

너무 이른 온톨로지. 지금은 메타데이터 RAG

공공부문 AI 신뢰 인프라 구축을 위한 현실적 접근
온톨로지 과잉 설계를 넘어
메타데이터 RAG 기반
실시간 현행화 · 출처 통제 · 정책 인텔리전스 (Policy Intelligence)



Metadata RAG 도입 → 실무 AI 에이전트 → 데이터 축적 → Bottom-up Ontology

스캐터엑스 (ScatterX)
공공 AI 신뢰 파트너
Trust · Security · Practical AI

「너무 이른 온톨로지. 지금은 메타데이터 RAG」

공공부문 AI 신뢰 인프라 구축을 위한 현실적 접근:

온톨로지 과잉 설계를 넘어, 메타데이터 RAG 기반
실시간 현행화·출처 통제·정책 인텔리전스로

 발간사: 강명수 스캐터엑스

대표이사 인사말

안녕하십니까, 스캐터엑스(ScatterX) 대표이사 강명수입니다.

최근 공공부문의 인공지능(AI) 도입을 위해 밤낮으로 헌신하시는 여러 기관장님과 실무 담당자님들의 노고에 깊은 존경의 마음을 전합니다. 행정의 효율을 높이고 국민에게 더 나은 서비스를 제공하기 위해 AI를 적극적으로 검토하시는 모습 현장에서 깊이 체감하고 있습니다.

그러나 현장의 목소리를 들을 때면 안타까운 마음이 드는 것도 사실입니다. 많은 공공기관이 AI 도입의 첫 단추로 '지식그래프'나 '온톨로지(Ontology)' 구축과 같은 대규모 프로젝트를 제안받고 계십니다. 개념과 관계를 정의하는 온톨로지는 훌륭한 기술임이 틀림없습니다. 하지만, 현재 공공기관의 데이터는 대부분 공직자 여러분의 PC 속에 개별 문서 형태로 흩어져 있습니다. 기초적인 데이터 정비가 완료되지 않은

상태에서 복잡한 온톨로지를 설계하는 것은, 튼튼한 기둥 없이 지붕부터 올리려는 시도와 같아 많은 시간과 예산의 소모를 동반할 수밖에 없습니다.

스캐터엑스는 AI 기술이 공공기관에 무거운 짐이나 예산의 부담이 되어서는 안 된다고 믿습니다. 화려한 기술적 수사보다는, 당장 내일 아침 주무관님의 책상 위에서 실질적인 도움을 줄 수 있는 '실용적인 AI'가 필요합니다.

우리는 거창한 온톨로지를 새로 짓는 대신, 이미 잘 정리되어 있는 정부와 지자체의 홈페이지, 법령 사이트의 구조를 활용하는 '메타데이터 RAG'를 제안합니다. 또한, 막대한 예산이 드는 대형 GPU 서버 대신, 합리적인 워크스테이션(RTX 4090) 한 대로 내부망에서 안전하게 데이터를 전처리하고, 국정원 보안 가이드를 준수하며 고성능 글로벌 LLM 을 활용하는 지혜로운 대안을 마련했습니다.

이 백서가 공공 AI 도입을 고민하시는 여러분께 작으나마 현실적이고 유용한 길잡이가 되기를 바랍니다. 스캐터엑스는 늘 겸손한 자세로, 공공 행정의 빈틈을 채우고 실무자들의 어려움을 덜어드리는 든든한 기술 파트너가 되겠습니다. 감사합니다.

Executive Summary

1. 공공 AI 도입의 현주소와 기술적 과제

최근 공공부문에서는 AI 도입을 위해 지식그래프와 온톨로지를 선제적으로 구축하려는 시도가 늘고 있습니다. 그러나 실무 현장의 데이터는 정형화된 형태보다는 공직자들의 PC 에 비정형 문서로 산재해 있어, 이를 즉시 구조화하는 데 어려움이 큼니다. 데이터가 충분히 정비되지 않은 상태에서의 하향식(Top-down) 온톨로지 설계는 예상보다 많은 예산을 요구하며, 실제 행정 업무에 적용되기까지 상당한 시일이 소요됩니다. 또한, 복잡한 정책을 분석하는 'AI 에이전트'를 원활히 구동하기 위해서는 거대언어모델(LLM)의 높은 추론 능력이 필수적이거나, 현재 공공에서 주로 검토되는

제한적인 모델들은 글로벌 최상위 모델 대비 추론 성능에서 아쉬운 지표를 보이고 있습니다.

2. 현실적 대안으로서의 메타데이터 RAG 도입

스캐터엑스는 이러한 고민을 덜어드리기 위해 '메타데이터 RAG' 아키텍처를 제안합니다.

공공 행정에서는 AI 가 '어떤 근거를 바탕으로 답변하는가'가 매우 중요합니다.

메타데이터 RAG 는 기관, 부서, 작성일, 시행일 등의 메타데이터를 기준으로 출처를 엄격히 필터링하여 환각(Hallucination) 현상을 최소화합니다. 새로운 데이터 구조를 처음부터 구축하는 대신, 이미 체계적으로 분류된 공공 웹사이트의 구조를 활용하며, '변경 감시 에이전트'를 결합해 조례 개정 사항을 자동 반영함으로써 행정 오류를 방지합니다.

