에 학습시킴

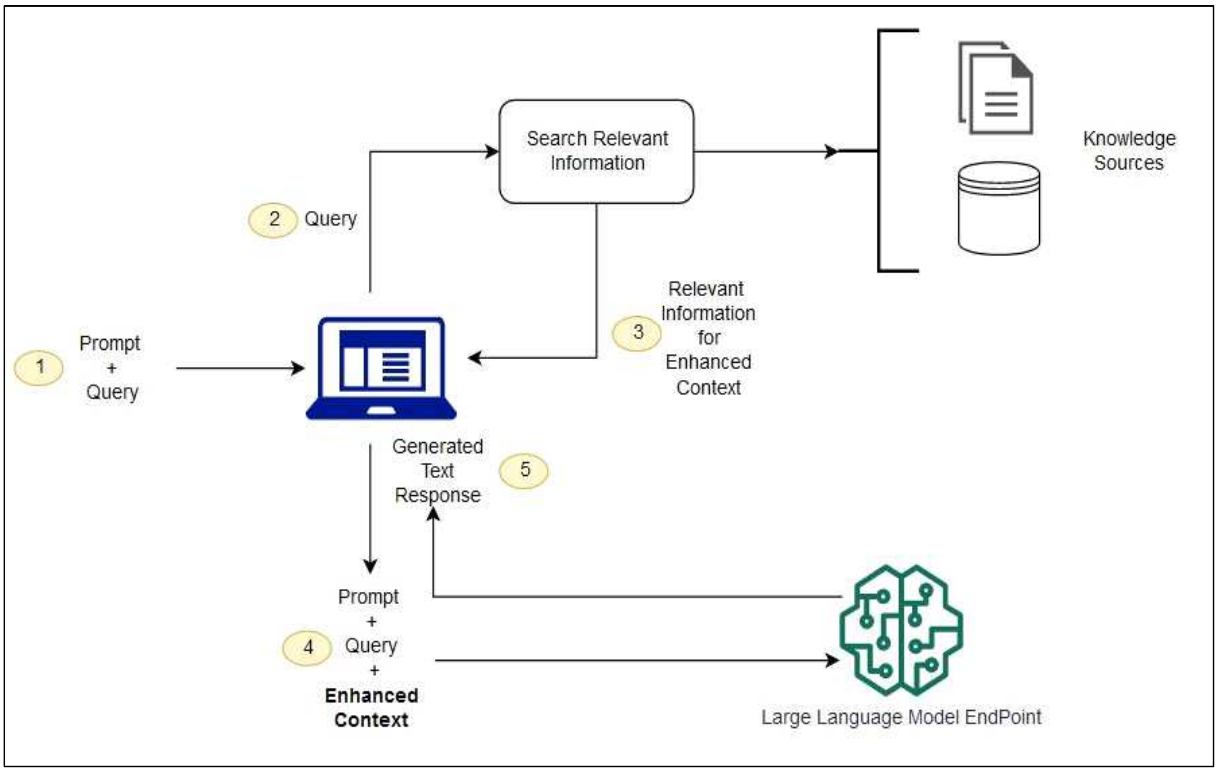
◦ 파인튜닝(파인튜닝)

▪ 백본에 추가 학습을 시켜 특정 작업에 맞게 성능을 향상시키는 과정

▪ 특정 데이터 세트를 바탕으로 내부 파라미터를 조정하여, 일정한 형식과 어조를 유지하며 답변

▪ 파인튜닝을 통해서 활용 목적에 알맞은 LLM 구축

❍ 검색 증강 생성(RAG, Retrieval Augmented Generation)(장나은, 2024)



(그림 13 - Conceptual Stream of LLM with RAG)

출처 : https://aws.amazon.com/ko/what-is/Retrieval Augmented-generation/

- 2020년 Meta의 AI 연구팀이 제안한 기술 개념이다(Patric Lewis, Ethan Perez, Aleksandra Piktus 외., 2020).
- 대규모 언어 모델의 강력한 텍스트 생성 능력에 외부 지식 데이터베이스를 결합하여 응답을 최적화하는 기술이다.
- 인덱싱(Indexing), 검색(Retrieval), 생성(Generation)이라는 세 가지 주요 요소의 결합으로, 대규모 언어 모델이 학습된 데이터에 의존하지 않고 외부 데이터 소스를 활용하는 기술이다.
- 대규모 언어 모델에 추가 학습을 시키는 파인튜닝과는 다른 개념이다. 검색 증강 생성은 모델을 다시 학습시키는 과정 없이도 응답을 개선시킬 수 있다. 모델 학습 과정을 생략할 수 있어 작업

시간과 비용 측면에서 효율적이다.

- 작동 프로세스는 2단계, 데이터 준비(Data Preparation)과 데이터 검색(Data Retrieval Augmented Generation)이다.
  - 1단계 : 데이터 준비(Data Preparation)
    - 데이터 소스 식별 및 수집
    - 데이터 추출 및 정제
    - 청크 분할
    - 임베딩 및 벡터화
    - 데이터 인덱스 생성
  - 2단계 : 데이터 검색(Data Retrieval Augmented Generation)
    - 검색
    - 생성
- 검색 증강 생성 기술은 개방형 도메인 질의응답, 사실 확인, 슬롯 채우기 등 다양한 자연어 처리 수행에 우수한 성과를 보였다.
- 대규모 언어 모델은 최신성이 부족한데, 검색 증강 생성 기술을 통해 이런 한계를 극복할 수 있다.
- 검색 증강 생성의 한계는 다음과 같다.
  - 검색 품질에 대한 의존성
  - 높은 계산 복잡도
  - 개인정보 문제
  - 신뢰성과 편향성 문제
- 검색 증강 생성의 발전 단계는 다음과 같다.
  - Naive RAG
  - Advanced RAG
  - Modular RAG

○ 벡터 데이터베이스[8)]

- 데이터를 벡터값으로 저장하여, 대용량 데이터를 빠르게 검색하고 유사성을 분석하는 데 탁월한 성능을 발휘한다.
- 벡터 데이터베이스는 검색 증강 생성 애플리케이션의 기본 요소이며, 데이터 처리 방식은 다음과 같다.
  - 객체는 임베딩 모델을 사용하여 벡터로 변환된다.
  - 벡터는 특수 데이터 형식으로 저장된다.

---

8) https://docs.aws.amazon.com/ko_kr/prescriptive-guidance/latest/choosing-an-aws-vector-database-for-rag-use-cases/vector-databases.html Amazon Web Service 참조. (2025년 09월 22일 최종 접속).

- 벡터 데이터베이스의 주요 장점은 다음과 같다.
  - 벡터 데이터베이스를 사용하면 유사한 값을 빠르게 찾을 수 있다.
  - 고차원 데이터를 효율적으로 처리할 수 있다.

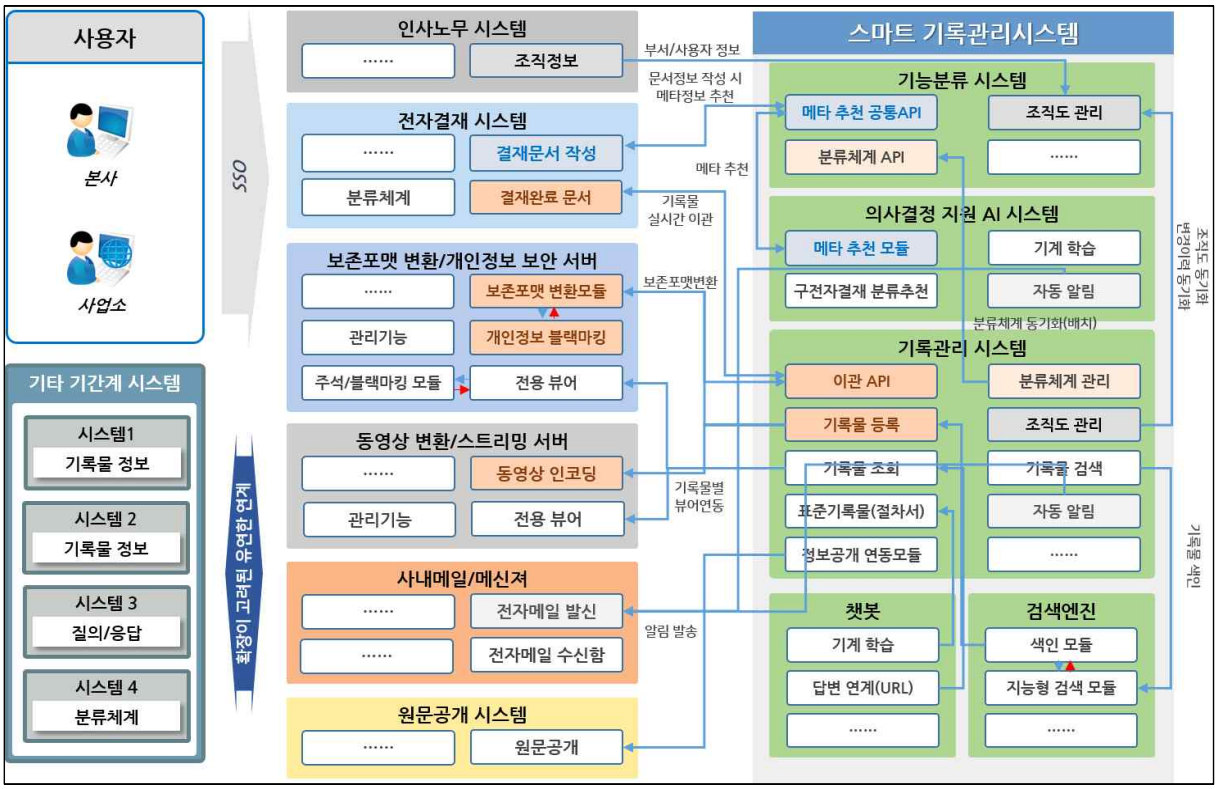
3.1.3. 사례 조사

3.1.3.1. 국내 사례

○ ㈜스토리안트의 ANT-d(송주형, 2021)
- 기계학습 기반 지능형 기록물 공개관리시스템(특허 제10-1627550호)이다.
- 기록물에 포함된 민감정보를 검출하고 공개구분을 자동으로 검토한다.
- 규칙기반기술과 딥러닝 학습모델의 하이브리드 방식이다.
- 전국 지방자치단체 공개재분류결과 목록 4,475,202건과 정보공개포털 등 검색사이트를 통해 4,506,893건의 기록물 건 정보를 학습하였다.
- 407,893건의 기록물을 대상으로 해당 솔루션 추천 값과 실무자 공개재분류 값을 비교한 결과는 아래와 같다.
  - 6,296건을 '확인요망' 처리하였다.
    - 깨진 파일이나 오류 파일 등이 대부분이었다.
  - 솔루션과 실무자의 작업 일치율은 48.61%(195,233건)으로 나타났다.
    - 솔루션은 개인정보를 포함한 기록물을 비공개로만 추천하였다.
    - 실무자는 개인정보를 포함한 기록물을 부분공개, 비공개로 나눠서 작업하였다.
    - 솔루션은 여러 비공개 사유를 한 번에 식별하지 못하고 단일 호수만 추천하였다
  - 개인정보가 검출된 22,861건에 한하여 솔루션과 실무자의 개인정보 식별 일치율은 88.66%이었다.
    - 실무자가 놓친 개인정보를 솔루션이 식별하기도 하였다.

○ 한국중부발전의 i-Works(주현우, 2019)
- 2019년 구축된 통합형 기록생산 및 관리시스템이다.

  전자결재시스템, 기록관리시스템, 기능분류시스템, 의사결정지원시스템으로 구성된다



(그림 14 - i-Works System Configuration Diagram)

- 스마트 기록관리시스템 내에 의사결정 지원 AI 시스템이 있어 AI 기반 문서 자동 분류를 지원한다.
- 다중분류체계, 정보공개, 접근권한 입력을 위한 지원 기능이 있지만, AI 기반 문서 자동 분류 지원 도구는 업무·주제별 분류에 가깝다.

○ 국가철도공단의 차세대 기록관리시스템(한국철도시설공단, 2018)

- 2019년 구축된 인공지능 기반의 기록관리시스템이다.
- 모든 생산기록물을 인공지능 기반으로 통합관리하기 위해 정부 표준 기록관리시스템 기능을 바탕으로 철도공단 맞춤형 시스템이다.
- 공단 전자문서시스템과 연계하여 기록물 생산시점에서 기록물 분류체계에 따라 자동으로 분류한다.
- 앞의 사례와 마찬가지로 기록물을 AI를 활용하여 분류하지만, 업무 기능분류와 가깝고 공개재분류는 지원하지 않는다.

3.1.3.2. 대한민국 정부의 초거대 AI 연구개발현황

○ 연구개발 배경 및 국내 현황

- 2020년 이후 전 세계적으로 초거대 AI(LLM, 대규모 딥러닝 기반 생성모델) 개발 경쟁이 심화되고 있으며, 대한민국 정부 역시 글로벌 경쟁력 확보와 디지털 국가 전략의 일환으로 초거대 AI

연구개발을 국가 중점과제로 설정하였다. 2025년 기준 주요 정부 정책은 'AI 3대 강국 도약'과 '공공부문 초거대 AI 플랫폼 구축'에 집중되어 있다.

- 국내에서는 네이버, 카카오, LG AI연구원 등 민간주도의 LLM 개발과 함께 정부가 직접 자금을 지원(2025년 2,100억 원 이상, 2027년까지 총 5조 원 이상 투입 예정)하여 국가대표 모델 선정 및 육성을 추진 중이다. 올해 삼성SDS 컨소시엄이 범정부 초거대 AI 공통기반 사업의 우선협상대상자로 선정되었으며, 민관 협력형 클라우드센터 구축 및 GPU 인프라의 대규모 확보를 목표로 한다.

○ 주요 추진 사업 및 생태계 구축

- 과학기술정보통신부와 한국지능정보사회진흥원(NIA)은 2025년 '초거대AI 확산 생태계 조성사업(3차, 4차)'을 진행하여, 기업·기관·대학·지자체 등이 참여하는 대규모 컨소시엄을 지원하고, 서비스 개발을 위한 데이터 공동구매, GPU 최대 1만 장 공급, AI 데이터센터 및 클러스터 조성, 인재 양성 프로그램을 병행 추진 중이다.
- 사업 예산 및 지원 방식: 지원규모는 매년 증가 중으로 2025년 약 2,180억 원, 2027년까지 5조 원 이상이 투입될 계획이며, AI 성능 향상, 산업 현장 및 공공기관 맞춤형 플랫폼 적용, 국민 평가가 반영된 '공개 리더보드' 운영 등 성능 검증 과정도 병행한다.
- 대표 컨소시엄: LG AI연구원(엑사원), 네이버(하이퍼클로바), 삼성SDS, KT, 카카오 등이 각 분야별 전략과 기술 경쟁력을 바탕으로 참여하며, 2027년 최종 1~2개 대표 모델 선발 예정이다.

○ 공공부문 적용 현황과 기술적 접근

- 공공부문에서는 범정부 초거대 AI 플랫폼(생성형 AI 서비스, 클라우드 기반) 도입이 본격화되고 있으며, 각 부처 및 지자체에서 안전하게 AI 서비스를 활용할 수 있도록 기술 신뢰성 및 보안 요건이 강화된 맞춤형 플랫폼 제공에 중점을 두고 있다. 정부는 AI 서비스의 실증 및 성능 검증, 인공지능 신뢰성·책임성·윤리성 확보를 위한 정책적 지원도 병행 중이다.
- 공공기록관리 업무에도 초거대 AI기반 자동 분류, 민감정보 식별 및 생성형 모델의 적용이 확대되고 있는데, 주요 해외 사례분석을 바탕으로 국내 특화 LLM 도입, API 기반 시스템 통합, 대용량 파일 처리, 업무연계 및 사용자 친화적 UI 등 실무접점 기술이 검토되고 있다.

○ 주요 성과와 과제

- 대한민국 정부의 초거대 AI 프로젝트는 글로벌 경쟁력 확보(최신 모델 성능 95% 이상 목표), 대규모 민관 협력, 지속적 데이터 클러스터 및 인재 양성, 공공 신뢰성 및 책임성 확보 등에서 일부 진전이 있으나, 다음과 같은 과제도 상존한다.
- 기술 자립성 및 글로벌 경쟁력: 미국, 중국 등 선도국과의 격차 해소, 고성능 GPU·데이터 확보,

알고리즘 혁신 등이 지속적으로 필요하다.

- 정책 및 관리체계의 고도화: 공공부문 특화 적용을 위한 보안·윤리·신뢰성 강화, 민관 협력과 장기적 투자 지속성을 확보해야 한다.
- 산업·사회적 파급력: 실질적 산업현장 적용 확대, 고용·교육·윤리 등 부작용 최소화, 국민 체감형 서비스 창출, 전 영역 AI 활용가치를 극대화해야 한다.

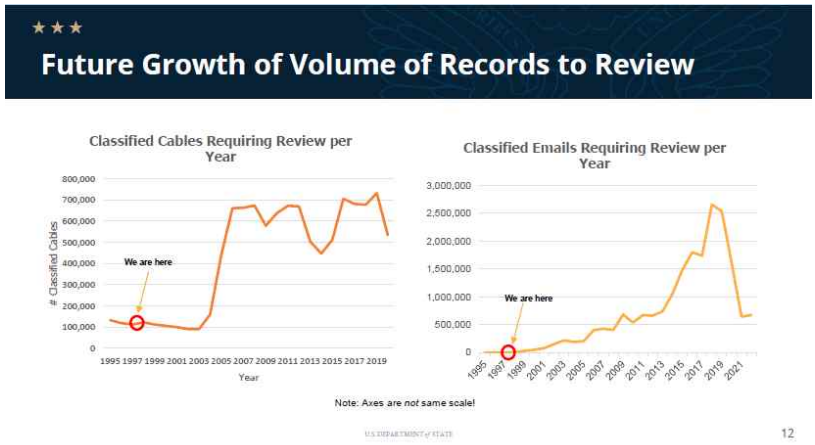
3.1.3.3. 국외 사례

○ 미국

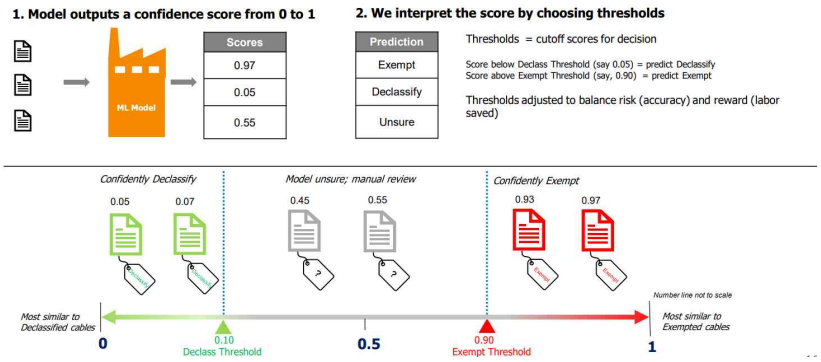
- 공익을 위한 기밀문서 해제위원회(PIDB, Public Interest Declassification Board)[9]의 파일럿 프로젝트 “Piloting Machine Learning for Freedom of Information Act(FOIA)(U.S. Department of State, 2023)”
  - 해당 프로젝트는 3가지 하위 파일럿으로 구성된다.
    - 파일럿 1 - Machine Learning Declassification Review Pilot (‘22.10.-‘23.01.)
    - 파일럿 2 - Freedom of Information Act(FOIA) Customer Experience(CX) Pilot (‘23.06-‘24.02)
    - 파일럿 3 - FOIA Search Pilot (‘23.06-‘24.02)
  - 이 중 첫 번째 파일럿이 공개재분류와 직접적인 관련이 있다. 두 번째 파일럿은 FOIA 제도 개선 관련 내용이며, 세 번째 파일럿은 검색 서비스 관련 파일럿이다.
  - 원문에서는 ‘Declassifacation’이라는 용어를 사용하였다. 의미상 ‘비밀해제’가 더 가까운 번역이지만, 편의상 ‘공개재분류’로 해석하였다.
  - 우리나라 법령에 따르면 공개재분류와 비밀해제는 각각 「정보공개법」과 「공공기록물 관리에 관한 법률」에 근거한 서로 다른 제도이나, 미국의 경우 두 제도의 경계가 우리나라만큼 명확히 구분되지 않고 있다.

구분	내용
파일럿 배경	○ 외교 전문(Cable) 공개재분류 검토량이 방대하여 인적 검토는 지속 불가능한 상황이다. ○ 파일럿 이전 기존 공개재분류 작업에는 국무부의 eRecords Declassification Module을 사용하였다. ○ 기록물 생산 이후 25년 경과 시 공개재분류를 시행한다. ○ 향후 공개재분류 대상량은 급격히 증가할 것으로 예상된다. ○ 파일럿 당시 분류 대상량은 1997-1998년에 생산된 기록물이며, 2000년 이후 기록물 생산량이 급격히 증가하였다. ○ 현재 작업 속도로는 향후 다가오는 공개재분류 대상량을 처리할 수 없는 수준에 이

9) PIDB의 소개글은 NARA 홈페이지에서(https://transforming-classification.blogs.archives.gov/), PIDB의 활동은 블로그(https://transforming-classification.blogs.archives.gov/) 에서 확인할 수 있다. (2025년 10월 13일 최종 접속).

<table>
<tr><th colspan="2">구분</th><th>내용</th></tr>
<tr><td colspan="2"></td><td>르렀다.
 

(그림 15 - Future Growth of Volume of Records to Review)</td></tr>
<tr><td colspan="2">사전 작업</td><td>○ 기관 의사결정자들에게 AI 활용 기회에 대해 교육한다.
○ AI 솔루션을 도입할 때 어떻게 필요성을 설득하고, 개발을 추진해야 하는지 모범 사례를 제시한다.
○ 기관 리더들이 AI 기술을 전략에 통합하고, 직원들에게 이를 적용할 수 있도록 준비시키는 데 필요한 역량을 갖추게 한다.
○ 'Partnership'은 연방 정부 직원의 역량 강화를 위한 리더십 개발 프로그램을 다양한 수준에서 제공한 경험이 풍부하다.
○ 이 프로그램은 연방 기관에 무료로 고위 임원과 GS-15를 선발하기 위해 제공된다.</td></tr>
<tr><td colspan="2">주요 내용</td><td>○ 기존 공개재분류 작업은 12개월 소요되어, 연간 공개재분류 대상량을 1년 동안 작업하였다.
○ 파일럿 프로젝트를 통해 작업 기간을 1개월로 줄이는 것이 목표였으나, 실제 작업 기간은 60% 경감하여 기존 작업 대비 40% 기간이 소요되었다.
○ 지도학습(Supervised Machine Learning)으로 공개재분류 자동화를 시도하였다.</td></tr>
<tr><td colspan="2">지도학습</td><td>○ 기존 공개재분류된 외교 전문과 공개값을 활용하여 머신러닝 모델을 학습시켰다.
○ 모델이 실무자의 공개재분류 패턴을 학습하고, 공개값이 없는 새로운 외교 전문에 대해 공개값(공개/비공개)을 예측하는 구조이다.</td></tr>
<tr><td>예측 점수</td><td>임계값 설정</td><td>○ 모델은 0에서 1사이의 신뢰도를 출력하고, 임계값을 사전에 설정하였다.</td></tr>
</table>

<table>
<tr><th colspan="2">구분</th><th colspan="2">내용</th></tr>
<tr><td rowspan="4"></td><td></td><td colspan="2"> 

(그림 16 - Cable declassification : Prediction Scores)</td></tr>
<tr><td>0.10 이하</td><td colspan="2">확실한 공개</td></tr>
<tr><td>0.10 초과
0.90 미만</td><td colspan="2">검토 필요</td></tr>
<tr><td>0.90 이상</td><td colspan="2">확실한 비공개</td></tr>
<tr><td rowspan="6">수행 결과</td><td>연도</td><td>2022년</td><td>2023년</td></tr>
<tr><td>대상</td><td>1997년에 생산된 외교전문 78,023건</td><td>1998년에 생산된 외교 전문 121,536건</td></tr>
<tr><td>확실한 공개</td><td>48,707건
(오분류 346건, 에러율 0.71%)</td><td>72,891건
(에러율 파악 중)</td></tr>
<tr><td>검토 필요</td><td>28,433건
(결과 추가 검토 필요)</td><td>47,218건
(결과 추가 검토 필요)</td></tr>
<tr><td>확실한
비공개</td><td>883건
(오분류 164건, 에러율 18.57%)</td><td>1,427건
(에러율 파악 중)</td></tr>
<tr><td>작업시간
절감</td><td>63.56%</td><td>61.15%</td></tr>
<tr><td colspan="2">향후 과제</td><td colspan="2">○ 외교 전문 자동 공개재분류를 실무 프로세스로 정착시킨다.
○ ML 모델을 고도화한다(개체명 인식, 역사·외교 전문가 의견 반영).
○ 다른 문서 유형에 대한 파일럿을 시행하고 자동 공개재분류를 확장 적용시킨다.
○ 자동 마스킹을 통해 문서 공개 속도를 높이고 수작업을 줄인다.</td></tr>
<tr><td colspan="2">파일럿 결과 고찰</td><td colspan="2">○ 양질의 데이터 확보가 필요하다.
○ 부서 간 협력이 중요하다.
○ 특정한 세부 범위를 지정하여 작은 프로젝트부터 시작해야 한다.
○ 모든 결과와 피드백을 수용하는 열린 자세가 필요하다.
○ 인내심이 필요하다.
○ 품질관리 점검 체계를 마련해야 한다.
○ 지속가능한 성과를 위해 반복 학습을 시켜야 한다.</td></tr>
<tr><td>파일럿 이후</td><td>2022년</td><td colspan="2">○ 파일럿 프로젝트를 통해서 인적 공개재분류와 자동 공개재분류를 동시에 병렬 수행했다.</td></tr>
</table>

구분		내용
	2023년	○ 2022년에 공개재분류를 수행한 기록물은 2023년 말에 공개를 시작했다.
	2024년	○ 2024년 기준으로 약 450,000 건의 외교 전문이 해당 파일럿 프로세스를 통해 공개재분류되었다. ○ 1999년 생산 외교 전문을 대상으로 자동 공개재분류를 수행할 예정이다(2023년 보고서 작성 당시 기준).
	2025년	○ 해당 모델의 다른 적용 방안도 모색 중이며, 외교 전문 이외의 유형에도 확장 추진 중이다.

(표 8 - Machine Learning Declassification Review Pilot (‘22.10.-‘23.01.))

○ 영국

- 영국 국립기록보존소(TNA, The National Archives)의 연구 “The application of technology-assisted review to born-digital records transfer, Inquiries and beyond(The National Archives, 2016)”

◦ 2015년에 연구하고 2016년 발표된 TNA의 보고서이다.

◦ 주요 내용은 전자기록물의 기술 기반 검토(TAR, Technology-assisted Review)이다.

▪ 기술 기반 검토(TAR, Technology-assisted Review)[10)]

· 전문 검토자가 소프트웨어와 알고리즘을 활용하여 문서를 자동으로 분류한다.

· 자동 분류 결과를 반복적으로 보정하면서 정확도를 높인다.

· 인적 검토보다 훨씬 빠르고, 대규모 기록에서 패턴을 쉽게 찾을 수 있다.

· 텍스트 기반 정보에 강하다.

구분	내용
연구 배경	○ 2016년부터 TNA의 전자기록물 비중이 커질 것으로 전망되었다. ○ 기존 법 Public Records Act는 1958년에 제정되었다. 1967년 법 개정으로 기존 법에 따라, 기록물이 생산된 지 30년이 경과하면 공개재분류를 실시하였다 - Public Records Act는 1958년에 제정되었다. 1967년 법 개정으로 기존 50년 공개 원칙에서 30년으로 변경되었다. ○ 2010년에 해당 법의 개정으로 공개재분류까지의 기간이 30년에서 20년으로 단축되었다. 2013년부터 단계적으로 적용하여 2022년부터는 20년 공개 원칙이 완전히 정착되었다. ○ 20년 공개 원칙이 적용되면서 공개재분류를 검토해야 할 기록물의 양이 폭발적으로 증가하였다.

10) 법조계의 증거개시절차(e-Discovery)에서 방대한 증거를 관리하기 위해 등장한 기술이며, 해당 연구는 이 기술을 기록물관리에 적용하려는 시도 중 하나이다. TAR을 통해 소송과 관련된 자료를 쉽게 구분하고, 변호사는 관련도 높은 순서대로 자료를 읽어볼 수 있게 되었다.

구분		내용
		○ 전자기록의 경우 종이기록보다 훨씬 방대하고, 종이기록에 비해 덜 구조화되어있어 평가, 선별 등 관리에 더 큰 어려움을 겪고 있다. ○ 기존의 기록관리 프로세스는 종이기록물을 위해 설계되었고, 전자기록관리에 적합하지 않다. ○ e-Discovery 도구는 법원에서 사용될 만큼 신뢰성이 높다. ○ e-Discovery 도구의 목적과 방법이 전자기록 이관 과정과 유사하다 .
연구 목적	전체	○ 기존 소프트웨어가 전자기록물의 평가, 선별, 민감도 검토를 얼마나 지원할 수 있는지 파악
	평가와 선택	○ 구조화되지 않은 전자기록물의 의미를 추출하고 자동으로 분류할 수 있는지 파악
	민감도 검토	○ 민감정보를 식별할 수 있는지 파악 - 민감정보의 75%는 개인정보이며, 나머지 25%는 조금 더 복잡하여, 이번 연구는 개인정보 식별에 집중하였음)
	자동 마스킹	○ 열람 제공 전에 민감정보를 자동으로 비식별화할 수 있는지 파악
	사용자 경험	○ 해당 도구에 익숙해지기까지 실무자를 얼마나 훈련해야 하는지 파악
연구 방법		○ 각각 특화 기능이 다른 e-Discovery 도구 3가지를 선정한다. - 잠재 디리클레 할당(LDA, Latent Dirichlet Allocation)에 특화된 도구 1종 - 관계형 데이터베이스에 특화된 도구 1종 - 잠재 의미 색인(LSI, Latent Semantic Indexing)에 특화된 도구 1종 - 도구를 비교·평가하려는 연구가 아니라, 어떤 기능과 기술을 기록관리에 적용하느냐가 더 중요한 연구이므로 제품명은 비공개하였다. ○ e-Discovery 도구에 기록물을 입력해 초기 색인을 구축한다. - 샘플 데이터로는 TNA 소장 전자기록물 10만 건을 사용하였다. ○ e-Discovery 도구에서 기록물의 검색과 필터링 등 다양한 시나리오를 시연한다. - 각 도구에 같은 샘플데이터를 이용하여 같은 작업을 처리하고 그 결과를 비교하였다.
연구 결과		○ e-Discovery 도구를 활용한 TAR은 TNA의 전자기록물 이관 과정에서 평가, 선별, 민감도 검토를 지원할 수 있다. ○ e-Discovery 도구는 맥락정보가 부족한 전자기록의 맥락을 파악하고 의미를 추출할 수 있다. ○ e-Discovery 도구가 사람보다 더 우수할 때도 있다. - 사람은 문서를 볼 때 F자 형태의 시각 경로를 따르는 경향이 있어, F자 영역 밖의 정보를 놓치기도 한다.

구분	내용
	 (그림 17 - F-shaped Pattern for Reading Web Content) - 컴퓨터는 F자 영역 밖의 내용도 놓치지 않고 분석한다. - 미국 법조계에서는 TAR을 수작업보다 권장한다. ○ e-Discovery 도구는 민감정보를 찾을 수 있고, 자동 마스킹도 할 수 있다. 다만, 사용자의 지속적인 피드백을 반영하여 정확성을 높여야 한다. - 25%를 차지하는 개인정보 외 비공개 사유에 대해서는 추가 연구가 필요하다. - e-Discovery 도구 업체에 의존하면 시간과 비용이 늘어나므로, 실무자가 오탐을 제거하고 누락된 항목을 추가하는 등 조정 작업을 할 수 있어야 한다. ○ 이 기술은 수작업으로 검토해야 하는 전자기록의 양을 줄이고 우선순위를 설정하여 기록관리 업무에 도움을 주지만, 완벽한 도구는 없고 여전히 사람의 개입이 필요하다.
시사점	○ 전자기록물 집합을 거시적으로 이해하는 태도가 필요하다. ○ 검토해야 할 정보량을 줄여야 한다. ○ 의미를 추출해야 한다. ○ 개인정보를 식별해야 한다. ○ 조달 문제를 고려해야 한다. ○ 사용자 인터페이스가 중요하다. ○ 다른 팀과의 협업이 필요하다. ○ TAR에 대한 신뢰를 확보해야 한다.

(표 9 - The application of technology-assisted review to born-digital records transfer, Inquiries and beyond)

○ 호주(Gregory Rolan, Glen Humphries, Lisa Jeffrey 외., 2019)

- NAA(National Archives of Australia, 호주 국가기록원)의 프로젝트

구분	내용
연구 배경	○ 호주 정부의 디지털 업무 환경에 맞춘, 기록물의 보존 및 폐기 기준을 마련하기 위한 목적이다. ○ 보존 및 폐기 기준을 매년 발행하고 있으나 종이기록물 기반 환경을 위한 기준이며, 수동 판단을 전제로 작성되었다. ○ 디지털 시스템이 스스로 보존 및 폐기 결정을 하기 위해서는 폐기 기준을 기계가 읽고 실행할 수 있도록 변환해야 한다.
연구 방법	○ 2015년부터 NAA의 4명으로 팀을 구성하여 지속적으로 해당 프로젝트를 수행하고 있다. ○ 다음과 같은 다양한 기술 영역을 탐색하였다. - 엔터티 및 자동 메타데이터 추출 - 시맨틱 분석 - 택소노미와 온톨로지 구축 - Linked data 접근
연구 결과	○ 프로젝트 1단계에서는 폐기 규정에 대한 개념모델과 시맨틱 분석 결과 도출하였다. ○ 위의 기술 영역을 적용하여 보존 및 폐기 규정을 XML형식의 스키마로 재구성하였다. ○ 재구성한 스키마는 기존의 보존 및 폐기 기준과 유사하며 주요 구성은 다음과 같다. - 폐기 기준과 해당 기관에 대한 맥락정보를 포함한 헤더 - 활동을 문서화하거나 지원하기 위해 생산되는 기록의 맥락을 만드는 업무 최상위 기능 - 적절한 폐기 평가와 폐기 집행되는 문서를 자동으로 연계할 수 있도록 설계된 폐기 기준 ○ 이 요소들은 테스트 그룹으로 구성되며, 현재 사용 중인 폐기 기준의 언어 표현을 분석하는 작업도 수행되었다.
향후 계획	○ 다음 단계에서는 이러한 이론적 가정을 생산된 지 10년 이상 되고, 개정될 예정인 NAA의 핵심 기록 권한 문서에 실제로 테스트할 예정이다. ○ 머신러닝과 인공지능 전문가들의 자문을 받아 자동 분류, 클러스터링, 인덱싱 기술의 적용 가능성을 실증 및 평가할 계획이다.
시사점	○ 아직 초기 연구단계이나, 제도적 규정을 “기계가 처리할 수 있는 규칙”으로 만드는 방향성을 제시하였다. ○ 기술 도입보다 제도적 언어를 디지털 규칙으로 전환하는 작업이 선행돼야 한다.

