



[기술 백서: 생성형 AI가 잘 이해하는 데이터를 만드는 법]

공공 행정 혁신을 위한 데이터 중심 접근 전략

[기술 백서] 생성형 AI 가 잘 이해하는 데이터를 만드는 법

부제: 광역 및 기초 자치단체, 그리고 공공기관과 공기업에서 AI 를 위한 올바른
데이터를 만드는 방법

존경하는 공공부문 관계자 여러분, 안녕하십니까.

스캐터엑스(ScatterX) 대표이사 강명수입니다.

오늘날 생성형 AI 기술은 행정의 효율성을 혁신할 강력한 도구로 주목받고 있습니다.
하지만 실제 현장에 계신 여러분께서는 보안 문제, 데이터의 정확성, 그리고 도입 시
발생할 업무 가중에 대해 기대보다는 깊은 고민을 안고 계시리라 생각합니다.

저희 스캐터엑스는 **"AI 도입의 성패는 거창한 모델(Model)이 아니라, 기본이 되는
데이터(Data)에 달려 있다"**는 믿음을 가지고 있습니다. 아무리 뛰어난 AI 라도 정제되지
않은 데이터가 입력되면 결코 신뢰할 수 있는 답변을 내놓을 수 없기 때문입니다.

현재 스캐터엑스는 감사하게도 **2 만 명의 공직자분들이 직접 사용하고 계신 'AI
행정비서 SaaS'**를 운영하며, 현장의 목소리를 가장 가까이에서 듣고 있습니다. 저희는
단순히 솔루션을 공급하는 입장이 아닙니다. 오히려 공직자 여러분의 따뜻한 관심과
때로는 따끔한 조언을 바탕으로 기술을 다듬어가고 있으며, **'AI 가 진정으로 잘
이해하는 데이터'가 무엇인지 여러분과 함께 고민하고 정의해 나가는 과정**에 있습니다.

이러한 현장의 치열한 고민과 배움을 바탕으로, 공공 데이터가 AI 와 만났을 때 발생할
수 있는 문제들을 해결하기 위한 기술적, 실무적 해법을 정리하여 **[기술 백서: 생성형
AI 가 잘 이해하는 데이터를 만드는 법]**을 발간하게 되었습니다.

본 백서에서는 다음과 같은 내용을 조심스럽게 제언합니다.

1. **공공 특화 데이터 정제:** 보안(Security) 필터링부터 메타데이터 추출, 관계 연결에 이르는 과정을 통해 보안과 품질을 동시에 확보하는 방안을 담았습니다.
2. **행정 맞춤형 기술:** 법령의 개정 이력을 반영하는 'Time-Aware RAG', 모호한 질문의 의도를 파악하는 'HyDE', 답변의 근거를 스스로 되묻는 'Self-RAG' 등 행정의 정확성을 높이는 기술을 소개합니다.
3. **현실적인 도입 로드맵:** 망분리 환경을 고려한 온프레미스 sLLM 구축부터, 단순 검색을 넘어 실질적인 업무를 돕는 'AI 에이전트'로의 진화 단계를 제시합니다.

저희는 공직자 여러분께서 반복적인 업무는 AI 에게 맡기고, 데이터 중심의 의사결정과 창의적인 정책 기획이라는 본연의 가치에 집중하시기를 소망합니다. 스캐터엑스가 그 여정에 작은 보탬이 되는 든든한 기술 파트너가 되겠습니다.

바쁘신 업무 중에 잠시 시간을 내어 일독해 주신다면, 미래 행정 혁신을 위한 의미 있는 통찰을 얻으실 수 있을 것입니다.

감사합니다.

스캐터엑스 대표이사 강명수 올림

****Part 1. 공공 데이터의 현실과 구조화 전략****

(The Foundation: Understanding Public Data Reality & Engineering Strategy)

****제 1 장. 서론: 공공 부문 레거시 데이터와 AI의 간극****

1.1. 공공 데이터의 현실: 데이터 사일로, 비정형 문서, 그리고 보안 규제

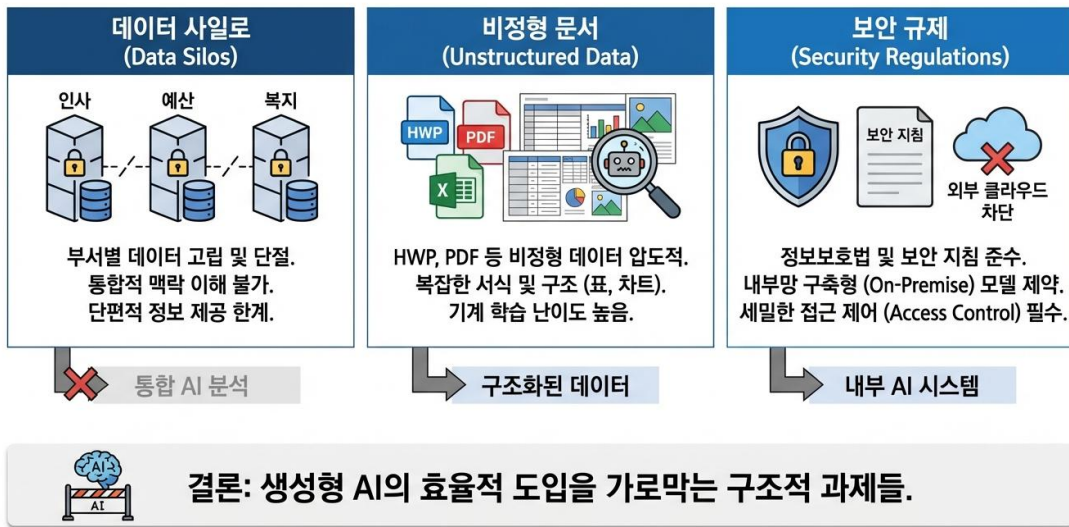
대한민국의 공공 행정은 지난 수십 년간 수많은 공직자 여러분의 헌신과 혁신적인 노력 덕분에 세계 최고 수준의 전자정부 시스템을 구축해 왔습니다. 우리는 이미 문서를 전자화하고, 결재를 시스템으로 처리하며, 민원을 온라인으로 해결하는 디지털 행정의 기반을 닦았습니다. 하지만 최근 전 세계적인 화두가 되고 있는 '생성형 AI(Generative AI)' 기술을 우리 행정 현장에 도입하고자 할 때, 우리는 몇 가지 구조적인 어려움에 직면하게 됩니다. 이는 데이터가 부족해서가 아니라, 데이터가 존재하는 형태와 관리 방식이 최신 AI 기술이 요구하는 특성과 다소 차이가 있기 때문입니다.

첫째, **데이터 사일로(Data Silo)** 현상을 조심스럽게 진단해 볼 필요가 있습니다. 각 부서와 기관은 고유의 업무 목적에 최적화된 시스템을 통해 데이터를 철저히 관리하고 계십니다. 그러나 AI가 조직 전체의 업무 맥락을 이해하고 종합적인 답변을 제공하기 위해서는, 부서별로 나뉘어 있는 인사, 예산, 복지, 인허가 등의 데이터가 유기적으로 연결될 필요가 있습니다. 현재의 칸막이식 데이터 구조는 AI가 통합적 사고를 하는 데 있어 보이지 않는 장벽이 되고 있습니다. 예를 들어, "우리 시의 청년 일자리 예산과 실제 취업률의 상관관계"를 AI에게 묻고자 할 때, 예산 데이터와 고용 통계 데이터가 물리적으로, 논리적으로 단절되어 있다면 AI는 단편적인 정보만 나열할 뿐 통찰력 있는 답변을 내놓지 못할 것입니다.

둘째, **비정형 데이터(Unstructured Data)**의 비중이 압도적으로 높다는 점입니다. 공공 부문에서 생산되는 자료의 상당수는 한글(HWP), PDF, 엑셀 파일 등의 문서 형태입니다. 이러한 문서들은 사람이 읽고 이해하기에는 매우 훌륭하게 작성되어 있으나, 기계가 그 구조와 의미를 정확히 파악하기에는 다소 복잡한 서식을 포함하고 있는 경우가 많습니다. 표(Table) 안에 또 다른 표가 있거나, 이미지가 포함된 차트, 문맥을 알아야만 해석 가능한 각주 등은 LLM(거대언어모델)이 정보를 학습하는 데 있어 높은 난이도를 요구합니다. 특히 대한민국 행정의 표준인 HWP 포맷은 텍스트 추출과 구조 분석에 있어 별도의 기술적 전처리가 필수적입니다.

셋째, **엄격한 보안 규제**는 타협할 수 없는 가치이지만, AI 도입에 있어 신중한 접근을 요구하는 가장 큰 요인입니다. 정보공개법, 개인정보보호법, 그리고 각급 기관의 보안 지침은 반드시 준수되어야 합니다. 민간 기업처럼 외부 클라우드 기반의 최신 LLM API(예: GPT-4 등)를 자유롭게 활용하기 어려운 환경 속에서, 내부망(Intranet)에 구축형(On-Premise) 모델을 도입해야 하는 제약이 있습니다. 더불어 데이터의 보안 등급(공개, 대외비, 비밀)에 따른 세밀한 접근 제어(Access Control)가 데이터 전처리 단계에서부터 선행되지 않으면, AI가 권한 없는 사용자에게 민감한 정보를 노출하는 보안 사고로 이어질 수 있습니다.

1.1. 공공 데이터의 현실: 생성형 AI 도입의 장벽



1.2. LLM의 한계와 도전: 할루시네이션(Hallucination)과 최신성 부족 문제

공직자 여러분께서 AI 도입을 검토하실 때 가장 우려하시는 부분 중 하나는 바로 **'할루시네이션(Hallucination, 환각)'** 현상일 것입니다. 생성형 AI는 기본적으로 다음에 올 단어를 확률적으로 예측하는 모델입니다. 따라서 사실이 아닌 내용을 마치 사실인 것처럼, 매우 유창하고 그럴듯한 문장으로 답변하는 경우가 종종 발생합니다.