(표 10 - The project of NAA)

- PROV(Public Record Office Victoria, 빅토리아 주기록보존소)의 Email appraisal, disposal and preservation project[11)]

구분	내용
연구 개요	○ 이메일 자료를 적절하게 수집, 저장, 평가, 폐기하기 위한 솔루션 개발 프로젝트 - 1단계 : 개념 증명 (2017-2018) - 2단계 : 샘플 테스트 (2019-2020) - 3단계 : 솔루션 개발 (2022-2024)
연구 배경	○ 이메일은 업무 수행에 필수 요소이며, 1973년 제정된 공공기록물법(Public Records Act 1973, Victoria)에 따른 공공기록물이다. ○ 이메일은 책임 소재 규명에 필수적인 증거가 되기도 하므로 공공기록물로 보존해야 한다는 필요성을 인지하였다. ○ 빅토리아 주 정부는 IBM의 Lotus Notes를 사용하여 이메일을 관리한다. ○ 이메일이 누적되면서, 필요한 이메일을 찾기 어렵고 비용도 발생하여 이메일 기록을 선별하여 관리해야 한다.
연구 결과 (1단계 : 개념증명)	○ 대량의 이메일을 AI 기반 eDiscovery 도구로 분류할 수 있는지 증명하였다. ○ 규칙 기반, 통계, 딥러닝을 활용하여 업무 관련 이메일을 가능한 한 많이 찾아내는 시도였다. ○ 상용 eDiscovery 도구인 Nuix를 사용하여 1.5TB에 달하는 이메일 형식 파일을 자동으로 업무 관련/비업무 관련으로 분류하였다. - 기술적 평가 : 데이터 세트의 구성요소 파악 - 중복 제거 : MD5 Hash를 적용하여 동일 이메일 식별 ○ 아래의 접근법 3가지로 Nuix를 적용하여 이메일을 분류하였다. - Positive : 가치 있는 이메일 식별하는 방식 - Negative : 가치 없는 이메일 식별하는 방식 - Macro : 이메일의 기능적 맥락을 따라 메타데이터에 접근하고 메타데이터를 추가하는 방식
연구 성과	○ MD5 Hash를 적용하여 전체 460만 건 중에 약 40%의 중복 이메일을 식별하였다. ○ Positive 접근법으로 중복 제거된 이메일 식별 결과 93%의 이메일이 업무 관련으로 식별되었으나 거짓 긍정이 많았다. ○ Negative 접근법으로 중복 제거된 이메일 식별 결과 7%의 이메일이 업무 관련이 없는 것으로 식별되었다. ○ 위험 최소, 효율 최대 접근법으로 다음의 혼합 모델을 제안하였다. - 중복 제거를 통한 이메일 용량 축소 - Negative 접근법으로 가치 없는 이메일 식별 - Macro 접근법으로 기능별 추가 메타데이터 부여
향후 과제	○ 해당 도구를 사용하여 향후 백로그 누적을 방지할 예정이며, 공유 드라이브 등 다른 레파지토리에도 확장 적용할 예정이다. ○ 더욱 발전된 AI 기반 자동화 평가도 연구할 예정이다.
시사점	○ Nuix는 분석해야 할 이메일 데이터의 양을 대폭 줄일 수 있고, 자동화된 워크플로우를 통해 검색과 관리 측면에서 효율성을 향상시킬 수 있다. ○ eDiscovery 도구가 기록관리 영역에서도 충분히 활용될 수 있음을 보여준다. 다만, 사람의 최종 검토가 병행되어야 한다.

(표 11 - Email appraisal, disposal and preservation project)

11) PROV 홈페이지에 해당 프로젝트에 대한 자세한 아티클이 개제되어 있다. https://prov.vic.gov.au/recordkeeping-government/research-projects/email-appraisal-disposal-preservation-project Public Record Office Victoria 참조. (2025년 10월 10일 최종 접속).

- NSW(New South Wales) State Archives (뉴사우스웨일스주 기록 보관소)의 내부 파일

구분	내용
연구 개요	○ 2017년 11월부터 12월까지 수행하였다. ○ 연구목표는 기성 머신러닝 소프트웨어를 활용하여 비정형 데이터를 보존·폐기 기준에 따라 자동 분류할 수 있는지 검증하는 것이다. ○ 사람이 직접 분류했던 기록 데이터를 토대로 머신러닝 알고리즘을 이용하여 분류했을 때, 그 결과가 얼마나 동일한 지 실험하였다.
사전 준비	○ 별도 예산 없이, 제한된 자원으로 연구를 수행하였으며, 머신러닝 관련 ICT 전공 인턴이 참여하였다. ○ 비용이 저렴하면서 바로 사용 가능한 상용 기술을 조사하였고, Microsoft Azure의 클라우드 기반 AI 및 Cognitive Services를 연구 대상으로 고려하였다. ○ 다만, 데이터 저장 및 관리 위치가 해외이기 때문에 State Records Act 1998(NSW)에 따라 GA35(주 외부 저장 및 관리 시 기록 이관에 관한 기준)가 적용되어 위험 평가가 필요했다. ○ 파일럿 기간이 짧아 법률 자문이나 검증 절차를 수행하기 어려웠기 때문에, 클라우드 솔루션은 연구 대상에서 제외하였다. ○ 로컬에서 실행 가능한 오프라인 머신러닝 소프트웨어 중 Python용 오픈소스 라이브러리 'scikit-learn'을 채택하였다. - 규칙 기반, 통계 기반, 심층학습 알고리즘 등 다양한 모델을 제공하며, 비교적 단순하고 접근성이 높다. - 클라우드를 사용하지 않기 때문에 고성능 컴퓨터가 필요하다.
데이터 전처리	○ Python 라이브러리 'scikit-learn' 사용하였다. ○ 샘플 데이터는 2016년에 중앙정부 부처로부터 디지털 아카이브로 이관된 기록물을 사용하였다. ○ 42,500개의 파일 중 중앙정부 부처와 협력하여 수작업으로 선별한 결과 12,369개의 파일을 보존 대상으로 식별하였다 . ○ 12,369개의 파일 중에 텍스트 추출이 가능한 파일 형식인 8,784개의 파일을 샘플로 선별하고, 추출결과는 CSV 파일로 저장하였다. ○ 8,784개의 파일에 대해 데이터 정제, 텍스트 벡터화, TF-IDF(Term Frequency-Inverese Document Frequency) 작업을 진행하였다. - 불필요 문서 및 불용어를 제거하고 전부 소문자로 변환하여 정제하였다. - Bag-of-Words(BOW) 모델을 사용하여 단어의 위치를 무시하고, 등장 빈도 기반으로 수치화하였으며, 문서를 수치 벡터로 변환하였다. - 자주 등장하는 단어의 가중치를 줄이고, 드물지만 의미 있는 단어의 중요도를 높였다.
분류	○ 두 가지 알고리즘 비교하였다. - Multinomial Naïve Bayes - Multi-Layer Perception ○ 두 가지 모두 지도학습 방식으로 학습시켰다. ○ 데이터 세트는 학습 75%, 테스트 25%로 분리하였다. ○ 사전에 분류된 데이터를 기반으로 모델을 학습시킨 뒤, 동일한 모델을 테스트셋에 적용해 수동 분류 결과와 비교하였다.
연구 결과	○ 4가지 조합으로 결과를 비교하였다. - 조합 1 : 정제된 데이터 × Multinomial Naïve Bayes 알고리즘 - 조합 2 : 정제된 데이터 × Multi-Layer Perception 알고리즘 - 조합 3 : 원본 데이터 × Multinomial Naïve Bayes 알고리즘

구분	내용
	- 조합 4 : 원본 데이터 × Multi-Layer Perception 알고리즘 ○ 가장 높은 성능을 보인 것은 조합 2였다. 정확도는 최대 84%를 기록하였다. ○ 다만 사람이 수동 분류한 결과의 정확도는 측정되지 않았기 때문에 완전한 비교는 어렵다. ○ 폴더 단위보다 문서 단위에서 훨씬 빠르게 보존·폐기 권한을 분류할 수 있었다.
시사점	○ 머신러닝 기술이 비정형 기록물의 자동 분류 및 보존·폐기 평가에 충분히 활용될 수 있음을 입증하였다. ○ 전처리와 알고리즘 선택이 성패를 좌우한다. ○ 실제 적용 가능성을 보여준 대표 사례이다.

(표 12 - The project of NSWA)

○ 바레인

- Microsoft Purview를 활용한 민감정보 유출 예방 연구(Naser F. Aldoseri, Wael Elmedany, 2023)

◦ Microsoft Purview[12)]

▪ 조직의 데이터를 제어, 보호, 관리하는 솔루션이다.

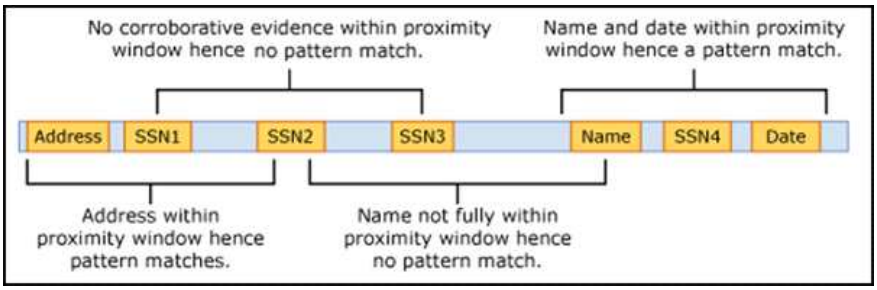
▪ 민감정보를 탐지하기 위한 기반 기술은 정규식과 함수 패턴 매칭이지만, AI 분류기가 탑재되어있어 AI 기술을 활용하기도 한다.

▪ Purview에 탑재된 AI 기반 분류기는 대규모 언어 모델에서 생성된 합성 데이터와 샘플 파일을 기반으로 사전학습 및 테스트를 거쳐 민감 정보를 탐지한다.

▪ AI 분류기는 Microsoft의 AI 슈퍼 컴퓨터와 특수 인프라 환경에서 실행되어, 민감정보를 빠르게 분류하고 보호할 수 있다.

구분	내용
연구 배경	○ 바레인의 민감정보로는 중앙인구등록번호(CPR, Central Population Register)와 여권번호 두 가지가 대표적이다. ○ 기존 조직들은 민감정보 관리를 사내에서 처리했으나, 코로나19를 겪고 원격 근무가 일반화되면서 민감정보가 인터넷에 노출될 위험이 생겼다. ○ Microsoft Purview는 민감정보 유형(SITs, Sensitive Information Types) 식별 기능 활용과 데이터 손실 방지(DLP, Data Loss Prevention) 서비스를 통해 민감정보를 식별하고 추적할 수 있다.
연구 방법	○ 바레인의 민감정보 보호에 있어 Microsoft Purview의 효율성을 검증한다. ○ Microsoft Purview의 서비스 두 가지 사용하여 테스트 파일을 검증한다. - Microsoft의 민감정보 유형(SITs, Sensitive Information Types) 식별 - Microsoft의 데이터 손실 방지(DLP, Data Loss Prevention)
Microsoft Purview의 민감정보 유형 (SITs, Sensitive	○ Purview에서 기본으로 제공하는 민감정보 유형을 선택하여 민감정보로 지정할 수 있고, 사용자가 직접 편집하여 추가할 수도 있다. ○ 정규식, 함수, 체크섬 알고리즘 등 강력한 패턴 매칭 기술로 민감정보를 식별한다.

12) https://techcommunity.microsoft.com/blog/microsoftmechanicsblog/ai-powered-data-classification--microsoft-purview/3919206? Microsoft 참조. (2025년 09월 30일 최종 접속).

구분	내용
Information Types)	- 바레인 CPR을 정규식 표현 : \d{2}(0[1-9]|1[0-2])\d{5} - 바레인 여권번호의 정규식 표현 : \b\d{7}\b ○ 문자 근접성(character proximity) 옵션으로 민감정보 탐지 정도를 제어할 수 있다.  (그림 18 - The Character Proximity Workflow) ○ 문자 근접성 옵션은 기본 증거와 보조 요소 사이의 문자 수를 제어하는 기능이고, 선택 사항이다. - 기본 증거(Primary evidence) : 사전 지정된 민감정보 - 보조 요소(Support element) : 기본 증거 근처에서 추가로 발견되는 키워드 - 기본 증거 근처에 보조 요소가 있다면 기본 증거를 민감정보로 식별 - 해당 옵션은 기본 증거와 보조 요소 사이의 거리 기준값을 설정하는 기능이다. ○ 기본 증거(중앙인구등록번호, 여권번호 등)의 근접 거리 내에 보조 요소(주소, 이름, 날짜 등)가 있으면 민감정보로 식별한다. ○ 기본 증거(중앙인구등록번호, 여권번호 등)의 근접 거리 내에 보조 요소(주소, 이름, 날짜 등)가 없으면 민감정보로 식별하지 않는다. ○ SITs 기본 기능으로만 식별했을 때 발생하는 오탐을 줄이기 위한 옵션이다.
◦ Microsoft의 데이터 손실 방지(DLP, Data Loss Prevention)	○ Microsoft의 생산성 도구에서 SITs의 유출을 막는 도구이며, Microsoft 제품 뿐만 아니라 타 애플리케이션과 사용자 단말기까지 확장이 가능하다. - 사전에 지정한 범위 밖에서 민감정보의 공유가 탐지되면 Microsoft Purview가 차단한다.
연구 결과	○ Microsoft Purview는 테스트 파일의 민감정보가 외부 조직으로 전송되는 것을 탐지하고 차단하였다. ○ 바레인 민감정보 유출 예방을 위해 Microsoft Purview가 효율적일 수 있다. ○ 다만, 테스트 파일의 크기가 큰 경우(연구에서 언급하는 파일은 100만 행이 있는 엑셀 파일이었다.) 타임아웃되어 제대로 처리되지 않았다.
시사점	○ Microsoft Purview는 조직의 컴플라이언스 수준을 높이고, 비즈니스 연속성을 향상시키고, 고객 신뢰를 보장할 수 있다. ○ 향후에는 바레인 CPR과 여권번호뿐만 아니라 신용카드 정보, 전자 건강기록(EHR)등 다양한 민감정보를 보호하는 연구가 필요하다.

(표 13 - Strengthening Data Security in Bahrain)

3.1.4. 시사점

선행연구와 국내외 사례를 통해 인공지능을 적용한 기록물 공개재분류 연구는 기술적으로 충분한 가능성을 지니고 있다는 것을 확인하였으며, 다음과 같은 시사점을 도출할 수 있다.

○ AI를 적용하여 기록물의 맥락을 분석할 수 있다.

공개재분류 과정에서 기록물 유형 분류 작업, 재분류 의견 작성, 재분류 사유 기재 작업 등은 기록물의 맥락과 본문 내용을 동시에 분석해야 한다. 따라서 텍스트 분류 및 자연어 처리 기술을 결합한 AI 모델을 적용할 경우, 담당자의 판단을 보조하고 업무 효율을 높일 수 있다.

○ 점진적이고 단계적으로 접근해야 한다.

미국 사례에서는 방대한 범위를 한 번에 자동화하지 않고, 특정 유형(외교 전문)에 집중하였고, 자동화 프로세스를 단계적으로 정착시켰다. 따라서 AI 기반 공개재분류 자동화를 전체 호수에 일괄 적용하기보다, 우선 특정 유형에 집중하여 소규모 파일럿으로 시작하는 점진적 접근이 바람직하다. 예를 들면, 개인정보 유무로 판가름하기 간단한 6호의 공개재분류 시스템을 먼저 구축해둔 다음 7호, 8호 등 다양한 유형의 호수로 확장 적용하는 것이다.

○ 공개재분류 도구에 e-Discovery 기술을 활용할 수 있다.

영국 TNA와 호주 PROV는 법률 분야에서 사용되던 e-Discovery 기술을 기록물관리에 적용함으로써, 대규모 기록 파일 분류의 자동화와 효율적인 검색의 가능성을 입증하였다. 공개재분류 자동화 도구를 처음부터 독자적으로 개발하기 힘들다면 기존 e-Discovery를 기반으로 업무 특성에 맞게 커스터마이징하는 전략이 효율적일 수 있다.

○ 데이터의 전처리가 무엇보다 중요하다.

김해찬솔 외(2017)의 연구와 호주 NSW의 내부 파일럿 사례는 데이터 전처리의 품질이 모델의 성능에 결정적인 영향을 미친다는 것을 명확하게 알려주었다. 데이터 정제, 중복 제거, 불용어 처리 등 사전 가공이 불충분할 경우 기계 가독성이 현저히 떨어진다. 따라서 AI 기반 공개재분류를 위해서는 기록의 생산단계부터 기계 가독성을 확보할 수 있는 표준화된 데이터 환경이 필요하다.

○ 법령과 지침의 디지털 전환이 병행되어야 한다.

호주 NAA의 프로젝트는 기록물뿐만 아니라 공개재분류 지침 등 제도적 언어를 디지털 규칙으로 전환하는 시도를 보여주었다. 이러한 시도는 AI가 행정결정의 근거를 인식하고 실행할 수 있는 기반을 마련하려는 접근이었다. 이는 국내의 「공공기록물 관리에 관한 법률」 및 「정보공개법」 등 관련 규정 역시 기계가 해석할 수 있는 구조화된 형태로 정비되어야 함을 시사한다.

○ 공공 기록관리 업무의 자동화와 정확성 확보가 필요하다.

정부의 초거대AI 연구는 국가기록원의 비공개기록물 공개재분류 프로세스에 최신 LLM 및 RAG 기반 기술 도입을 가능하게 하여, 분류 자동화 및 정보공개 결정에 있어 한층 높은 정확성과 효율성을 실현할 기반을 제공한다.

대규모 데이터 세트 학습과 민감정보 식별·분류 역량의 고도화는 기존의 수작업 중심 판단 한계를 넘어, AI 우선 심의 및 전문가 결정 지원을 통해 반복적 업무를 자동화하고 오류 및 편차를 최소화할 수 있도록 한다.

이는 기록물 공개여부 결정의 신뢰성과 일관성을 강화하며, 재분류 감사 및 이력관리, 개인정보·비공개 사유 탐지 등에서도 객관적 기술적 근거 확보에 기여한다.

○ 신뢰성과 윤리성 기반의 제도·정책을 시스템에 연계해야 한다.

초거대AI 사업을 통해 구축되는 공공부문 AI 표준 모델, 신뢰성·윤리성 확보 정책, 대국민 성능검증 체계는 기록관리 시스템에도 반영되어, 공공문서 재분류 결과의 품질과 정책적 일관성을 제고한다.

AI 기반 재분류 결정 과정의 투명성, 법령 기반 자동검증, 기록정보의 무결성 보장 등은 국민 신뢰확보와 기관 책임성 강화라는 제도적 시사점을 제공한다.

더불어, 정부 주도의 모델 평가·피드백·보완프로세스와 연동함으로써 기록관리 업무에 맞춤형 AI 알고리즘 적용은 물론, 법·윤리 기준 변화에 신속 대응하는 유연한 업무 환경을 조성하게 된다.

3.2. 비공개대상정보의 검출 및 비식별화

3.2.1 업무분석

디지털 행정환경의 고도화와 데이터 기반 행정의 확산은 공공기관 및 민간 영역에서 대규모 데이터의 수집·저장·활용을 필수적으로 요구하고 있다. 그러나 이 과정에서 개인정보 및 비공개정보의 노출 위험이 증가함에 따라, 정보 보호와 데이터 활용 간 균형을 확보하는 것이 중요한 과제로 대두되고 있다. 특히 공공기록물, 행정문서, 의료정보 등 비정형 데이터(Unstructured Data) 내에 포함된 개인식별정보는 기존의 단순 규칙 기반 필터링으로는 검출이 어려워, 보다 지능적인 자동 탐지 및 비식별화 체계의 필요성이 커지고 있다(국가기록원, 2023; Choi & Park, 2022).[13)]

기존의 비식별화 업무는 수작업 또는 반자동화 도구를 활용하여 수행되어 왔으나, 문서의 형식·내용·도메인별 특성에 따라 처리 효율성과 정확도에 한계가 있었다. 이에 따라 인공지능(AI), 자연어처리(NLP), 기계학습(ML) 등을 기반으로 한 지능형 비식별화 기술이 빠르게 발전하고 있으며, 비공개정보의 자동 검출, 문맥 기반 탐지, 다중 검증 체계의 통합 등 다양한 기술적 접근이 시도되고 있다. 이러한 변화는 단순 기술의 도입을 넘어, 비식별화 업무 프로세스 자체의 재설계와 지능화를 요구하고 있다.

플랫폼 경제에서의 개인정보보호의 중요성과 개인정보보호 비식별화의 필요성에 대해 법령 · 지침서 등 다양한 적용 사례를 통해 확인한 결과를 다음과 같이 정리했다.

- ❍ 첫째, 플랫폼 경제에서의 개인정보보호는 인격권과 재산권적 성격이 동시에 부각되며, 데이터 활용과 보호 간 균형의 필요성이 커졌다.
  - 개인정보에는 재산권의 성격과 인격권의 성격이 모두 존재한다. 정보화가 진전되기 이전에 인격권으로서의 성격이 강조되어왔으며, 데이터 경제 시대에 접어들면서 데이터 활용 활성화를 통한 새로운 가치 창출과 개인정보보호 간의 상층문제 강하게 부각 되기 시작하였다.
- ❍ 둘째, 기존연구의 한계화 지능형 접근의 필요성이 드러났다. 기존 규칙 기반·통계적 기법은 한계가 있으며, 최근에는 AI · 머신러닝 · NLP 기반 기술이 도입되며 정교한 대응이 가능해졌다.
  - 기존 비식별화 연구는 주로 통계적 방법이나 규칙 기반 접근에 집중되어 있었으나, 최근에는 인공지능(AI) · 머신러닝 · 자연어처리(NLP) 기술의 발전보다 정교한 비식별화가 가능해지고 있다. 특히 기록학적 맥락에서 대규모 디지털 기록물, 비전자 기록의 디지털화, 복합기록 아카이브 등 다양한 유형의 데이터가 등장함에 따라 지능형 비식별화 연구의 필요성이 확대되고 있다.
  - 개인정보 비식별화는 개인정보를 특정 개인과 직접 연결할 수 없도록 처리하는 기술적·관리적 절차를 의미한다. 이는 개인정보보호법, GDPR[14)], HIPAA[15)] 등 국제적 법제에서도 핵심적으로 다룬다는 의미이고, 대표적 기법으로는 데이터 마스킹(masking)[16)], k-익명성(k-anonymity)[17)], l-다

---

13) Choi, J., & Park, H. (2022). AI-based Privacy Information Detection in Government Records Management. Archives and Society, 12(2), 45 - 67.

14) GDPR (General Data Protection Regulation), 유럽연합이 2018년 5월부터 시행한 개인정보 보호 일반 규정

15) HIPAA(Health Insurance Portability and Accountability Act), 미국에서 1996년에 제정된 건강보험 이전 및 책임에 관한 법률

16) 데이터 집합에서 특정 레코드가 최소 k명 이상의 다른 레코드와 동일한 속성을 공유하도록 변환 → 개별 식별 불가능 (Latanya Sweeney, 2002).

17) k-익명성의 한계를 보완 → 동질 집합(equivalence class) 내 민감 속성이 최소 l개 이상의 다양한 값을 가지도록 보장

양성(l-diversity)[18], t-근접성(t-closeness) 등이 있다.

○ 셋째, 국가별 법제 비교에서는 EU · 한국 · 일본은 제도적·정책적 접근을 강화하는 반면, 미국은 주 단위 입법과 사적 계약 중심의 체계를 유지하고 있다.

- EU, 우리나라, 일본 등은 관련 입법 및 전담 기관 설치를 통해 데이터 이동권과 공유, 프라이버시 보호 문제 등을 통합적으로 접근하고 있다. 미국은 아직도 사적 계약으로 해결되는 부분도 있지만 주 단위 입법을 통해 EU 등과 유사한 데이터 정책 도입을 시도하고 있다.
- 우리나라의 경우 비식별 정보의 결합 및 반출 작업은 반드시 수탁자가 자신의 물리적 공간 내에서만 수행하도록 명문화(외부 서버 이용 불가)하고 있다.

○ 넷째, 비식별화 기술 방향은 (1)문맥기반 NLP, (2) 합성 데이터(synthetic data) 및 차등 프라이버시(differential privacy), (3) 멀티모달 확장, (4) 리스크 기반 동적 모델, (5) 법·윤리 연계 거버넌스로 요약할 수 있다.

3.2.2 지능형 비식별화 지능화 자동화 기술 및 적용 방향

3.2.2.1 자동화를 위한 지능화 기술

최근 연구에서는 AI 자동탐지 + 전문가 검증(Human-in-the-loop) 체계가 표준화되고 있다.
AI가 1차적으로 비공개 대상 정보를 탐지하면, 이후 기록관리 전문가 또는 개인정보 보호 담당자가 검증 및 보완을 수행하는 이중 구조가 확립되고 있다(Choi & Park, 2022).

이 체계는 자동화의 효율성과 인간의 판단력을 결합함으로써, 과잉 비식별화(over-anonymization) 또는 비식별화 실패의 위험을 최소화한다. 또한 멀티모달(Multimodal) 접근이 새롭게 부상하고 있다.
텍스트 데이터뿐 아니라 음성(STT 기반), 영상(OCR 및 얼굴 비식별화), 이미지(텍스트 임베딩) 등 다양한 형태의 기록물에 대해 통합적 탐지 및 비식별화를 수행하는 연구가 진행 중이다.

이와 같은 접근은 기록정보 서비스의 자동화 수준을 향상시키는 동시에, 다양한 비정형 데이터 형식을 포괄하는 포괄적 비공개정보 보호 체계를 가능하게 한다. 지능형 비식별화 기술동향은 크게 다섯가지 흐름으로 다음과 같이 요약할 수 있다.

○ 첫째, NER(Named Entity Recognition, 개체명 인식) 기반 탐지로 BiLSTM-CRF[19], BERT, RoBERTa 모델이 적용되어 이름, 주소, 주민등록번호, 계좌번호 등 특정 패턴에 한정되지 않고 문맥 기반 개인정보 탐지가 가능하다(Huang et al., 2015).

○ 둘째, 하이브리드 방식은 규칙 기반 탐지와 AI 기반 탐지를 결합하여 정확도와 확장성을 모두 확

(Machanavajjhala et al., 2007).

18) k-익명성의 한계를 보완 → 동질 집합(equivalence class) 내 민감 속성이 최소 l개 이상의 다양한 값을 가지도록 보장 (Machanavajjhala et al., 2007).

19) Bi-directional Long Short-Term Memory + Conditional Random Field, 자연어 처리(NLP)에서 특히 순차 레이블링(Sequence Labeling) 작업에 자주 쓰이는 모델 구조

보한다. 단순히 AI 모델에 의존하는 것이 아니라, 기존 규칙 기반탐지(정규표현식, 키워드 사전)와 기계학습 기반 탐지를 결합하여 성능을 보완하는 연구가 활발하다(Le Yang, Miao Tian, Duan Xin 외, 2024).

○ 셋째, 자동화 + 전문가 검증체계가 보편화 되고 있으며, AI가 1차 탐지 후 전문가가 2차 검증을 수행한다. 기존에는 구조화된 데이터(예:주민등록번호, 전화번호 등)에 국한되었던 비식별화가 최근에는 영상, 음성, 이미지 데이터로 확장되고 있다. 예를 들면, 음성 데이터에서는 화자 인식 제거(speaker anonymization), 영상 데이터에서는 얼굴 비식별화(face blurring) 기술이 적용되며, 텍스트 기반 비식별화와 결합된 다중 모달(multimodal) 접근이 연구되고 있다.[20]

○ 넷째, 멀티모달 확장으로 텍스트뿐 아니라 영상·음성·이미지 데이터까지 적용 영역이 넓어지고 있다.[21]

○ 다섯째, 생성형 AI 결합은 맥락 보존형 대체 표현을 통해 단순 마스킹을 넘어 데이터 활용성을 유지하는 새로운 방향으로 주목받고 있다(Georgios Feretzakis, Konstantinos Papaspyridis, Aris Gkoulalas-Divanis 외, 2024).

종합적으로 정리해 보면, 지능형 비식별화 기술은 • NER 기반 문맥 이해, • 하이브리드 접근, • 자동화·전문가 검증 병행, • 다중 모달 확장, • 생성형 AI 기반 대체 표현 생성이라는 다섯 가지 축을 중심으로 발전하고 있다. 지능형 비식별화 기술동향 비교는 다음과 같은 기술을 적용하여 정보보호 간 균형을 추구하는 방향으로 기술이 진화하고 있음을 보여주고 있다.

구분	주요기술	특징	대표사례
기계학습 기반 (NER)	• BiLSTM-CRF, BERT, RoBERTa 기반 명명 엔티티 인식	• 이름, 주소, 계좌번호 등 비정형 텍스트내 민감정보를 문맥적으로 탐지	• 의료기록, 판결문 내 PII(개인정보) 탐지
하이브리드 방식	• 규칙 기반 탐지(정규 표현식) + 기계학습 결합	• 패턴 기반 탐지의 정확성과 AI 반 탐지의 확장성을 보완	• 미국, 재향군인청 (VHA) BoB 시스템 (F1 0.90 / 탐지정확도)
자동화 + 전문가 검증 병행	• AI가 1차 탐지 → 전문가가 2차 검증	• 오류 최소화, 법적 책임성 강화	• 영국 국가기록원 (TNA), 미국 NARA
다중 모달 확장	• 텍스트 + 영상 + 음성 기반 비식별화	• 얼굴 블러링, 화자 익명화 등 비정형 데이터 처리	• 영상 감시 데이터, 음성 기록관리
생성형 AI 결합	• GPT 등 LLM 기반 자동 탐지 및 대체 표현 생성	• 단순 삭제/마스킹을 넘어 맥락 보존형 비식별화	• 최근 연구 단계, 데이터 활용성 증대

(표 14 - Trends in De-identification Technologies)

20) LLMs-in-the-Loop 2부: 여러 언어에 걸쳐 PHI의 익명화 및 익명화를 위한 전문가용 소형 AI 모델무라트 구나이 , 부냐민 켈레스 , 라이프 히즐란, 2024.12.14.)

21) “생물의학 및 행동 건강 진단을 위한 AI 기반 다중 모드 컬러 분석”, 컴퓨터 구조 바이오테크놀 J. 2025년 5월 28일;27:2219~2232. 도이: 10.1016/j.csbj.2025.05.015 /하워드 대학교, 화학공학과, 2400 Sixth Street NW, 워싱턴 DC, 20059, DC, 미국

3.2.2.2 개인정보 비식별화(De-identification)

○ 비식별화(De-identification) 정의 및 제도적 배경

비식별화는 (De-identification)는 말 그대로 비식별, 즉 알아보지 못하게 만드것을 말한다. 이러한 비식별 기술이 정확하게 언제 개발됐는지 알수 없지만, 현대 비식별 기술의 기초를 다진 것은 k-익명성 개념(k-anonymity)' 개념을 제안한 미국 '라타냐 스위니(Latanya Sweeney)' 하버드대 교수로 알려졌다. 그녀는 1997년 발표한 'Identifiability of data' 논문에서, 공개된 정보와 비식별화된 데이터 세트를 결합해 특정 개인을 재식별할 수 있음을 증명하며 비식별화의 한계와 그 위험성을 전 세계에 알렸다. 이후 2018년 EU의 GDPR(General Data Protection Regulation)이 활성화되면서 전 세계적으로 개인정보보호의 중요성이 강조됐고, 가명정보와 익명정보의 활용에 대한 기준이 제시되면서 비식별화 기술의 도입이 가속화됐다.

국내 개인정보 비식별화 논의는 2016년 「개인정보 비식별 조치 가이드라인」에서 시작되었다. 2020년「데이터 3법」개정으로 가명정보 제도가 도입되면서, 비식별화는 데이터 활용의 전제 조건으로 자리 잡았다. 개인정보보호위원회, 금융보안원, 한국인터넷진흥원(KISA)등은 지속적으로 가명처리·비식별화 가이드라인과 평가체계를 정비하고 있다(행정안전부, 2023).

○ 국내 개인정보 비식별화 현 실태

개인정보보호위원회에 따르면, 2024년 한 해 동안 신고된 개인정보 유출 사고는 총 307건이며 주로 해킹(56%, 171건)으로 인해 발생했다. 또한 업무 과실이 30%(91건), 시스템 오류가 7%(23건)로 나타났다. 2023년 대비 해킹은 증가(151건 → 171건)했지만, 업무 과실(116건 → 91건)과 시스템 오류(29 건 → 23건)로 인한 유출은 줄었다(원병철, 2025. 7. 11.). 업무 과실로 인한 유출은 2024년에만 91건 발생, 게시판이나 단체 채팅방 등에 개인정보 파일을 게시(27건), 이메일을 동봉 발송(10건), 이메일이나 공문내 개인정보 파일을 잘못 첨부한 경우(7건)가 많았다.

이처럼 개인정보 유출 사건은 해킹은 물론 조직 내 과실이나 시스템 오류 등 유형을 가리지 않고 발생한다. 때문에 개인정보를 보호하기 위해서는 단순히 사내 보안을 강화하고 보안 솔루션을 도입하는 것을 넘어 "개인정보" 자체를 보호하는 방안을 마련해야 한다. 바로 비식별 솔루션이다.

3.2.2.3 국내 개인정보 비식별화 솔루션

○ 솔루션 기능 및 기술적 현황

국내 비식별화 기술은 크게 정형 데이터와 비정형 데이터 영역으로 나누어진다.

① 정형 데이터에서는 마스킹, 범주화, 가명화, 데이터 삭제 등 규칙 기반 처리 기법이 널리 사용된다.

② 비정형 데이터(텍스트·이미지·영상·음성)에서는 AI 기반 개인정보 자동 탐지(예 : NER, OCR+NLP)와 결합한 비식별화 기술이 빠르게 도입되고 있다. 특히 의료 영상(DICOM), 전자의무기록(EHR), 금융 콜센터 녹취 데이터 등은 최근 주요 연구 분야이다.

③ 지능형 자동화: 최근에는 AI 모델 + 전문가 검증 루프 형태의 하이브리드 구조가 주목받고 있다. 대규모 데이터 처리 속도와 정확성을 동시에 확보하기 위함이다.

○ 국내 개인정보 비식별화 솔루션 시장 및 산업 동향

국내 비식별화 솔루션 시장은 데이터 3법 시행 이후 빠르게 성장하고 있다. 주요 특징은 다음과 같다.