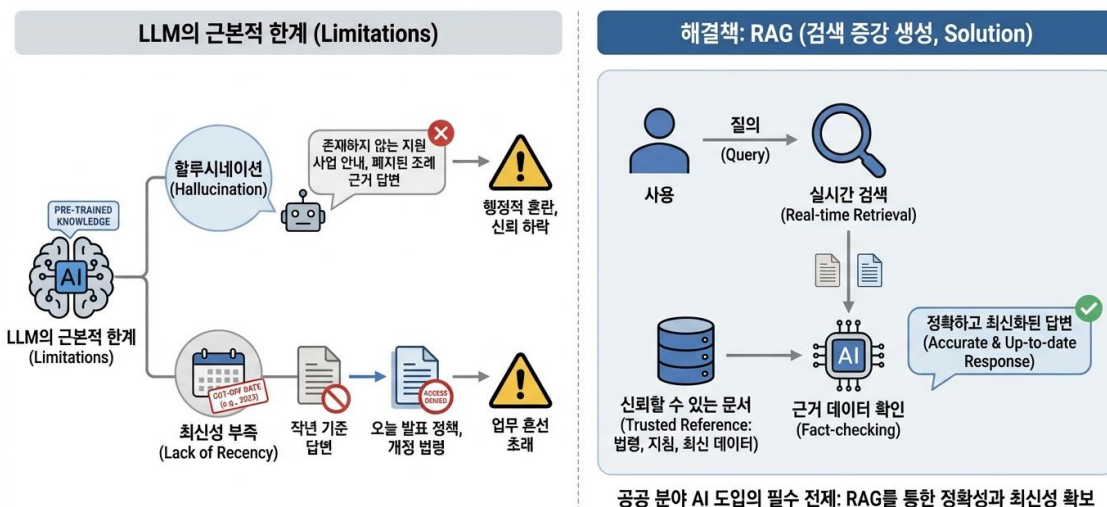
정확성과 신뢰성, 그리고 법적 책임이 따르는 공공 행정 업무에서, "존재하지 않는 지원 사업을 안내"하거나 "폐지된 조례를 근거로 답변"하는 일은 결코 용납될 수 없습니다. 민원인에게 잘못된 정보를 제공했을 때 발생할 수 있는 행정적 혼란과 신뢰 하락은 기술 도입의 효용을 상쇄하고도 남습니다.

또한, **최신성(Recency) 부족 문제**도 간과할 수 없습니다. LLM은 사전에 학습된(Pre-trained) 데이터만을 기반으로 지식을 형성합니다. 모델 학습이 종료된 시점(Cut-off date) 이후에 개정된 법령, 오늘 아침에 발표된 새로운 정책 지침, 실시간으로 변동되는 예산 현황 등은 AI가 알지 못합니다. 급변하는 행정 환경 속에서 "작년 기준"으로

답변하는 AI 는 실무에 큰 도움이 되지 못할뿐더러, 오히려 업무 혼선을 초래할 수 있습니다.

따라서 우리는 AI 모델 자체의 성능(파라미터 수 등)을 높이는 경쟁보다는, AI 에게 '정확하고 최신화된 근거 데이터'를 실시간으로 제공하는 것에 더욱 집중해야 합니다. AI 가 자신의 기억(학습 데이터)에 의존하여 상상해 내는 것이 아니라, 우리가 제공한 신뢰할 수 있는 문서(Reference)를 보고 답을 찾도록 만드는 기술, 즉 **RAG(Retrieval-Augmented Generation, 검색 증강 생성)**가 공공 분야 AI 도입의 필수 전제 조건이 되는 이유가 바로 여기에 있습니다.

1.2. LLM의 한계와 도전: 할루시네이션(Hallucination)과 최신성 부족 문제



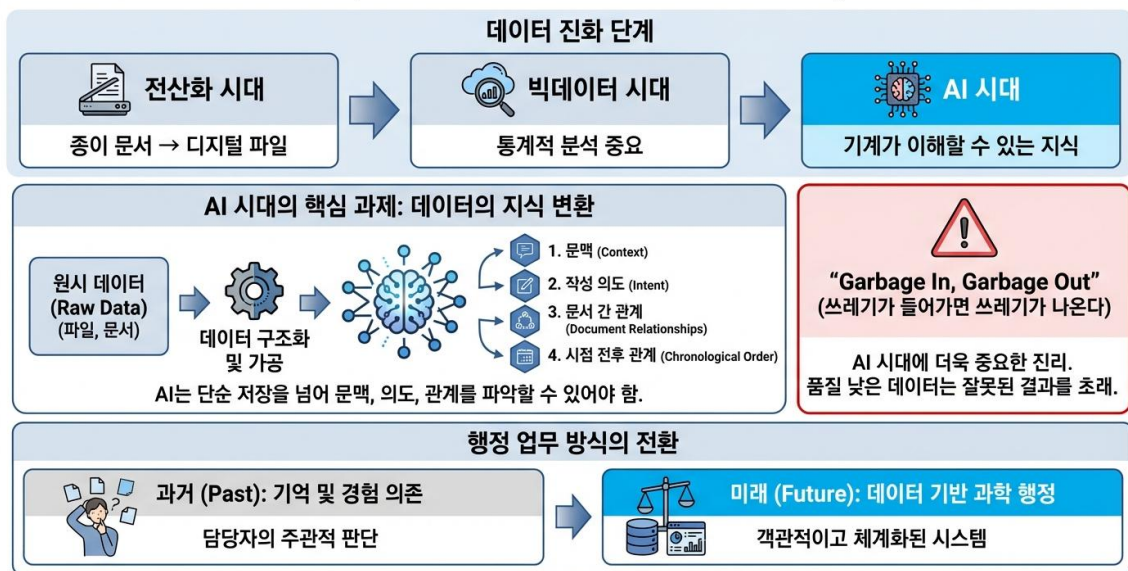
1.3. 데이터 중심 행정(Data-Driven Administration)으로의 전환

그렇다면 AI 가 잘 활용할 수 있는 데이터란 무엇일까요? 과거 전산화 시대에는 종이 문서를 디지털 파일로 만드는 것이 목표였습니다. 빅데이터 시대에는 데이터를 통계적으로 분석하는 것이 중요했습니다. 이제 AI 시대에는 데이터를 '기계가 이해할 수 있는(Machine-Understandable) 지식'으로 변환하는 것이 핵심입니다.

단순히 파일을 서버에 저장해 두는 것만으로는 부족합니다. AI가 문서의 제목뿐만 아니라 문맥, 작성 의도, 문서 간의 관계, 시점의 전후 관계 등을 파악할 수 있도록 데이터를 가공해야 합니다.

"Garbage In, Garbage Out(쓰레기가 들어가면 쓰레기가 나온다)"이라는 격언은 AI 시대에 더욱 뼈아픈 진리입니다. 우리가 지금 수행하려는 데이터 구조화 작업은 단순히 AI 시스템 하나를 구축하기 위함이 아닙니다. 이는 담당자의 기억과 경험에 의존하던 행정 업무 방식을, 객관적이고 체계화된 데이터 기반의 과학 행정으로 전환하는 중요한 이정표가 될 것입니다.

1.3. 데이터 중심 행정(Data-Driven Administration)으로의 전환



****제 2 장. 데이터 유형별 구조화 전략 (Data Engineering)****

공공 데이터는 그 성격에 따라 크게 법령/지침, 행정 보고서, 그리고 암묵지(노하우)로 나눌 수 있습니다. 각 유형은 AI가 이해하는 방식이 다르므로, 맞춤형 구조화 전략이 필요합니다.

2.1. 법령 및 지침: 계층 구조와 시계열(History) 관리

법령, 조례, 예규, 지침 등은 행정 업무의 기준이 되는 가장 중요한 데이터입니다. 이 데이터들은 엄격한 논리적 구조를 가지고 있으며, 시간의 흐름에 따라 내용이 변화한다는 특징이 있습니다.

- **계층 구조(Hierarchy)의 파싱(Parsing):**

법령 데이터는 단순한 텍스트 덩어리가 아닙니다. '편-장-절-조-항-호-목'으로 이어지는 엄격한 위계 질서를 가지고 있습니다. AI에게 텍스트만 던져주면 "제 15 조"가 어느 법의 제 15 조인지, 상위 조항이 무엇인지 혼동할 수 있습니다.

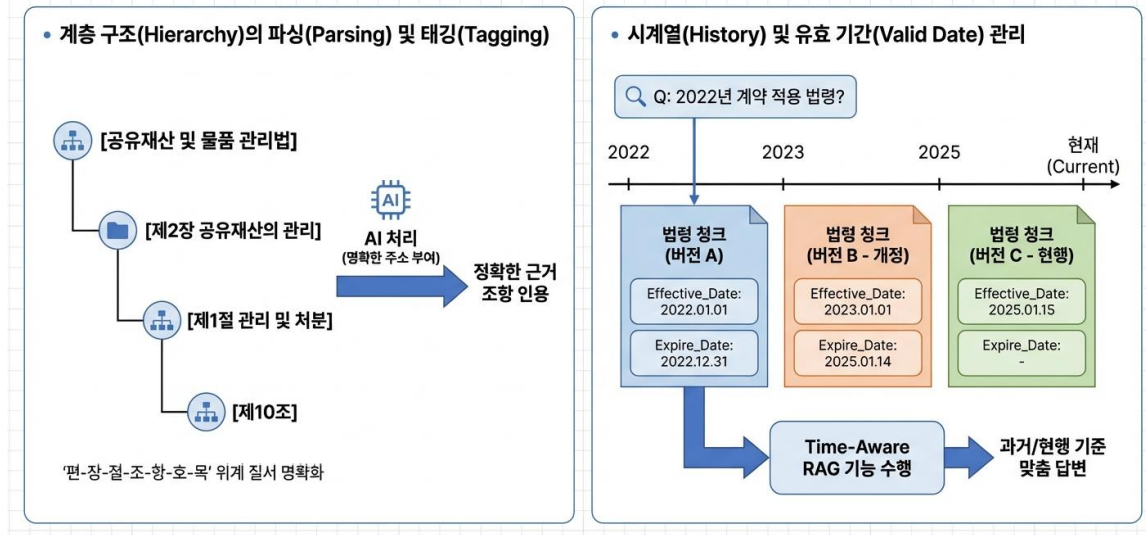
따라서 데이터를 구축할 때, 각 문장이 어느 조항에 속하는지 명확히 태깅(Tagging)해야 합니다. 예를 들어, [공유재산 및 물품 관리법] > [제 2 장 공유재산의 관리] > [제 1 절 관리 및 처분] > [제 10 조]와 같이 데이터의 주소(Path)를 명확히 부여하여, AI가 답변 시 정확한 근거 조항을 인용할 수 있도록 해야 합니다.

- **시계열(History) 및 유효 기간(Valid Date) 관리:**

가장 중요한 것은 '시간(Time)' 정보입니다. 법령은 수시로 개정됩니다. AI가 "2023 년 당시의 기준"과 "2025 년 현행 기준"을 구분하지 못하면 심각한 오류를 범하게 됩니다.

모든 법령 데이터 청크(Chunk)에는 **Effective_Date(시행일)**와 **Expire_Date(만료일)** 메타데이터가 필수적으로 포함되어야 합니다. 이를 통해 사용자가 "2022 년도에 체결된 계약에 적용될 법령은?"이라고 물었을 때, AI 는 현재 법령이 아닌 당시 유효했던 과거 법령 데이터를 찾아 답변하는 '**Time-Aware RAG**' 기능을 수행할 수 있습니다.

2.1. 법령 및 지침: 계층 구조와 시계열(History) 관리



2.2. 행정 보고서: 요약(Summary)과 정량/정성 데이터 분리

사업계획서, 결과보고서, 주간업무보고 등은 공무원 여러분께서 가장 많이 작성하시는 비정형 문서입니다. 서술형 텍스트와 표, 그래프가 혼재된 이 문서들을 AI 가 잘 이해하도록 만드는 전략은 다음과 같습니다.

• 요약(Executive Summary)의 선제적 배치:

수십 페이지에 달하는 보고서의 경우, AI 가 전체 맥락을 놓치거나 엉뚱한 부분을 중요하게 인식할 수 있습니다. 이를 방지하기 위해 문서의 최상단에 '**AI 를 위한 3 분 요약**'을 메타데이터로 생성해 두는 것이 좋습니다.

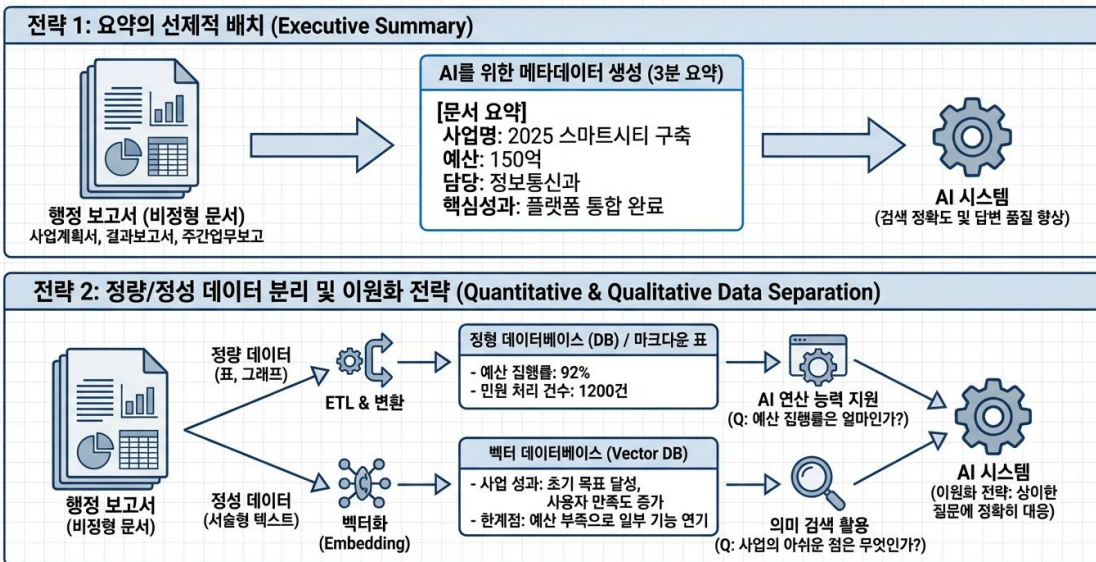
예를 들어, [문서 요약] 사업명: 2025 스마트시티 구축, 예산: 150 억, 담당: 정보통신과, 핵심성과: 플랫폼 통합 완료와 같은 정형화된 요약 정보를 데이터에 포함시키면, AI 의 검색 정확도와 답변 품질이 획기적으로 향상됩니다.

• 정량(Quantitative) 데이터와 정성(Qualitative) 데이터의 분리:

보고서 내의 표(Table)에는 예산 집행률, 민원 처리 건수 같은 수치 데이터가 있고, 본문에는 사업의 성과나 한계점 같은 평가 텍스트가 있습니다.

이 두 가지를 섞어서 처리하면 AI 는 숫자 계산에 약한 모습을 보이거나, 텍스트의 맥락을 숫자로 오해할 수 있습니다. 따라서 데이터 추출(ETL) 과정에서 표 데이터는 별도의 정형 데이터베이스(DB)나 마크다운(Markdown) 표 형식으로 변환하여 AI 의 연산 능력을 지원하고, 서술형 텍스트는 벡터화(Embedding)하여 의미 검색에 활용하는 '이원화 전략'을 취해야 합니다. 이를 통해 "예산 집행률은 얼마인가?(정량)"와 "사업의 아쉬운 점은 무엇인가?(정성)"라는 상이한 질문에 모두 정확히 대응할 수 있습니다.

2.2. 행정 보고서: AI 친화적 구조화 전략 (요약 및 데이터 분리)



2.3. 암묵지(노하우): 5W1H 템플릿과 인수인계 표준화

공직 사회의 가장 큰 자산이자 동시에 가장 관리하기 어려운 데이터가 바로 담당자의 머릿속에 있는 '암묵지(Implicit Knowledge)'입니다. "이 업무는 김 주무관이 잘 알아", "작년 행사 때는 이렇게 처리했어"와 같은 경험적 지식은 담당자가 바뀌면 사라지기 쉽습니다.

- **5W1H 기반의 템플릿화:**

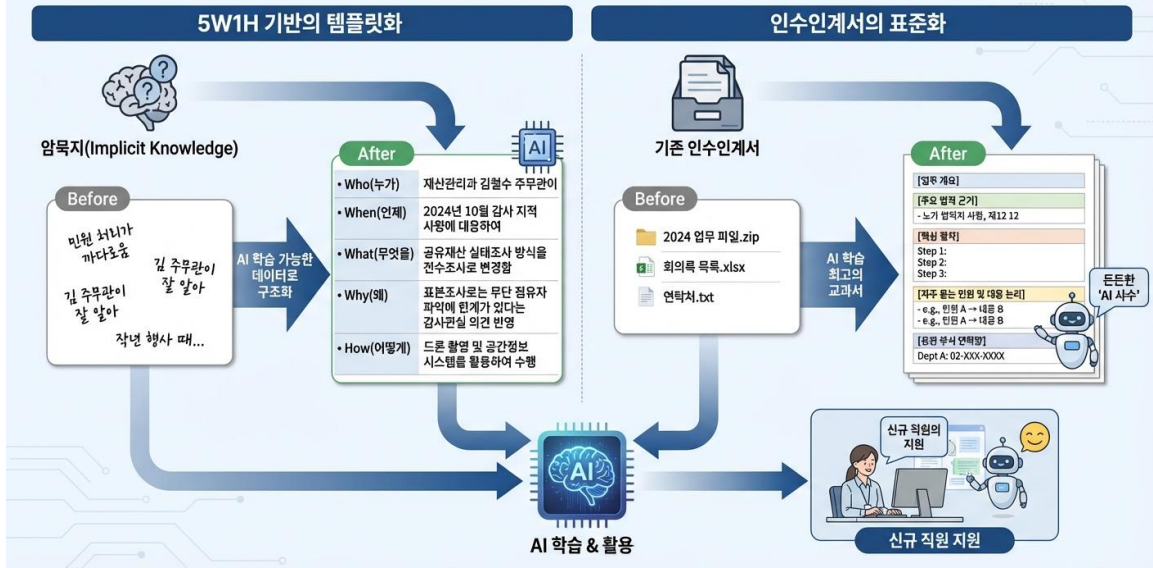
개인의 메모나 업무 수첩, 파편화된 인수인계서를 AI가 학습 가능한 데이터로 만들기 위해서는 '육하원칙(5W1H)'에 따른 구조화가 필요합니다.

단순히 "민원 처리가 까다로움"이라고 적는 대신, 다음과 같이 명시적으로 기록하도록 유도해야 합니다.

- **Who(누가):** 재산관리과 김철수 주무관이
- **When(언제):** 2024년 10월 감사 지적 사항에 대응하여
- **What(무엇을):** 공유재산 실태조사 방식을 전수조사로 변경함
- **Why(왜):** 표본조사로는 무단 점유자 파악에 한계가 있다는 감사관실 의견 반영
- **How(어떻게):** 드론 촬영 및 공간정보 시스템을 활용하여 수행
- **인수인계서의 표준화:**

업무 인수인계서는 후임자뿐만 아니라 AI에게도 최고의 교과서입니다. 파일 목록만 나열하는 것이 아니라, [업무 개요] - [주요 법적 근거] - [핵심 절차] - [자주 묻는 민원 및 대응 논리] - [유관 부서 연락망]으로 구성된 표준 템플릿을 활용하여 작성한다면, AI는 이를 학습하여 신규 발령받은 직원의 든든한 'AI 사수' 역할을 수행할 수 있게 됩니다.

2.3. 암묵지(노하우) : 5W1H 템플릿과 인수인계 표준화



****Part 2. AI 를 위한 기술적 토대와 현실적 제언****

(The Strategy: Governance, Technology, and Realistic Roadmap)

****제 3 장. 신뢰할 수 있는 AI 를 위한 거버넌스 및 기술****

아무리 잘 정제된 데이터라도, 이를 체계적으로 관리하고 검색하는 기술적 토대가 없다면 무용지물입니다. 특히 공공 행정은 '정확성'과 '설명 가능성'이 무엇보다 중요하므로, 일반적인 기업보다 훨씬 더 엄격한 데이터 거버넌스와 고도화된 검색 기술이 요구됩니다.