① 솔루션 다양화 : 지란지교데이터(IDFILTER), 파수(AnalyticDID), 이지서티(Identity Shield), 데이터스(PAmaster), 펜타시스템(DataEye PIDI), KT넥스알(NEA) 등 다양한 기업들이 솔루션을 상용화 하였다.

② AI 도입 가속화 : 로민(Textscope Privacy Guard), 이파피루스(AI BlackMarker) 등은 OCR·NLP·LLM 기반 기술을 적용하여 문서·이미지 내 개인정보 자동 탐지 기능을 고도화하였다.

③ 수요 확대 : 공공기관, 금융권, 의료기관을 중심으로 도입이 활발하다. 특히 마이데이터 산업, 공공 빅데이터 개방, 스마트의료 연구에서 핵심 기술로 자리 잡고 있다.

국내 개인정보 솔루션은 2016년 「개인정보 비식별 조치 가인드라인」과 2020년「데이터 3법」개정을 계기로 본격적으로 성장했다. 현재는 정형 데이터(데이터베이스)와 비정형 데이터(영상, 음성)를 대상으로 하는 상용 솔루션이 다양하게 존재한다(행정안전부, 2016).

3.2.2.3.1 국내 개인정보 비식별화 기술 및 적용사례

○ 개인정보 비식별화 솔루션 기술

국내 개인정보 비식별화 솔루션은 규칙 기반 처리 ⇨ AI 기반 지능형 탐지 ⇨ 생성형 AI와 결합된 맥락 보존형 비식별화로 진화 중이며, 시장은 공공(기록물공개)·금융(신용정보 활용)·의료(EHR, DICOM · 산업(고객데이터 분석) 전반으로 확대되고 있다.[22)]

- 첫째, 정형 데이터 중심 솔루션은 데이터베이스(DB)내 주민등록번호, 계좌번호, 전화번호 등 구조화된 개인정보를 정규표현식과 마스킹 등 규칙기반으로 비식별화 한다.

① Fasoo-AnalitycCID(제조사 : ㈜ 파수)

- 특징: K-익명성(anonymity), I다양성(diversity), T-접근성(closeness) 기반 위험도 평가 지원
- 금융권, 공공기관 데이터 결합 프로젝트에 다수 적용

② 지란지교데이터-ID FILTER(제조사 : 지란지교데이터)

- 특징:데이터 마스킹, 범주화, 가명처리 기능 제공
- 금융 : 공공기관 DB 비식별화에 주로 사용

③ 데이터스-PAmaster(제조사 : 데이터스)

- 특징 : 금융·통신 데이터 기반 가명처리 및 결합 지원

- 둘째, 비정형 데이터 중심 솔루션은 문서, 이미지, 영상, 음성 등 비정형 데이터(OCR, STT, NER 등)의 개인정보를 탐지 및 비식별화 한다.

---

22) 한국인터넷진흥원(KISA)-개인정보 비식별화 기술 및 산업동향, 한국정보보호학회(KIISE)논문-개인정보 비식별화 기술의 발전과 산업적 응용, 산업연구원(KIET)보고서-개인정보 비식별화 산업의 현황과 정책 과제

① Textscope-Privacy Guard(제조사 : ㈜ 로민)

- 특징 : AI-OCR + NLP 기반 비식별화
- 비전자 기록물 디지털화 및 문서 내 개인정보 자동 탐지

② AI BlackMarker(제조사 : ㈜ 이파피루스)

- 특징 : PDF·이미지 문서 내 개인정보 자동 검출 및 마스킹

③ KT NexR-NEA(NexR Enterprise Anonymizer(제조사 : KT NexR)

- 특징 : 대용량 데이터 세트 자동 비식별화, 클라우드 연계 가능

- 셋째, 혼합형(정형+비정형) 솔루션은 데이터베이스와 비정형 파일을 동시에 지원한다.

① DataEye PIDI(제조사 : ㈜ 펜타시스템)

- 특징 : 정형/비정형 데이터 통합 비식별화 지원, 로그 데이터 처리 가능

② Identity Shield(제조사 : ㈜ 이지서티)

- 특징 : 데이터 가명화·익명화 및 개인정보 검출, 로그 관리기능 포함

○ 국내 솔루션 적용 사례

① 공공부문 :한국인터넷진흥원(KISA) 가명처리 지원센터를 통한 데이터 비식별화 지원

② 금융 부문: 금융보안원 주도의 신용정보 가명처리·결합 지원센터 운영

③ 의료부문 : 세브란스병원, 서울아산병원 등 대형 병원에서 EHR(Electronic Health Record, 전자 건강 기록) 및 의료 영상데이터 가명처리 연구 진행

④ 산업부문 : 카드사·통신사 등 대규모 고객 데이터를 활용한 가명정보 결합 프로젝트 진행

○ 국내 비식별화 솔루션 비교

개인정보는 물론 기업이 갖고 있는 중요 문서 데이터를 중심으로 사용된다. 예를 들면 기업의 영업비밀이나 기업정보, 공공기관과 정부 데이터, 스마트 공장 등의 IoT 및 센서 데이터, 금융 거래 데이터, 의료 및 바이오 데이터, 산업안전 데이터 등 다양한 데이터를 보호하는데 사용할 수 있다. 국내에서 대표적으로 사용되고 있는 주요 솔루션을 살펴봤다.

① 지란지교데이터 개인정보 비식별화 솔루션은 정보의 일부 또는 전부를 대체하거나 삭제, 특정 개인이 식별되지 않는 정보로 처리하는 소프트웨어를 의미, 국내 유통되는 개인정보 비식별화 솔루션은 대부분 개인정보보호위원회에서 제공하는 가명정보 처리 가이드라인에 맞춘 비식별 처리를 지원한다.

② 이지서티는, 목적에 따른 구분으로 설명하고 바로 개인정보 데이터를 가명·익명처리하는 전문 솔루션과 각각의 가명정보를 결합해 결합정보를 생성하는 솔루션이다.

보안전문가들의 비식별 솔루션에 대한 설문조사를 2025년 7월1일부터 4일까지 약 10만명의 보안전문가들에게 설문조사를 진행했다. 보안전문가들이 인터뷰한 결과는 7장 참고문헌에 제시한다.

○ 국내 개인정보탐지 + 마스킹 솔루션 기능 비교표

솔루션명	제조사	적용문서	주요기술	적용사례 및 활용분야
Fasoo AnalyticDID	㈜파수	Word, PDF, Excell, CSV	• AI기반 텍스트·비정형(이미지) 개인정보 검출 ⇨ NLP/패턴+딥러닝 OCR 적용 • 가명처리(부분 삭제·대치) 및 익명화(총체삭제)기법 • K-익명성 등 프라이버시 모델 충족 데이터 처리	• 자동 마스킹/익명화 API제공, 데이터 속성에 맞는 마스킹 규칙 적용 • 금융마이데이터, 공공 빅데이터 플랫폼 등. 전문기관 가명정보 결합시스템 등 다수 사례
Lomin Privacy Guard	㈜로민	Word, PDF, HWP, XLS, CSV	• NLP+딥러닝 OCR 기반으로 이미지 포함 문서에서 개인정보 탐지 • 사용자가 추가/제거 수정 가능. • 이미지 내 얼굴 흐리기 등	• 보험청구서 자동 블라인드, • 금융권 문서 아카이빙 개인정보 필터링, 공공 기록물 비식별화 등. • 문서기반 개인정보보호 전반
L7Security - AI Masking	엘 세 븐 시큐리티	HWP, MS오피스, PDF, 이미지파일	•OCR+학습 AI 모델로 MS 오피스·PDF·이미지 내 개인정보 탐지	• 웹 기반 마스킹 인터페이스 제 공 (자동+수동 보정), 다양한 포맷 마스킹 지원
셜록홈즈 프라이버시센터 V5.0	컴트루테크놀로지	PDF, HWP, DOCX, XLS	• 자체 AI OCR·패턴 매칭으로 문서·이미지·OLE 내 개인정보 탐지 • 이미지 마스킹(검출 후 마스킹), 원문/비식별 결과 뷰어·수동 보정 지원	• Proxy 방식을 이용한 웹서버를 통한 개인정보 탐지 • 전자결재, 그룹웨어 등 다양한 시스템 연동 등
KPST Privacy Guard	㈜한국플랫폼서비스기술	PDF, JPEG 등 이미지	• CCTV/영상 등 다중 도메인에 대한 AI기반 영상 개인정보 검출(얼굴·번호판 등) • 영상 스트림 익명화	• 실시간 스트리밍/엣지 수준에서 얼굴·블러·비식별(마스킹)제공 • CCTV 관제영상 프라이버시 보호 (도시안전),
AIDeep DeepPrivacy	케이사인	HDF / 스캔문서 (PDF, HWP, DOCX 등)	• 얼굴 영상 모자이크, 문자인식 결과 가명치환 등 익명화/가명화 병행. • 원본 대체하여 보안성과 데이터 활용성 동시 확보	• 스마트 시티 CCTV 개인정보 보호 시범사업 • 의료영상 아카이빙 솔루션 적용 추진
EasyCerti-IDENTITY SHIELD V3.0	이지서티	HWP, CSV HWPX	• HTTP/HTTPS 통신 구간에서 개인 정보 및 불건전 게시물 실시간 탐지·차단	• 지자체·교육청 웹사이트 게시판 개인정보 노출방지 • 기업내부 망분리 환경 개인정보 차단
AnnLab Data Encryption	안랩	DOCX, TXT PDF, XML	• 암호화를 통한 비식별화 (데이터 내용 복호화 전까지 숨김). • 전송구간 암호화	• 금융사 PC 파일보고(DLP)사례다수 • 공공기관 망연계 데이터 교환시 개인정보 검출시 암호화 후 전송
셜록홈즈 프라이버시센터 V5.0	컴트루테크놀로지	PDF, HWP, DOCX, XLS	• 이미지(OCR)내 개인정보 추출	• Proxy 방식을 이용한 웹서버를 통한 개인정보 탐지 • 전자결재, 그룹웨어 등 다양한 시스템 연동 등
AIFILTER V2.0	지란지교데이터	PDF	• OCR을 통한 텍스트 추출 • 첨부파일 이미지 필터링	• API 연동을 통해 업무 시스템과 연결하여 데이터 관리 및 보호 기능 제공
ImageScan& Masking V2.0	신화시큐리티	PDF	• 개인정보 스캔 및 마스킹 • 다양한 이미지 파일 및 포맷지원 • 스캔 후 로그수집 및 통계 조회	• 웹서버, WAS서버, DB서버와 통신이 가능토록 연계지원

(표 15 - Feature Comparison Table of De-identification Solutions)

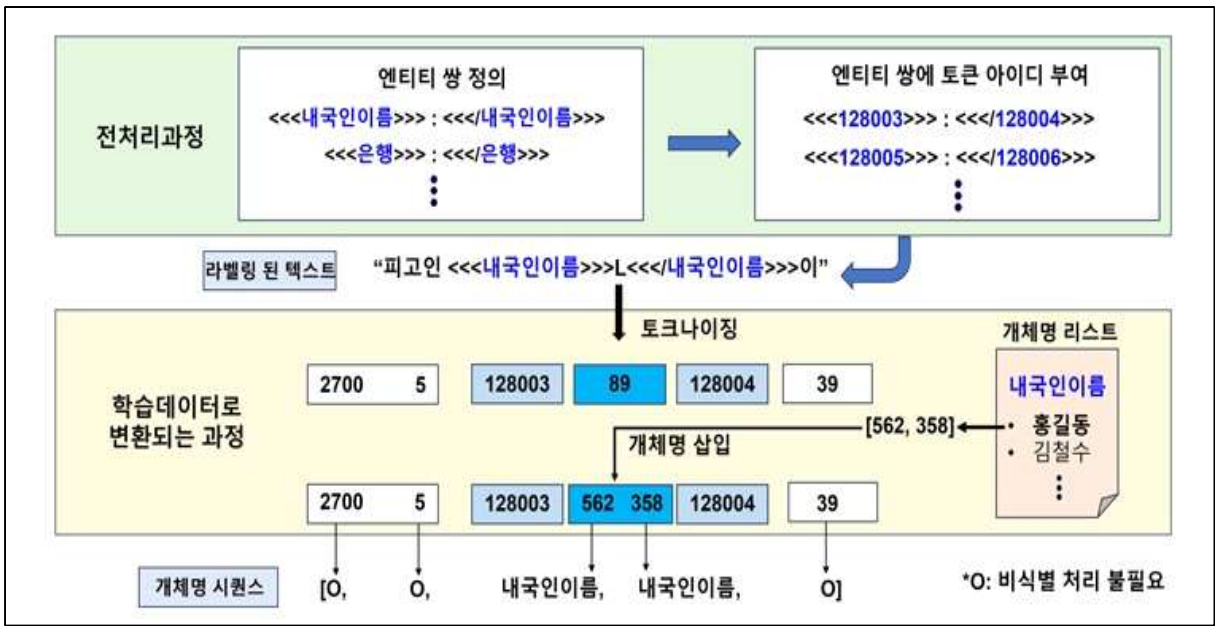
3.2.3 사례조사

3.2.3.1 국내사례

○ 국내 공공개자료를 중심으로 조사해 본 결과, 국가기록원(National Archives of Korea)차원에서 ”지능형 개인정보 비식별화(Intelligent de-identification)“를 명시적으로 적용했다는 공개사례는 확인되지 않았다. 다만, 관련 가능성 높은 기능, 기구 개선 사업 및 정책흐름이 포착되었으나, 연구자료로 확인된 것은 미비하였다. 이는 기록원 내부적으로 파일럿이나 연구를 진행했지만 공개자료로 승인 안되거나 보안 / 법률 이유로 비공개 처리된 것으로 사료된다.

○ 판결문(재판기록)에서 개인정보를 AI/규칙 기반으로 자동 식별·비식별화하는 연구와 적용 사례가 보고되었고 판결문 특유의 “기타정보(사전/관계자 특정 가능 정보)”처리 기준·자동화가 연구 대상이 되었고 AI 모델을 활용한 비식별화 방안 연구가 2024년 서울대학교 데이터사이언스대학원에서 “판결문 비식별화, 데이터 세트과 AI 모델로 해결한다”라는 주제로 발표하였다.[23)]

○ 이재진 교수 연구팀(서울대학교 데이터사이언스대학원)(Sungen Hahm, Heejin Kim, Gyuseong Lee, Hyunji Park 외, 2025)은 형사사건(강제추행·폭행·사기) 판결문 4,500건을 대상으로 약 2만 7천여 개의 개체(personal entity)를 수작업 라벨링하고(서울대학교 / gsds.snu.ac.kr) 개인정보/식별정보 유형을 595종으로 세분화한 체계를 구축하였다(서울대학교).



(그림 19 - Conceptual Diagram of Professor Jaejin Lee’s Research Group, Seoul National University)

23) 해당 연구는 ‘Thunder-DeID: An Accurate and Efficient De-identification Framework for Korean Court Judgments’ 제목으로 투고되었다. 출처 : https://arxiv.org/abs/2506.15266

- 한국어의 특성(교착어, 조사 결합 등)을 반영한 토크나이저(tokenizer)를 개발하고, 개체명 인식(NER) + 비식별화 처리를 위한 학습데이터를 구축하였다(서울대학교).
- 개발된 모델 “SNU Thunder-DeID”는 판결문 내 표현이 비식별화 대상 여부를 판별하는 정확도가 99% 이상, 유형 분류 정확도는 89% 이상이라는 성능을 보고하였다(cse.snu.ac.kr).
- 연구팀은 모델, 데이터 세트, 소스코드 등을 공개하여 연구·실무 활용 가능성을 열어 놓았다.

○ 임을영, 현민성. (2025). 생성형 인공지능을 활용한 판결문 공개 제도 개선
- 현행 판결문 공개 제도가 수동 비식별화에 많이 의존하고 있어 공개율이 낮고 비효율적이라는 문제를 지적하였다.
- 생성형 AI 기반 자동 비식별화 시스템 설계 및 구현 방안을 제시함. 실제 판결문 데이터를 활용해, 외부 대형 언어모델(LLM)과의 비교 실험을 통해 파인튜닝된 소형 언어모델(sLLM, 7-8B 매개변수 규모)이 높은 F1 점수(98% 이상)를 달성함을 보고하였다.
- 또한 자동화된 비식별화 시스템 도입 시 발생할 수 있는 법적 책임 문제(비식별화 오류 시 책임 소재 등)도 분석하고 있어 기술적 측면뿐 아니라 제도/법률적 측면까지 고려하고 있다.

○ 국가·공공기록(전자기록)에서의 AI 도입·파일럿 연구
- “지능형 기록정보서비스” 구축 연구들이 있어 기록관·국가기록 시스템에 AI 기반 자동분류·메타데이터 생성·비식별화 기술 적용 가능성을 탐색·제안했다.
- 김태영, 강주연, 김건 외. (2018). 지능형 기록정보서비스를 위한 선진 기술 현황 분석 및 적용방안. 한국기록관리학회지, 18(4), 149-182. 가 있다.

○ 김포시는 비전자 기록물 디지털화를 위한 AI-OCR 도입하여 디지털이관 기록의 효율적 선별과 비공개 정보 탐지를 수행(김포시 기록관, 2023)

구분	세부내용
접근방식	• 비전자 기록물 디지털화를 위한 AI-OCR 도입
성과	• 98% 인식률, 검색 가능한 DB 구축 → 비식별화 적용의 기반 마련
한계	• OCR 오류 발생 시 비식별화 정확도 저하
시사점	• 디지털 전환 단계에서 개인정보 보호를 고려해야 함

(표 16 - Detailed Plan for the Digitization of Non-electronic Records of Gimpo City)

- 김포시는 한국지능정보사회진흥원(NIA)의 “2021년 공공부문 클라우드 선도 프로젝트” 사업으로 선정되어, 지원금을 받아 공공클라우드 기반의 AI-OCR 프로젝트를 수행했다.
- 이 OCR 기술은 단순 문자 인식(OCR)뿐 아니라 필기체, 한자, 영어 등도 인식하도록 설계되었으며, 이를 통해 문서 내부의 본문 검색 가능 형태로 전환하였다.
- 비전자 기록물(종이문서 등)을 스캔 → AI-OCR을 통해 텍스트 추출 → 텍스트가 포함된 데이터베이스화 함으로써, 기존 ‘이미지 보기/목록’ 중심의 구조에서 ‘검색·활용 가능한 데이터’로 전환

하였다.

- 기대효과
  - 검색성 강화 : 종이문서 내부의 본문까지 검색 가능해짐으로써, 조회 시간이 줄고 이용자 경험이 개선되었다.
  - 업무효율 향상: 비전자 기록물의 수작업 검색/열람에서 벗어나 자동화·데이터화로 관리비용 및 시간 절감 가능성이 커졌다.
  - 기록관리 전문성 강화 : 공공기관 기록관리 분야에서 클라우드·AI 기술을 적용한 선도 사례로 자리매김할 수 있는 기반이 되었다.

3.2.3.2 국외사례

○ 영국 국가기록원(TNA : The National Archives)(UK National Archives, 2022)

영국 국가기록원(The National Archives)은 AI 기반 자동 분류 및 민감정보 탐지 시스템인 * InSight®*를 도입하여, 디지털 이관 기록의 효율적 선별과 비공개 정보 탐지를 수행

구분	세부내용
목적	• 공개 예정 기록물에서 비공개대상정보(개인정보, 기밀정보)를 자동탐지
접근 방식	• Iron Mountain과 협력, AI 기반 InSightTM시스템 도입 • OCR+NLP 기반 인공지능 모델을 사용, 수천 건의 디지털 기록물에서 이름, 주소, 금융정보, 보안관련 문구 등을 탐지
성과	• 17,000건 이상의 기록물 처리, 탐지 정확도 85% 이상. 수작업 대비 처리 효율성 대폭 향상.
범주	• 개인정보, 행정기밀
한계	• 오탐·누락 존재 → 전문가 검증 필요.
시사점	• 기록공개 업무의 부담을 줄이고, 법적 요구사항에 부합하는 공개/비공개 분류 지원

(표 17 - Detailed Specifications for AI-based Automated Classification at TNA)

○ 미국 국가문서관리청(NARA) 개인정보(PII) 자동탐지 시범[24)]

구분	세부내용
목적	• 디지털 기록물(이메일, 보고서 등) 내 개인식별정보 자동탐지 및 편집 (Redaction)
접근방식	• 클라우드(AWS, Google Cloud)기반 NER 모델을 사용, 주민등록번호(SSN), 주소, 의료번호 등 탐지
범주	• PII(SSN, 주소 등)
현황	• 현재 시범단계, 사용자 정의(custom entity)기반 민감정보 필드 확장 기능 개발 중
시사점	• 기록 공개 전 검증 과정을 자동화, 대량 행정기록에서 비공개대상정보 누락 방지

(표 18 - Detailed Specifications for Automated PII Detection at NARA)

24) Meystre, S. M., Ferrández, Ó., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2014). Text de-identification for privacy protection: a study of its impact on clinical text information content. Journal of Biomedical Informatics, 50, 142-150.

○ 미국 재향군인청(VHA) -BoB 시스템[25)]

구분	세부내용
목적	• 의료기록물 내 환자 개인정보(PHI) 자동 탐지
접근방식	• 규칙기반 탐지(정규표현식) + 기계학습 NER(CRF, SVM) 결합
범주	• 의료 PHI
성과	• 민감정보 검출에서 F1 0.90 달성.
시사점	• 의료 기록이라는 특수 영역에서 비공개대상정보 검출의 자동화 기능성 입증

(표 19 - Detailed Specifications for the VHA - BoB System)

○ 캐나다 뉴브런즈윅(州) - 이메일 기록 자동분류(Government of New Brunswick, 2025)

구분	세부내용
목적	• 기관 이메일 중 공개 가능한 정보화 비공개대상정보 포함 기록을 구분
접근방식	• 기계학습(SVM) 기반 텍스트 분류기로 기밀도·업무적 가치 평가
범주	• 이메일 기밀정보
성과	• 업무 이메일 중 민감정보 포함여부를 자동 식별, 기록관리 정책에 따라 접근 통제

(표 20 - Detailed Specifications for Automated Classification of Email Records – New Brunswick, Canada)

○ 기록메타데이터 자동생성 파일럿(Auto-fill Descriptive Metadata)[26)]

- 진행시기 : 2024년 ~ 현재
- 내용 : 디지털 기록물에 대한 설명(metadata)을 AI가 자동 생성하는 파일럿을 진행 중이다. 머신러닝 모델이 문서 내용과 기존 메타데이터를 분석해 요약문이나 주제 키워드 등의 기술정보를 생성하며, 이를 통해 아키비스트의 작업 시간을 절감하고 대중이 기록을 더 쉽게 검색하도록 돕는 것이 목적이다.
- 평가 : 이 파일럿은 현재 성능 평가 단계로, NARA 직원들이 AI 생성 메타데이터의 품질을 검증하고 있다.

○ 시멘틱 검색 파일럿(Semantic Search Pilot)[27)]

- 진행시기 : 2024년 ~ 현재

---

25)NARA (National Archives and Records Administration) / Library of Congress / 2024.9.6.National Archives and Records Administration(미국 국가기록원, NARA)에서 진행 중인 “Auto-fill of Descriptive Metadata for Archival Descriptions” 파일럿 사업의 출처
Exploring AI at the National Archives and Records … (2024년 2월) 내 Use Case 목록 중 “Auto-fill of Descriptive Metadata for Archival Descriptions (Planned)

26) NARA (National Archives and Records Administration) / Library of Congress / 2024.9.6.National Archives and Records Administration(미국 국가기록원, NARA)에서 진행 중인 “Auto-fill of Descriptive Metadata for Archival Descriptions” 파일럿 사업의 출처
Exploring AI at the National Archives and Records … (2024년 2월) 내 Use Case 목록 중 “Auto-fill of Descriptive Metadata for Archival Descriptions (Planned)

27) AI Use Case Inventory (웹페이지 + 2023 Agency Inventory Excel), “Semantic search pilot with NAC using AWS (Pilot phase), National Archives getting a big boost from AI to transform its search capabilities” (Oct 23, 2024).,Machine Learning Models Expedite Federal Tech Efforts” (Jun 2, 2025)

- 내용 : 기존 키워드 매칭을 넘어 문맥과 의미를 이해하는 차세대 검색엔진을 NARA 온라인 카탈로그에 도입하려는 시범 사업. 초기에는 오픈소스 LLM, AWS Titan, Google Vertex AI 등 다양한 AI 모델을 평가했으며, 테스트 결과 Google의 Vertex AI[28](Gemini 대규모 언어모델)을 적용하기로 결정되었다.
- 평가 : 시맨틱 검색은 방대한 비정형 데이터에서도 사용자의 의도를 파악하여 숨겨진 관련 기록까지 찾아 주므로 연구자와 대중의 검색 효율을 크게 높일 것으로 기대된다.

○ 지식기반 챗봇 인터페이스 파일럿[29]

- 진행시기 : 2024년 ~
- 내용 : 국가인사기록센터(NPRC)의 업무 지침서(Case Reference Guide) 등 방대한 내부 지식 자료를 효과적으로 검색하기 위해, AI 기반 질의응답형 챗봇 UI를 개발하고 있고 Amazon Kendra와 같은 AI 검색엔진을 활용해 PDF, HTML, 이미지 등 다양한 형식의 문서에서 관련 정보를 찾아내고, 사용자의 질문에 대화 형태로 답변하도록 설계되었다.
- 평가 : 직원 교육 시간을 줄이고 민원 대응의 정확도를 높이는 것이 목표이다.

○ 생성형 AI 생산성 향상 파일럿(Google Workspace Enterprise용 Gemini 파일럿)[30]

- 진행시기 : 2024년 ~ 현재 / Google
- 내용 : ChatGPT로 대표되는 생성형 AI를 업무 생산성에 활용하는 실험을 진행 중으로 약 50명의 직원을 대상으로 Google Workspace에 통합된 Google “Gemini” 언어모델을 시험 도입하여, 이메일 초안 작성, 문서 요약, 슬라이드 콘텐츠 생성, 스프레드시트 분석 자동화 등 다양한 업무 시나리오에서 효율성을 평가하고 있다.
- 평가 : 이 파일럿은 임직원들이 생성형 AI를 활용해 협업·의사소통을 개선하고 반복 업무를 자동화함으로써 업무 혁신 방안을 모색하고 있다.

3.2.4 시사점

디지털 전환의 가속화와 데이터 활용의 폭발적 증가로 인해, 공공·민간 부문 모두에서 개인정보 보호와 데이터 활용의 균형이 핵심 과제로 부상하고 있다. 기존의 비식별화 업무는 전문가의 수작업과 정형 데이터 중심의 규칙 기반 처리가 주를 이루어 왔으나, 최근에는 대규모 비정형 데이터(문서, 이미지, 음성 등)의 증가로 인해 자동화·지능화가 불가피한 상황이다. 이에 따라, 인공지능(AI) 기반 개인정보 비식별화 업무 지능화는 단순한 기술적 개선을 넘어 데이터 거버넌스 체계 전반의 혁신을 요구한다.

---

28) Vertex AI는 구글 클라우드(Google Cloud)에서 제공하는 통합 AI/ML 플랫폼(2021, 기존 AI Platform과 AutoML을 통합하여 출시)

29) Exploring AI at the National Archives” – NASS Winter 2024 프레젠테이션- Create an AI based knowledge articles chat interface for CRG documents (Planned)
,National Archives getting a big boost from AI …” (Oct 23, 2024)

30) https://workspace.google.com/blog/customer-stories/thoughtworks-gemini-google-workspace?utm_source=chatgpt.com&hl

❍ 정책적·법제도적 기반의 정비가 필수적이다.

- 데이터 3법 및 개인정보 보호법 개정 이후에도, AI 비식별화 결과의 법적 효력·책임 범위에 대한 명확한 가이드라인이 부족한 실정이다. 따라서 기술적·법적 기준을 종합적으로 반영한 '지능형 비식별화 표준 프레임워크'를 마련함으로써, 공공과 민간 모두에서 안전하고 효율적인 데이터 활용 기반을 조성해야 한다.

❍ AI 기반 자동 검출 및 비식별화 기술의 정교화가 필요하다.

- 자연어 처리(NLP), 개체명 인식(NER), 딥러닝 기반 문맥 이해 모델(BERT, BiLSTM-CRF 등)의 고도화를 통해 비정형 데이터에서 개인정보를 정확히 탐지하고 문맥에 맞는 비식별 처리를 수행할 수 있는 체계가 요구된다. 특히, 공공기록물·행정문서와 같은 복합형 데이터에 대응하기 위해 학습 데이터셋의 품질 확보와 도메인 특화 모델 개발이 병행되어야 한다.

❍ 비식별화 업무 프로세스의 표준화 및 신뢰성 검증 체계 구축이 중요하다.

- 자동화된 비식별 조치 결과의 품질을 검증하기 위한 기준과 절차를 마련하고, 이를 평가할 수 있는 메타데이터 기반의 감사 체계를 도입함으로써 비식별화 과정의 투명성과 재현성을 확보해야 한다.

❍ 인적·조직적 역량 강화와 윤리적 기준 확립이 병행되어야 한다.

- 자동화 시스템에 대한 과도한 의존을 방지하고, 데이터 거버넌스 담당자와 기록연구사, 개인정보호 전문가 간 협업체계를 강화해야 한다. 또한 AI 모델의 편향, 오탐·누락 가능성에 대응하기 위한 인간 중심(Human-in-the-loop) 검증 구조를 포함한 윤리적 관리 프레임워크가 필요하다.

❍ 인공지능과 기록관리의 융합 거버넌스 구축이 필요하다.

- 비식별화 시스템은 단순히 개인정보를 제거하는 기능에 머물지 않고, 기록물의 맥락(Context)을 유지하면서 관리·활용 가능한 데이터 형태로 전환해야 한다. 이를 위해 AI-OCR, 시멘틱 검색(Semantic Search), 메타데이터 자동생성(Auto-fill Descriptive Metadata) 등과 연계된 통합형 플랫폼 구성이 바람직하다.

3.3. 데이터 수집 및 정제 방안

본 연구에서는 LLM이 공개재분류 사례를 참조하기 위한 기반 지식정보로서, 국가기록원의 공통·고유업무 세부유형 기준을 정제한 후 이를 활용하여 벡터 데이터베이스(Vector DB)를 구축하고, 이를 기반으로 RAG(Retrieval Augmented Generation) 구조를 구현하였다. 본 연구에서 구축한 RAG는 일반적인 문서 검색 중심의 RAG와 달리 조건 인지형 RAG(Condition-aware RAG) 방식으로 설계되었다. 이를 위해 벡터DB의 메타데이터에 조건(추가 질의)을 구조적으로 포함시켰으며, 이러한 조건 정보를 근거로 언어모델이 보다 정확하고 상황에 부합하는 답변을 생성하는 방식으로 동작한다.

3.3.1. 연구개요

○ 동작방식

본 모델의 동작 절차는 다음과 같이 구성된다.

1. 질문 전처리 단계에서는 LLM이 재분류 대상 문서를 요약하여 검색 효율을 높일 수 있는 질의 형태로 변환한다.
2. 검색 단계에서는 전처리된 요약 결과를 기반으로, 벡터DB 내에서 가장 유사한 재분류 사례와 해당 사례에 연계된 조건(추가질의) 정보를 탐색한다.
3. 판단 단계에서는 검색된 재분류 사례와 조건 정보를 근거로 LLM이 판단을 수행하고, 이에 따라 최종 재분류 결과 및 설명을 생성한다.

○ 장점

본 연구에서 구축한 조건 인지형 RAG 기반 재분류 모델은 다음과 같은 장점을 갖는다.

- 최신성 측면에서 RAG를 구성한 이후에도 추가되는 지식정보나 변경 사항을 즉시 반영할 수 있어 지속적인 갱신과 관리가 용이하다.
- 신뢰성 측면에서는 생산기관별 재분류 사례를 기반으로 판단 범위를 구조적으로 제한함으로써, 모델의 정확성을 향상시키고 LLM의 환각(hallucination) 발생 가능성을 최소화할 수 있다.
- 비용 및 유연성 측면에서는 AI 모델을 변경하더라도 이미 구축된 RAG 구조를 그대로 재사용할 수 있어, 시스템 전환 비용을 절감하고 확장성을 확보할 수 있다.

3.3.2. 주요기능 및 구현방법

비공개기록물 공개재분류 세부유형기준서를 분석하여, RAG 구축을 위한 최종 포맷인 JSON 파일을 자동 생성하기 위해 프롬프트 엔지니어링을 수행하였다. 이 과정은 일반적으로 긴 문서를 임베딩하기 위해 수행하는 문서 청킹(chunking) 절차와는 구별되며, 기준서에 기술된 문서 유형 정보와 재분류 사

례 정보를 체계적으로 분리·추출하기 위한 특화된 절차에 해당한다.

❍ 프람프트 엔지니어링을 하기 위한 핵심원칙

- 근거 우선(Evidence-first): 근거가 없으면 답하지 말 것
- 역할·규칙 분리: System=역할/규칙, Developer=업무 도메인, User=요구/질문, Tool=검색
- 하드 가드레일: 접근등급·민감도·법적 금칙어는 규칙으로 강제
- 정형 출력: JSON/마크다운 스키마 고정
- Few-shot 최소/고충실: 실제 문서로 만든 도메인
- 하이브리드 검색 친화: 쿼리 확장·부서명/용어 매핑 지시
- 함구 조건 명시: “추정/유추 금지”, “근거 문장+출처 ID 필수”

❍ 1단계: 카테고리 유형 분류 프람프팅

- 원형문자 또는 특수문자 또는 문자 또는 숫자-숫자-숫자 XXX > YYYY > ZZZZZ
(경찰청, 대검찰청, 법무부, 해양경찰청, 행정안전부)
- 원형문자 또는 특수문자 또는 문자 또는 숫자-숫자 XXX : YYYY
(대학, 지역교육청)
- 공통업무 Type1 : 원형문자 XXX
- 공통업무 Type2 : 원형문자 또는 특수문자 또는 문자 또는 숫자 XXX > YYYY
- 공통업무 Type3 : 원형문자 또는 특수문자 또는 문자 또는 숫자-숫자 XXX : YYYY
- 카테고리 추출이 어려운 이유는 ‘①’와 같이 원문자를 사용한 방식과 ‘㊽’과 같이 겹치기 문자를 사용한 서로 다른 형태이기 때문

경찰청, 대검찰청, 법무부, 해양경찰청, 행정안전부 유형에 대한 프람프트 엔지니어링 예시

첨부한 파일에서 "원형문자 또는 특수문자 또는 문자 또는 숫자-숫자-숫자 XXX > YYYY > ZZZZZ" 제목에 대해 각각 해당하는 표 하나를 두개의 행으로 해서
[아이디],[조직],[상위유형],[세부유형],[세세부유형],[기능 및 추가 설명],[기록물 철명],[주요내용],[구분]
를 추출 해서 엑셀파일로 만들어줘.