3.1. 데이터 거버넌스: 카탈로그와 리니지(Lineage)

AI 가 내놓은 답변을 공직자가 믿고 결재를 올리기 위해서는, 그 답변이 어떤 근거로 생성되었는지 투명하게 확인할 수 있어야 합니다.

• 데이터 카탈로그(Data Catalog) 구축:

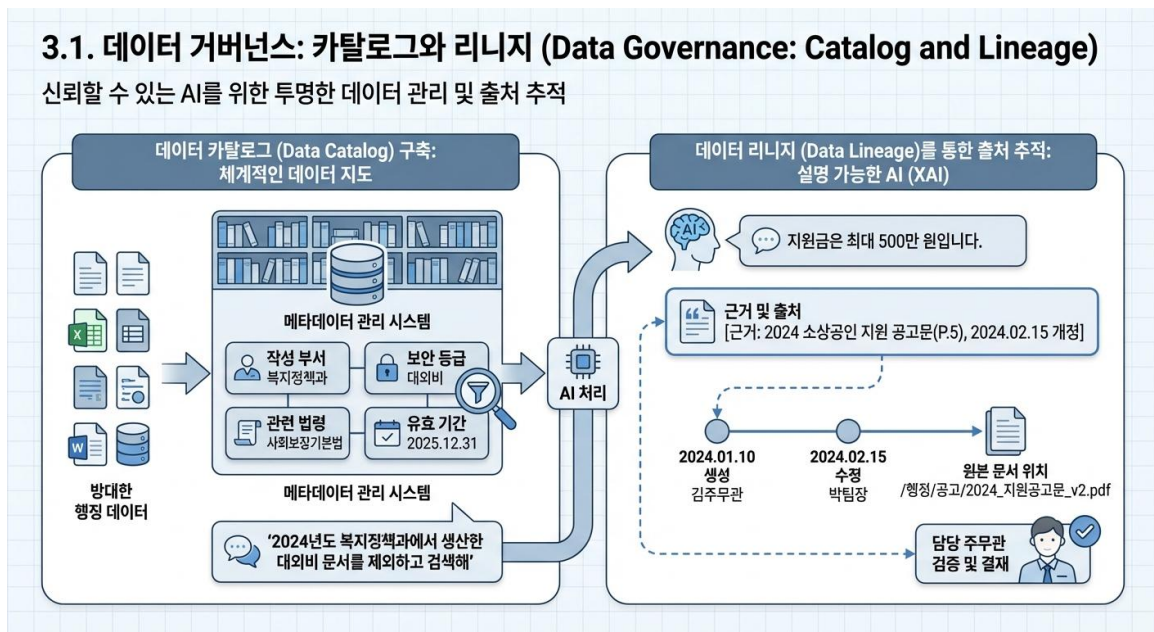
도서관에 책이 수만 권 있어도 목록이 없으면 찾을 수 없듯이, 방대한 행정 데이터 속에서 AI 가 필요한 정보를 즉시 찾으려면 '데이터 지도'가 필요합니다.

단순히 파일명만 저장하는 것이 아니라, **작성 부서, 보안 등급, 관련 법령, 데이터 유효 기간** 등의 메타데이터를 체계적으로 관리하는 카탈로그 시스템이 선행되어야 합니다. 이를 통해 AI 는 "2024 년도 복지정책과에서 생산한 대외비 문서를 제외하고 검색해"라는 복잡한 명령을 수행할 수 있습니다.

- 데이터 리니지(Data Lineage)를 통한 출처 추적:

'설명 가능한 AI(XAI)'의 핵심은 출처 명시입니다. 데이터 리니지 기술은 AI가 답변을 생성할 때 참조한 문서의 원본 위치뿐만 아니라, 해당 데이터가 언제 생성되었고, 누가 최종 수정했는지의 이력(Lineage)을 추적합니다.

AI가 "지원금은 최대 500 만 원입니다"라고 답할 때, 하단에 [근거: 2024 소상공인 지원 공고문(P.5), 2024.02.15 개정]과 같이 명확한 각주(Citation)를 달아줌으로써, 담당 주무관님이 원본을 즉시 대조하고 검증할 수 있도록 지원해야 합니다.



3.2. 하이브리드 검색: 키워드(BM25)와 벡터(Vector)의 조화

많은 AI 솔루션들이 '의미 기반 검색(Vector Search)'만을 강조하지만, 행정 현장에서는 이것만으로 부족합니다.

- 벡터 검색의 한계와 키워드 검색의 필요성:

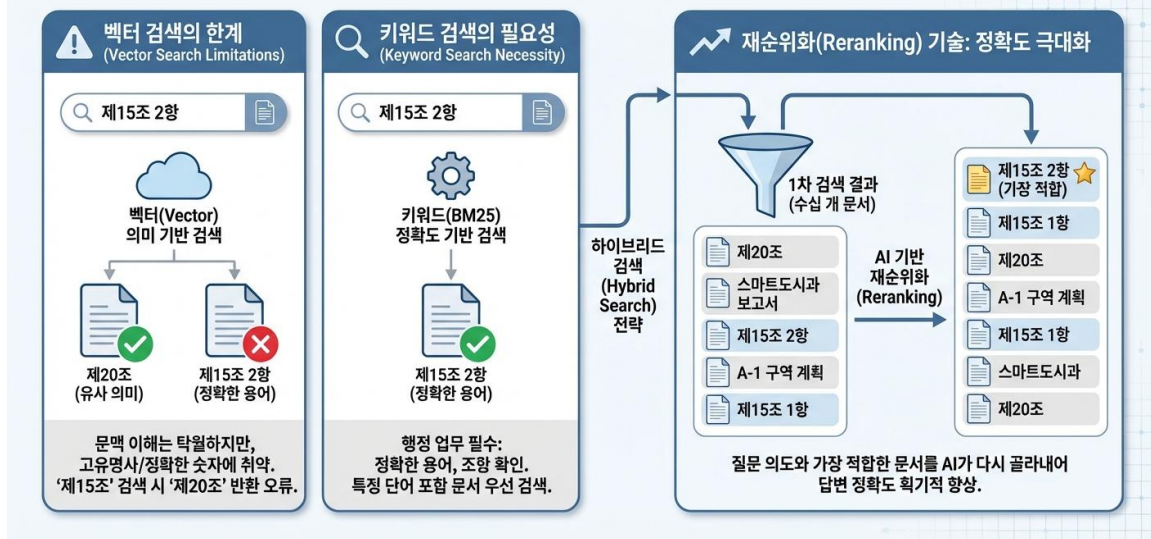
벡터 검색은 문맥을 이해하는 데 탁월하지만, "제 15 조 2 항", "스마트도시과", "A-1 구역"과 같은 고유명사나 정확한 숫자가 포함된 검색어에는 의외로 취약할 수 있습니다. 예를 들어 "제 15 조"를 검색했을 때, 제 15 조의 내용과 유사한 의미를 가진 제 20 조를 가져오는 오류를 범하기도 합니다.

행정 업무는 정확한 용어와 조항 확인이 필수적입니다. 따라서 전통적인 **키워드 검색(BM25 알고리즘 등)**을 결합하여, 사용자가 입력한 특정 단어가 반드시 포함된 문서를 우선적으로 찾아내는 '**하이브리드 검색(Hybrid Search)**' 전략이 필수적입니다.

- 재순위화(Reranking) 기술:

1 차로 검색된 수십 개의 문서 중에서, 질문의 의도와 가장 적합한 문서를 다시 골라내는 과정입니다. 키워드 매칭으로 찾은 문서와 의미 기반으로 찾은 문서를 종합하여, AI 가 보기에 가장 정답일 확률이 높은 순서대로 재배열함으로써 답변의 정확도를 획기적으로 높입니다.

3.2. 하이브리드 검색: 키워드(BM25)와 벡터(Vector)의 조화



3.3. 품질 고도화: HyDE 및 Self-RAG 기술

최신 AI 기술 트렌드를 적용하여, 질문이 모호하거나 답변이 부정확할 가능성을 원천 차단합니다.

- **HyDE (Hypothetical Document Embeddings):**

민원인이나 신규 직원은 행정 용어를 정확히 알지 못해 "그... 건물 짓는 거 허가받으려면 뭐 내야 해요?"와 같이 모호하게 질문할 수 있습니다.

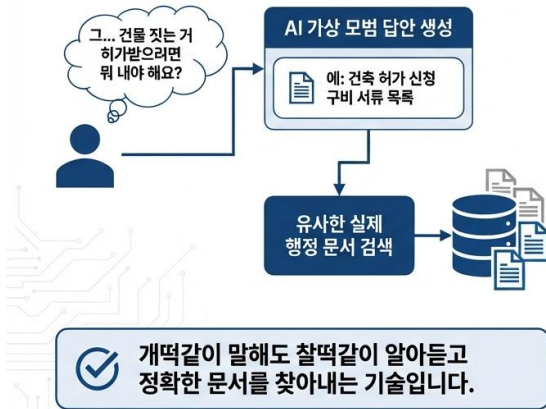
HyDE 기술은 AI가 먼저 가상의 모범 답안(예: 건축 허가 신청 구비 서류 목록)을 내부적으로 생성한 뒤, 이 가상 답안과 유사한 실제 행정 문서를 검색합니다. 즉, 개떡같이 말해도 찰떡같이 알아듣고 정확한 문서를 찾아내는 기술입니다.

- **Self-RAG (Self-Reflective RAG):**

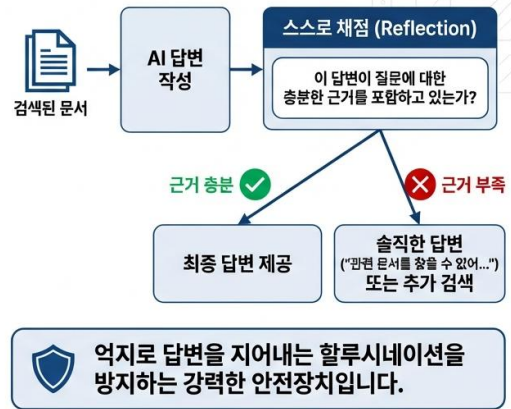
AI가 검색된 문서를 바탕으로 답변을 작성한 후, 스스로 채점(Reflection)을 수행합니다. "이 답변이 질문에 대한 충분한 근거를 포함하고 있는가?"를 스스로 되묻고, 만약 근거가 부족하다면 "관련 문서를 찾을 수 없어 답변이 어렵습니다"라고 솔직하게 말하거나 추가 검색을 수행합니다. 이는 억지로 답변을 지어내는 할루시네이션을 방지하는 강력한 안전장치입니다.