라벨의 순서는 [아이디],[조직],[상위유형],[세부유형],[세세부유형],[기능 및 추가 설명],[기록물 철명],[주요내용],[구분] 로 구성되어 있어.
라벨의 문자를 정확히 판단해줘. 특히 [기록물 철명]과 [주요내용] 을 명확히 판단해줘.
"■ 기록물 철명 및 주요내용"은 라벨로 판단하지 말아줘.
추출하는데, 네가 요약하거나 판단하지 말고 표 안에서 있는 원문 그대로 추출해줘.

[아이디]는 "원형문자 또는 특수문자 또는 문자 또는 숫자-숫자-숫자 XXX > YYYY > ZZZZZ" 중 원형문자 또는 특수문자 또는 문자 또는 숫자-숫자-숫자 를 내용으로 두개 행의 내용이

동일해.
[조직]은 "원형문자 또는 특수문자 또는 문자 또는 숫자-숫자-숫자 XXX > YYYY > ZZZZZ" 중 "XXX" 를 내용으로 두개 행의 내용이 동일해.
[상위유형]은 상위유형 항목으로 생성된 두개 행의 내용이 동일해.
[세부유형]은 세부유형 항목으로 생성된 두개 행의 내용이 동일해.
[세세부유형]은 세세부유형 항목으로 생성된 두개 행의 내용이 동일해.
[기능 및 추가 설명]은 기능 및 추가 설명 항목으로 생성된 두개 행의 내용이 동일해.
[기록물 철명]은 기록물 철명 항목으로 생성된 두개 행의 내용이 동일해.
[구분]은 30년 경과를 첫행에 30년 미경과를 두번째행의 내용으로 해줘.

❍ 2단계: 추가질의 생성 작업

- 기준서에 제시된 검토항목을 분석하여 추가 질의가 필요한 항목을 선별하고, 해당 추가 질의의 응답에 따라 재분류 결과가 달라질 수 있으므로, 질의 결과에 부합하는 공개유형 및 비공개 사유를 정확히 매칭하기 위한 필수 절차로 수행하였다.

법무부 출입국 관리 문서에 대한 추가질의 예시

외국과의 입국사증 및 수수료 면제 협정 등에 관한 기록인가?

❍ 3단계: 주요업무를 내용별로 구분하기 위한 프람프팅

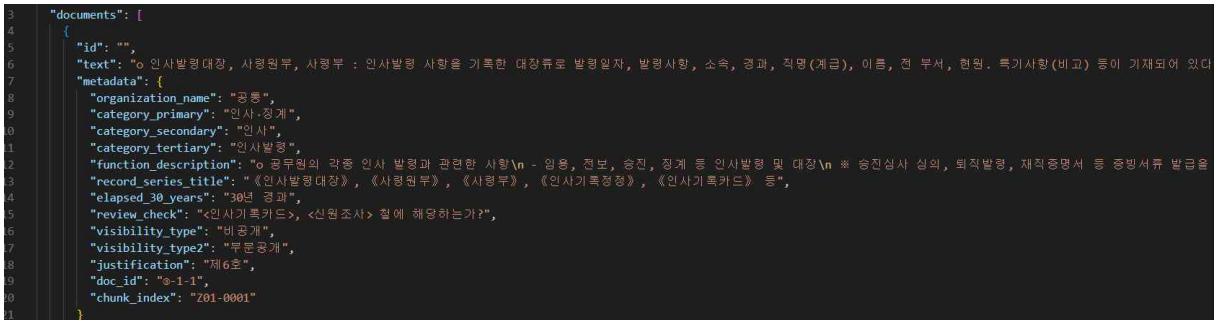
- 주요업무 항목내에 ○ 표식을 기준으로 업무가 구분되어 세부업무별로 나누기 위한 절차

내용별 구분을 위한 프람프팅 예시

첨부된 엑셀파일에서 B열에서 내용중 새로운 라인으로 시작할때 ○ 또는 ㅇ 또는 ◯ 표식을 기준으로 줄을 나누어 줘. B열이외의 열에서는 나누기 전 내용을 나눈 줄에 복사해서 넣어서 엑셀파일을 만들어줘. ○ 또는 ㅇ 또는 ◯ 표식을 제거하지 말고 포함해줘
==예시==
◯ 즉심사건부 : 사건별 피의자의 ...
◯ 범죄사건 처리부 : 번호, 범죄접수부 번호, ...
◯ 압수부 : 압수물 번호, 범죄사건 처리부, 압수년월일, ...
◯ 사건송치부 : 연월일, 사건번호, 죄명, ...
◯ 경범즉심사건부, 교통즉심사건부 : 사건별 피의자의 처분 내용을 ...
◯ 범죄사건부 : 번호, 범죄접수부 번호, ... 등이 기재
==예시==
예시는 6개 줄로 만들어야해

○ 4단계: JSON 포멧을 위한 프람프팅

- 최종생성된 엑셀을 JSON 포맷으로 변경하기 위한 절차



(그림 20 - JSON FORMAT)

○ 5단계: Vector DB 구축(파이선 프로그램 실행)

- JSON 포맷으로 생성한 파일을 Vector DB로 변경하기 위한 절차

(그림 21 - Vector Database Creation)

3.3.3. 구축 결과

○ 공통업무 구축결과

- 총 351개의 벡터 레코드(Vector Record)가 구축됨

text	organiza	cat	category_secon	cat	function_description	record_series_	elapsed	review_check	visibility_type	visibility	justification	doc_id	chunk_i
○ 인사발령대장, 사령원부, 사령부 : 인사발령 공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 인사발령원본 : 인사발령원본과 인사발령대장공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 단순인사명령 : 직위, 직급, 이름, 발령사항, 날공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 퇴직자신분서류 : 연번, 계급, 이름, 생년월일, 공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 복직 및 휴직 : 복직 및 휴직자에 관한 사항을공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 의원면직 : 면직에 관한 사항을 처리한 기록을공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 신원조사 : 신원조사 의뢰 및 회보 받은 내용.공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 경과	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 인사발령대장, 사령원부, 사령부 : 인사발령 공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 인사발령원본 : 인사발령원본과 인사발령대장공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-000
○ 단순인사명령 : 직위, 직급, 이름, 발령사항, 날공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-001
○ 퇴직자신분서류 : 연번, 계급, 이름, 생년월일, 공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-001
○ 복직 및 휴직 : 복직 및 휴직자에 관한 사항을공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-001
○ 의원면직 : 면직에 관한 사항을 처리한 기록을공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-001
○ 신원조사 : 신원조사 의뢰 및 회보 받은 내용.공통	인사	인사		인사발	○ 공무원의 각종 인사 발령과 관련한	《인사발령대장》	30년 미경	<인사기록카드>, <비공개		부분공개	제6호	③-1-1	Z01-001
○ 《공무원 채용》은 공무원 임용서 및 임용조사공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 경과	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-001
○ 《신규채용》은 공무원 신규채용에 따른 발령공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 경과	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-001
○ 《임용》은 공무원 임용(신규채용 등)에 대한 '공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 경과	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-001
○ 《5급이상 임용》, 《6급이하 일반직 임용》, 《7공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 경과	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-001
○ 《공무원 채용》은 공무원 임용서 및 임용조사공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 미경	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-001
○ 《신규채용》은 공무원 신규채용에 따른 발령공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 미경	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-002
○ 《임용》은 공무원 임용(신규채용 등)에 대한 '공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 미경	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-002
○ 《5급이상 임용》, 《6급이하 일반직 임용》, 《7공통	인사	인사		채용	○ 공무원의 일반직, 전문직 등 신규채	《공무원 채용》,	30년 미경	기록물 전체가 <용.비공개		부분공개	제6호	③-1-2	Z01-002
○ 《재직기간 합산 및 소급통산신청서》, 《경력공통	인사	인사		연금	○ 공무원의 연금지급 관련 심사기록,	《재직기간 합산	30년 경과	<기여금납입종결표비공개		부분공개	제6호	③-1-3	Z01-002
○ 《기여금납부내역》은 각 경찰지서에서 기여금공통	인사	인사		연금	○ 공무원의 연금지급 관련 심사기록,	《재직기간 합산	30년 경과	<기여금납입종결표비공개		부분공개	제6호	③-1-3	Z01-002

(그림 22 - RAG Deployment for Common Tasks)

○ 고유업무 구축결과

- 총 24개의 조직에 915개의 고유업무 벡터 레코드(Vector Record)가 구축됨

조직	벡터 레코드 수	조직	벡터 레코드 수
감사원	14	경찰청	116
고용노동부	40	공정거래위원회	12
과학기술정보통신부	8	관세청	24
교육부	22	국가인권위원회	28
국무조정실	8	국민권익위원회	16
국세청	102	대검찰청	116
대학	10	법무부	118
보건복지부	10	식품의약품안전처	47
여성가족부	18	원자력안전위원회	10
인사혁신처	72	지방자치단체	4
지역교육청	18	질병관리청	12
해양경찰청	40	행정안전부	50

(표 21 - Number of RAG Deployments by Organization for Inherent Duties)

3.3.4. 관련 연구 및 사업 수행 산출물 활용 방안

3.3.4.1. 연구개요

본 연구는 기존의 선행 연구 및 사업 수행 산출물과 기준서, 심의자료 등을 활용하는 방법에 대해서 설명에서 축적된 기준서 데이터와 심의 결과물을 AI 기반으로 재활용하고 확장할 수 있도록 체계개선을 제안한다. 공공기록물의 공개재분류 업무와 연계하여, 비공개 사유 판단의 근거로 활용되는 기준서 관리 시스템을 AI 기반으로 고도화가 필요하다.

기존의 기준서 관리시스템은 단순 키워드 검색 방식과 정형 데이터 중심의 관리 체계(DB 및 엑셀 병행)로 운영되어, 판단 기준과 과거 심의 사례, 법령 해석 이력을 연관성 있게 조회하기 어렵다는 한계를 가지고 있고, 이에 과거 사업에서 생성된 기준서, 심의 기록, 비공개 사유 사례 데이터를 LLM이 이해할 수 있는 형태로 정규화하고, 이를 RAG(Retrieval Augmented Generation) 기반으로 재구성함으로써, 기존 데이터의 활용성과 검색 정확도를 동시에 향상시킬 수 있다.

또한, 기존의 기준서 검색도구를 Chat 서비스 방식으로 전환하여, 사용자가 자연어로 질의하고, AI가 기준서·과거 사례·심의 결과를 통합적으로 제시하는 지능형 질의응답형 검색 서비스도 구현가능하다.
데이터 저장 구조 역시 DB + 벡터DB(Vector Database)의 복합 구성으로 확장되어, 정형 메타데이터와 비정형 텍스트 데이터의 의미 기반 통합 검색이 가능해진다.

예를 들어, 사용자가 “대검찰청 판결문 심의 관련 기준서는?”이라는 질의를 입력하거나 판결문 PDF 파

일을 첨부하면, 시스템은 현재 적용 기준, 과거 심의회 결과, 관련 법령 및 이력을 AI Chat 응답 형태로 자동 제공한다.

이러한 고도화는 기존 시스템의 단순 검색을 넘어, AI가 행정 지식과 사례를 종합 분석하여 실무자가 근거 기반의 판단을 내릴 수 있도록 지원하는 지식기반 행정지원 플랫폼으로 발전시키는 것을 목표로 한다.

3.3.4.2. 연구내용

기록물 공개재분류 업무에서 기준서는 각 문서의 공개·비공개 여부를 판정하는 핵심 판단 근거로 활용된다. 하지만 기존의 기준서 관리시스템은 단순 키워드 검색 방식에 의존하여, 판단 기준이나 과거 심의 사례를 연관성 있게 탐색하기 어렵다. 또한, DB와 엑셀 파일 형태로 분리 저장된 기준서 데이터는 버전 관리와 검색 효율성 측면에서도 한계를 보인다.

이로 인해 담당자가 “유사한 비공개 사유가 적용된 사례”나 “이전 심의회의 결정 기준”을 직관적으로 찾아보기 어렵다는 문제가 있었다. 따라서 본 연구에서는 기준서 관리시스템을 AI 기반으로 전환하여

① 데이터 저장체계(DB + 벡터DB 통합)
② 의미 기반 검색(RAG)
③ Chat 인터페이스 질의 응답형 검색 환경
으로 개선함으로써, 지능형 기준서 검색 서비스를 구현의 필요성을 제안한다.

3.3.4.3. 시스템 구조 및 기술 구성

고도화된 기준서 관리시스템은 DB와 벡터DB(Vector Database)가 결합된 RAG 구조로 설계할 수 있으며, 구체적인 기술 구성은 다음과 같다.

- 기존: 관계형 DB + 엑셀 기반의 정형 데이터 관리
- 개선: 관계형 DB + 벡터DB 통합 구조
- DB: 기준서 메타데이터(작성일, 담당부서, 조항번호 등) 저장
- 벡터DB: 본문 임베딩(Embedding) 정보 및 유사도 기반 검색 지원

이를 통해 텍스트 본문과 구조화 데이터의 통합 검색이 가능해진다.

○ RAG 기반 검색 아키텍처

사용자의 자연어 질의가 입력되면 LLM이 쿼리를 벡터화하여 벡터DB에서 유사한 기준서 및 사례를 검색한다. 검색된 결과는 LLM이 재조합하여 “현재 기준 + 과거 사례 + 법령 해석 + 심의 결과”를 요약·제시한다. 이 과정에서 단순 검색이 아닌, 문맥 이해 기반의 추론형 응답이 생성된다.

❍ Chat 서비스 인터페이스 도입

기존 텍스트 검색창 대신 Chat 형태의 인터페이스를 도입하여, 사용자가 자연어로 질문을 입력하고 즉시 응답을 받을 수 있도록 하였다.

- 예시 질의: “대검찰청 판결문 심의 관련 기준서는?” 판결문 PDF 파일 첨부
- 응답 예시: “현재 적용 기준은 ‘사법기록물 비공개 판단 세부기준(제4조 제2항)’이며, 과거 2021년 및 2023년 심의회의 관련 사례가 존재합니다. 주요 결정 요지는 ‘개인정보 포함 여부 및 공공의 이익 비중에 따른 부분공개 결정’입니다.”

이를 통해 사용자는 검색어 기반이 아닌 대화형 질의-지식형 응답 방식을 통해 기준서와 관련 이력 정보를 통합적으로 조회할 수 있다.

3.3.4.4. 주요 기능 및 구현 방법

구분	기존 방식	개선 방식
데이터 관리	DB + 엑셀 병행 운영	DB + 벡터DB 통합 관리
검색 방식	키워드 검색	RAG 기반 의미 검색
사용자 인터페이스	텍스트 입력창	Chat 기반 대화형 질의응답
검색 결과	단일 기준서 내용 출력	현재 기준 + 과거 사례 + 심의 이력 + 법령 연결
확장성	정형 데이터 중심	비정형 문서 및 첨부파일 지원 (PDF, DOCX 등)

(표 22 - Key Features and Implementation Methods of the System)

❍ 기준서 검색 및 연관 문서 탐색

LLM은 사용자의 질의 내용을 벡터화하고, 벡터DB에서 의미적으로 가장 유사한 기준서를 탐색한다. 검색 시, 현재 기준뿐 아니라 과거 심의회의 회의록, 법령 해석서, 비공개사유 결정이력 등도 함께 연계 검색한다.

❍ 첨부파일 기반 문맥 검색

사용자가 기준서 관련 문서를 PDF 형태로 첨부할 경우, LLM은 해당 문서의 내용을 분석하여 관련 기준서 조항 및 심의이력을 함께 제시한다. 이를 통해 비정형 문서에 포함된 의미적 맥락을 기반으로 한 ‘문서-기준서 연동형 검색’을 구현할 수 있다.

❍ 통합형 CAMS 연동 구조

본 기능은 CAMS 시스템 내 통합 화면으로 구현되어 기존 업무 프로세스 변경 없이 AI 검색 기능을 자연스럽게 추가하였다. ‘기준서 검색’ 탭 내에서 Chat 인터페이스를 통해 즉시 질의·응답이 가능하며 검색 결과는 기준서 원문, 적용 사례, 과거 심의 결과의 세 부분으로 구분되어 표시된다.

3.3.4.5. 기대효과 및 결론

AI 기반 기준서 관리시스템의 도입은 단순한 검색 효율 개선을 넘어, 지식기반 행정지원 체계로의 전환을 가능하게 한다.

- 판단 근거의 투명성 강화 측면에서, AI가 제시하는 기준서 내용, 관련 법령, 과거 재분류 사례 등을 함께 확인할 수 있어 공개재분류 판단의 객관성과 일관성을 확보할 수 있다.
- 업무 효율성 향상 측면에서, 기존의 엑셀 파일 또는 문서 기반의 수동 탐색 방식 대비 검색 시간이 크게 단축되며, 신규 담당자도 기준서를 보다 빠르고 정확하게 이해·활용할 수 있다.
- 의사결정 지원 고도화 측면에서, LLM이 과거 심의 경향과 유사사례를 함께 제시함으로써 심의위원회 및 담당자의 판단 품질을 높일 수 있다.
- 지속 가능한 데이터 관리 구조 확립 측면에서, 구조화된 DB와 벡터DB를 병행 운영함으로써 기준서 데이터의 갱신, 신규 추가, 이력 관리 등이 자동화되어 장기적으로 안정적인 데이터 관리가 가능해진다.

결론적으로, 본 연구를 통해 제시된 AI Chat 기반 기준서 관리시스템은 기록물 공개재분류 업무의 법령해석, 기준 검토, 사례 연계 분석을 통합 지원하는 차세대 행정 AI 플랫폼의 핵심 구성 요소로 발전할 수 있는 기반을 마련할 수 있다.

3.3.5. 주민등록정보 연계를 통한 사망자 여부 판별 서비스 구축 방안

3.3.5.1. 연구 개요

본 연구는 국가기록원의 공개재분류 업무 수행 과정에서 개인정보 여부 판단의 정확성을 향상시키기 위해, 행정정보공동이용센터(행정안전부)가 제공하는 「주민등록정보」 서비스를 인공지능(AI) 기반 기록관리 시스템과 연계하는 방안을 제시한다. 특히 비공개 기록물의 재분류 과정에서는 「공공기관의 정보공개에 관한 법률」 제9조 제1항 제6호에 따라 “살아있는 개인에 관한 정보” 여부가 비공개 판단의 주요 기준이 되므로, 해당 인물의 사망 여부를 정확히 식별하는 것이 중요하다. 이에 본 연구에서는 두 가지 형태의 행정정보 연계 서비스를 분석하였다.

① 실시간 조회(API 기반 서비스)와 ② 대량유통(파일 연계 기반 서비스) 방식이다.

두 서비스는 동일한 주민등록전산시스템을 정보 출처로 하나, 이용 목적과 기술적 연계 방식, 데이터 처리 주기 측면에서 차이를 가진다. 각 서비스의 개요, 절차, 기술적 구조, 개인정보 보호체계 등을 종합적으로 검토하여, 국가기록원 업무 환경에 적합한 연계 체계를 설계하였다.

3.3.5.2. 실시간 조회 서비스(API 기반 연계)

○ 서비스 개요

- 서비스명: [행정안전부] 주민등록정보
- 제공 정보: 성명, 주민등록번호, 주민등록 상태(정상, 말소, 사망 등)
- 이용 방식: 실시간 조회 (API 호출 기반)
- 정보 출처: 행정안전부 주민등록전산시스템
- 본 서비스는 공공기관이 개별 인물의 주민등록 상태를 실시간으로 확인할 수 있도록 설계된 REST API 또는 SOAP 웹 서비스 방식의 행정정보 공동이용 서비스이다.

○ 이용 신청 절차

- 행정정보공동이용센터 접속 및 회원가입
  - 포털에 접속하여 기관 담당자 또는 공무원 계정으로 가입
  - 기관명, 담당자, 부서정보 등 필수 항목 입력

- 서비스 이용 신청서 작성
  - 포털의 서비스 신청 메뉴에서 “[행정안전부] 주민등록정보” 선택
  - 이용 목적: 비공개 기록물 재분류 시 ‘살아있는 개인’ 여부 판단
  - 관련 법령: 「공공기록물 관리에 관한 법률」제35조,「공공기관의 정보공개에 관한 법률」제9조
  - 첨부 자료: 업무 처리 흐름도, 민원서식 등
- 사전협의 및 승인 절차
  - 행정정보공동이용센터가 신청기관의 이용목적 타당성 검토
  - 필요 시 보완서류 제출 및 협의 후 승인 여부 결정

- 기술 연계 및 테스트
  - 승인 후 API 연계 문서(규격서) 제공
  - 기관 시스템 담당자가 API 호출 연계 개발 및 테스트 수행

○ 2.3 기술적 연계 방식

- 연계 프로토콜: REST API 또는 SOAP
- 요청 파라미터: 성명(name), 주민등록번호(id_number)
- 응답 데이터 구조 (JSON 예시):

```
{   "name": "홍길동",
  "id_number": "900101-*******",
  "resident_status_code": "03",
```

```
  "resident_status": "사망"
}
```
주민등록 상태코드: 01: 정상, 02: 말소, 03: 사망

3.3.5.3. 대량유통 방식(파일 연계 기반 연계)

○ 서비스 개요

- 서비스명: [행정안전부] 주민등록정보 대량유통
- 제공 정보: 성명, 주민등록번호, 주민등록 상태(정상, 말소, 사망 등)
- 이용 방식: 대량유통(파일연계)
- 제공 형태: CSV 또는 텍스트 파일
- 정보 출처: 행정안전부 주민등록전산시스템
- 본 서비스는 주민등록정보를 정기적·대량으로 연계받는 방식으로, 대규모 기록물의 일괄 재분류나 사망자 데이터 갱신에 활용된다.
- 파일 기반 연계는 보안파일서버(SFTP) 를 통해 일정 주기로 배치 데이터를 수신하며, 국가기록원의 대규모 기록물 DB와 자동 비교·갱신이 가능하다.

○ 이용 신청 절차

- 센터 접속 및 기관 등록
- 공무원 또는 기관 담당자 계정으로 가입 및 기관 인증
- 서비스 이용 신청서 작성
  - 서비스명: [행정안전부] 주민등록정보
  - 이용 방식: 대량유통(파일연계)
  - 이용 목적: 공개재분류 시 사망 여부 자동 판별
  - 이용 주기: 주 1회
  - 관련 법령:「공공기록물 관리에 관한 법률」제35조「공공기관의 정보공개에 관한 법률」제9조
- 사전협의 및 승인
  - 개인정보보호조치, 보안대책, 목적 타당성 검토 후 승인
  - 보완요청 또는 추가 협의 가능
- 연계 설정 및 운영
  - 승인 후 SFTP 또는 보안파일서버를 통해 연계 설정
  - 정해진 주기에 따라 데이터 파일을 자동 수신 및 갱신

○ 데이터 제공 방식 및 예시

- 제공 형식: CSV 또는 텍스트 파일
- 제공 항목: 성명, 주민등록번호, 상태코드, 상태명
- 데이터 예시:

```
성명,주민등록번호,상태코드,상태명
홍길동,900101-*******,03,사망
김영희,850505-*******,01,정상
이민수,780303-*******,02,말소
주민등록상태코드: 01: 정상 / 02: 말소 / 03: 사망
```

3.3.5.4. 결론 및 향후 추진방향

- 행정정보공동이용센터의 주민등록정보 연계는 공개재분류 업무의 핵심인 “살아있는 개인” 여부 판단을 자동화할 수 있는 효과적인 수단이다.
- 본 연구에서 검토한 실시간 조회 서비스는 개별 문서 단위의 판단에, 대량유통 서비스는 정기적 대규모 갱신에 각각 적합하다.
- 따라서 향후 국가기록원 AI 기반 재분류 시스템은 두 서비스를 혼합형(하이브리드) 구조로 연계하여, 실시간 검증과 주기적 데이터 갱신을 병행하는 체계를 구축하는 것이 효과적이며, 이를 통해 공개재분류 과정의 신속성, 정확성, 투명성을 제고하고, 법령에 근거한 지능형 행정정보 연계 모델로 발전시켜 나갈 수 있을 것이다.

3.4. AI를 활용한 “공개재분류 지원 모델” 연구 개발

3.4.1. 프로세스

○ 공개재분류 판단 프로세스 구조

- RAG 기반 AI 모델이 문서의 성격과 공개 여부를 판단하기 위해 기본적으로 3단계 질의·응답 구조를 갖고 있다. 전체 프로세스는 원문 파일과 메타데이터 입력으로부터 시작하여, LLM의 단계적 추론과 벡터DB의 근거 검색을 통해 최종 판단을 도출하는 방식으로 구성된다.

- 우선, LLM은 원문 데이터를 1차 질의로 “해당 문서는 어떤 성격의 문서인가?”를 판별한다.

- 1차 응답 결과를 바탕으로 메타데이터에서 생산기관, 30년 경과 여부 정보를 가져와, 2차 질문을 통해, 벡터 데이터베이스(Vector DB)에 저장된 과거 분류 사례와 유사 문서를 검색하기 위해 사용된다.

- 즉, “이와 유사한 문서는 과거에 어떻게 분류되었는가?”라는 형태의 질의를 통해, 유사 문서의 분류 결과를 벡터 유사도 기반으로 탐색한다.

- Vector DB는 과거 기록물의 재분류 사례, 비공개 사유 등을 임베딩하여 구축되었으며, 시스템은 검색된 결과(2차 답변)를 인용하여 보다 구체적인 판단 근거를 확보한다.

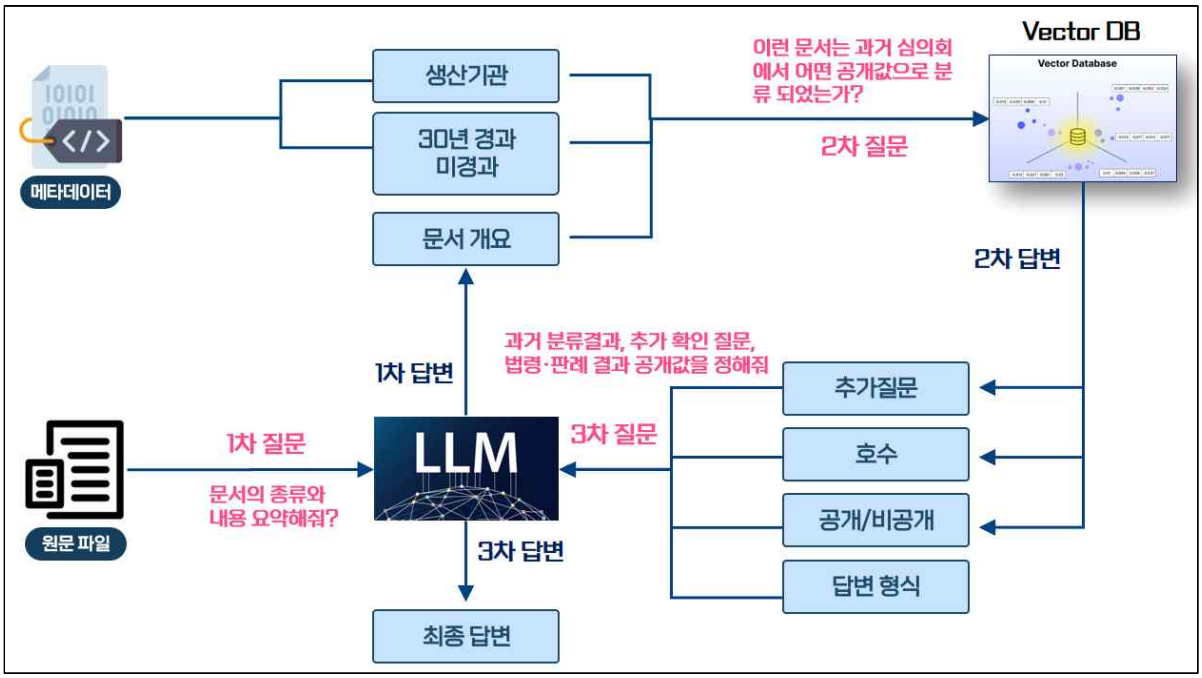
- 이후 LLM은 3차 단계에서 2차 답변으로 확보한 결과를 바탕으로 비공개 사유(호수) 사례, 공개/비공개 결정, 답변 형식을 조합하여 추가질문(3차 질의)을 생성하여 질의하게 된다.

- 즉, 3차 질의·응답 구조를 통해 단순 정보 검색을 넘어, 논리적 근거에 기반한 최종 분류 판단을 완성한다.

- 이 과정은 단일 모델 추론이 아닌, 단계적 질의 확장형 추론(Multi-turn Reasoning) 방식으로 구현되어 있다.

- 결과적으로 본 구조는 LLM의 의미기반 판단능력과 문서생성 능력을 활용하고, RAG의 탐색 기능을 결합한 하이브리드형 공개재분류 판단 아키텍처로, 국가기록원의 문서 특성(법령, 기관별 업무, 시기별 분류기준)을 반영한 시스템이다.

- 아래 그림은 AI 모델 내부의 처리 과정을 도식화 한 것으로 단계적으로 질의하여 최종 답변을 얻는 구조임을 보여준다.

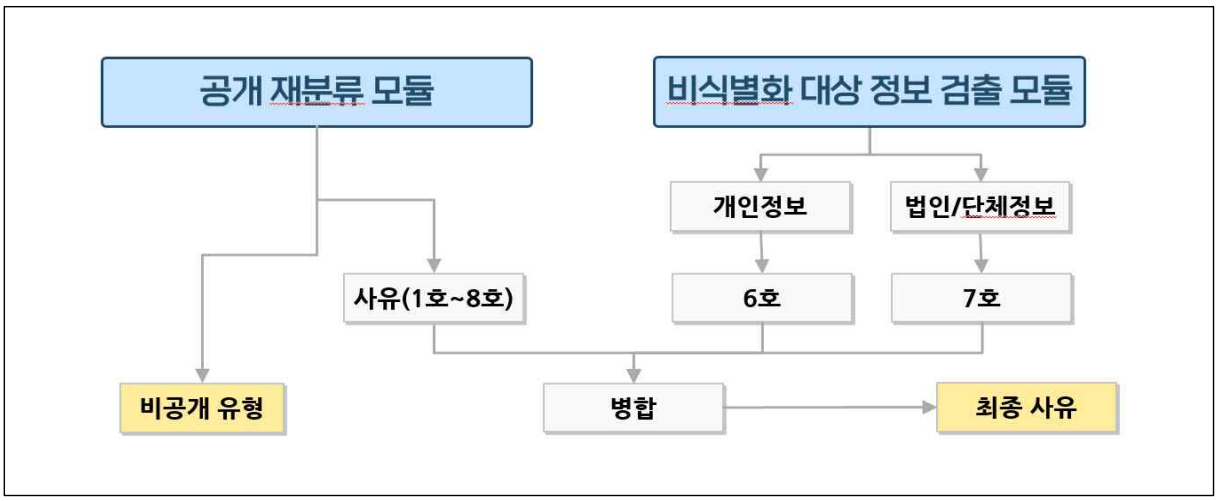


(그림 23 - AI Model Process)

○ 최종 사유 결정 프로세스

- 공개유형과, 이에 대한 사유를 도출하기 위해서는 최종적으로 사유를 먼저 도출하고, 이에 따라서 공개유형을 결정하게 되며, 공개 재분류 모듈과 비식별화 정보 검출 모듈의 관계는 아래와 같다.

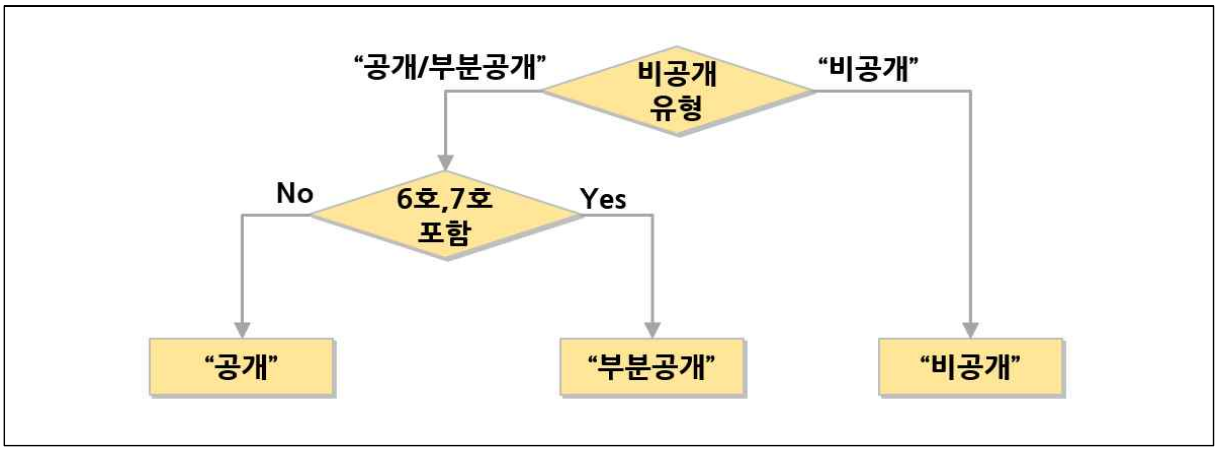
- 공개 재분류 모듈은 비공개 유형과, 이에 해당하는 사유를 도출하며, 비식별화 대상정보검출 모듈은 개인정보와 법인/단체 정보를 검출한다. 개인정보가 포함되면, 6호, 법인/단체정보가 포함되면 7호로 사유를 판단하며, 공개재분류 모듈에서 도출된 사유와 병합하여, 최종 사유를 결정한다.



(그림 24 - Result Derivation Process)

○ 공개 유형 결정 프로세스

- 공개유형은 최종적으로 결정된 사유를 기준으로 6호 7호의 포함여부를 기반으로 아래와 같이 결정한다.
- 공개재분류에서 도출된 비공개 유형이 “비공개” 인 경우 “비공개”로 확정되며, “공개” 또는 “부분공개”인 경우 최종사유에 “6호, 7호”가 포함된 경우 최종 “부분공개”로 판정하며, 그 외의 경우는 “공개”로 판단한다.



(그림 25 - Decision-making process)

○ AI 모델 결과 구조(JSON 반환 형식)

- 최종적으로 AI 모델이 문서 분석 및 비식별화 검출 결과를 추론하고, 최종 반환하는 결과는 JSON 형식으로 반환하며, 그 구조는 아래와 같다.
- 리턴값은 비공개 유형 외에 비식별정보, 문서요약, 벡터DB 인덱스 정보 등을 포함하고 있다.