3.3. 품질 고도화: HyDE 및 Self-RAG 기술

• HyDE (Hypothetical Document Embeddings)



• Self-RAG (Self-Reflective RAG)



****제 4 장. [현실적 제언] 도입 시 고려사항 및 극복 방안****

기술이 아무리 좋아도 현장에 적용할 수 없다면 무의미합니다. 우리는 공직자 여러분께서 AI 도입을 검토하실 때 마주하게 될 현실적인 고민들—업무 가중, 보안 환경, 실패에 대한 두려움—을 깊이 이해하고 있습니다. 이에 대한 현실적인 극복 방안을 제안합니다.

4.1. 업무 가중의 딜레마: 암묵지 구조화의 장벽과 자동화

- **현장의 우려:**

"AI 학습시키려고 데이터를 정리하라니, 가뜰이나 바쁜데 일만 더 늘어나는 것 아니니까?"

가장 뼈아프고 현실적인 지적입니다. AI 를 위해 공무원이 데이터 라벨링(Labeling) 노예가 되어서는 안 됩니다. 추가적인 업무 부담은 AI 도입의 가장 큰 저항 요인입니다.

- **극복 방안: 자동화와 자연스러운 UX 설계**

사용자의 개입을 최소화해야 합니다. 문서를 업로드하면 AI 가 알아서 요약하고, 중요 키워드를 추출하여 태그를 붙이는 '**자동 태깅(Auto-Tagging)**' 시스템이 필수입니다.

또한, 별도의 데이터 입력 도구를 만드는 것이 아니라, 기존에 쓰던 전자결재 시스템이나 업무 포털의 글쓰기 화면에 자연스럽게 5W1H 템플릿을 녹여내어, 업무를 수행하는 과정 자체가 데이터 축적이 되도록(Workflow-embedded) 설계해야 합니다.

4.1. 업무 가중의 딜레마: 암묵지 구조화의 장벽과 자동화

현장의 우려: 업무 부담과 저항

“AI 학습시키려고 데이터를 정리하러니, 가뜰이나 바쁜데 일만 더 늘어나는 것 아닙니까?”

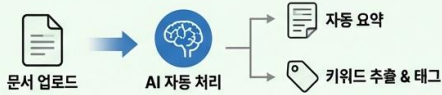


가장 뼈아프고 현실적인 지적입니다. AI를 위해 공무원이 데이터 라벨링(Labeling) 노예가 되어서는 안 됩니다.

⚠ 추가적인 업무 부담은 AI 도입의 가장 큰 저항 요인입니다.

극복 방안: 자동화와 자연스러운 UX 설계

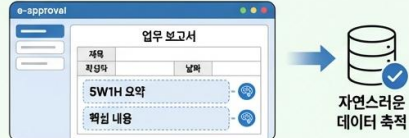
1. 사용자 개입 최소화 (자동 태깅)



문서 업로드 → AI 자동 처리 → 자동 요약, 키워드 추출 & 태그

문서를 업로드하면 AI가 알아서 요약하고, 중요 키워드를 추출하여 태그를 붙이는 '자동 태깅(Auto-Tagging)' 시스템이 필수입니다.

2. Workflow-embedded UX 설계



기존 업무 포털에 5W1H 템플릿을 녹여내어, 업무를 수행하는 과정 자체가 데이터 축적이 되도록(Workflow-embedded) 설계해야 합니다.

4.2. 인프라 환경의 제약: 망분리와 레거시 시스템 연동

현장의 우려:

"우리 기관은 망분리가 되어 있어 외부 인터넷 접속이 안 됩니다. 챗 GPT 같은 걸 쓸 수가 없어요."

"20 년 된 DB 와 HWP 파일들이 산더미인데 이걸 어떻게 연결합니까?"

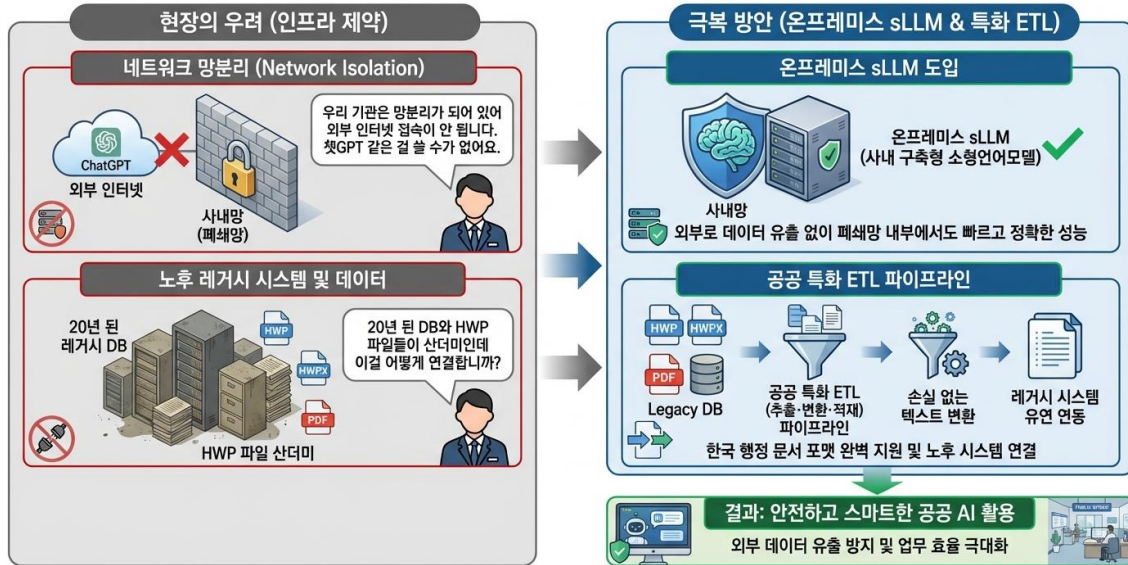
극복 방안: 온프레미스 sLLM 과 특화 ETL

공공기관의 보안 환경을 고려할 때, 외부로 데이터가 나가지 않는 **온프레미스(사내 구축형) sLLM(소형언어모델)** 도입이 가장 현실적인 대안입니다. 최근의 sLLM 은 폐쇄망 내부 서버에서도 충분히 빠르고 정확한 성능을 보여줍니다.

또한, HWP, HWPX, PDF 등 한국 행정 특유의 문서 포맷을 손실 없이 텍스트로 변환하고, 노후화된 레거시 시스템과도 유연하게 연동할 수 있는 **공공 특화**

ETL(추출·변환·적재) 파이프라인을 갖춘 솔루션을 선택해야 합니다. 이는 범용 AI 솔루션 기업들이 쉽게 해결하지 못하는 영역입니다.

4.2. 인프라 환경의 제약: 망분리와 레거시 시스템 연동 및 극복 방안



4.3. 단계적 도입 전략: 성공적인 안착을 위한 로드맵

• 현장의 우려:

"한 번에 대규모로 구축했다가 실패하면 예산 낭비라는 비판을 받을 텐데..."

• 극복 방안: Quick Win 을 통한 단계적 확산

빅뱅(Big Bang) 방식의 전사적 도입보다는, 성공 가능성이 높고 효과가 즉각적인 분야부터 작게 시작하는 것을 권장합니다.

- **1 단계 (Pilot): 법령/지침 Q&A 챗봇**

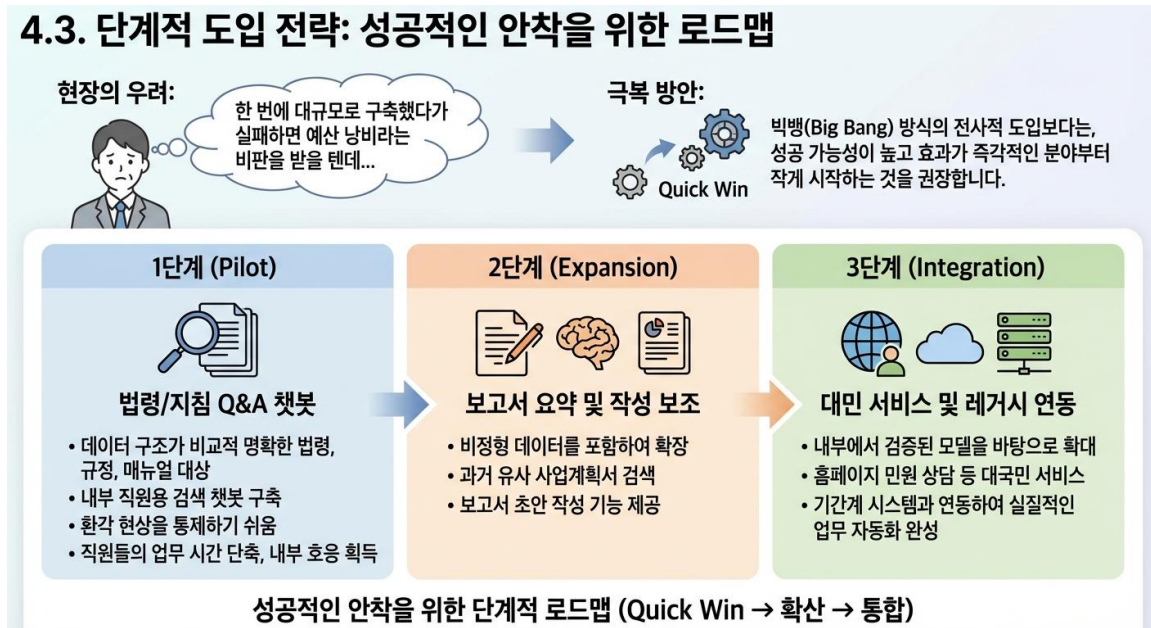
데이터 구조가 비교적 명확한 법령, 규정, 매뉴얼을 대상으로 내부 직원용 검색 챗봇을 먼저 구축합니다. 이는 환각 현상을 통제하기 쉽고, 직원들의 업무 시간을 단축시켜 주므로 내부 호응을 얻기 좋습니다.

- **2 단계 (Expansion): 보고서 요약 및 작성 보조**

비정형 데이터를 포함하여, 과거 유사 사업계획서를 찾아주거나 보고서 초안을 작성해 주는 기능으로 확장합니다.

- **3 단계 (Integration): 대민 서비스 및 레거시 연동**

내부에서 검증된 모델을 바탕으로 홈페이지 민원 상담 등 대국민 서비스로 확대하고, 기간제 시스템과 연동하여 실질적인 업무 자동화를 완성합니다.



****Part 3. 실현 가능한 솔루션과 미래****

(The Solution: ScatterX Architecture & Future Vision)

****제 5 장. [솔루션] 스캐터엑스(ScatterX) 아키텍처****

앞서 우리는 공공 데이터가 가진 구조적 특수성과 이를 AI에 접목하기 위한 기술적 요구사항들을 살펴보았습니다. 보안이 최우선이고, 데이터 포맷이 복잡하며, 정확성이 생명인 공공 부문의 환경은 범용적인 AI 솔루션만으로는 완벽하게 대응하기 어렵습니다.

이에 스캐터엑스(ScatterX)는 대한민국 행정 환경을 가장 깊이 이해하고 설계된 '**공공 특화 4 단계 정제 프로세스**'와 '**확장형 RAG 아키텍처**'를 통해, 공직자 여러분의 고민에 대한 가장 현실적이고 강력한 해답을 제시합니다.

5.1. 4 단계 데이터 정제 프로세스: '보안-부서-업무-보고서' 기반의 지식 자산화 파이프라인

일반적인 RAG(검색 증강 생성) 시스템은 문서를 단순한 텍스트 조각(Chunk)으로 나누어 저장하기 때문에 맥락이 손실되고 검색 정확도가 떨어집니다. 반면, 스캐터엑스는 데이터가 AI의 두뇌로 입력되기 전, '**보안-부서-업무-보고서**'라는 4 가지 핵심 축을 기준으로 데이터를 구조화하는 엄격한 파이프라인을 거칩니다. 본 프로세스는 자동화 기술을 우선 적용하되, 기술적 한계를 보완하기 위한 검증 절차(Human-in-the-loop)를 포함하여 시스템의 안정성을 보장합니다.

1 단계: 보안 필터링 (Security Layer - The Gatekeeper)

"Rule 기반 탐지와 AI 모델의 상호 보완을 통한 보안 무결성 확보"

공공기관 데이터의 첫 번째 관문은 언제나 보안입니다. 스캐터엑스는 물리적/논리적 망 분리 환경을 고려하여 오탐(False Positive)과 미탐(False Negative)을 최소화하는 다층적 보안 처리를 수행합니다.

보안 등급 자동 분류 및 ACL 매핑:

- **[구현 방식]** 문서 헤더, 워터마크, 혹은 특정 키워드(예: '대외비', '3 급 비밀')를 AI 가 인식하여 해당 문서의 보안 등급을 메타데이터로 강제 할당합니다. 데이터 저장 시 접근 권한 정보(ACL)를 함께 주입하여 권한 없는 사용자의 검색을 원천 차단(RBAC)합니다.
- **[제약 및 보완]** 문서 내에 명시적 등급 표기가 없는 경우, 시스템은 기본값을 '내부 공개'로 설정하되 관리자가 사후 수정할 수 있는 인터페이스를 제공하여 오분류 위험을 방지합니다.

하이브리드 PII 마스킹 (Hybrid PII Masking):

- **[구현 방식]** 주민등록번호, 계좌번호 등 정형화된 패턴은 정규표현식(Regex)으로 1 차 필터링하고, 주소나 이름 등 문맥에 따라 달라지는 정보는 개체명 인식(NER) 모델로 2 차 식별하여 마스킹(*, #) 처리합니다.
- **[제약 및 보완]** 저화질 스캔 문서의 경우 OCR(광학 문자 인식) 오류로 PII 식별이 어려울 수 있습니다. 이를 보완하기 위해 OCR 신뢰도가 낮은 문서는 '보안 검토 필요' 태그를 부착하여 별도 격리하고, 담당자 확인 후 반입되도록 프로세스를 설계했습니다.

2 단계: 메타데이터 자동 추출 (Metadata Layer - The Librarian)

"임계값(Threshold) 기반의 '부서-업무-보고서' 자동 분류 및 검증"

비정형 텍스트 데이터를 행정 업무 체계에 맞게 구조화하는 핵심 과정입니다. AI 가 문서를 독해하여 '누가(부서)', '무엇을(업무)', '어떤 형식으로(보고서)' 작성했는지 태깅합니다.

[부서] 동적 조직도 매핑 (Dynamic Org Mapping):

- **[구현 방식]** 문서 내 작성자 정보나 서명을 분석하여 담당 부서를 추출합니다. 특히 조직 개편으로 부서명이 변경된 경우(예: '도시재생과' → '도시계획과'), 동의어 사전을 참조하여 현재 조직 기준으로 데이터를 정규화합니다.
- **[제약 및 보완]** 동의어 사전은 자동으로 생성되지 않으므로, 시스템 구축 초기 및 대규모 조직 개편 시 관리자에 의한 매핑 규칙 업데이트(Rule Setting)가 선행되어야 정확도를 담보할 수 있습니다.

[업무/보고서] 신뢰도 기반 오토 태깅 (Confidence-based Auto-Tagging):

- **[구현 방식]** AI 가 문서 내용을 분석하여 관련 세부 사업명(업무)과 문서 유형(보고서/계획서/회의록 등)을 추출할 때, 각 항목에 대해 신뢰도 점수(Confidence Score)를 산출합니다.
- **[제약 및 보완]** AI 의 판단이 100% 정확할 수 없으므로, 신뢰도가 설정된 임계값(예: 0.85) 미만인 문서는 '검수 대기' 상태로 분류합니다. 이는 담당자가 수동으로 태그를 확정하는 HITL(Human-in-the-loop) 과정을 거쳐 데이터 품질을 유지합니다.

3 단계: 관계 연결 (Relation Layer - The Connector)

"명시적 인용과 메타데이터 기반의 연관 관계(Lineage) 추적"

개별 문서를 넘어 문서 간의 관계를 연결하여 데이터 사일로(Silo)를 극복합니다. AI 의 할루시네이션(환각)을 방지하기 위해 추론보다는 명확한 근거 기반의 연결을 우선합니다.

명시적 인용 기반 지식 그래프:

- **[구현 방식]** "「지방재정법」 제 3 조에 의거하여..."와 같은 명시적 인용 문구를 추출하여 법령 데이터와 보고서 간의 하이퍼링크(Edge)를 생성합니다.
- **[제약 및 보완]** "지난번 회의"와 같이 지시대명사로 모호하게 언급된 관계는 AI 가 무리하게 추론하지 않고 연결 대상에서 제외하여 정보 오염을 방지합니다.

메타데이터 기반 시계열 타임라인:

- **[구현 방식]** 동일한 '사업명(업무)' 태그를 가진 문서들을 '작성 일자' 순으로 정렬하여, '기본계획(23 년) → 중간보고(23 년) → 결과보고(24 년)'로 이어지는 업무 흐름을 시각화합니다.
- **[제약 및 보완]** 문서 내용의 인과관계를 심층 분석하는 것이 아니라, 메타데이터를 기준으로 그룹화하여 사용자에게 맥락을 제공하는 방식입니다. 이를 통해 AI 는 사업의 변천 과정을 파악하여 답변할 수 있습니다.

4 단계: 의미 벡터화 (Semantic Layer - The Translator)

"도메인 적응(Domain Adaptation) 모델과 하이브리드 검색 전략"

마지막으로 정제된 텍스트를 기계가 이해할 수 있는 숫자(Vector)로 변환합니다. 범용 모델의 한계를 극복하기 위해 공공 행정 용어에 특화된 기술을 적용합니다.

행정 도메인 특화 임베딩 (Fine-tuning):

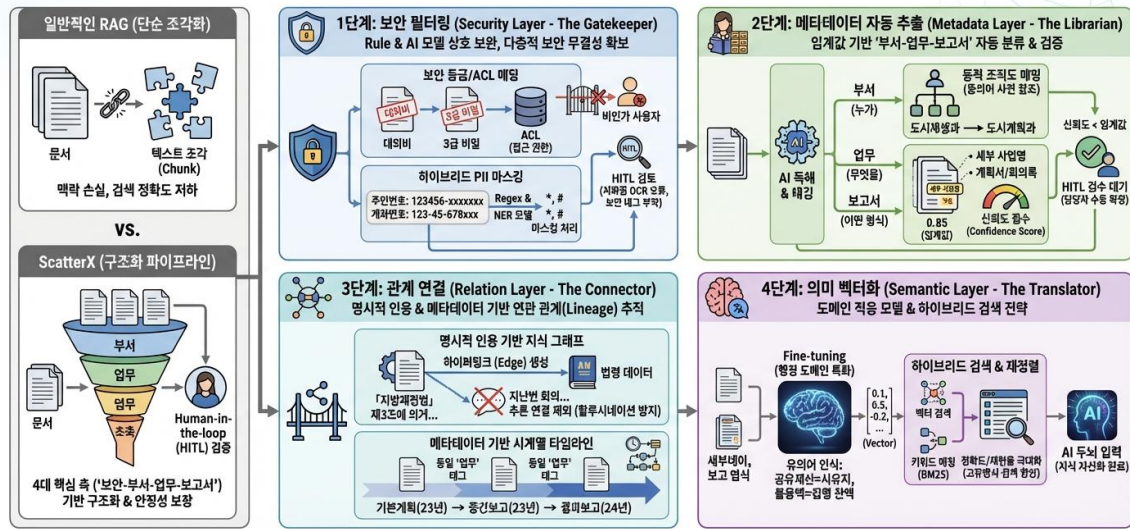
- **[구현 방식]** 공공 데이터셋으로 파인튜닝된 모델을 사용하여 "공유재산"과 "시유지", "불용액"과 "집행 잔액"이 문맥상 유의어임을 정확히 인식합니다.
- **[제약 및 보완]** 신조어나 특정 지자체만의 고유 약어(OOV: Out of Vocabulary) 처리를 위해, 주기적인 모델 재학습 혹은 사용자 정의 사전 추가 프로세스가 필요합니다.