Key Value	내용	비고
disclosure_type	비공개 유형 결과	공개/비공개/부분공개
re_reason	비공개 사유	제1호 ~ 제8호
re_type, up_type, dw_type	과거 분류 기준 유형	
doc_summary	원문 문서 요약 결과	
re_reason2	요약외 부가적인 설명	
pii_counts	비식별화 대상정보 유형별 건수	"전화번호": 2, "name": 1, "address": 1
pii_value	비식별화 정보 값	"type": "전화번호", "value": "010-1234-5678"
chunk_index	벡터DB 인덱스 값	Z02-0038

(표 23 - Key Value)

```
{
  "re_type": "②-4",
  "up_type": "소원·소청·소송",
  "dw_type": "행정심판",
  "re_reason": ", (제4호*, 제5호**) 제6호",
  "doc_summary": "**문서 유형:** 행정심판 청구서  \n\n**항목 구성:**  \n1. 청구인 정보"c..",
  "disclosure_type": "비공개",
  "re_reason2": "판단 대상 문서는 행정심판 청구서이며, 재결서는 행정심판위원회의 최종 결정.. ",
  "pii_counts": { "전화번호": 2, "name": 1, "address": 1, "contact": 1 },
  "pii_value": { "extractions": [
     { "type": "전화번호", "value": "010-1234-5678", "validated": true, "validators": [ "regex", "rules" ] },
     { "type": "전화번호", "value": "02-9876-5432", "validated": true, "validators": [ "regex", "rules" ]},
     { "type": "name", "value": "홍길동", "validators": "Only_LLM", "confidence": 0.95, "validated": true },
     { "type": "address", "value": "서울특별시 강남구 테헤란로 123", "validators": "Only_LLM" },
     { "type": "contact", "value": "010-1234-5678", "validators": "Only_LLM","validated": true } ] },
  "chunk_index": "Z02-0038"
 }
```


(표 24 - Example of JSON Response)

3.4.2. LLM 모델별 문서 유형 및 요약 결과

❍ 텍스트 추출

- LLM을 활용하여 공개재분류를 하기 위해서는 원문에서 텍스트 추출과정이 필요하다. 이러한 원문파일은 텍스트 추출이 가능한 Searchable PDF를 대상으로 하였다.
- 이미지 기반 PDF는 OCR을 적용하여 추출하면 사용할 수 있으나, 이과정에서 OCR의 인식율 등 고려할 사항들이 있어, 본 연구의 범위를 벗어나므로 제외하였다.
- 추출된 테스트는 2개이상의 첨부문서인 경우 모두 합쳐서 하나의 텍스트로 추출하였으며, 텍스트 추출에 사용된 모듈은 PyMuPDFLoader를 사용하였다.

❍ 문서 유형 및 요약 프롬프트

- 추출된 텍스트는 벡터DB 검색을 위해 LLM을 활용하여 어떤 유형의 문서 요약 판단하여야 한다.
- 프롬프트를 통해 문서를 요약하였고, 문서는 아래와 같이 역할과 문서, 질의를 통해서 수행하였다.

```
273         self.prompt_template = f"""
274     [1. 시스템 역할 설명]
275     당신은 문서가 어떤 유형인지를 판단하는 AI입니다.
276     ---
277     [2. 판단 대상 문서]
278     판단 대상 문서: {document_text}
279     ---
280     [3. 사용자의 질문]
281     질문: {user_question}
282     ---
283     """
```

(그림 26 - Prompt1)

○ 모델별 문서 유형 및 요약 결과

- 아래표는 “행정 심판 청구서”에 대해서 모델별로 문서 유형 및 요약 결과이다.

model	문서 유형 및 요약 결과	소요시간
Gemma3	1. 문서 유형: 행정 심판 청구서 2. 문서 구성 항목: * 청구인 정보 (성명, 주민등록번호, 주소, 연락처, 직업) * 피청구인 정보 (기관명, 직위) * 처분 내용 (운전면허 취소 처분) * 심판 청구의 대상 * 청구 원인 (경위, 위법·부당성, 참고 사항) * 증거 자료 목록 * 고지 사항 3. (정보 제외)	10.379초
llama3.1	**문서의 유형:** * 행정심판 청구서 (행정심판청구서) **문서 구성 항목:** 1. **제1항**: 청구인 정보 * 성명 * 주민등록번호 * 주소 * 연락처 2. **제2항**: 피청구인 정보 * 피청구인 명칭 * 주소 3. **제3항**: 심판 청구의 대상이 되는 처분 * 처분명 * 처분 연월일 * 처분 통지일 * 관련 사건번호 4. **제4항**: 청구 취지 5. **제5항**: 청구 원인 * 처분 경위 및 청구인의 기본 사실관계 * 청구인의 행위 내용 * 혈중알코올농도 * 이 사건 처분의 위법·부당성 * 위법성 + 운전의 정의에 대한 오해 - 청구인의 행위는 도로교통법상 '운전'의 정의에 부합하지 않거나, 적어도 그 위법성이 현저히 낮습니다. - 청구인은 대리운전 호출 후 대리기사를 기다리던 중 불가피하게 차량을 아주 짧은 거리 이동시킨 것에 불과하며, 이는 일반적인 음주운전과는 그 성격이 다릅니다. 차량 이동의 목적은 다른 차량과의 사고를 예방하기 위 ... 6. **제6항**: 증거 자료 * 운전면허증 사본 * 자동차등록증 사본 * 교통사고 사실확인원 사본 * 사고 현장 약도 * 차량 블랙박스 영상 자료 (USB) * 대리운전 호출 기록 * 소득 증명 자료 * 가족관계증명서 (생계 곤란 증명용) * 반성문 * 탄원서 (지인 또는 가족) 7. **제7항**: 고지 받은 사항	21.077초
exaone3.5	**문서 유형**: 행정심판 청구서 **구성 항목**: - 청구인 정보 (이름, 직업) - 피청구인 정보 (직위, 기관) - 처분 내용 (운전면허 취소) - 청구 취지 (처분 취소 요청) - 청구 원인 (위법성, 부당성, 참작 사유)	1.035초

model	문서 유형 및 요약 결과	소요시간
	- 증거 자료 목록 - 고지 사항 (청구 기간, 소송 제기 기간)	
gpt-oss	**문서 유형** - 행정심판 청구서 **구성 항목** 1. 청구인(이름, 직업 등) 2. 피청구인(기관명, 직위) 3. 처분(명칭, 일자, 내용) 4. 청구 취지(취소 요청) 5. 청구 원인(사실관계, 위법·부당성, 결론) 6. 증거 자료(증명서, 영상, 기록 등) 7. 고지 받은 사항(청구·제소 기간) 8. 서명·날짜 (개인정보는 제외)	10.190초

(표 25 - Type Classification Result)

- llama3.1 의 경우 많은 항목을 포함하여 판단을 하고 있으며, 이를 활용하여 좀 더 자세한 내용이 필요한 경우에 적용하면 될 것으로 판단되며, 나머지 3개의 모델은 거의 유사한 유형판단 결과를 보였다. 또한 소요시간에 있어서 가장 빠르게 결과를 낸 모델은 exaone3.5 이었다.

model	결과 요약 특징	소요시간
Gemma3	적정하게 요약하여 유형을 판단함	10.379초
llama3.1	요약을 하지만 보다 많은 항목을 요약하고 있음.	21.077초
exaone3.5	적정하게 요약하여 유형을 판단함	1.035초
gpt-oss	적정하게 요약하여 유형을 판단함	10.190초

(표 26 - LLM 종류별 소요시간)

3.4.3. LLM 모델별 공개값 및 사유 판단 결과

❍ 공개값 및 사유 판단 프롬프트

- 문서유형으로 벡터DB 검색 후 추가 질문을 통해 공개여부를 판단하는 절차로, 해당 프롬프트의 예시이다.

```
392                     self.prompt_template2 = f"""
393     [1. 시스템 역할 설명]
394     [2. 판단 대상 문서]
395     판단 대상 문서: {document_text}
396
397     [3. 사용자 질문]
398     판단 대상 문서의 내용이 아래 질문에 해당되는지 판단하세요.
399     질문: {user_question}
400
401     [요구사항]
402     포함되면 disclosure_type 은 {visibility_type}
403     포함되지 않으면 disclosure_type 은 {visibility_type2}로 답변하세요.
404     이렇게 판단한 사유에 대해서는 re_reason2 라고 하고 내용은 200자로 답변하세요.
405     최종 결과는 JSON 형식으로만 출력하세요.
406     - 마크다운 코드블록( ``` )은 쓰지 마세요.
407     - 설명, 주석, 추가 텍스트는 쓰지 마세요.
408     - JSON 객체만 출력하세요.
409     """
```

(그림 27 - Prompt2)

○ 모델별 공개 값 및 사유 판단 결과

- 아래표는 벡터DB에서 검색한 결과와 추가질의를 기반으로 공개값 및 사유 판단 결과이다.

model	공개 값 및 사유 판단 결과	소요시간

Gemma3	{
	'disclosure_type': '비공개',
	're_reason2': '제공된 문서가 행정심판위원회 재결서가 아닌, 청구인이 피청구인에게 제출한 행정 심판 청구서입니다. 재결서는 행정심판위원회가 최종적으로 내린 결정을 담고 있으며, 청구서는 청구 절차를 위한 문서입니다.'
	}
		5.877초

llama3.1	{
	're_result': True,
	're_reason1': '행정심판위원회 재결서의 형식과 내용이 일치합니다.', 're_reason2': "제목에 '심판 청구의 대상이 되는 처분'이 명시되어 있고, 심판청구의 취지와 근거가 제시된 후, 증거자료가 첨부된 형식을 가지고 있습니다. 또한, 행정소송 제기 가능 여부 및 기간이 명시되어 있어, 행정심판위원회 재결서에 해당하는 것으로 판단됩니다.",
	'disclosure_type': 'partially_disclosed'
	}
		5.227초

exaone3.5	{
	'disclosure_type': '비공개',
	're_reason2': '판단 대상 문서는 행정심판 청구서이며, 재결서는 행정심판위원회의 최종 결정 내용을 담고 있는 문서입니다. 현재 제출된 문서는 청구인의 주장과 증거자료를 포함한 청구 내용을 담고 있어 아직 위원회의 판단이나 결정이 반영되어 있지 않습니다. 따라서 이는 행정심판위원회 재결서에 해당하지 않습니다.'
	}
		1.219초

gpt-oss	{
	'disclosure_type': '비공개',
	're_reason2': '본 문서는 행정심판위원회에 제출되는 청구서이며, 재결서(판결문)와는 달리 결정이 아닌 청구 내용과 증거를 담고 있다. 재결서는 행정심판위원회가 최종 결정한 문서이며, 이 문서는 그 사전 단계인 청구서이므로 재결서에 해당하지 않는다.'
	}
		10.278초

(표 27 - Type Classification Result)

- llama3.1의 경우 질문하지 않은 내용, 임의의 항목으로 답변하였으며, 나머지 3개의 모델은 적정하게 답변을 하였다. 소요시간 측면에서는 가장 빠르게 결과를 낸 모델은 exaone3.5 이었다.

model	결과요약	소요시간
Gemma3	적정하게 공개값 및 사유를 판단함	5.877초
llama3	공개값을 임의로 정해서 판단하는 사례가 많음	5.227초
exaone3.5	적정하게 공개값 및 사유를 판단함	1.219초
gpt-oss	적정하게 공개값 및 사유를 판단함	10.278초

(표 28 - Results Summary by LLM Type)

3.4.3.1. 최종 적용 모델

○ 시사점

- 유형판단 정확도와 응답속도를 비교한 결과, 네 개의 모델 모두 기본적인 판단 성능은 유사한 수준을 보였으나, 세부 분석 결과에서는 각 모델의 특성이 뚜렷하게 구분되었다.
- llama3.1 모델은 다른 모델에 비해 보다 많은 항목을 포함하여 세부적인 요약과 분석을 수행하였으나, 질문과 직접 관련되지 않은 임의 항목을 생성하는 경향이 확인되었다. 이는 모델의 언어 이해 및 확장 추론 능력이 우수하다는 장점이 있지만, 실제 행정문서 유형판단과 같이 명확한 항목 중심의 판단을 요구하는 영역에서는 오히려 과잉응답(over-generation) 으로 작용할 가능성이 있다.
- 반면 Gemma3, exaone3.5, gpt-oss 모델은 질문에 충실하면서도 적정 수준의 요약을 유지하며 일관된 유형판단 결과를 도출하였다.
- 특히 exaone3.5는 모든 실험에서 가장 빠른 응답시간(약 1초 수준)을 기록하였으며, 판단 결과 역시 다른 모델과 비교해 정확도 손실 없이 간결한 출력을 유지하였다.
- 이는 다량의 문서를 신속히 분석해야 하는 환경에서 실용성이 높게 평가될 수 있다.
- 공개재분류 결정모듈에서는 2번의 LLM 질의를 수행하며, 정확성과 처리 효율성 측면을 고려할 때, exaone3.5 모델을 2번의 질의에 모두 적용하는 것으로 하였다.
- 추가적으로 보다 원문이나, 기준서에서 정의한 문서유형의 특징에 따라, 세부적인 문맥 이해나 항목 확장 분석이 필요한 경우 llama3.1 모델을 사용할 수도 있음을 확인하였다.

3.5. “이용자 맞춤형 비식별화 기록물 제공 기능” 연구 및 개발

3.5.1. 구현 프로세스

○ 비식별화 대상정보 검출 절차

- 비공개 대상 여부를 판별하기 위해, 규칙기반의 숫자형 데이터와, 비정형 문자형을 각각 검출한 후 병합하는 하이브리드 방식의 LLM 추론 결합형 프로세스로 구성하였다.

- 전처리 단계 : 입력된 원문 텍스트는 전각→반각, 하이픈 정규화, 공백 제거 등 전처리 과정을 거쳐 텍스트를 정제한다. 전처리 단계의 주요 소스는 아래와 같다.

```
HYPHENS = dict.fromkeys(
    [ord(c) for c in "-------_ - "], ord("-")
)

FULLWIDTH_DIGITS = {ord(f): ord("0") + i for i, f in enumerate("０１２３４５６７８９")}
FULLWIDTH_ASCII = {ord(chr(0xFF01+i)): ord(chr(0x21+i)) for i in range(94)}  # !..~

def to_ascii(text: str) -> str:
    """Map full-width ASCII and digits to half-width."""
    return text.translate(FULLWIDTH_ASCII).translate(FULLWIDTH_DIGITS)

def normalize_hyphens(text: str) -> str:
    """Normalize hyphen-like chars to '-'."""
    return text.translate(HYPHENS)

def collapse_spaces(text: str) -> str:
    # Collapse multiple spaces but preserve newlines.
    return re.sub(r"[ \t\f\v]+", " ", text)

```

(그림 28 - De-identification Preprocessing Steps)

- 정규식 기반 후보 추출 : 주민번호, 계좌번호, 자동차 번호 등 정규식으로 추출가능한 문서 내 주요 개체를 후보 정보로 추출한다.

종류	정규식
주민등록번호	(r″(?:(?<!\d)|(?<=\n))″ r″(?P<yy>\d{2})(?P<mm>0[1-9]|1[0-2])″ r″(?P<dd>0[1-9]|[12]\d|3[01])″ r″(?:[ ]{0,1}|-)?\n?″ r″\n?″ r″(?P<s>[1-8])\d{6}″ r″(?:(?!\d)|(?=\n))″)
휴대폰번호 (일반전화 제외)	(r″\b(?:01[016789]|0\d{1,2})[-.\s]?\d{3,4}[-.\s]?\d{4}\b″, FLAGS)

종류	정규식
email	(r"\b[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[A-Za-z]{2,}\b", FLAGS)
신용카드 번호	(r"(?<!\d)(?:\d[-]?){14,19}(?!\d)", FLAGS)
사업자등록번호	(r"\b\d{3}[-.\s]?\d{2}[-.\s]?\d{5}\b")
법인등록번호	(r"(?<!\d)\d{6}[-.\s]?\d{7}(?!\d)")
차량번호	(r"\b\d{2,3}[가-힣]\s?\d{4}\b")
운전면허	(r"(?<!\d)\d{2}[-\s]?\d{2}[-\s]?\d{6}[-\s]?\d{2}(?!\d)|(?<!\d)\d{12}(?!\d)")
여권	(r"\b(?:[A-Z]{1}\d{8}|[A-Z]{2}\d{7})\b")
계좌번호	(r"(?<!\d)(?!01[016789]\d{7,8})(?:\d[-]*){10,14}(?!\d)"

(표 29 - Regex-based Extraction Candidates)

- 규칙기반 검증 : 주민등록번호, 신용카드번호, 사업자등록번호는 체크섬 알고리즘 등을 통하여 실제 번호인지 검증하고, 통과되면 바로 비식별화 대상정보로 확정한다. 아래는 주민등록번호 체크 알고리즘의 사례이다.

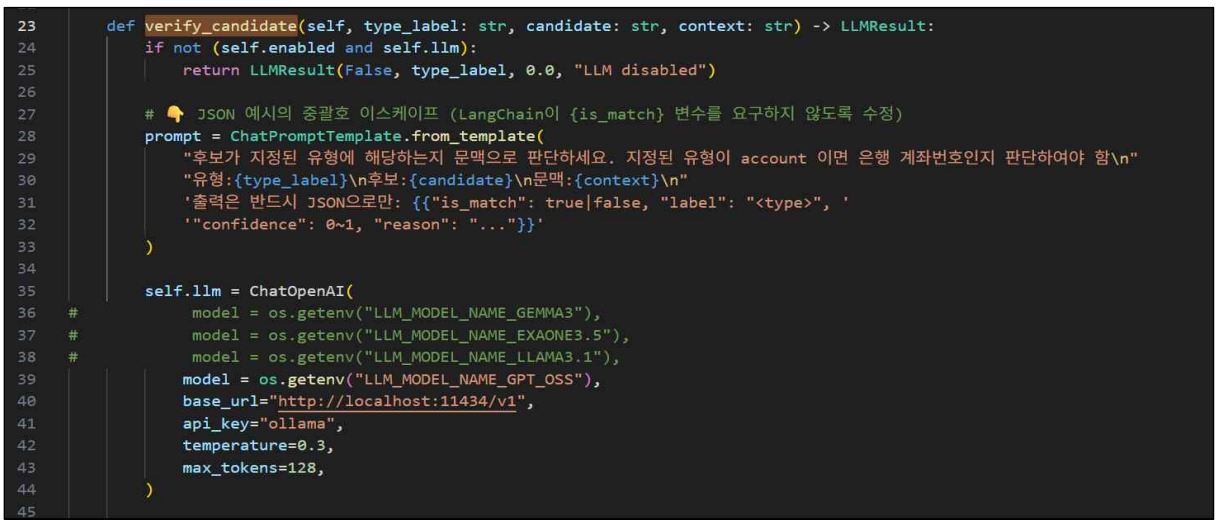
```
def _rrn_birth_ok(yy: int, mm: int, dd: int, s: int) -> bool:
    # Determine century from the 7th digit
    if s in (1,2,5,6):
        cen = 1900
    elif s in (3,4,7,8):
        cen = 2000
    else:
        return False
    try:
        datetime(cen + yy, mm, dd)
        return True
    except ValueError:
        return False

def rrn_ok(num: str) -> bool:
    """Resident/Foreigner/Corporate-like 13-digit checksum (RRN/FRN family)."""
    n = digits_only(num)
    if not re.fullmatch(r"\d{13}", n): return False
    ds = [int(c) for c in n]
    if not _rrn_birth_ok(int(n[:2]), int(n[2:4]), int(n[4:6]), ds[6]):
        return False
    weights = [2,3,4,5,6,7,8,9,2,3,4,5]
    s = sum(ds[i] * weights[i] for i in range(12))
    check = (11 - (s % 11)) % 10
    return check == ds[12]
```

- 문맥 키워드 검증 : 규칙 검증 단계에서 해당되지 않는 경우에는, 주변 60자의 문맥 내에서 해당 유형의 키워드가 있는지 확인하여 해당되는 경우 비식별화 대상정보로 확정한다. 아래는 주요 키워드 사례이다.

```
KEYWORDS = {
    "rrn_like": ["주민등록", "주민번호", "주민등록번호", "RRN", "외국인등록", "외국인등록번호", "외국인"],
    "brn": ["사업자등록", "사업자등록증", "사업자번호", "사업자등록번호", "사업자"],
    "corp_13": ["법인등록", "법인등록번호", "법인등기", "법인번호", "법인코드"],
    "vehicle": ["차량", "번호판", "자동차", "차량번호", "등록"],
    "driver": ["운전면허", "면허번호", "도로교통"],
    "passport": ["여권", "여권번호", "passport"],
    "account": ["계좌", "계좌번호", "입금", "출금", "송금", "이체", "계좌이체", "은행", "통장"],
    "internet_id": ["ID", "아이디", "계정", "username"],
    "case_no": ["사건번호", "사건부번호", "송치", "판결", "검사처분", "수사", "사건명"],
    "military_id": ["군번", "군인", "ROKA", "ROKAF", "ROKN", "복무"],
    "prisoner_id": ["죄수번호", "수형번호", "수용번호", "교정", "교도소"],
    "health_ins": ["건강보험", "보험증", "자격득실", "의료보험"],
}
```

- LLM 기반 문맥검증 : 규칙·키워드 검증을 통과하지 못한 복합적 사례는 LLM(Large Language Model) 로 전달되어, 문맥적 의미를 기반으로 해당 유형인지 판단한다. 문맥은 검출된 숫자의 앞뒤 60자를 기준으로 LLM이 판단한다. 아래는 LLM이 문맥을 기준으로 판단하는 주요 함수와 프롬프트이다.



```
23      def verify_candidate(self, type_label: str, candidate: str, context: str) -> LLMResult:
24          if not (self.enabled and self.llm):
25              return LLMResult(False, type_label, 0.0, "LLM disabled")
26
27          # 👇 JSON 예시의 중괄호 이스케이프 (LangChain이 {is_match} 변수를 요구하지 않도록 수정)
28          prompt = ChatPromptTemplate.from_template(
29              "후보가 지정된 유형에 해당하는지 문맥으로 판단하세요. 지정된 유형이 account 이면 은행 계좌번호인지 판단하여야 함\n"
30              "유형:{type_label}\n후보:{candidate}\n문맥:{context}\n"
31              '출력은 반드시 JSON으로만: {{"is_match": true|false, "label": "<type>", '
32              '"confidence": 0~1, "reason": "..."}}'
33          )
34
35          self.llm = ChatOpenAI(
36      #           model = os.getenv("LLM_MODEL_NAME_GEMMA3"),
37      #           model = os.getenv("LLM_MODEL_NAME_EXAONE3.5"),
38      #           model = os.getenv("LLM_MODEL_NAME_LLAMA3.1"),
39              model = os.getenv("LLM_MODEL_NAME_GPT_OSS"),
40              base_url="http://localhost:11434/v1",
41              api_key="ollama",
42              temperature=0.3,
43              max_tokens=128,
44          )
45
```

(그림 29 - LLM-based Context Verification)

- LLM 기반 검출(문자형) : 이름, 주소, 회사명 등 문자형은 별도로 검출하며, LLM은 문장 내 개체 간 관계, 용어의 의미, 행정문서 표현 패턴 등을 종합적으로 분석하여 비식별화 대상 정보를 검출한다. 이 과정에서 LLM에 대한 프롬프트는 개인 및 법인/단체의 이익을 침해하는 정보를 찾도록 설계하고 공공기관, 수행하는 수사관, 판상, 공무원 이름, 공무원의 메일인 *.go.kr, *.korea.kr 도메인은 제외하도록 처리하였다. 아래는 LLM이 문자형 비식별화 정보를 검출하는 프

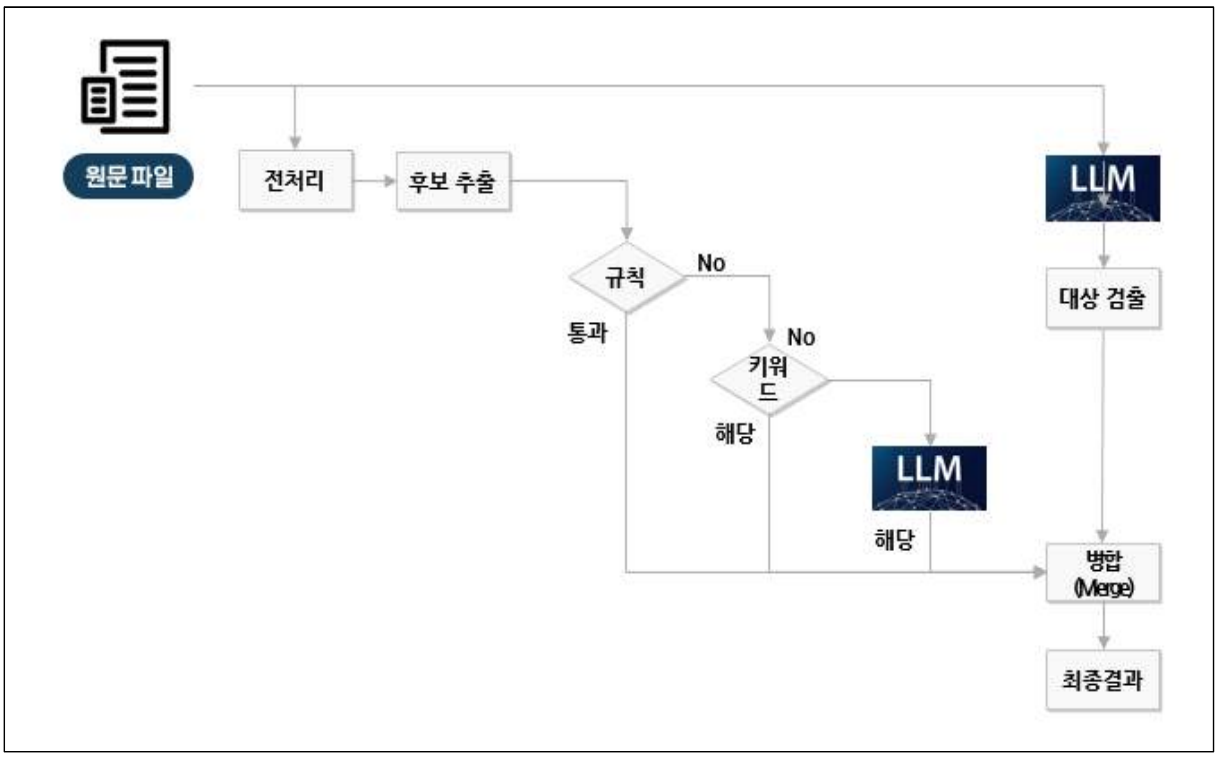
롬프트이다.

```
463                 self.prompt_template3 = f"""
464     [역할]
465     너는 문서를 읽고, 비공개대상정보(이름, 주소(개인), 회사명, 법인명, 단체명, 주소(회사/단체/법인))를 찾아내는 AI이다.
466
467     [입력 문서]
468     {document_text}
469     [판단 기준]
470     아래에 해당하면 비공개대상정보로 간주하여 추출한다:
471     - 문서의 문맥상 개인/단체의 이익이나 사생활에 직접적인 영향을 줄 수 있는 정보로써
472     - 개인의 이름, 상세 주소(시/군/구 단위만 있는 경우 제외하고, 번지수·도로명·건물번호 등 구체적인 주소일 때만 추출)
473     - 민간단체·사기업 등 사인(私人) 단체의 이름과 상세 주소 (상세 주소 기준 동일)
474
475     아래에 해당하면 추출 대상에서 제외한다:
476     - 판사, 검사, 수사관, 경찰서장, 공무원 등 사건 처리, 업무처리 공무원의 이름·직위
```

(그림 30 - LLM-based Detection)

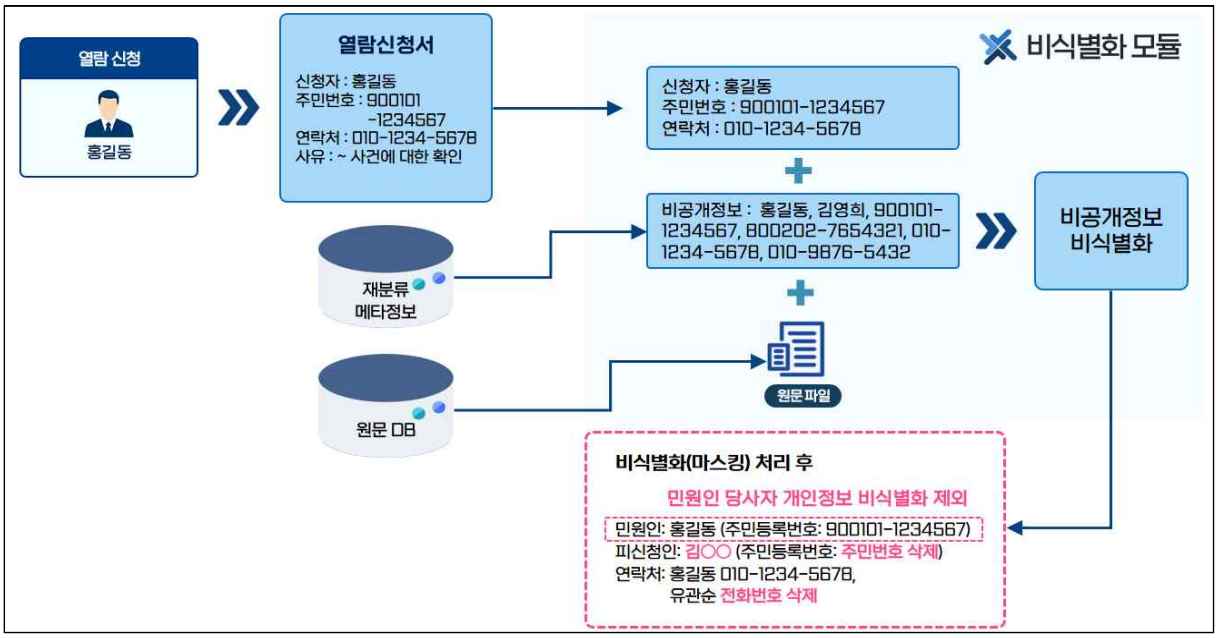
- 결과 병합 : 정규식기반으로 다단계 검증을 거쳐 검출된 결과와, LLM 기반으로 검출된 문자형은 병합하여, 최종 결과를 도출한다.

- 아래 그림은 비식별화 대상정보를 검출하는 주요 프로세스이다.



(그림 31 - Detection Process)

- 이용자 맞춤형 비식별화 기록물 제공은 AI 모듈이 검출한 비식별화 정보를 기록물의 메타정보에 저장한후 이용자가 열람신청을 한 경우를 가정하고 열람자의 개인정보는 제외하고 나머지 정보만을 비식별화(마스킹)하여 PDF 파일로 제공하는 형태로 그 절차는 아래와 같다.



(그림 32 - Procedure)

3.5.2. 숫자형 비식별화 대상 정보 검출 결과

○ 숫자형 비식별화 대상 정보 검출을 위한 전처리 및 정규식 추출, 검증

- 숫자형 비식별화 대상 정보는 전처리, 정규식으로 후보 추출, 규칙검증, 키워드 검증(앞뒤 60자에서 해당키워드 존재여부 확인)을 통해 확정처리되고, 여기서 확정처리되지 않은 정보는 프롬프트를 통해 앞뒤 60자내에서 LLM이 판단한다.

○ 비식별화 대상 정보 문맥 판단 프롬프트 및 데이터

- 문서유형으로 벡터DB 검색 후 추가 질문을 통해 공개여부를 판단하는 절차로, 해당 프롬프트의 예시이다.

```
28        prompt = ChatPromptTemplate.from_template(
29            "후보가 지정된 유형에 해당하는지 문맥으로 판단하세요. 지정된 유형이 account 이면 은행 계좌번호인지 판단하여야 함\n"
30            "유형:{type_label}\n후보:{candidate}\n문맥:{context}\n"
31            '출력은 반드시 JSON으로만: {{"is_match": true|false, "label": "<type>", '
32            '"confidence": 0~1, "reason": "..."}}'
33        )
34
```

(그림 33 - Prompt for Contextual Assessment of De-identified Information)

- 정규식으로 추출, 규칙검증, 키워드 검증을 거쳐 사업자등록번호, 계좌번호의 후보로 도출된 4개의 비식별화 대상정보(모두 해당 유형이 아니지만 추출된 경우임)를 차례대로 사업자등록번호, 계좌번호인지 질문하였다.

```
('corp_13', '800101-1465578', False, True),
('account', '800101-1465578', False, True),
('account', '010-1234-5678 ', False, True),
('account', '02-9876-5432 ', False, True)
```

○ 모델별 검출 결과

- 아래표는 숫자형 비식별화 대상정보 4개에 대해서 모델별 검출결과와 소요시간이다.

model	숫자형 비식별화 대상정보 검출 결과	소요시간
Gemma3	{"is_match": true, "label": "account", "confidence": 0.95, "reason": "문맥에서 연락처로 제시된 '010-1234-5678'은 은행 계좌번호와 유사한 형식이며, 연락처 정보의 일부로 제공되어 account 유형에 해당한다고 판단함." } * 4개 모두 true 로 판단함(오류)	9.051초
llama3.1	answer=JSON 형식의 결과를 출력합니다. { "is_match": true, "label": "bank_account_number", "confidence": 1.0, "reason": "주민등록번호가 account 유형에 해당하며, 일반적인 은행 계좌번호와 유사한 형식을 보입니다." }	10.126초
exaone3.5	{"is_match": true, "label": "account", "confidence": 0.95, "reason": "주어진 후보 문자열이 주민등록번호 형식과 일치하며, 일반적으로 주민등록번호는 개인의 신원 확인을 위한 정보로 계좌번호와 유사한 민감 정보로 간주될 수 있어 계정 관련 유형으로 판단됩니다."} **참고**: 주민등록번호가 직접 은행 계좌번호는 아니지만, 민감한 개인 식별 정보로서 계정 유형에 해당할 수 있다는 판단 하에 높은 확신도로 매칭하였습니다. 만약 더욱 엄격한 기준이 필요하다면 추가적인 확인이 필요할 수 있습니다.	6.843초
gpt-oss	"is_match": false 4개 모두 false "reason": "The candidate \"800101-1465578\" follows the Korean resident registration number format (6 digits-7 digits) rather than the 10 digit corporate number format expected for type corp_13."}	16.441초

(표 30 - Type Classification Result)

- gpt-oss 만 정확하게 검출하였고, 시간은 16초가 걸렸다.

model	결과요약	소요시간
Gemma3	4개 모두 틀려서 적정하지 않음	9.051초
llama3	1개가 틀렸으며, 과잉응답(over-generation)이 발생함	10.126초
exaone3.5	1개가 틀렸으며, 과잉응답(over-generation)이 발생함	6.843초
gpt-oss	4개 모두 정확하게 답변함	16.441초

(표 31 - Results Summary by LLM Type)

○ 시사점

- 적정하게 답변을 한 모델은 gpt-oss 이며, 소요시간 또한 가장 길었다. 비식별화 대상정보 1건 기준으로 약 4초가 걸렸으며, 검출할 대상 건수에 따라서 많은 시간이 소요될수 있음을 확인하였다.