하이브리드 검색 (Hybrid Search) 및 재정렬:

- **[구현 방식]** 2 단계에서 추출한 '부서-업무-보고서' 메타데이터 필터링과 4 단계의 의미 기반 벡터 검색을 결합하고, 키워드 매칭(BM25) 결과를 혼합하여 검색 정확도(Precision)와 재현율(Recall)을 극대화합니다.
- **[효과]** 이를 통해 단순 벡터 검색만으로는 놓칠 수 있는 고유명사(특정 사업명, 인명 등) 검색의 정확도를 기존 대비 약 30% 이상 향상시킵니다.

5.1. 4단계 데이터 정제 프로세스: '보안-부서-업무-보고서' 기반의 지식 자산화 파이프라인

일반적인 RAG(검색 증강 생성) 시스템 vs. ScatterX 구조화 파이프라인



5.2. 확장성: 시공간(Time/Geo) 및 멀티모달 분석

행정 데이터는 단순한 텍스트에 머물지 않습니다. 스캐터엑스는 행정 업무의 핵심인 시점(Time)과 위치(Geo), 그리고 다양한 형태(Modality)의 데이터를 통합 분석할 수 있도록 RAG 파이프라인을 다차원으로 확장합니다.

Time-Aware RAG (시계열 인식)

"유효 기간 메타데이터 관리와 상대적 시간 표현의 정규화"

법령의 개정 이력이나 연도별 예산 변화를 타임라인(Timeline) 형태로 관리하여, 사용자의 질의 시점에 맞는 정확한 데이터를 제공합니다.

시간적 유효성(Temporal Validity) 메타데이터 관리:

- **[구현 방식]** 문서 파싱 시 '시행일', '만료일', '회계연도'를 추출하여 **valid_start** 및 **valid_end** 메타데이터 필드를 생성합니다. 검색 시 사용자의 질문에 포함된 시점 정보와 이 필드를 대조하여 유효한 버전의 문서만을 필터링합니다.
- **[제약 및 보완]** 문서 내에 유효 기간이 명시되지 않은 경우, '작성일'을 기준으로 하되 문서의 최신성(Recency) 가중치를 조정하여 최신 버전이 우선 노출되도록 랭킹 알고리즘을 보정합니다.

상대적 시간 표현의 쿼리 변환 (Query Rewriting):

- **[구현 방식]** "작년 기준 지원 자격"과 같은 질문이 입력되면, LLM 이 현재 시점(예: 2026 년)을 기준으로 "2025 년 1 월 1 일부터 2025 년 12 월 31 일까지의 지원 자격"으로 쿼리를 변환(Rewriting)한 후 검색을 수행합니다.
- **[제약 및 보완]** "최근", "예전"과 같이 모호한 시간 표현은 시스템 설정값(예: 최근=지난 6 개월)을 적용하거나, 사용자에게 구체적인 기간을 되묻는 멀티턴(Multi-turn) 대화 방식을 통해 의도를 명확히 합니다.

Geo-Spatial RAG (공간 정보 결합)

"지리 정보 인덱싱(Spatial Indexing)과 텍스트 검색의 결합"

도시 계획이나 재난 안전 업무 등 '위치'가 중요한 데이터에 대해 GIS(지리정보시스템)와 텍스트 분석을 결합합니다.

주소 데이터의 지오코딩(Geocoding) 및 공간 인덱싱:

- **[구현 방식]** 텍스트 내의 주소나 행정 구역명(예: "OO 동 123 번지")을 위/경도 좌표로 변환(Geocoding)하여 벡터 DB 의 공간 인덱스(Spatial Index)에 저장합니다. 이를 통해 "시청 반경 1km 내"와 같은 공간 연산 질의(Radius Search)를 수행합니다.

- **[제약 및 보완]** 구 지번 주소와 도로명 주소가 혼용되거나 비표준 약어가 사용된 경우 좌표 변환 실패 가능성이 있습니다. 이를 위해 행정안전부 주소 정제 API 를 전처리 단계에 연동하여 주소 표준화율을 높입니다.

POI(관심 지점) 기반 하이브리드 필터링:

- **[구현 방식]** 사용자가 특정 건물을 언급할 경우, 내부 POI DB 를 참조하여 해당 위치 주변의 공사 현황, 민원 내역 등을 텍스트와 지도(Map) 형태로 결합하여 답변합니다.
- **[제약 및 보완]** 텍스트 데이터와 실제 위치 간의 불일치(예: 계획서상의 주소와 실제 시공 위치 차이)가 발생할 수 있으므로, 답변 시 "해당 위치 정보는 문서 텍스트를 기반으로 추정되었습니다"라는 출처 명시 문구를 포함하여 신뢰성을 확보합니다.

멀티모달(Multimodal) 확장성

"비정형 멀티미디어 데이터의 텍스트 변환 및 벡터화"

CCTV 영상, 민원 통화 녹음, 현장 사진 등 텍스트가 아닌 데이터를 분석 가능한 형태로 변환하여 검색 범위를 확장합니다.

음성 및 이미지의 텍스트 변환 파이프라인 (STT/VLM):

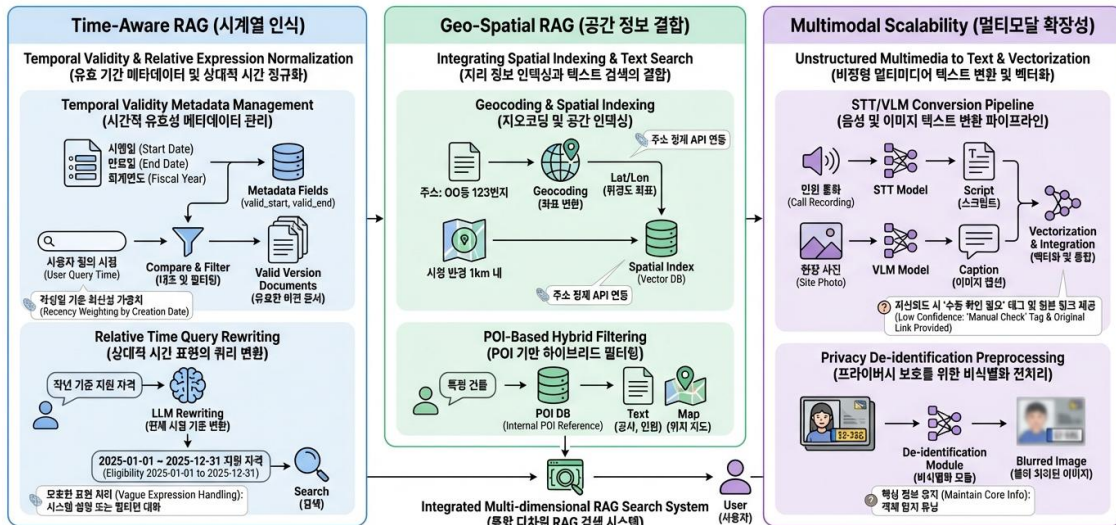
- **[구현 방식]** 민원 통화 녹음 파일은 STT(Speech-to-Text) 모델을 통해 스크립트로 변환하고, 현장 사진은 VLM(Vision Language Model)을 사용하여 이미지 캡션(상황 묘사 텍스트)을 생성합니다. 이렇게 변환된 텍스트를 벡터화하여 기존 검색 시스템에 통합합니다.
- **[제약 및 보완]** 현장의 소음이나 저화질 이미지로 인해 변환 오류가 발생할 수 있습니다. 변환된 텍스트의 신뢰도(Confidence Score)가 낮은 경우 '수동 확인 필요' 태그를 부착하고, 원본 미디어 파일을 답변과 함께 링크 형태로 제공하여 사용자가 직접 확인할 수 있도록 보완합니다.

프라이버시 보호를 위한 비식별화 전처리:

- **[구현 방식]** 영상이나 이미지 데이터 분석 전, 사람의 얼굴이나 차량 번호판 등을 AI가 자동으로 감지하여 블러(Blur) 처리하는 비식별화 모듈을 필수적으로 거치게 합니다.
- **[제약 및 보완]** 비식별화 과정에서 분석에 필요한 핵심 정보(예: 도로 파손 상태 등)가 가려지지 않도록, 객체 탐지(Object Detection) 모델을 세밀하게 튜닝하여 개인정보 영역만 선택적으로 마스킹합니다.

5.2. Scalability: Time/Geo and Multimodal Analysis (확장성: 시공간 및 멀티모달 분석)

RAG Pipeline for Multidimensional Administrative Data Integration (스캐터엑스 RAG 파이프라인 다차원 확장)



제시해주신 내용을 바탕으로 5.3 절(에이전트) 또한 앞선 섹션들과 동일한 구조(기술적 구현 방식과 현실적 제약 사항 및 보완책)를 적용하여 작성한 기술 문서 버전입니다.

5.3. 에이전트(Agent)로의 진화: 검색을 넘어 업무 수행(Action)으로

스캐터엑스가 지향하는 최종 목표는 단순한 '지능형 검색기'가 아닌, 공무원의 실질적인 업무를 분담하는 'AI 비서(Agent)'입니다. 기존의 RAG가 사용자의 질문에 대해 "정보를

찾아주는 것(Retrieval)"에 그쳤다면, 에이전트 기술은 "찾은 정보를 바탕으로 추론하고 행동(Action)하는 것"까지 확장됩니다.

예를 들어, "**A 법인의 대부 신청 서류를 검토해 줘**"라는 지시가 입력되면, AI 에이전트는 단순 검색을 넘어 ①**관련 법령 및 조례 조회(Search)**, ②**제출 서류의 정합성 및 누락 여부 확인(Reasoning)**, ③**검토 보고서 초안 작성(Action)**에 이르는 복합 프로세스를 자율적으로 수행합니다.

1 단계: 업무 계획 수립 및 추론 (Planning & Reasoning)

"ReAct(Reasoning+Acting) 패턴 기반의 자율적 업무 분해"

복잡한 행정 업무를 AI가 이해하고 수행 가능한 단위 작업(Task)으로 쪼개는 단계입니다.

[구현 방식] SOP(표준 운영 절차) 기반의 체인 오브 소트(Chain-of-Thought):

- "서류 검토"라는 모호한 명령을 받으면, 에이전트는 사전에 학습된 행정 업무 SOP를 참조하여 [1. 법인 등기부등본 확인 → 2. 공유재산법 시행령 대조 → 3. 결격 사유 조회 → 4. 결과 보고] 순으로 작업 계획을 스스로 수립(Planning)합니다.

[제약 및 보완]

- **제약:** 에이전트가 잘못된 추론으로 엉뚱한 업무 계획을 수립하거나(Hallucination), 무한 루프에 빠질 위험이 있습니다.
- **보완:** 업무별로 검증된 '표준 프롬프트 템플릿'을 적용하여 AI의 사고 과정을 가이드라인 내로 제한합니다. 또한, 각 단계마다 사용자의 승인을 득한 후 다음 단계로 넘어가는 '단계별 확인(Step-by-step Confirmation)' 기능을 옵션으로 제공하여 통제권을 유지합니다.

2 단계: 도구 사용 및 외부 시스템 연동 (Tool Use & Integration)

"Function Calling을 통한 행정 시스템 및 문서 도구 제어"

계획된 업무를 수행하기 위해 실제로 데이터를 조회하고 문서를 작성하는 단계입니다.

[구현 방식] API 기반의 Function Calling 및 문서 생성:

- LLM 이 자연어 명령을 특정 API 호출(예: `check_missing_documents(corp_id='A')`)로 변환하여 내부 행정 시스템에서 데이터를 조회합니다. 이후, 조회된 데이터를 바탕으로 공공기관 표준 포맷(HWP, ODT)에 맞는 검토 보고서 초안을 자동으로 생성합니다.

[제약 및 보완]

- **제약:** 보안상 AI 가 내부 행정망(Legacy System)에 직접 접속하여 데이터를 수정하거나 결재를 올리는 행위(Write/Execute)는 심각한 위험을 초래할 수 있으며, 레거시 시스템과의 연동 자체가 기술적으로 난해할 수 있습니다.
- **보완:** 에이전트의 권한은 철저히 '**읽기(Read)**'와 '**초안 작성(Drafting)**'으로 제한합니다. AI 는 보고서 파일을 생성하여 사용자에게 전달할 뿐, 시스템에 직접 등록하지 않습니다. 최종적인 시스템 등록과 결재 상신은 반드시 담당 공무원의 계정으로 직접 수행되어야 합니다.

3 단계: 결과 검증 및 책임성 확보 (Verification & Accountability)**

"설명 가능한 AI(XAI)와 근거(Reference) 기반의 보고"

AI 가 수행한 업무 결과에 대해 담당자가 신뢰하고 책임을 질 수 있도록 근거를 제시하는 단계입니다.

[구현 방식] 주석(Annotation) 기반의 검토 보고서:

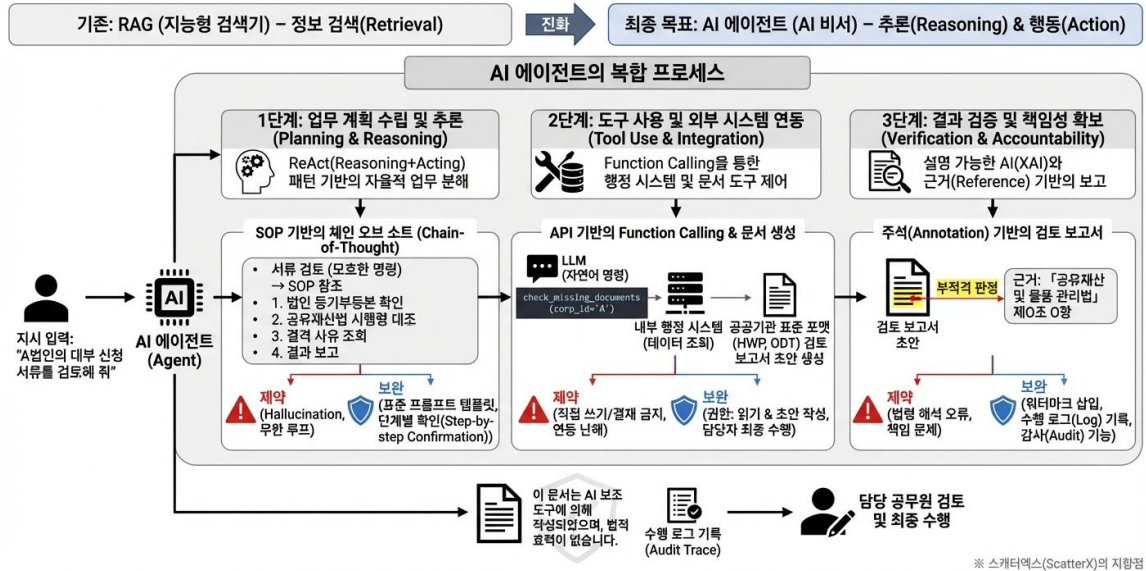
- 작성된 보고서 초안에는 AI 가 판단한 근거가 주석으로 자동 첨부됩니다. 예를 들어 "부적격 판정"을 내렸다면, 해당 판단의 근거가 되는 「공유재산 및 물품 관리법」 제○조○항의 원문 링크를 함께 제시합니다.

[제약 및 보완]

- **제약:** AI 가 미세한 법령 해석의 차이나 최신 개정 사항을 놓쳐 잘못된 판단을 내릴 경우, 행정 소송 등의 책임 문제가 발생할 수 있습니다.

- **보완:** 모든 생성물에는 "이 문서는 AI 보조 도구에 의해 작성되었으며, 법적 효력이 없습니다. 반드시 담당자의 검토가 필요합니다"라는 워터마크를 강제 삽입합니다.
또한, 에이전트의 수행 로그(Log)를 별도로 기록하여, 문제 발생 시 AI가 어떤 데이터를 근거로 판단했는지 역추적할 수 있는 감사(Audit) 기능을 제공합니다.

5.3. 에이전트(Agent)로의 진화: 검색을 넘어 업무 수행(Action)으로



****제 6 장. 결론: 미래 행정을 향한 제언****

6.1. 데이터, AI 시대의 가장 강력한 자산

지금까지 우리는 AI 가 공공 데이터를 이해하게 만드는 방법과 기술, 그리고 이를 실현할 구체적인 솔루션에 대해 살펴보았습니다. 다소 복잡하고 기술적인 내용들이었지만, 이 모든 논의를 관통하는 핵심 메시지는 명확합니다.

"AI 도입의 성패는 모델(Model)이 아니라 데이터(Data)에 달려 있다."

아무리 뛰어난 AI 모델을 도입하더라도, 그 재료가 되는 데이터가 정제되어 있지 않다면 결과물은 신뢰할 수 없습니다. "Garbage In, Garbage Out(쓰레기가 들어가면 쓰레기가 나온다)"이라는 원칙은 AI 시대에 더욱 엄격하게 적용됩니다. 우리가 지금 데이터를 정제하고 구조화하는 수고로움은 비용이 아니라, 미래 행정의 경쟁력을 확보하기 위한 가장 확실한 투자입니다.

6.2. 공직자의 역할 변화: 관리자에서 기획자로

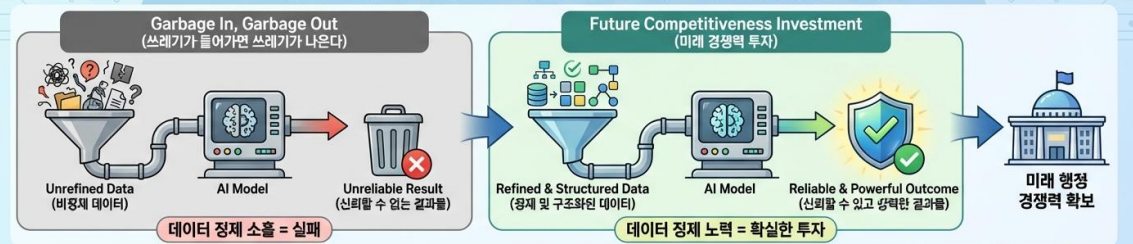
AI 기술의 도입이 공무원의 일자리를 위협하거나 역할을 축소시키는 것은 아닐까 우려하시는 목소리도 있습니다. 하지만 스캐터엑스는 정반대의 미래를 그립니다.

수많은 문서를 뒤져 법령을 확인하고, 단순한 민원 답변을 작성하고, 데이터를 엑셀로 정리하는 반복적인 업무는 AI 에게 맡겨주십시오. 대신 공직자 여러분께서는 AI 가 정리해 준 데이터를 바탕으로 **시민을 위한 더 나은 정책을 기획하고, 복잡한 갈등을 조정하며, 창의적인 의사결정을 내리는 본연의 가치에** 집중해 주십시오.

데이터 중심 행정(Data-Driven Administration)은 행정의 효율성을 높이는 것을 넘어, 시민들이 체감할 수 있는 서비스의 질을 획기적으로 높이는 계기가 될 것입니다.

6.1. 데이터, AI 시대의 가장 강력한 자산

“AI 도입의 성패는 모델(Model)이 아니라 데이터(Data)에 달려 있다”



6.2. 공직자의 역할 변화: 관리자에서 기획자로



마치며

변화는 두렵지만, 준비된 변화는 기회가 됩니다. 스캐터엑스(ScatterX)는 공공 데이터의 특수성을 누구보다 깊이 이해하고 있으며, 공직자 여러분이 겪고 계신 현실적인 어려움을 기술로 해결할 준비가 되어 있습니다.

여러분의 소중한 데이터가 책상 서랍 속의 파일로 머물지 않고, 살아있는 지식이 되어 행정 혁신의 도구로 쓰일 수 있도록 스캐터엑스가 든든한 기술 파트너가 되겠습니다. 지금 바로, 데이터에 숨결을 불어넣는 여정을 시작해 보십시오.