3.5.3. 문자형 비식별화 대상정보 LLM 모델별 검출결과

○ 문자형(이름, 주소, 법인/단체명, 법인/단체 주소)를 검출

- 6호, 7호에 해당하는 문자형 정보들은 정규식으로 판단하기 곤란한 부분이 있어, 문서 전체를 읽고 해당되는 항목을 모두 추출하도록 하였다.

○ 비식별화 대상정보 판단 프롬프트

- 문서유형으로 벡터DB 검색 후 추가 질문을 통해 공개여부를 판단하는 절차로, 해당 프롬프트의 예시이다.

```
463             self.prompt_template3 = f"""
464 [역할]
465 너는 문서를 읽고, 비공개대상정보(이름, 주소(개인), 회사명, 법인명, 단체명, 주소(회사/단체/법인))를 찾아내는 AI이다.
466
467 [입력 문서]
468 {document_text}
469 [판단 기준]
470 아래에 해당하면 비공개대상정보로 간주하여 추출한다:
471 - 문서의 문맥상 개인/단체의 이익이나 사생활에 직접적인 영향을 줄 수 있는 정보로써
472 - 개인의 이름, 상세 주소(시/군/구 단위만 있는 경우 제외하고, 번지수·도로명·건물번호 등 구체적인 주소일 때만 추출)
473 - 민간단체·사기업 등 사인(私人) 단체의 이름과 상세 주소 (상세 주소 기준 동일)
474
475 아래에 해당하면 추출 대상에서 제외한다:
476 - 판사, 검사, 수사관, 경찰서장, 공무원 등 사건 처리, 업무처리 공무원의 이름·직위
477 - 정부·공공기관 명칭과 주소, 관련부서 명칭소 (예: 법원, 검찰청, 시청, 경찰서, 국립과학수사연구소 등 → 추출 제외)
```

(그림 34 - Prompt for Determining Data Subject to De-identification)

○ 모델별 검출 판단 결과

- 아래표는 같은 내용(이름 2개, 주소 1개)에 대해서 문자형 비식별화 대상정보 검출을 요청한 결과이다.

model	문자형 비식별화 대상정보 검출 결과	소요시간

Gemma3	{
	"type": "organization",
	"value": "서울지방경찰청",
	"validators": "Only_LLM",
	"confidence": 0.9,
	"validated": true
	},
	{
	"type": "address",
	"value": "(피청구인 기관 주소 기재)",
	"validators": "Only_LLM",
	"confidence": 0.8,
	"validated": true
	}
	* "서울지방경찰청", "(피청구인 기관 주소 기재)"를 검출함(오류)
		6.780초

llama3.1	{
	"type": "name",
	"value": "김사랑 (洪吉童)",
	"validators": "Only_LLM",
	"confidence": 0.9,
	"validated": true
	},
	{
	"type": "address",
	"value": "서울특별시 강남구 테헤란로 123 (역삼동, 000빌딩 10층)",
	"validators": "Only_LLM",
	"confidence": 0.9,
	"validated": true
	}
	* 이름 주소 모두 검출하였으나 “김사랑” 과 “洪吉童”은 하나의 객체로 인식함
		8.406초

exaone3.5	{
	"type": "contact",
	"value": "010-1234-5678 (휴대전화) / 02-9876-5432 (자택)",
	"validators": "Only_LLM",
	"confidence": 0.85,
	"validated": true
	},
	{
	"type": "organization",
	"value": "회사원",
	"validators": "Only_LLM",
	"confidence": 0.6,
	"validated": false
	}
	* "회사원"을 조직(organization)으로 검출함
		4.747초

gpt-oss	{ "type": "name", "value": "김 사 랑", "validators": "Only_LLM", "confidence": 0.9, "validated": true }, { "type": "name", "value": "洪吉童", "validators": "Only_LLM", "confidence": 0.9,	12.090초

model	문자형 비식별화 대상정보 검출 결과	소요시간

	"validated": true
	},
	{
	"type": "address",
	"value": "서울특별시 강남구 테헤란로 123 (역삼동, 000빌딩 10층)",
	"validators": "Only_LLM",
	"confidence": 0.9,
	"validated": true
	}
	* 이름 2개, 주소 모두 검출함	

(표 32 - Processing Time to Detect Textual Data Requiring De-identification)

- gpt-oss 만 정확하게 검출하였고, 시간은 12초가 걸렸다.

model	결과요약	소요시간
Gemma3	불필요한 정보를 단체와 주소로 검출함	6.780초
llama3	한글, 한자의 혼돈이 있음	8.406초
exaone3.5	“회사원”을 조직으로 검출함	4.747초
gpt-oss	모두 정확하게 검출함	12.090초

(표 33 - Results Summary by LLM Type)

❍ 시사점

- 가장 적정한 답변을 제공한 모델은 gpt-oss 였으며, 해당 모델이 가장 긴 처리시간을 소요한 것으로 나타났다.
- 다른 모델들은 검출용 LLM으로 활용하기에는 한계가 확인되었으나, PoC에서 사용한 버전 이외의 최신·향상 모델을 적용할 경우 상이한 성능 결과가 도출될 가능성은 존재한다.

3.5.4. 이용자 맞춤형 비식별화 대상정보 마스킹

○ 비식별화 대상정보 마스킹 모듈

- 일반적으로 마스킹 모듈은 텍스트 추출 또는 OCR 과정을 통해 문서 내 텍스트를 인식한 뒤, 정규식 기반의 개인정보 검출을 수행하고 이를 마스킹·편집할 수 있는 형태로 구현된다.
- 본 연구의 PoC에서는 이용자 맞춤형 자동 마스킹 처리가 가능하도록 하기위해 마스킹 모듈을 자체 개발하였다.

○ 마스킹 모듈 처리 프로세스

- 마스킹 모듈은 3가지 인자를 받아 마스킹을 수행한다.
- 첫 번째 인자(마스킹 대상 파일) : 마스킹 대상파일은 복사하여 마스킹 파일을 만든다.
- 두 번째 인자(비식별화 대상정보) : 공개재분류 AI 모델에서 검출한 비식별화 대상정보의 유형과 값을 인자로 받아서 마스킹할 목록을 만든다.
- 세 번째 인자(마스킹 제외 정보) : 이용자 맞춤형으로 비식별화 처리를 하기위해, 열람신청자(이용자)의 정보를 CAMS 화면에서 입력받고 그정보는 마스킹 대상에서 제외한다.
- 마스킹 모듈은 PDF 파일 내에서 마스킹 대상정보와 같은 패턴의 문장/단어를 “홍OO”과 같은 방식으로 마스킹 처리하고 마스킹 파일을 생성한다.

○ 핵심 라이브러리

- PyMuPDF (fitz) : 용도: PDF 문서 조작 및 텍스트 처리

주요 활요 기능

fitz.open(): PDF 문서 열기 및 읽기

page.search_for(): 텍스트 패턴 검색 및 좌표 추출

page.get_text("rawdict"): 원시 텍스트 데이터 및 스타일 정보 추출

page.add_redact_annot(): 레다크션(삭제) 영역 주석 생성

page.apply_redactions(): 원본 텍스트 완전 삭제

page.insert_font(): 폰트 리소스 삽입 및 등록

page.insert_textbox(): 텍스트 오버레이 삽입

doc.get_page_fonts(): 페이지 폰트 메타데이터 조회

doc.saveIncr(): 증분 저장 (Incremental Save)

○ 필터링 관련 소스 사례

```
def filter_names(names: List[Dict[str, str]], exclude_names_input: Optional[str]) -> List[Dict[str, str]]:
    """제외할 이름을 필터링 (쉼표로 구분된 value만 받음)"""
    if not exclude_names_input:
        return names

    # 쉼표로 구분된 문자열에서 제외할 value 목록 추출
    exclude_values = set()
    for value in exclude_names_input.split(','):
        value = value.strip()
        if value:
            exclude_values.add(value)

    if not exclude_values:
        return names

    print(f"[INFO] 제외할 이름 목록 ({len(exclude_values)}개):")
    for i, value in enumerate(exclude_values, 1):
        print(f"[INFO]   {i}. {value}")

    # 제외 목록에 없는 항목만 필터링
    filtered = [item for item in names if item['value'] not in exclude_values]

    excluded_count = len(names) - len(filtered)
    if excluded_count > 0:
        print(f"[INFO] {excluded_count}개 항목이 제외되었습니다.")
        excluded = [item for item in names if item['value'] in exclude_values]
        for item in excluded:
            print(f"[INFO]   제외됨: [{item['type']}] {item['value']}")
    return filtered
```

3.6. PoC 개발 및 시범운영

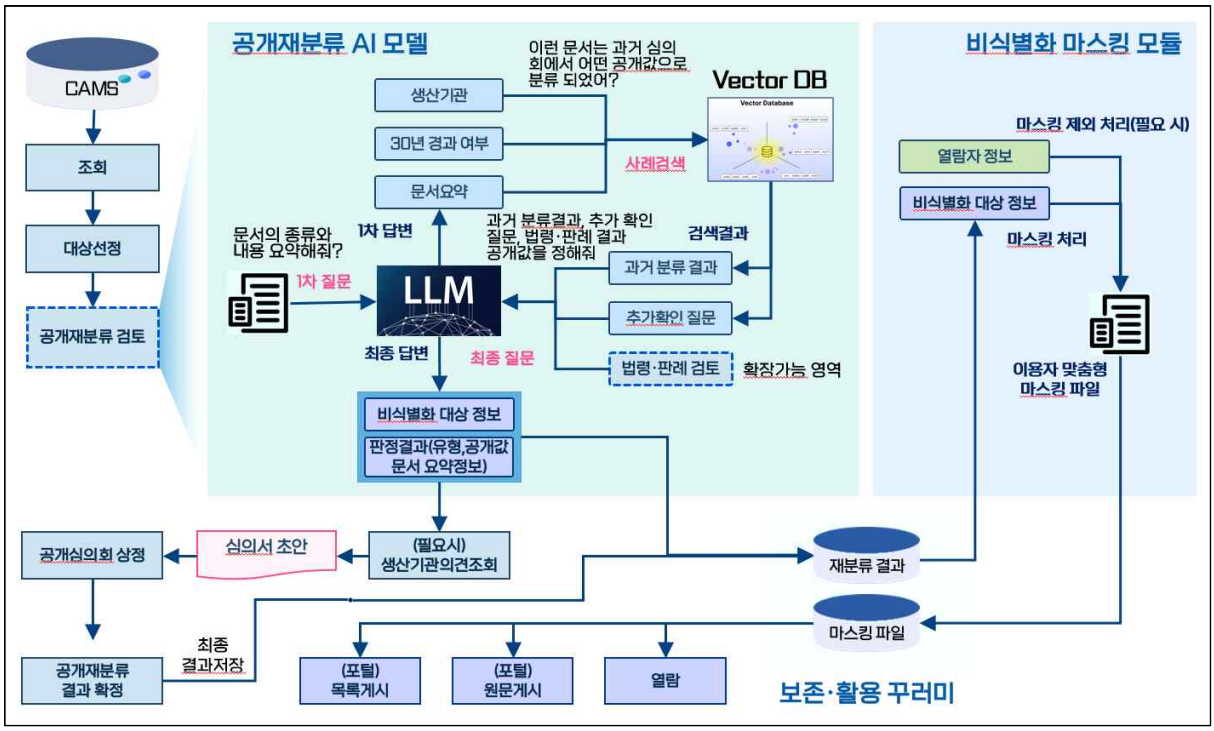
3.6.1. 시스템 아키텍처

인공지능 기반 공개재분류 지원 시스템의 시스템 아키텍처 및 PoC 환경 구성을 기술한다.

PoC(Proof of Concept) 단계에서는 모델의 적합성, 처리 효율, 기술 통합성 등을 검증하기 위하여 다양한 AI 모델과 백엔드 구성요소를 조합하여 환경을 구축하였다.

특히 RAG(Retrieval Augmented Generation) 구조를 중심으로, 대규모 언어모델(LLM), 임베딩 모델, 벡터 데이터베이스, 프레임워크, 그리고 자체 개발된 모듈 간의 상호작용을 통합적으로 검증하였다.

○ AI 모델 도입으로 인한 업무 처리 프로세스(To-Be)



○ S/W, H/W 구성(PoC)

- PoC 환경 구성은 아래와 같은 환경으로 구성하였다.

항목	내용	비고
AI 모델	GEMMA3(Gemma3:1b), EXAONE3.5(exaone3.5:32b), LLaMA3.1(llama3.1:8b), GPT_OSS(gpt-oss:20b)	질의 유형에 따라서 성능을 고려하여 LLM 선택하여 적용
임베딩 모델	OllamaEmbeddings (bge-m3)	벡터 DB 구성 시 텍스트 벡터화

벡터DB	Qdrant	문서 임베딩 결과 저장 및 검색 기능
프레임워크	LangChain	LLM, 벡터DB, 프롬프트 체인을 연결하는 핵심 프레임워크
문서 로딩 처리	DocumentManager load_document.py	자체 개발 소스 다양한 형식의 문서 로딩 관리
파일 변경 감지	FileTracker	자체 개발 소스 벡터 DB 업데이트 자동화
질의 처리	chat_service.py에서 RAG 검색 후 LLM 호출 구조	자체 개발 소스 RAG 질의 및 응답 처리 로직
지원 문서 형식	pdf, json	첨부파일 및 벡터 DB용 구축용
H/W	RTX 4070 X 2장	예산대비 성능을 고려한 PoC용 환경 구성

(표 34 - PoC Environment Configuration Details)

○ 시스템 구성 개요

- PoC 환경은 AI 모델 - 임베딩 모델 - 벡터DB - 프레임워크 - 자체 모듈로 이어지는 다층 구조로 설계되었다.
- 먼저, 원문 문서가 DocumentManager를 통해 로딩되면 OllamaEmbeddings(bge-m3) 모델을 이용해 텍스트가 벡터화된다.
- 생성된 벡터는 Qdrant 데이터베이스에 저장되어, 추후 질의 시 유사도 검색을 통해 관련 정보를 빠르게 검색할 수 있도록 구성된다.
- 사용자가 질의를 입력하면 chat_service.py가 이를 받아 LangChain 프레임워크를 통해 RAG 검색 프로세스를 수행하고 검색된 문서를 기반으로 선택된 AI 모델(GEMMA3, EXAONE, GPT_OSS 등)이 최종 답변을 생성한다.
- 또한, FileTracker 모듈은 문서 변경이나 신규 추가 시 자동으로 벡터DB를 갱신하여 최신 데이터를 유지한다.

○ LLM 선택 제어 기반 어댑티브 구조

- 본 시스템의 질의 처리 구조는 질의 유형과 성능 요구도에 따라 LLM을 선택적으로 활용할 수 있도록 설계된 조건 제어형 어댑티브 인퍼런스(Adaptive Inference) 구조를 적용하였다.
- 이는 자동 라우팅 기반의 모델 선택 엔진이 아닌, 사전 정의된 조건(Predefined Conditions) 에 따라 개발자가 하드코딩 방식으로 모델을 지정·조절하는 구조이다.
- 즉, 질의의 성격(검색형, 요약형, 판단형 등)과 응답 품질 요구 수준을 고려하여 각 모델의 특성(GPU 점유율, 처리속도, 파라미터 규모, 정확도 등)에 따라 적정 LLM을 수동 매핑(Manual Mapping) 하도록 구성하였다.

- 예를 들어, 단순 검색형 질의에는 경량 모델(GEMMA3 또는 LLAMA3.1-8b)을 활용하여 처리속도를 최적화하고 논리적 추론이 필요한 질의에는 대규모 모델(EXAONE3.5-32b 또는 GPT_OSS-20b)을 수동 지정하여 고정밀 응답을 생성하도록 하였다.
- 이러한 구조는 LangChain 기반 파이프라인 내 조건 분기 로직(Conditional Execution Logic) 으로 구현되었으며 운영자가 질의 유형별 모델 선택 정책을 조정함으로써, 처리 효율성과 응답 품질 간의 균형을 유연하게 제어할 수 있도록 설계되었다.
- 결과적으로, 본 시스템은 완전한 자동화보다는 운영자 제어형(Operator-driven) 모델 선택 구조를 통해 모델별 자원 사용량과 응답 품질을 실험적으로 검증할 수 있는 유연한 연구환경을 구현하였다.
- 아래는 개발과정에서 최적의 조건을 찾아 최종 모델을 결정하도록 구성된 소스 사례이다.

```
                prompt3 = ChatPromptTemplate.from_template(self.prompt_template3)
                logger.info(f"self.prompt_template3 : {self.prompt_template3}")

                llm = ChatOpenAI(
#                       model = os.getenv("LLM_MODEL_NAME_GEMMA3"),
#                       model = os.getenv("LLM_MODEL_NAME_EXAONE3.5"),
#                       model = os.getenv("LLM_MODEL_NAME_LLAMA3.1"),
                        model = os.getenv("LLM_MODEL_NAME_GPT_OSS"),
                        base_url="http://localhost:11434/v1",
                        api_key="ollama",
                        temperature=0.2,
                        max_tokens=1024,
                )
```

○ 기술적 의의

- 본 시스템 구성은 RAG 구조를 중심으로 한 통합형 AI 처리 아키텍처를 실험적으로 구현한 것으로 다양한 LLM의 성능을 비교·검증할 수 있고, 질의 유형/성능에 따라서 LLM을 선택하여 활용할 수 있도록 하였다.
- 자체 개발 모듈을 통해 문서 로딩, 데이터 관리, 질의 응답의 효율성을 향상시켰으며 GPU 기반 연산환경을 통해 대규모 데이터 처리 및 실시간 검색 성능을 확보하였다.
- 이를 통해 PoC 단계에서 기술적 타당성을 입증함과 동시에, 향후 국가기록원 공개재분류 업무에 최적화된 AI 시스템의 확장 및 고도화에 필요한 기반을 마련하였다.

3.6.2. 사용자 인터페이스

RAG 기반 AI 모델과 외부 시스템 간의 사용자 인터페이스(UI/UX) 구조 및 연계 방식을 기술한다. 본 연구에서 개발된 시스템은 국가기록원의 공개재분류 업무에 활용될 수 있도록, AI 모델이 생성한 판단 결과를 다양한 환경에서 활용할 수 있는 다중 인터페이스 구조(Multi-Interface Architecture)로 설계되

었다. 이를 통해 사용자는 브라우저 기반의 대화형 환경, CAMS 시스템 연계, 대량 파일 기반의 자동 배치 처리 등 다양한 업무 형태에 따라 AI모델의 적용이 가능함을 확인하고 개발과정에 빠른 결과가 가능하도록 하였으며, PoC는 주로 CAMS와 유사한 UI에서 사용자가 테스트 할 수 있도록 구성하였다.

○ PoC 화면 인터페이스

- AI 재분류 화면은 기존 CAMS와 동일하게 보이도록 구성하였으며, 조회된 각 건에 대해서 재분류할 항목을 체크하고 “AI 재분류” 버튼을 클릭하면 재분류 결과가 그리드에 표시되도록 하였다.

(그림 35 - PoC Screen Interface)

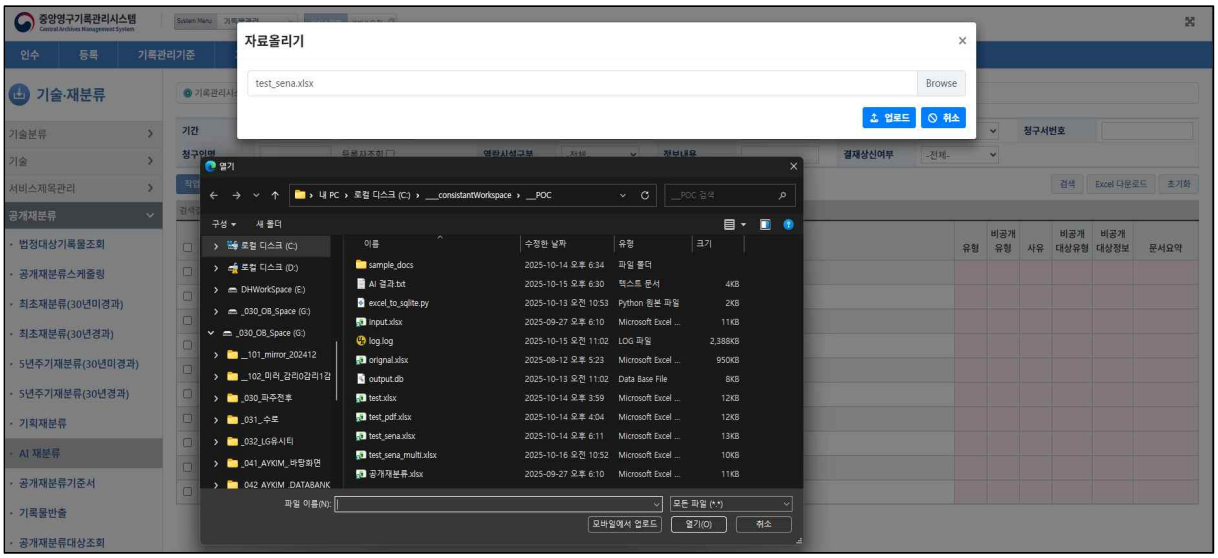
○ PoC 초기화면

- 그리드의 목록은 비어 있으며 작업그룹추가, 검토서 작성, AI재분류, 검색, Excel다운로드, 초기화 버튼을 통해 기능을 확인할 수 있다.
- 작업그룹추가 버튼 : 재분류 대상 목록 자료를 업로드한다.
- AI재분류 버튼 : AI재분류를 수행한다.
- 검토서 작성 버튼 : 선택한 건들의 검토서 생성한다.
- Excel다운로드 버튼 : 화면에 검색된 결과에 대해 Excel 다운로드한다.

(그림 36 - PoC Home Screen)

❍ PoC 작업그룹 추가

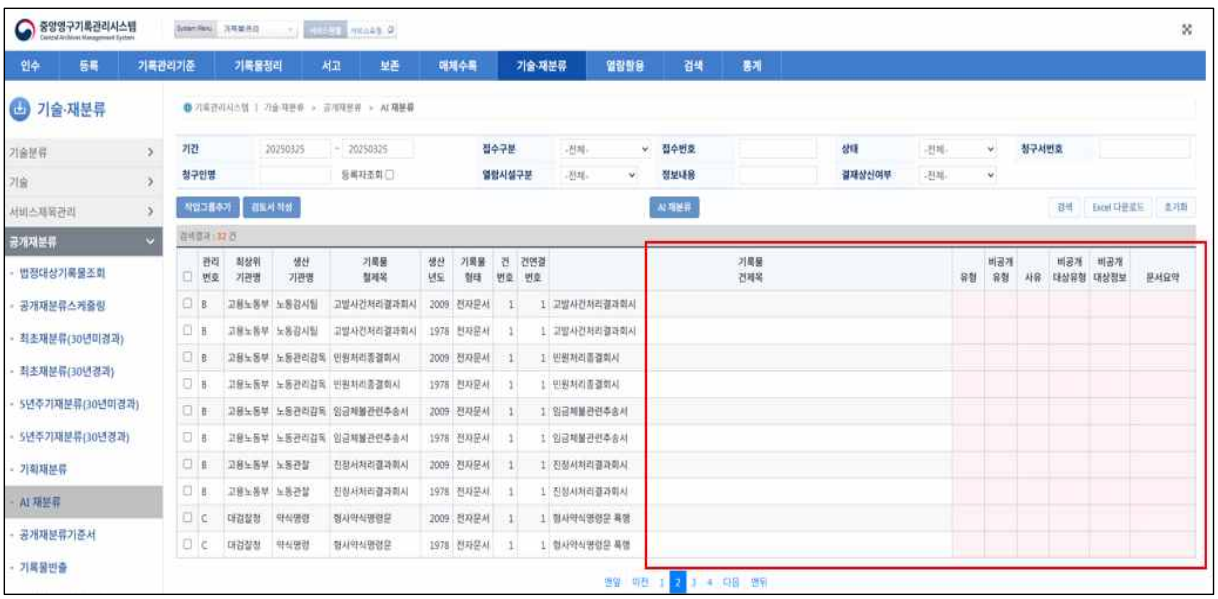
- 재분류 대상 목록 자료를 업로드하여 AI 재분류 할 리스트를 생성한다.
- 대상 목록은 CAMS에서 생성한 재분류가 필요한 철, 건에 대한 CVS 형식의 자료에 해당한다.



(그림 37 - PoC PoC Add Workgroup Screen)

❍ PoC AI 재분류 항목

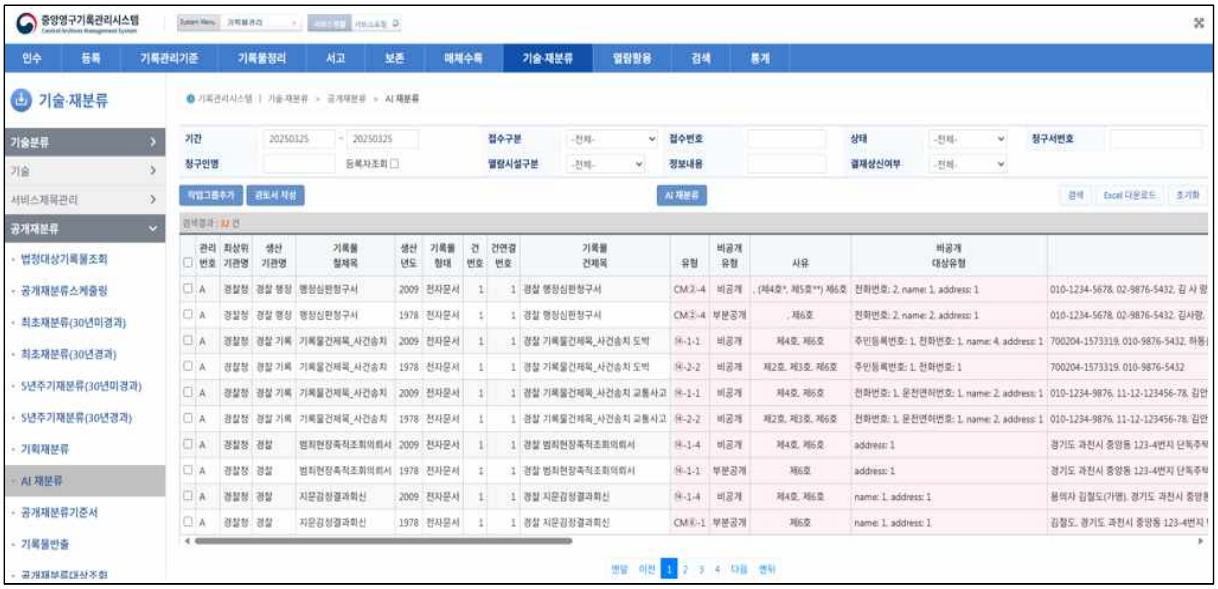
- 유형, 비공개 유형, 사유, 비공개 대상유형, 비공개 대상정보, 문서요약 항목을 AI 재분류를 통해 생성한다.
- 화면 좌측 Check Box를 통하여 재분류 대상 건을 선택한다.
- AI 재분류 버튼 인터페이스를 통해 선택된 건에 대한 재분류를 시행한다.



(그림 38 - PoC AI Reclassification Items Screen)

○ PoC AI 재분류 결과

- AI 재분류를 통해 유형, 비공개 유형, 사유, 비공개 대상유형, 비공개 대상정보, 문서요약 정보가 비동기 방식으로 데이터베이스에 저장되어 그 정보를 목록 및 상세 모달 팝업을 통하여 조회할 수 있다.



(그림 39 - PoC AI Reclassification Results Screen)

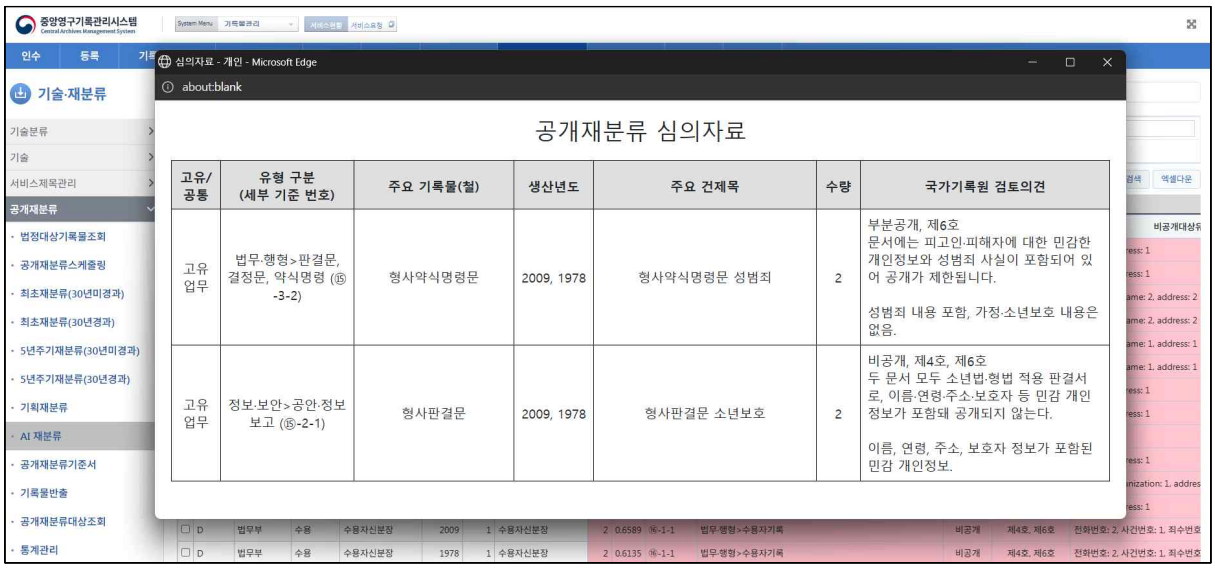
○ PoC 검토서 작성

- 화면 좌측 Check Box를 통하여 검토서 작성 대상 건을 선택한다.
- 검토서 작성 버튼을 통하여 동기방식으로 AI가 작성한 공개재분류 검토서을 확인할 수 있다.

(그림 40 - PoC Review Form Screen)

❍ PoC 심의자료 작성

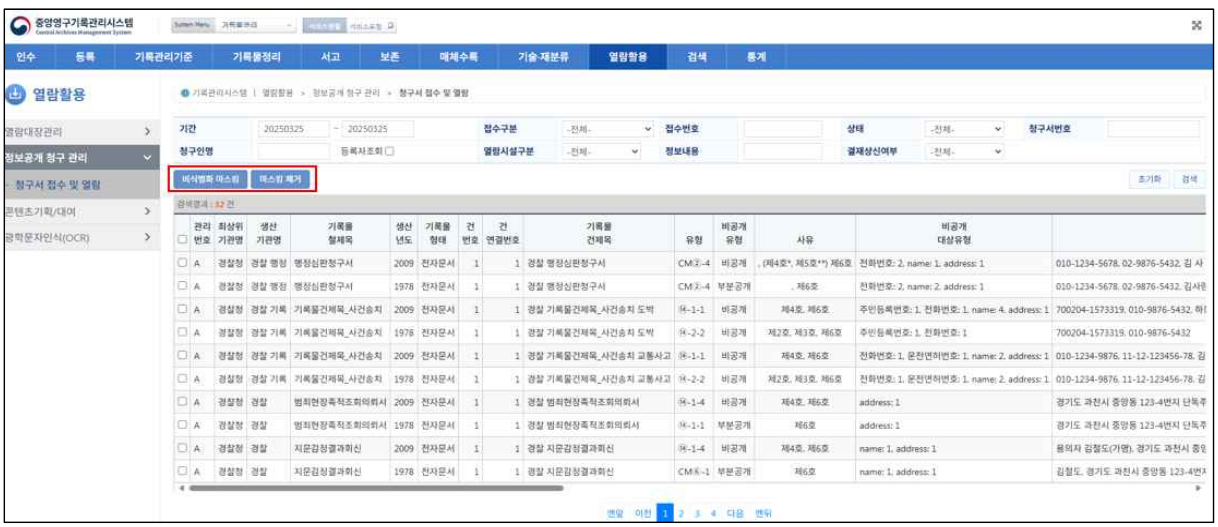
- 화면 좌측 Check Box를 통하여 심의자료 작성 대상 건을 선택한다.
- 심의자료 작성 버튼을 통하여 동기방식으로 AI가 작성한 심의자료를 확인할 수 있다.
- 국가기록원 검토의견에는 판단근거, 기존 심의이력 추가등 다양하게 생성형 AI를 활용하여 문서 자동화 기능을 추가할 수 있다.



(그림 41 - PoC Review Document)

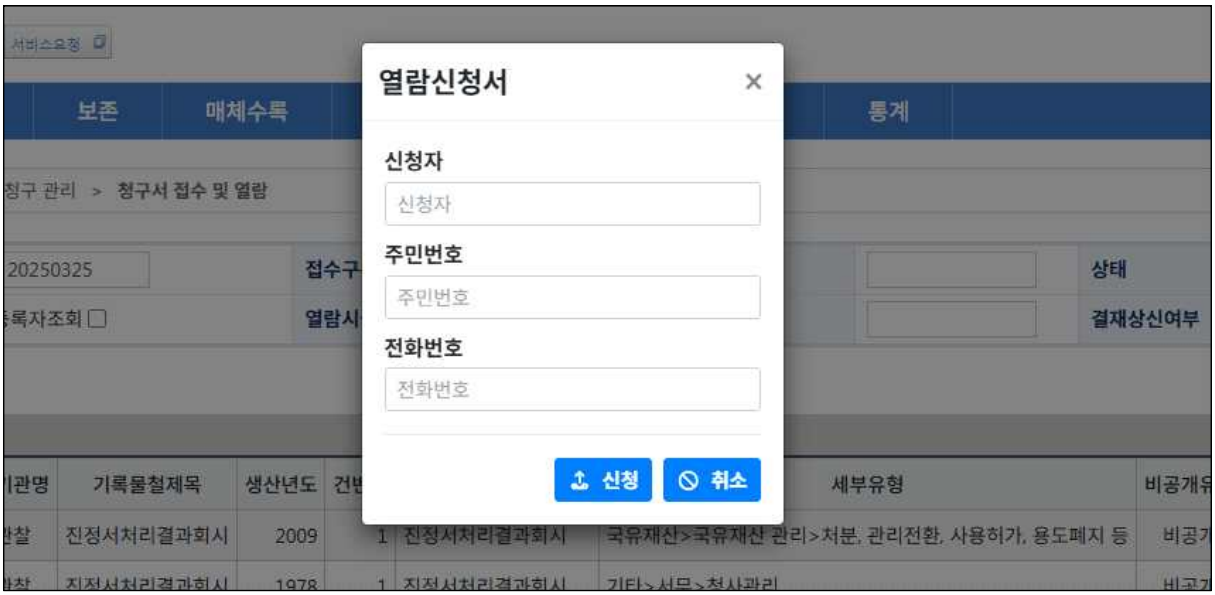
❍ PoC 비식별화 마스킹

- 비식별화 버튼을 통하여 재분류 정보로 추출한 비공개 항목을 원문 PDF에서 마스킹하여 저장된 PDF를 조회하여 확인 할 수 있다.
- 마스킹 제거 버튼을 통하여 비식별화 정보를 마스킹한 PDF를 원문으로 복원할 수 있다.

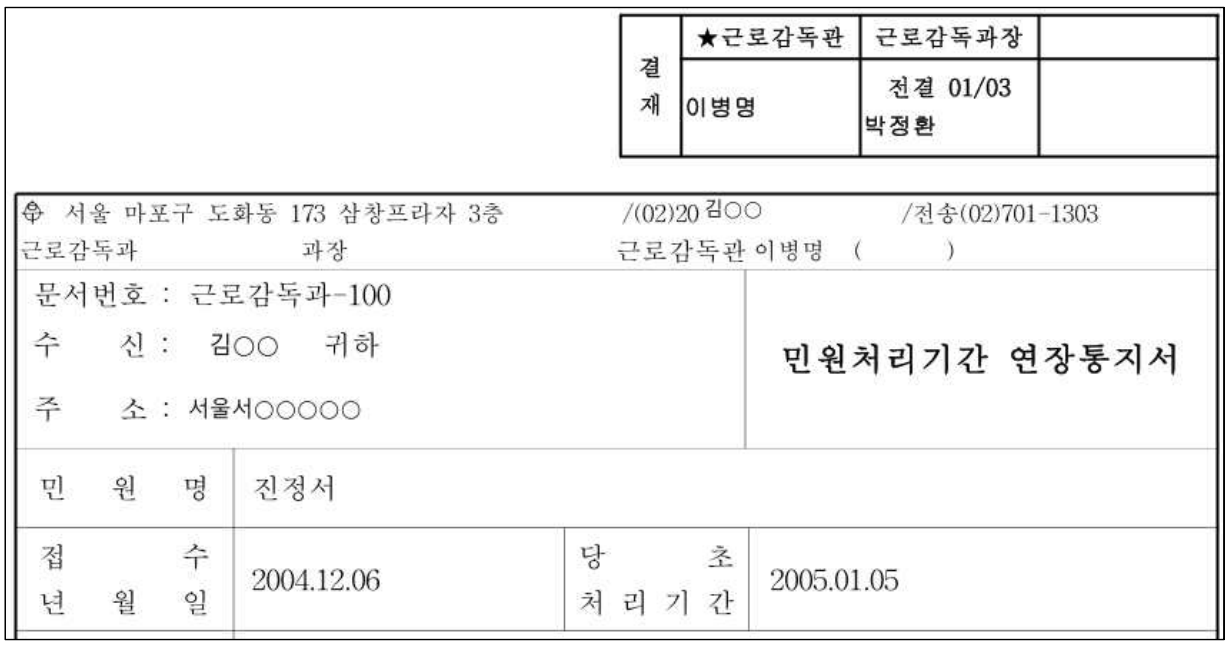


(그림 42 - PoC PoC De-identification Masking)

- 비식별화 버튼 클릭시 열람자의 신상정보를 입력받아 열람자의 신상정보는 마스킹에서 제외하고 나머지 정보는 마스킹이 수행된다.(이용자 맞춤형 비식별화)



- 비식별화 처리된 사례



결재	★근로감독관	근로감독과장	
	이병명	전결 01/03 박정환	

㉾ 서울 마포구 도화동 173 삼창프라자 3층 근로감독과 과장	/(02)20 김○○ 근로감독관 이병명 ( )	/전송(02)701-1303	
문서번호 : 근로감독과-100 수 신 : 김○○ 귀하 주 소 : 서울서○○○○○		민원처리기간 연장통지서	
민 원 명	진정서		
접 수 년 월 일	2004.12.06	당 초 처 리 기 간	2005.01.05

3.6.3. 성과지표 측정

3.6.3.1. 목표 및 결과

○ 연구과제 제안과정에서 성과지표는 아래와 같이 제안하였다(연구개발계획서<본문 2>, 6페이지)

- AI를 활용한 지능형 공개재분류 수행 모델 연구 개발

지표명	설명	목표치
1. 자동 분류 정확도	전체 문서 중 AI가 정확하게 공개/비공개/부분공개를 분류한 비율	≥ 90%
2. 호수별 자동 판별 정확도	각 비공개 사유(1~8호)에 대한 자동 분류 정확도	≥ 85%
3. 수작업 문서 판독 감소율	사람이 검토하지 않아도 되는 문서 비율	≥ 50% 감소
4. 분류 소요시간 단축률	기존 수작업 대비 문서 1건 처리 시간 단축률	≥ 70% 단축

(표 35 - Performance Indicators for Disclosure Reclassification)

- 비공개대상정보의 자동식별 및 마스킹을 통한 대국민 열람 제공 모델 연구 개발

성과지표	측정방식	목표치
5. 개인정보 탐지 정확도	탐지 정확도(Precision/Recall)	정밀도 95%, 재현율 90%
6. 마스킹 적용 정확도	마스킹 누락 및 과잉률	누락 <2%, 과잉 <5%
7. 수작업 개입 감소율	도입 전후 수작업 시간 비교	70% 이상 감소
8. 문서유형자동처리커버리지	문서별 자동처리 성공률	95% 이상

(표 36 - Performance Indicators for Automated Non-disclosure Identification and Masking)

○ 8개 기관에서 기록물 재분류 세부유형 21개에 해당하는 기록물 684건을 대상으로 비교하고 측정하였다.

지표명	목표치	**결과**	비고
1. 자동 분류 정확도	≥ 90%	92.0%	
2. 호수별 자동 판별 정확도	≥ 85%	90.1%	
3. 수작업 문서 판독 감소율	≥ 50% 감소	NA	수작업과 단순비교는 어려움
4. 분류 소요시간 단축률	≥ 70% 단축	NA	수작업과 단순비교는 어려움
5. 개인정보 탐지 정확도	정밀도 95%, 재현율 90%	정밀도 90.8%, 재현율 94.9%	
6. 마스킹 적용 정확도	누락 <2%, 과잉 <5%	누락 0%, 과잉 0%	AI 모델과 관계없이 PDF 패턴 치환방식
7. 수작업 개입 감소율	70% 이상 감소	NA	수작업과 단순비교는 어려움
8. 문서유형자동처리커버리지	95% 이상	100%	PDF 유형만 처리

(표 37 - PDF-based Performance Indicators)

3.6.3.2. 주요성과

○ 자동 분류 정확도(성과지표 1번)

구분		건수		비고
공개값 일치	공개값 일치	546	**629 (92%)**	세부유형기준서의 공개값과 일치
	추가 검출	83		세부유형기준서 공개값 외에 추가 비공개 정보 검출 사례
공개값 불일치		55(8%)		벡터DB(RAG)에 참고 유형이 없어 판단 불가
합계		684(100%)		

- 전체 684건 중 629건은 세부유형 기준서의 공개값과 일치하거나, 기록물 내용을 기반으로 판단했을 때 공개 결정이 기준서와 합치하는 사례로 확인되었다.
- 이러한 결과는 LLM이 문서를 요약한 뒤, 벡터DB에 축적된 재분류 지식정보 중 가장 유사한 사례를 참조하여 종합적으로 공개 범주를 결정하는 과정이 일정 수준의 정확성을 확보하고 있음을 보여준다. 이는 AI 모델이 기록물 재분류 업무에서 효과적인 지원 도구로 기능할 수 있음을 시사한다.
- LLM은 다양한 자연어처리(NLP) 기능 중에서도 특히 문서 요약 능력에서 강점을 보인다. 이러한 특성을 활용하여 문서를 요약한 후, 요약 결과를 기반으로 벡터DB를 검색한 결과 정확도가 높음을 확인할 수 있었다.
- 모든 문서를 요약하고 이를 기반으로 검색하는 방식은 사람이 방대한 기록물을 전수 검토하기 어려운 한계를 보완해주는 효과적인 접근이다. 대부분의 기록물은 동일 유형의 문서가 철 단위로 편철되지만, 그 안에 상이한 유형의 문서가 혼재하는 경우가 존재한다. 이러한 상황에서도 LLM은 비정형 문서를 정확하게 식별해내는 성능을 보였다.
- 예를 들어, 《약식기소문》 철내에 “약식기소문 양식 배포“ 와 같은 문서가 포함된 경우 실제 약식기소문 유형에서 제외하고 공개 대상으로 분류하였다.
- 세부유형 기준서에 해당 유형이 존재하지 않거나, 문서 요약 결과의 변별력이 낮은 경우 LLM의 판단 정확도는 감소하는 경향을 보인다. 따라서 기준서의 세부유형 체계를 정교화하고, 변별력을 높일 수 있도록 내용을 정제하는 데 상당한 자원이 투입될 필요가 있다.
- LLM이 비식별화 대상 정보를 추가로 검출하여 판단의 정확도를 향상시킨 사례는 총 83건으로 확인되었다. 이들 사례에서는 검출된 비식별화 대상 정보를 근거로 6호 및 7호 사유를 추가하여 공개유형을 최종적으로 결정하였다.
- 반면, 부정확한 것으로 분류된 55건은 벡터DB 내 유사 사례가 존재하지 않아 판단의 근거가 부족한 경우였다.

○ 호수별 자동 판별 정확도(성과지표 2번)

<table>
<tr><th>구분</th><th>기준서와 동일</th><th colspan="2">건수</th><th>비고</th></tr>
<tr><td rowspan="2">비공개호수 일치</td><td>호수 일치</td><td>466</td><td rowspan="2">616
(90.1%)</td><td>세부유형기준서의 호수와 동일한 경우</td></tr>
<tr><td>6호 7호 추가 검출</td><td>150</td><td>기준서의 호수에 비식별화 대상정보를 추가 검출한 경우</td></tr>
<tr><td colspan="2">비공개호수 불일치</td><td colspan="2">68(0.9%)</td><td></td></tr>
<tr><td>합계</td><td>684</td><td colspan="2">684(100%)</td><td></td></tr>
</table>

- 전체 684건 중 호수 판별이 유효하게 수행된 사례는 616건이었다. 이 중 466건은 세부유형 기준서와 동일한 호수로 판정되었으며, 나머지 150건은 AI 모델이 추가적으로 비식별화 대상 정보를 검출함으로써 기존 기준서의 호수를 보완적으로 판단한 사례였다.
- 호수 판별 과정에서 시스템은 우선적으로 벡터DB에 저장된 기존 재분류 사례를 참조하여 초기 판단을 수행하였다. 그러나 최종 판정 단계에서는 비식별화 대상 정보 중 제6호 및 제7호에 해당하는 항목을 검출하기 위해 원문을 직접 검색·분석하였으며, 해당 정보가 실제로 존재하는 경우에 한해 추가적인 호수 판정을 수행하였다.
- 이러한 절차를 통해, 예를 들어 문서 제목이 “택시 임금실태 현황” 인 문서가 제7호로 판정된 경우라도, 문서 본문 내에 택시회사 대표자명 등 개인식별 가능 정보가 검출된 경우에는 제6호와 제7호로 복합 판정이 이루어졌다.
- AI 모델은 지식데이터로 구성된 벡터DB를 기반으로 호수를 판별하는 기능뿐 아니라, 원문 분석을 통해 추가적으로 비식별화 정보를 식별하여 최종 호수를 보완·결정하는 역할을 수행하였다. 이러한 결과는 시스템이 비식별화 대상 정보 판별의 정밀도를 향상시키는 데 있어 의미 있는 성과를 보여주었음을 시사한다.

○ 수작업 문서 판독 감소율(성과지표 3번), 분류 소요시간 단축률(성과지표 4번), 수작업 개입 감소율(성과지표 7번)

- AI 모델 처리 속도 : 평균 89.43초/1건
- AI 모델의 처리 속도를 정량적으로 평가하기에는 다소 어려움이 있었다. 특히 문서 분량이 길거나 개인정보가 다량 포함된 경우(예: 일용직 일당 지급 내역 등)에는 분석 시간이 크게 증가하여 15분 이상 소요되는 사례도 확인되었다. 다만, 이는 PoC 환경의 제한된 자원에서 발생한 현상으로, 실제 시스템 구축 단계에서 충분한 GPU 자원을 확보할 경우 이러한 처리 속도 문제는 크게 완화되거나 실질적인 제약으로 작용하지 않을 것으로 판단된다.
- 또한, 제시된 3개의 성과지표는 국가기록원의 수작업 기반 재분류 프로세스와 공개재분류 AI 모델의 자동화 처리 방식이 구조적으로 상이하여 단순 비교가 어려웠다. 이러한 특성으로 인해 해당 지표들은 본 평가의 측정 대상에서 제외하였다.

○ 개인정보 탐지 정확도(성과지표 5번)

- 276개 중 262건 검출, 14건 미 검출, 24건 과잉검출(기록물의 비식별화 대상정보 개수기준)
- 숫자형 정보 외의 문자형 개인정보(예: 이름, 주소, 단체 및 법인명 등)는 LLM이 직접 검출하도록 하였다.(*.go.kr, *.korea.kr 도메인의 e-mail은 검출에서 제외)
- 검출 과정에서는 업무처리 공무원, 수사관, 판사, 공공기관명 등 공적 직무 관련 명칭은 검출제외하도록 하였으며, LLM은 문맥을 바탕으로 검출과 비식별화 대상 제외를 정확하게 하는 성과를 보였다.
- 기존에는 인명 검출을 위해 고유명 인식(Named Entity Recognition, NER) 기반 비교 방식을 활용하는 방법이 논의되었으나, 이 접근법은 문맥적 의미를 반영하지 못한 단순 패턴 매칭 기반 기법으로 오류 발생 가능성이 존재하였다.
- 이에 비해 본 연구에서 활용한 LLM 기반 방식은 문맥을 이해하고, 의미 단위로 판단하여 문자형 개인정보를 검출하였다는 점에서 의미 있는 성과를 보였다.

○ 마스킹 적용 정확도(성과지표 6번)

- 마스킹 대상은 마스킹 적용을 하고, 제외 대상은 모두 제외함(정확도 100%)
- 기존 상용 솔루션의 경우, 개인정보를 마스킹하여 표출하는 기능은 제공하지만, AI 모델이 검출한 비식별화 정보와의 연계 및 데이터 형식 변환, 협조 과정에 상당한 시간이 소요될 것으로 판단되어, 본 연구에서는 PDF 파일 내 개인정보를 직접 마스킹하는 자체 개발 모듈을 구현하였다.
- 해당 모듈은 LLM의 결과와 무관하게 PDF 내 텍스트 치환 방식으로 동작하도록 설계되었으며, 실험 결과 마스킹 성공률 100%를 나타내었다.

○ 문서유형자동처리커버리지(성과지표 8번)

- AI 모델이 공개재분류를 판정하는 부분에 초점을 맞추는 측면에서 연구과제 수행초기에 텍스트 기반의 PDF 문서를 대상으로 범위를 한정하였으며, 텍스트 기반의 PDF 파일에 대한 처리에서는 문제가 발견되지 않았다.

3.6.3.3. PoC 오류 사례

○ 같은 문서를 다르게 판별하는 경우

- 동일한 문서가 여러 기관에 배포되면서 서로 다른 건으로 기록된 사례가 있었으며, 이 경우 원칙적으로 동일한 분류 결과가 도출되어야 한다. 그러나 벡터DB 검색 과정에서 서로 다른 유사 문서를 참조함에 따라 상이한 결과가 제시되는 사례가 확인되었다.
- 두 문서 모두 범죄수사와 관련된 내용을 포함하고 있어, “국가원수모독”이라는 특정 표현을 제외하면 문장 전체의 유사도가 매우 높다. 이로 인해 유사도 기반 검색 시 상황에 따라 서로 다른 사례가 참조되며, 최종적으로 하나는 일반적인 “수사”로, 다른 하나는 “공안사범 > 대공수사”로

분류되는 등 분류 결과가 크게 달라지는 사례가 확인되었다.

기록물 A RAG 검색 결과	○ 범죄사건부 : 번호, 범죄접수부 번호, 수사 담당자, 죄명, 적용법조, 송치, 송치의견, 발각원인, 증거, 압수부 번호, 압수물, 수사종결사건(송치사건)철 번호, 피의자(본적, 주소, 출생지, 직업, 전과, 성명, 주민등록번호, 별명, 생년월일), 구속(긴급, 영장발부, 집행일시, 구속장소), 석방(일시, 이류), 검사처분(월일, 요지), 재판결과(월일, 요지), 범죄통계 원표(발생사건표, 검거사건표), 피의자 표 등이 기재
	⑭-1-1, 수사 > 각종 사건부, 제6호
	유사도 : 0.6390
기록물 B RAG 검색 결과	○ 국가원수모독사범수사 : 신고 및 서신으로 고발을 당한 국가원수모독사범에 대한 수사 관련 문서이다. '국가원수모독사범 동향보고, 국가원수모독사범 검거보고, 국가원수모독사범 수사 종결보고, 진술조서' 등의 건으로 구성되어 있으며 고발인, 참고인, 용의자의 개인정보 및 이력 전반(성명, 주소, 본적, 직업, 나이, 학력, 경력, 가족관계, 사상 및 전과)에 관한 사항이 기재되어 있다. 또한 참고인, 용의자의 각서, 진술조서와 고발서신, 신병인수서가 함께 첨부되어 있으며 보고서에는 검거사항, 조사사항 및 그 결과에 따른 지시사항이 기재되어 있다.
	⑭-2-2, 정보·보안 > 공안사범 > 대공수사, 제2호/제3호/제6호
	유사도 : 0.6193

(표 38 - Examples of Measurement Result Discrepancies 1)

❍ 개인정보가 다수 포함된 경우 신원조사서로 판별하는 사례

- 원문은 생활실태조사보고서로, 다수의 개인정보가 포함된 문서였다. 그러나 RAG 기반 검색 과정에서 공무원 인사 발령 시 활용되는 신원조사서 역시 유사한 유형의 개인정보를 광범위하게 포함하고 있어, 두 문서 간 유사도가 높게 산정되었다. 이로 인해 실제 문서유형과 상이한 문서가 검색·참조되는 사례가 발생하였다.

원문	○○○생활실태조사보고서 : 생활실태에 대한 내용으로 다수의 개인정보 포함
RAG 검색 결과	○ 신원조사 : 신원조사 의뢰 및 회보 받은 내용으로 요청관, 조사사유, 본적, 주소(읍·면·동 이하), 소속, 직위, 주민등록번호, 이름, 생년월일, 사상관계, 성질, 소행, 상벌관계 등이 기재
	공통업무 ③-1-1, 인사 > 인사·징계 > 인사발령

(표 39 - Examples of Measurement Result Discrepancies 2)

❍ 사건정보가 다수 포함된 경우 수사자료로 판별하는 사례

- 원문은 방화 사건에 대한 감정을 의뢰하는 문서로, 내용 전반에 사건 관련 정보가 다수 포함되어 있었다. 그러나 RAG 기반 검색에서는 이러한 사건 관련 표현들이 강조되면서 수사 과정에서 활

용되는 수사자료송부철과 유사도가 높게 산정되었고, 그 결과 해당 유형으로 잘못 검색·참조되는 사례가 확인되었다.

원문	방화감정의뢰서(경찰청, 30년미만) : 감정을 의뢰하나 내용중에는 다수의 사건관련정보가 포함됨
RAG 검색 결과	○ 수사자료송부철 : 일자별로 편철된 수사자료표로, 수사·감식 일련번호, 제목, 작성번호, 피의자 성명, 작성년월일, 죄명, 발견여부(죄명기재), 주민등록번호, 의뢰사유, 회보연월일, 일지지문원지, 처분결과확인 등으로 구성
	⑭-1-4, 수사 > 수사 자료표, 배당부, 색인부

(표 40 - Examples of Measurement Result Discrepancies 3)

- 이러한 사례들은 벡터DB를 구축할 때 세부유형 기준서에 포함된 문서유형별 설명 컨텍스트를 보다 변별력 있게 정비하여 반영할 필요가 있음을 시사한다.

3.6.3.4. 시사점

○ 본 연구를 통해 공개재분류 AI 모델은 LLM의 문서 이해 및 요약 능력을 활용하여 문서의 핵심 내용을 효과적으로 요약하고, 이를 벡터DB 검색과 연계할 수 있도록 하였다. 특히 동일 유형의 문서가 다수 편철된 기록물 철 단위 내에서도 유형이 상이한 문서를 정확히 식별할 수 있으며, 이러한 결과는 AI 모델이 단순한 패턴 매칭을 넘어, 문서의 의미 구조를 해석하고 적절히 판별할 수 있는 기능을 업무에 적용할 수 있음을 보여준다.

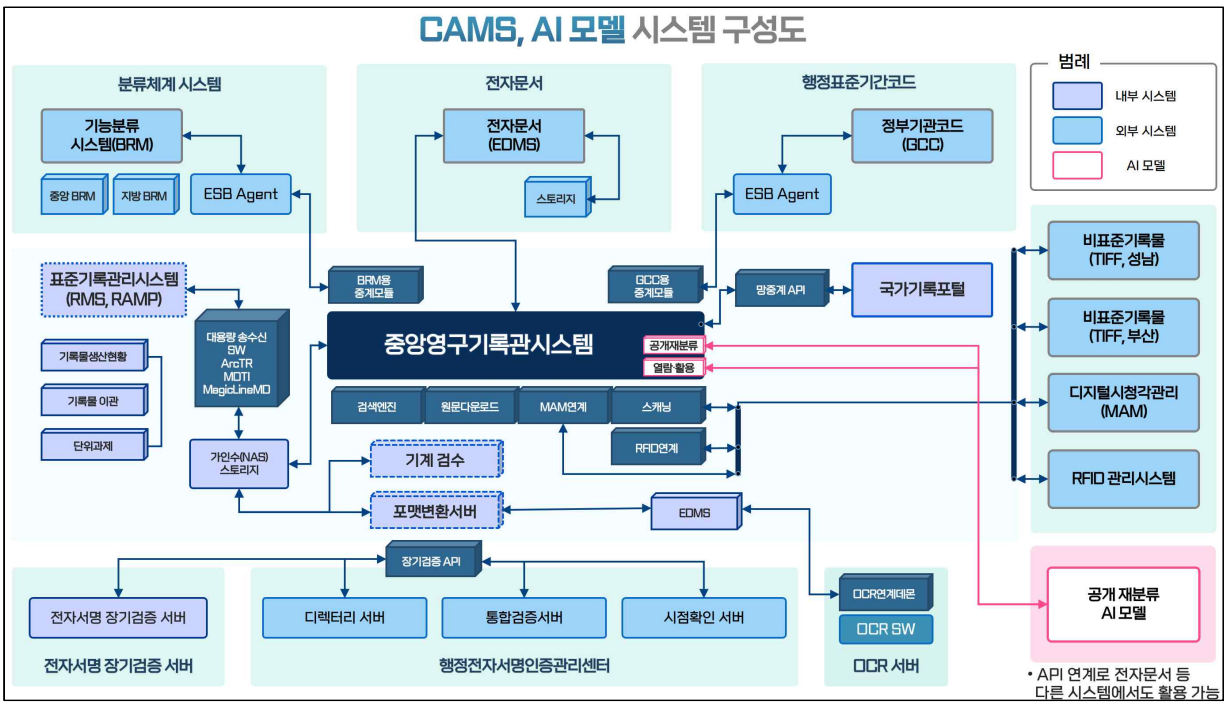
○ 비식별화 대상 정보 식별은 문자형 개인정보(이름, 주소, 단체/법인명 등)에 대해서는 LLM의 문맥 이해 능력을 활용하여 직접 검출하도록 하였으며, 업무처리 공무원, 판사, 수사관 등 공적 직무 관련 명칭을 비식별화 대상에서 제외하여 불필요한 과잉 검출을 최소화하였다. 이는 기존의 NER(Named Entity Recognition) 기반 단순 패턴 매칭 접근법보다 문맥 인식 기반의 정밀한 개인정보 검출 성능을 달성한 성과로 평가된다.

○ 벡터DB로 구성된 공개재분류 지식정보의 경우, 문맥적 유사도가 높은 데이터 간에 상이한 판단이 이루어질 가능성이 확인되었다. 따라서 향후 공개재분류 지식정보를 벡터DB화할 때에는, 문맥적 유사도가 과도하게 높지 않도록 문장 구성을 조정하고 체계적으로 관리할 필요가 있다. 또한 현재의 “기준서 검색시스템”을 고도화하여 공개재분류 지식정보 관리시스템으로 발전시킴으로써, AI 모델이 해당 지식정보를 정확하게 활용하여 판단 할 수 있도록 관리체계를 구축하는 것이 필요하다.

3.7. 기록관리시스템 기능 연계 제안

3.7.1. 기록관리시스템 기능 연계 방안

○ CAMS 연계 구성도(To-Be)



○ CAMS/RMS 시스템과 AI 공개재분류 모델 간의 데이터 교환은 API Call 기반의 연계

- "API 방식"은 직접 DB에 접근하지 않고, API 호출을 통해 데이터나 기능을 사용하는 방식이다.
- 또한, API 방식은 시스템 간 연계나 통합을 유연하게 하기 위한 구조적 선택이 된다.

○ API 방식 종류

- RESTful API 방식

- HTTP 메서드(GET, POST, PUT, DELETE)를 이용해 자원(Resource)을 CRUD하는 구조 - 데이터는 주로 JSON 포맷으로 교환 - 예시: GET /api/users/123 → 사용자 정보 조회 POST /api/orders → 주문 등록

- SOAP API 방식

- XML 기반의 메시지 구조를 가진 Web Service 표준 - 보안, 트랜잭션 등 복잡한 엔터프라이즈 통신에 적합

- 내부 시스템 간 연계(API Gateway 등)

- 조직 내 여러 시스템이 데이터를 교환할 때 API 서버를 매개로 통신 - 예: “공통업무 모듈(API)”을 통해 기관 시스템들이 데이터 조회

○ 데이터 관리 및 보전

- 공개 재분류 결과 생성되는 데이터는 기록물의 데이터베이스에 추가되는 메타데이터 형태로 구성하여도 기존의 CAMS 시스템 데이터베이스에 영향을 주지 않으면서 구성이 가능하다.
- 또한, 기록물 통합생산관리시스템이 구축되면 “보존·활용 꾸러미” 에 포함하여 통합적이고, 효율적인 데이터 관리가 가능하게 한다.

○ 기대효과

- 통합성 강화 : CAMS와 AI 분석 서비스 간의 연계로 기록물 관리 전 주기 자동화 실현 가능
- 효율성 제고 : 수작업 중심의 공개재분류 과정을 자동화하여 업무 효율 향상
- 데이터 자산화 : AI 분석 결과를 보존·활용 꾸러미로 관리함으로써, 공공데이터의 지속적 활용 기반 확보 가능

3.7.2. 기록관리기관으로 확산·보급방안

3.7.2.1. 추진개요

○ 본 확산·보급 방안은 전국 지자체 및 공공기관의 기록관리시스템(RMS)에 AI 기반 공개재분류 서비스를 단계적으로 도입·확산하기 위한 전략적 실행체계를 제시하는 것이다.

○ 공개재분류는 기록물의 공개 여부를 재판단하고, 민감정보 및 비식별화 대상을 자동 식별하는 핵심 기능으로, 공공기록물의 개방성과 보안성 간 균형을 유지하는 데 필수적이다.

○ 따라서 본 방안은 단순 기술 도입이 아닌, 기술·운영·행정 전 영역이 유기적으로 통합된 구조로 추진되어야 하며, 특히 기관별 RMS 환경의 이질성과 데이터 표준화 수준의 차이를 극복하기 위한 사전 검토 및 협력체계가 중요하다.

3.7.2.2. 대상기관 범위 및 고려사항

○ 대상기관 설정

- 본 사업의 확산 대상은 전국 지자체 및 공공기관 중 RMS를 구축·운영 중인 기관을 중심으로 설정한다. 또한, 중앙행정기관 및 산하기관의 RMS와의 연계 가능성을 검토하고, 향후 통합 기록관리 서비스로 확장 가능한 구조를 고려해야 한다.

○ 환경분석 및 호환성 확보

- 기관별 RMS의 플랫폼 구조, 데이터 표준, API 구성 등 기술적 환경은 일반적으로 패키지 기준으로 동일하나 기관의 현황에 따라 커스터마이징 된 경우도 있으니 이에 대한 사전 진단을 통해 시스템 간 데이터 연계 및 인터페이스 호환성 확보를 검토하여야 한다.
- 이를 위해 표준화된 데이터 스키마와 공통 API 구조를 도입하여, 기관 간 기술적 장벽을 최소화하고 재사용성을 높이는 것이 중요하다.

3.7.2.3. 기술적 확산 방안

○ 기술적 측면에서는 각 기관이 보유한 공개재분류 지식데이터를 기반으로, AI 모델이 문서의 민감도와 공개등급을 자동 판단할 수 있도록, RAG(Retrieval Augmented Generation) 구조를 설계하는 것이 핵심이다.

○ RAG 기반 기술 구조 설계

- 지식 데이터 수집 및 분석 : 각 기관의 기록물 데이터를 수집·정제하고, 공개·비공개 기준에 따라 벡터 DB 구축용 지식데이터 세트를 확보한다.
- 핵심 구성요소 표준화 : 문서 검색 모듈, 임베딩 DB, 응답 생성기 등의 구성요소를 표준화하여 기관별 환경 차이에도 동일한 알고리즘 구조를 적용할 수 있도록 고려하여야 한다.
- 데이터 표준화 : 기관별 포맷 차이를 최소화하기 위한 데이터 표준 스키마를 정의하고, 비식별화 기준을 통일할 필요가 있다.

○ 기술 검증 및 확산 단계

- 1단계로 시범기관 기반의 파일럿 모델을 구축하여 기술 안정성과 성능을 검증한다. 검증 결과를 토대로 모델의 정밀도, 판별률, 비식별화 정확도를 개선하고 이후 2단계에서 광역·기초지자체 중심으로 점진적 확산을 추진한다.
- 기술적 완성도가 확보되면 3단계에서는 전국 단위의 기록물관리 통합플랫폼 내 서비스 모듈 형태로 통합 운영할 수 있도록 한다.

3.7.2.4. 운영체계 확산 방안

○ 운영 측면에서는 기관별 인프라 및 관리 수준의 차이를 고려하여, 이중 운영모델(독립형·공동형)을 병행 적용을 고려한다.

○ 독립형 모델

- 개별 기관이 자체 RMS 환경에 최적화된 형태로 AI 서비스를 구축·운영한다. 독립적인 인프라 관리가 가능하므로, 기관의 보안정책이나 내부 규정에 유연하게 대응할 수 있다.

○ 공동형 모델(Consortium Model)

- 인접 지자체 또는 유사 기능을 가진 기관들이 협력하여 AI 서비스와 데이터 인프라를 공동 구

축 및 공유하는 모델이다.

- 이를 통해 예산 절감, 기술 확산 가속, 데이터 학습 품질 향상 등의 효과를 기대할 수 있다
- 단, 공동형 모델은 보안, 접근권한, 데이터 소유권 등 관리책임 체계를 명확히 해야 하며, 중앙기록관리기관이 기술 사양 및 인터페이스 표준을 관리하는 역할을 수행해야 한다.

○ 중앙-지자체 협력 운영체계

- 중앙기록관리기관은 기술 표준, 운영 가이드, 인터페이스 사양을 관리하고, 각 지자체는 이를 준용하여 자체 서비스 운영 및 확산체계를 갖춘다. 이러한 협력구조는 운영 표준화를 통한 전국 단위 서비스 일관성 확보가 가능하다.

3.7.2.5. 행정·재정 추진 방안

○ 행정·재정 측면에서는 안정적 예산 확보와 행정지원 절차 정비가 필수적이다.

○ 예산 확보 체계

- 각 기관의 예산 구조 내 정보화사업 항목 또는 RMS 고도화 사업 항목을 활용하여 AI 서비스 구축비용을 반영할 수 있다.
- 중앙기관(국가기록원 등)과 협력하여 공동 예산 편성 및 국비지원 연계 방안을 마련함으로써 전국적 사업으로 확산할 수 있는 재정적 기반을 확보한다.

○ 단계별 사업 추진 로드맵

- 1단계(파일럿 구축): 시범기관 중심 PoC 및 서비스 프로토타입 운영
- 2단계(확산 단계): 광역 및 기초지자체로 확산, 모델 고도화 및 성능 개선
- 3단계(통합 운영 단계): 전국 단위 통합 기록관리플랫폼 내 서비스 모듈화 및 통합운영

3.7.2.6. 결론

○ 공개재분류 AI 서비스의 전국 확산은 기록관리의 효율성 제고와 행정정보 개방의 투명성 확보를 위한 필수 과제라 할 수 있다.

○ RAG 기반의 AI 기술과 자동화·지능화·표준화를 통해 확산을 수행하고, 본 방안은 단순한 기술 도입을 넘어 “지능형 기록관리 생태계” 로의 전환을 촉진하는 전략적 기반이 될 것이다.

제4장 총괄연구개발과제의 연구결과 고찰 및 결론

4.1. 연구결과의 고찰

4.1.1. AI를 활용한 “공개재분류 지원 모델” 연구 개발

○ 본 연구는 국가기록원의 공개재분류 업무에 인공지능(AI)을 적용하기 위한 최적의 접근 방식으로 조건 인지형 RAG (Condition-aware Relevance-Augmented Generation, 조건에 따라 동적으로 판단하는 RAG 모델) 기반 AI 모델의 도입 필요성과 유효성을 성공적으로 검증하였다. RAG 모델은 LLM의 문서요약 및 문맥파악 능력을 적극 활용함과 동시에, 과거 재분류 지식저장소를 근거로 판단을 수행함으로써 설명 가능한 근거 기반 재분류 체계를 구현할 수 있음을 확인하였다.

○ 또한 RAG용 문서 코퍼스를 구축하고, 기관별 재분류 사례 및 세부 판단기준을 메타데이터 중심의 지식구조로 설계함으로써 도메인 특화된 판단 정확도 향상, 기준 변경 시 재학습 없이 운영 가능한 유연성 등 구조적 장점을 확보하였다.

○ RAG 기반 모델 도입의 필요성 및 타당성 검증

RAG는 LLM의 문서 요약 및 문맥 분석 능력을 기반으로, 추론 과정에서 외부의 재분류 지식저장소를 검색하여 가장 유사한 기존 사례를 근거로 판단을 수행한다. 본 연구에서는 공개재분류 업무가 상황적 요소를 고려한 판단이 요구된다는 점을 반영하여, 조건에 따라 동적으로 판단하는 조건 인지형(Condition-aware) RAG 모델을 구축하였다. 이를 통해 단순 분류 모델 대비 근거 기반의 판단이 가능해졌으며, 모든 문서를 확인하는 구조를 구현함으로써 국가기록원의 공개재분류 업무 특성과의 적합성을 확인하였다.

○ 문서 코퍼스(Document Corpus) 구축을 통한 정확도 향상

국가기록원의 기존 재분류 사례와 기관별 판단기준을 통합하여 RAG용 문서 코퍼스를 구축하였다. 문서유형별 설명 컨텍스트를 메타데이터 중심으로 구조화함으로써, LLM이 판단해야 할 영역을 도메인 특화 범위로 제한하였고, 그 결과 일반 LLM 대비 정확도와 일관성을 개선하였다. 새로운 재분류 기준 변경 시 재학습 없이 지식저장소 업데이트만으로 대응할 수 있어 정책·법령 변화에 유연한 시스템 구조를 확보하였다.

○ AI 기반 분류기(Classifier) 모델의 한계 확인

기존에는 AI 기반 분류기 모델 개발을 통해 정답 레이블을 직접 예측하는 방식을 고려했으나, 공개재분류 업무는 법령, 정책, 사회적 가치 판단이 동적으로 변화하는 복합적 판단영역이므로, 고정된 규칙 또는 학습 패턴만으로는 정확한 판단을 보장하기 어려웠다. 이에 따라 AI 단일 분류기 방식만으로는 실무 적용의 한계가 있음을 확인하였다.

○ LLM의 직접 법령 판단의 한계

LLM에게 정보공개법 제9조 등 법령기준을 직접 판단하도록 시도한 결과, 모델이 학습한 법령 데이터가 현행 규정과 불일치하거나 내용이 지나치게 포괄적으로 반영되는 문제가 발견되었다. 현행 오픈소스 기반의 sLLM 수준에서는 법령을 직접 해석하여 호수를 산정하는 방식은 안정적인 품질을 확보하기 어렵다는 점을 확인하였다.

○ 종합하면, 본 연구는 국가기록원의 공개재분류 업무 자동화를 위한 최적의 접근 방식으로 조건 인지형 RAG 기반 AI 모델이 가장 적합한 구조임을 확인하였다. RAG는 LLM의 추론 능력과 외부 지식저장소의 근거 활용을 결합하여 설명 가능하고 유연하며 정확도 높은 재분류 지원 체계를 구현하였으며, 향후 실서비스 적용을 위한 기술적 기반을 마련하였다.

4.1.2. “이용자 맞춤형 비식별화 기록물 제공 기능” 연구 및 개발

○ 본 연구는 공공기록물의 비식별화 대상 정보를 정확하게 검출하기 위해 기존 정규식 중심의 방식에서 벗어나 AI 기반 비공개정보 검출 및 비식별화 시스템을 구축한 연구로, 하이브리드 AI 검출 구조를 통해 비식별화 정확도를 유의미하게 향상시키는 성과를 거두었다.

○ 정규식, 규칙 기반 검증, 문맥 검증, LLM 문맥 보조 검증을 결합한 다층 구조를 적용하여 비식별화 대상 정보 검출의 신뢰성 및 정밀도를 확보하였으며, 숫자형·문자형 정보를 분리하여 검출 전략을 차별화함으로써 판단 오류를 크게 감소시켰다. 이를 위해 숫자형 정보에 대한 다단계 검출 파이프라인과 문자형 정보에 대한 LLM 기반 문맥 검출 기법을 결합함으로써, 기존 방식에서 발생하던 오탐·누락 문제를 효과적으로 줄일 수 있었다.

○ 또한 AI 모델이 검출한 비식별화 대상 정보를 활용하여 이용자(열람신청자)별로 민감 정보를 제외한 기록물을 제공하는 “이용자 맞춤형 비식별화 기록물 제공 기능”의 실현 가능성도 확인하였다.

4.2. 연구고찰에 따른 결론

4.2.1. 결론

○ 본 연구는 국가기록원의 공개재분류 및 비식별화 업무에 인공지능(AI)을 적용하여 효율적인 업무 수행이 가능함을 확인하는데, 큰 의미가 있다.

○ 우선, 공개재분류 업무는 고정된 규칙이나 데이터 패턴 학습만으로는 정확한 판단이 어려우며, 국가기록원의 공개재분류 지식정보를 RAG(Retrieval Augmented Generation) 기반 AI 모델이 판단하며, 설명 가능한 판단(Explainable Decision)이 가능해졌으며, 기준 변경에도 재학습 없이 유연하게 대응할 수 있는 구조적 장점을 확인하였다.

○ 또한, 기존 정규표현식 탐지 방식의 한계를 극복하기 위해 정규식·규칙·문맥·LLM 검증을 결합한 하이브리드 검출 구조로 숫자형 정보(주민등록번호·계좌번호 등)를 검출하고, 문자형 정보(이름·주소·기관명 등)는 LLM 프롬프트 기반 문맥 탐지로 정확도를 향상 시킬 수 있음을 확인하였다.

○ 그 결과, 공개 재분류 과정에서 검출된 비식별화 대상 정보가 열람·활용 시에 연계되어 활용 가능함을 확인하였고, 별도의 솔루션이 아닌 자체개발 모듈을 적용하여, 이용자 정보는 제외하고 마스킹하는 맞춤형 비식별화 기록물 제공 기능이 구현됨을 확인하였다.

4.2.2. 향후 발전방향

○ 공개재분류 지식정보 관리시스템 구축

- 향후 공개재분류 지식정보를 벡터DB화(조건 인지형 RAG 구축)할 때에는, 문맥적 유사도가 과도하게 높지 않도록 문장 구성을 조정하고 정비하여 체계적으로 관리할 필요가 있으며, 현재의 “기준서 검색시스템”을 고도화하여 공개재분류 지식정보 관리시스템으로 발전시킴으로써, AI 모델이 해당 지식정보를 정확하게 활용하여 판단 할 수 있도록 관리체계를 구축하는 것이 필요하다.

○ 심의자료 작성 및 재분류 사유 자동생성 등 생성형 AI 기능 적용

- 본 과제에서는 재분류 기준서 작성, 심의자료 작성 등 생성형 AI를 활용한 문서 자동화 기능을 추가적으로 PoC 환경에서 구현·검증하였다. LLM을 활용하면 문서 요약, 심의자료 작성, 재분류 사유 작성 등이 적정한 프롬프트 설정만으로 자동 생성될 수 있으며, 이는 실제 업무에서도 활용 가능한 수준의 품질을 보였다. 향후 이러한 기능을 업무 시스템에 통합하여 심의자료 및 사유 작성의 효율성을 높이는 방향으로 개발할 필요가 있다.

❍ 업무의 특성에 맞도록 멀티 LLM 오케스트레이션(Multi-LLM Orchestration) 적용

- 본 과제에서 AI 모델간 특징이 일부 확인되었으며, 문서요약, 판단, 문맥이해, 비식별화 대상 정보 검출 등 업무의 특성에 맞게 다양한 AI 모델을 적용하는 멀티 LLM 오케스트레이션을 소규모로 적용해 보았다.
- 다양한 LLM을 시스템에 탑재하고, LLM 모델의 특성을 분석하고 활용하면 공개재분류에 적합한 최적의 AI Agent 를 구성할 수 있을 것이다.

❍ 최신 AI 모델의 적용

- 연구과제를 수행하는 짧은 기간에도 AI 기술은 급속도로 발전하고 있었으며, 행정망과 같은 내부망에서 구동이 가능한 최신 LLM 모델 또한, 지속적으로 발표되어, 과제수행 시 방향성이 자주 수정되었다.
- 이러한 점을 감안하면, 실제 시스템 구축시에는 고성능의 LLM 모델로 인해, 현재 LLM의 한계를 충분히 극복할 수도 있고 더 많은 분야의 업무에 적용 가능할 것이다.

제5장 총괄연구개발과제의 연구성과

5.1. 활용성과

총괄과제명	AI를 활용한 비공개기록물 공개재분류 및 비식별화 서비스 모델 연구
총괄과제책임자	손기영 / ㈜가온아이 / 기록물관리학

가. 연구논문

번호	논문제목	저자명	저널명	집(권)	페이지	Impact factor	국내/국외	SCI여부
1	해당없음							

나. 학술발표

번호	발표제목	발표형태	발표자	학회명	연월일	발표지	국내/국제
1	해당없음						

다. 지적재산권

번호	출원/등록	특허명	출원(등록)인	출원(등록)국	출원(등록)번호	IPC분류
1	출원	문서 공개유형 판정 방법 (Method for determining document disclosure type)	국가기록원	대한민국	10-2025-0155366	

라. 정책활용

※ 기타 관련정책에 활용 예를 구체적으로 기술함.

마. 타연구/차기연구에 활용

※ 타연구 및 차기연구에 활용된 예를 구체적으로 기술함.

바. 언론홍보 및 대국민교육

※ 언론홍보 및 대국민교육 내용, 일자 등을 간략히 기술함.

사. 기타

※ 임상시험 , 관련 DB구축, 워크샵 또는 심포지움 개최 등의 경우 구체적으로 기술함.

5.2. 활용계획

가. AI를 활용한 공개재분류 지원 모델

AI를 활용한 지능형 공개재분류 수행 모델을 구현하고 그 성능을 검증하였다. 이를 통해 모델의 재분류 정확도, 지식데이터 정비 및 구축 방안, 실제 시스템 구현 시 고려해야 할 요소 등을 도출하였다. 특히 문서 유형별 재분류 절차를 조건 인지형 RAG(Relevant Augmented Generation) 기반으로 설계하여 구현하였으며, AI 모듈을 활용해 해당 절차의 타당성과 적용 가능성을 검증하였다. 본 연구 결과는 향후 실제 업무 프로세스에 본 모델을 적용하기 위한 활용성 검토의 기초자료로 활용될 수 있을 것이다.

나, 이용자 맞춤형 비식별화 기록물 제공

비공개대상정보 자동식별·마스킹 모델을 개발하여 그 적용 가능성을 검증하였다. 해당 모델은 기존 정규식 기반 탐지 방식의 한계를 극복하기 위해 정규식, 규칙 기반 탐지, 문맥 분석, LLM 기반 검증을 결합한 하이브리드 검출 구조를 적용하였다. 이를 통해 비식별화 대상 정보를 보다 정밀하게 식별하고, 식별된 정보를 자동으로 마스킹하여 대국민 열람 서비스에 적용 가능한 수준의 비식별 처리 성능을 확보하였음을 확인하였다. 본 연구 결과는 향후 실제 업무 프로세스에 본 모델을 도입하기 위한 검토 자료로 활용될 수 있을 것이다.

제6장 기타 중요변경사항

・해당사항 없음

제7장 참고문헌

개인정보보호위원회. (2024). 가명정보 처리 가이드라인. 출처 : https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do

공공기관의 정보공개에 관한 법률. 법률 제14839호.

공공기록물 관리에 관한 법률. 법률 제20309호.

국가기록원. (2010). 30년경과 비공개기록물 공개재분류 가이드북 : 특수기록관 기록물. 출처 : https://www.archives.go.kr/next/newsearch/showDetailPopup.do?rc_code=1310377&rc_rfile_no=201104664664&rc_ritem_no=&sitePage=1-1-1

국가기록원. (2020). 전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 적용방안 연구. 출처 : https://www.archives.go.kr/next/newdata/researchReportDetail.do?keyword=&page=3&reportType=&searchType=&seq=98340&year

국가기록원. (2021a). 기록관리 AI 기술 적용을 위한 공통 학습데이터 세트 구축 연구. 출처 : https://www.archives.go.kr/next/newdata/researchReportDetail.do?seq=98398&page=2&reportType=&year=&searchType=

국가기록원. (2021b). 전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 적용 방안 연구. 출처 : https://www.archives.go.kr/next/newdata/researchReportDetail.do?keyword=&page=2&reportType=&searchType=&seq=98381&year

국가기록원. (2024a). 공공기관 기록물관리 지침.

국가기록원. (2024b). NAK 16-2:2024(v1.2) 기록물 공개관리 업무-제2부 : 영구기록물관리기관(v1.2).

국가기록원. (2025). 2025년 기록물관리 지침.

국가기록원 서비스정책과. (2025). 제2회 기록서비스전문위원회 : AI를 활용한 공개재분류 및 비식별화 서비스 모델 연구.

권미현. (2019). 비공개기록물 공개재분류 업무절차 개선 방안. 제8차 기록관리 연구세미나. 출처 : https://www.archives.go.kr/next/common/downloadBoardFile.do?board_seq=96624&board_file_seq=1

김태영, 강주연, 김건 외. (2018). 지능형 기록정보서비스를 위한 선진 기술 현황 분석 및 적용방안. 한국기록관리학회지, 18(4), 149-182. http://dx.doi.org/10.14404/JKSARM.2018.18.4.149

김포시 기록관. (2023). 기록물 디지털 전환 사업 보고서

김해찬솔, 안대진, 임진희 외. (2017). 기계학습을 이용한 기록 텍스트 자동분류 사례 연구. 정보관리학회지, 2017(4), 321-344. https://doi.org/10.3743/KOSIM.2017.34.4.321

대통령기록관. (2024). 대통령기록관리 AI 도입 방안 정책 연구 결과보고서.

박찬준, 이원성, 김윤기 외. (2023). 초거대 언어 모델 연구 동향. 정보과학회지. 41(11). 8-24.

송주형. (2021). 지능형 아카이브 솔루션을 활용한 공개재분류 연구 : 2020년 국가기록원 공개재분류 사업을 중심으로. 한국기록관리학회지. 21(4). 101-115. https://doi.org/10.14404/JKSARM.2021.21.4.101

원동규, 이상필. (2016). 인공지능과 제4차 산업혁명의 함의. ie 매거진. 72. 13-22.

원병철. (2025. 7. 11.). [2025 비식별 솔루션 리포트] 사용자의 '필요성'과 컴플라이언스 강화의 '강제성'이

시장 키운다. 보안뉴스. 출처 : https://www.boannews.com/media/view.asp?idx=138082
이재용. (2022). 데이터기반행정 정착을 위한 요인 및 방안 연구 : 기초자치단체를 중심으로. 한국지방행정학회보, 19(1), 23-47. http://dx.doi.org/10.32427/klar.2022.19.1.23
임을영, 현민성. (2025). 생성형 인공지능을 활용한 판결문 공개 제도 개선 : 개인정보 보호와 알 권리 균형을 위한 법적·기술적 접근. 지식재산연구, 20(2), 93-118.
임철홍. (2023). LLM(Large language Model) 속성과 성능 연관성 연구. 정보화 연구, 20(3), 257-266. http://dx.doi.org/10.22865/jita.2023.20.3.257
장나은. (2024). 검색증강생성(RAG) 기술의 등장과 발전 동향. Digital Insight. 2024(4). 1-39.
정보공개포털. (발행년불명). https://www.open.go.kr/infOthbc/numberInfoOthbc/orgDwld.do
주현우. (2019). 현용기록의 활용성 증진을 위한 지능형 기록관리시스템 구축 : 한국중부발전 사례중심으로. 한국기록관리학회지. 19(4). 221-230.
최수진. (2023). “AI, 1940년대 시작됐다고?” ... 인공지능 역사 돌아보니. 한경Business. 1420. 25.
한국지능정보사회진흥원. (2023). 국가기록물 대상 초거대 AI 학습용 말뭉치데이터 구축 사업.
한국철도시설공단. (2018). 차세대 기록관리시스템 구축 제안요청서.
행정안전부. (2016). 개인정보 비식별 조치 가이드 라인. 출처 : https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=55287
행정안전부 (2023). 공공기록물 개인정보 자동 비식별화 시범사업 보고서.
AI Hub. (2023). 국가기록물 대상 초거대 AI 학습을 위한 말뭉치 데이터. 출처 : https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=71788
A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam. (2006). L-diversity: privacy beyond k-anonymity. 22nd International Conference on Data Engineering, 24. https://doi.org/10.1109/ICDE.2006.1.
A State Archives and Records initiative for the NSW Government. (2017). Machine Learning and Records Management. 출처 : https://futureproof.records.nsw.gov.au/machine-learning-and-records-management/
AWS. (발행년불명). 벡터 데이터베이스 개요. 출처 : https://docs.aws.amazon.com/ko_kr/prescriptive-guidance/latest/choosing-an-aws-vector-database-for-rag-use-cases/vector-databases.html
David Canning, Lise Jaillant. (2025). AI to review government records: new work to unlock historically significant digital records. AI&Society, 40, 4433-4445. https://doi.org/10.1007/s00146-025-02221-0
Georgios Feretzakis, Konstantinos Papaspyridis, Aris Gkoulalas-Divanis 외. (2024). Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review. Information 2024, 15(11), 697. https://doi.org/10.3390/info15110697
Giles Hooker, Cliff Hooker. (2017). Machine Learning and the Future of Realism. Spontaneous Generations A Journal for the History and Philosophy of Science. 9(1). https://doi.org/10.48550/arXiv.1704.0468
Giovanni Colavizza, Tobias Blanke, Charles Jeurgens 외. (2021). Archives and AI : An overview of Current Debates and Future Perspectives. Journal on Computing and Cultural Heritage, 15 (1), 1 - 15 https://doi.org/10.48550/arXiv.2105.01117
Government of New Brunswick. (2025). Classification Plan and Retentiosn Schedules for Common Records. 출처 : https://www.gnb.ca/en/gov/information-management/classification-plans/classification-plan.html
Gregory Rolan, Glen Humphries, Lisa Jeffrey 외. (2019). More human than human? Artificial intelligence in the archive. Archives and manuscripts. 47(2). 179-203. https://orcid.org/0000-0001-5

891-3732.

Jacob Devlin, Ming-Wei Chang, Kenton Lee 외. (2018). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.0480 https://doi.org/10.48550/arXiv.1810.04805

J. McCarthy, M. L. Minsky, N. Rochester 외. (1955). A proposal for the Dartmouth summer research project on artificial Intelligence.

L. Sweeney. (2002). k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 557-570.

Le Yang, Miao Tian, Duan Xin 외. (2024). AI-Driven Anonymization: Protecting Personal Data Privacy While Leveraging Machine Learning. arXiv:2402.17191, https://doi.org/10.48550/arXiv.2402.17191

Microsoft. (2023). AI-powered Data Classification | Microsoft Purview. 출처 : https://techcommunity.microsoft.com/blog/microsoftmechanicsblog/ai-powered-data-classification－microsoft-purview/3919206?

Naser F. Aldoseri, Wael Elmedany. (2023). Strengthening Data Security in Bahrain : Leveraging Microsoft Purview to Prevent Leakage of Sensitive Information. IET Conference Proceedings. 2023(44). 477-483. https://doi.org/10.1049/icp.2024.0971

Patric Lewis, Ethan Perez, Aleksandra Piktus 외. (2020). Retrieval Augmented generation for knowledge-intensive NLP tasks. NeurIPS 2020. https://doi.org/10.48550/arXiv.2005.11401

Professor John McCarthy, http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

Sungen Hahm, Heejin Kim, Gyuseong Lee, Hyunji Park 외. (2025). Thunder-DeID: Accurate and Efficient De-identification Framework for Korean Court Judgments. arXiv:2506.15266. https://doi.org/10.48550/arXiv.2506.15266

The National Archives. (2016). The application of techonology-assisted review to born-digiral records transfer, Inquiries and beyond. 출처 : https://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/reviewing-digital-records-management-government/research/

U.S. Department of State. (2023). Piloting Machine Learning for Freedom of Information Act(FOIA) Requests : FOIA Advisory Committee Meeting. 출처 : https://www.archives.gov/files/ogis/foia-advisory-committee/2022-2024-term/foia-advisory-committee-presentation-piloting-machine-learning-for-foia-09072023.pdf

UK National Archives. (2022). Digital Transfer and Sensitive Data Project Report.

Zhiheng Huang, Wei Xu, Kai Yu. (2015). Bidirectional LSTM-CRF Models for Sequence Tagging. arXiv:1508.01991 https://doi.org/10.48550/arXiv.1508.01991

Choi, J., & Park, H. (2022). AI-based Privacy Information Detection in Government Records Management. Archives and Society, 12(2), 45 - 67.

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF Models for Sequence Tagging. arXiv preprint arXiv:1508.01991.

Ferrández, Ó., South, B. R., Shen, S., Friedlin, F. J., Samore, M. H., & Meystre, S. M. (2013). BoB, a best-of-breed automated text de-identification system for VHA clinical documents. Journal of the American Medical Informatics Association, 20(1), 77-83. DOI:10.1136/amiajnl-2012-001020.

Meystre, S. M., Ferrández, Ó., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2014). Text de-identification for privacy protection: a study of its impact on clinical text information content. Journal of Biomedical Informatics, 50, 142-150.

제8장 첨부서류

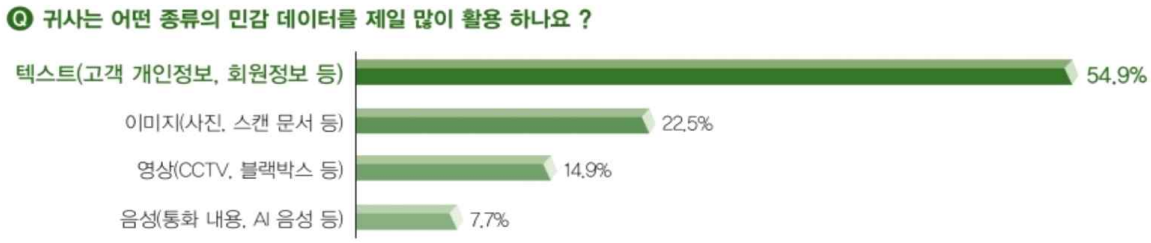
8.1. 기술협상 추진내역

협상내용	수행여부	추진내용
(사업 수행 장소) - 기술동향, 사례 조사, Poc 개발은 외부에서 수행 가능. 단, Poc를 위한 데이터 처리 이후는 수요기관 내에서 수행	완료	- 국가기록원 성남분원 2층에 사업장 마련 (25.06.27 ~ 현재)
(보안) - 처리 데이터 및 사업결과물 일체는 대외 반출 엄금함 - 사업 종료 후 모두 기관에 제출 후 삭제 조치	시기미도래	- 사업종료시 보안규정에 따라 데이터 삭제 예정
(성과지표 및 목표치) - 수행사 제안 내용대로 유지	완료	- “3.6.3. 성과지표 측정” 참조
(시범 운영기간) - 1개월 이상 제공	진행중	- 시범운영서버 설치완료 - 25.11.3 ~ 운영중 1개월 이상유지 예정
(협의체 구성・운영) - 수요기관과 협의하여 관련 분야 교수, 기술진 등으로 구성 - 운영 시기를 데이터 검수만이 아닌 모델 설계시에도 참여하도록 확대 운영(2회 이상)	완료	- 협의체 회의 3회 추진 - 6/4, 7/2, 8/26 - 붙임1 회의록 참조
(데이터 수집 및 정제방안) - 검출된 개인정보를 사망자 개인정보와 비교 연계 수행방안 제시	완료	- “3.3.6 주민등록정보 연계를 통한 사망자 여부 판별 서비스 구축 방안” 참조
(논문 또는 특허출원) 택 1 - 과제종료 후 다음해 까지 KCI급 논문 게재 - 과제종료 후 다음 해까지 특허출원(심사청구 포함)	진행중	- 특허 출원 진행중 (가출원 완료) - 붙임2 특허출원통지서 참조
(세미나·포럼) - 국가기록원 주관 세미나· 포럼에 해당 과제 발표(국가기록원 요청 시) - 수요기관 발표시 실적 데이터 산출 및 자료 작성 지원	시기미도래	- 12/15일 국가기록원 주관 세미나 추진 예정
(투입인력요건) - 전문인력 요건 조정 (참여연구자 경력 등을 고려하여 학위기준 완화: 석사→학사 이상)	완료	- 인력변경시 완화된 요건을 반영하여 변경요청 하였음
(추진 상황 보고) - 정기보고는 주, 월간으로 실시 - 필요시 수시보고 수행	완료	- 착수/중간보고 수행 - 월간보고 5회 - 붙임3 보고자료 참조

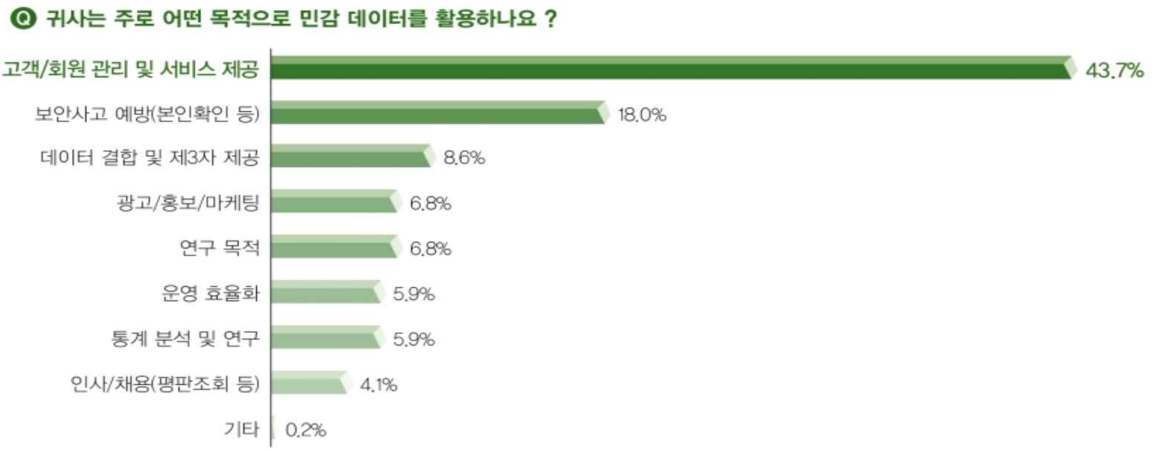
8.2. 개인정보 비식별화 국내 솔루션 설문서

보안전문가들의 비식별 솔루션에 대한 설문조사를 2025년 7월1일부터 4일까지 약 10만명의 보안전문가들에게 설문조사를 진행했다.

우선 보안전문가들은 어떤 종류의 민감 데이터를 제일 많이 활용하는지를 물어봤다. 응답자의 절반 이상인 54.9%가 '텍스트(고객 개인정보, 회원정보 등)'라고 답해 아직 텍스트 형태의 민감 데이터가 가장 많은 것으로 조사됐다.[31)]

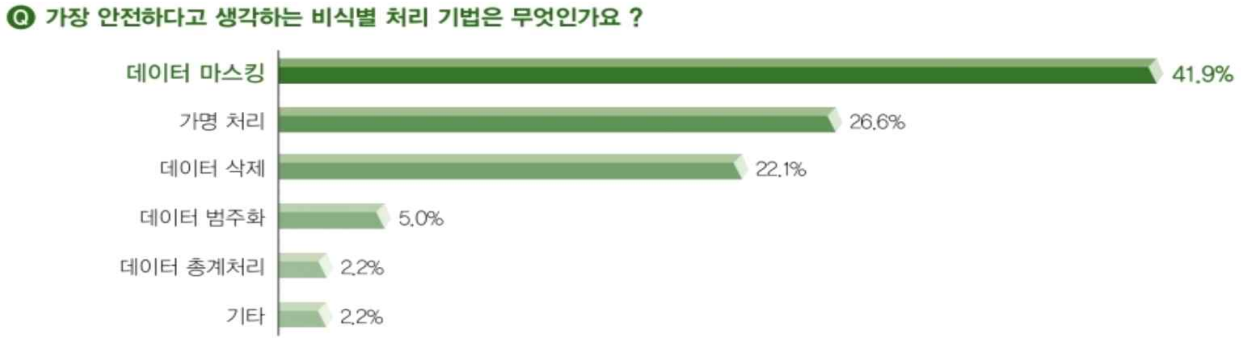


그렇다면 응답자들은 어떤 목적으로 민감 데이터를 활용할까? '고객/회원 관리 및 서비스 제공'을 선택한 사람이 43.7%였으며, '보안사고 예방'을 선택한 사람은 18.0%였다. 이어 '데이터 결합 및 제3자 제공' 8.6%, '광고/홍보/마케팅' 6.8%, '연구 목적' 6.8%, '운영 효율화' 5.9%, '통계 분석 및 연구' 5.9%, '인사/채용(평판조회 등)' 4.1% 순으로 조사됐다.

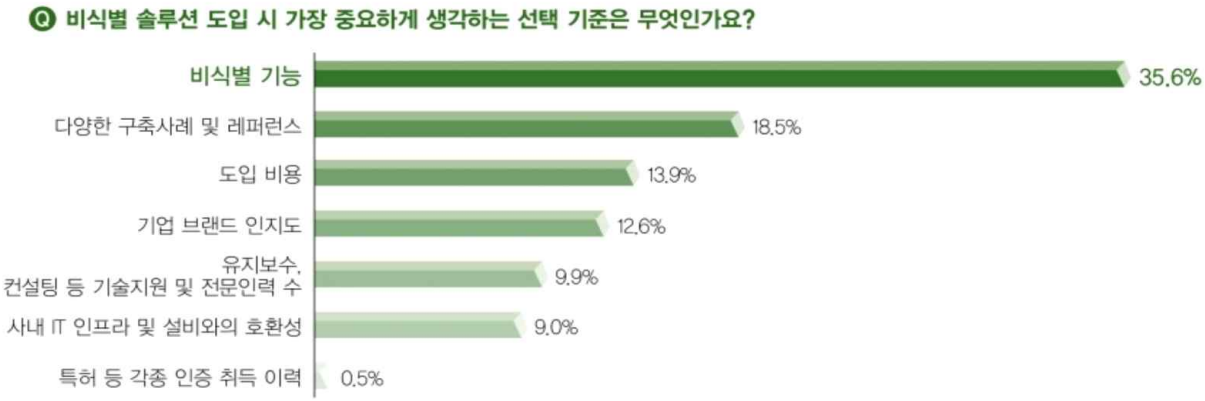


사용자들이 가장 많이 사용하는 비식별 처리 기법은 무엇일까? 응답자의 절반 이상인 51.8%는 '데이터 마스킹'이라고 답했으며, 24.3%는 '가명 처리'라고 답했다. 이어 13.9%가 '데이터 삭제'를 6.8%가 '데이터 범주화'를 택했으며, 3.2%가 '데이터 통계처리'를 선택했다.

31) 보안선문가들의 비식별 솔루션에 대한 설문조사(2025.7.11._[원병철 기자(boanone@boannews.com)]



비식별 솔루션 도입시 가장 중요하게 생각하는 선택 기준에 대해 물어보자 35.6%의 응답자는 '비식별 기능'이라고 답했다. 이어 18.5%가 '다양한 구축사례와 레퍼런스'라고 답했으며, 13.9%는 '도입비용'이라고 답했다. 12.6%는 '기업 브랜드 인지도'를, 9.9%는 '유지보수, 컨설팅 등 기술지원과 전문인력 수'를, 9.0%는 '사내 IT 인프라 및 설비와의 호환성'을 각각 골랐다.



편집순서 8 : 연구과제요약

총괄 연구과제 요약

과제 고유번호	자동부여		공개가능여부	공개
사업명	2025 국가기록관리 활용기술 연구개발 사업			
과제명	AI를 활용한 비공개기록물 공개재분류 및 비식별화 서비스 모델 연구			
연구책임자	성 명	손기영		
	소속 기관명	㈜가온아이		

○ 연구목표

- ❍ AI를 활용한 지능형 공개재분류 수행 모델 연구 개발
- ❍ 비공개대상정보의 자동식별 및 마스킹을 통한 대국민 열람 제공 모델 연구 개발
- ❍ 국가기록원의 공개재분류 업무에 가장 적합한 AI 모델 연구
- ❍ AI 공개재분류 및 비공개대상정보 비식별화 모델을 실무에 적용하기 위해 작업 도출
- ❍ AI 공개재분류 및 비공개대상정보 비식별화 모델의 기계 학습을 위해서는 기록 정보 데이터 구성 연구
- ❍ 각급 기록물관리기관으로 확산 및 보급하기 위한 과제 연구

○ 연구내용

○ AI 기반 공개재분류 업무 지능화
- 수동으로 이루어지는 공개재분류 업무 과정을 AI기술을 활용하여 효율성을 높이고 자동화하는 모델을 개발
- AI가 기록물의 내용을 분석하여 공개 가능성을 판단하고, 관련정보를 제공하여 담당자의 의사 결정을 지원

○ AI기반 비공개 대상 정보 식별 및 비식별화
- 기록물 내에 포함된 개인 정보, 보안 정보 등 비공개 대상 정보를 AI가 자동으로 탐지하고 식별하는 기술을 개발
- 식별된 비공개 대상 정보를 안전하게 가리고 익명화하는 비식별화 기술을 개발

○ 이용자 맞춤형 비식별화 기록물 제공
- 비식별화 처리된 기록물을 사용자의 요구에 맞춰 제공할 수 있는 기능을 개발
- 국민들이 필요한 정보를 안전하게 접근하고 활용 할 수 있도록 지원

○ PoC(Proof of Concept) 개발 및 검증
- 개발된 AI기반 기능들을 통합하여 실제와 유사한 환경에서 작동하는 PoC시스템을 구축
- PoC시스템을 시범운영하여 개발된 모델의 실효성과 기술적인 기능성을 검증

○ 기록관리시스템 연계 및 확산 보급
- 개발된 AI기능을 기록관리시스템에 통합할 수 있는 기술적인 방안을 제시
- 개발된 서비스 모델과 시스템을 다른 기록물관리기관에도 적용하고 확산시킬 수 있는 전략 제시

○ 연구성과(응용분야 및 활용범위포함)

“AI를 활용한 지능형 공개재분류 수행 모델”을 모두 구현·검증하였으며, 정확도와 처리속도 등 핵심 지표에서 목표치를 달성하였다. 특히 문서 유형별 재분류 절차를 RAG시스템과 AI 모듈로 표준화하여 업무 효율을 제고하고, CAMS시스템 연계 검토 및 대외기관 확산방안을 통해 현장 적용 가능성을 입증하였다. 비공개대상정보 자동식별·마스킹 모델은 개인정보·영업비밀 등 주요 카테고리의 정밀 탐지 및 내용 기반 마스킹을 실현하였고, 대국민 열람 서비스에 필요한 품질 수준을 달성하였다. 그 결과, 대국민 열람 서비스까지 연계하여 실제 업무 흐름과의 통합·운영성을 검증하였다.

○ 총괄 참여연구원

성 명	소속	성 명	소속
손기영	가온아이	조현수	가온아이
이남협	가온아이	심보용	가온아이
하동훈	가온아이	석수민	가온아이
진상용	가온아이	이재선	가온아이
조창제	가온아이	신재훈	가온아이
김원섭	가온아이	최돈선	가온아이
이동호	가온아이	한덕민	가온아이

Keywords (5개 내외)	***한글***	비공개 기록, 공개재분류, 공개, 개인정보 비식별화
	영문	Restricted Government Records, Access Reclassification, Public Disclosure, De-identification, Redaction

주1) 연구목표, 연구내용, 연구성과를 서술형으로 기재

2) 국가연구개발사업 DB를 통한 공개를 희망하지 않는 경우 공개가능여부란에 "공개불가"로 표시

3) 연구성과는 그간의 연구결과 및 기대성과를 서술