3. 국정원 보안 가이드 2.0 준수 및 온프레미스 전처리 아키텍처

고성능 글로벌 LLM 을 공공 업무에 도입할 때 가장 면밀히 검토해야 할 부분은 데이터 보안입니다. 스캐터엑스는 2025 년 12 월 10 일 발표된 '**국정원 AI 보안 가이드 2.0**'을 충실히 준수합니다. 대규모 예산이 수반되는 고가의 GPU 랙 서버 대신, 합리적인 수준의 엔비디아 RTX 4090 GPU 워크스테이션을 기관 내부망에 구축합니다. 민감한 비정형 데이터(도면, 민원 이미지, 회의 녹음 등)는 내부망의 AI OCR 및 STT 엔진을 통해 텍스트로 변환되고, AI DLP 시스템이 개인정보를 안전하게 비식별화 처리합니다. 이렇게 보안 처리를 거친 데이터만을 글로벌 LLM API 와 연동하여 안전하게 활용합니다.

4. 실무를 지원하는 12 종의 공공 특화 AI 에이전트

메타데이터 RAG 와 안전한 LLM 환경의 결합은 실질적인 정책 지원 플랫폼의 기반이 됩니다. 스캐터엑스는 국민신문고 민원 사전 분석, 내부 회의 자동 비서, 생활인구 기반 정책 설계, 재난 예방 에이전트 등 공직자의 실무 부담을 덜어드릴 수 있는 12 종의 특화 에이전트를 통해 행정 서비스의 질을 높이는 데 기여합니다.

5. 안정적인 공공 AI 진화 로드맵

도입 초기부터 대규모 온톨로지를 구축하기보다는, 메타데이터 RAG 를 선제적으로 도입하여 실무 에이전트를 운영하는 것이 예산과 시간 측면에서 합리적입니다. 이

과정을 통해 정제된 데이터가 자연스럽게 축적되며, 향후 이를 바탕으로
상향식(Bottom-up) 온톨로지를 점진적으로 구성해 나가는 것이 시스템의 실효성을
높이는 현실적인 방안입니다.

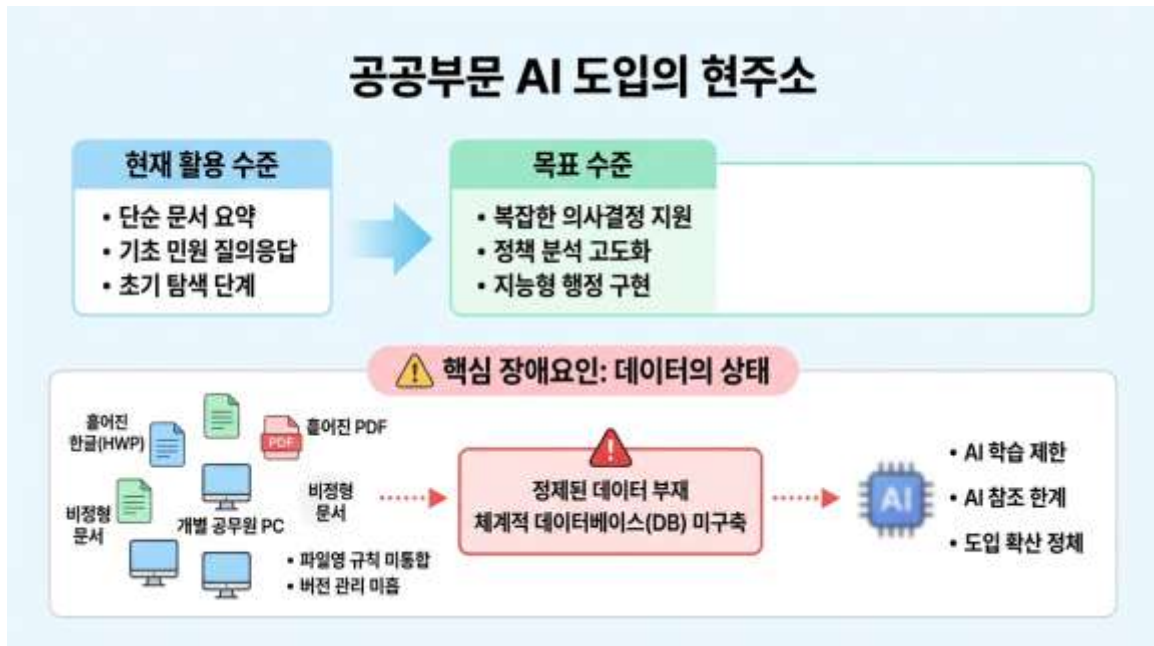
I. 서론: 왜 온톨로지는 지금 공공에 적합하지 않은가

1. 공공부문 AI 도입의 현주소

현재 공공부문의 인공지능 도입은 활발한 논의에도 불구하고, 실제 업무 적용에
있어서는 초기 탐색 단계에 머물러 있습니다. 대부분의 활용 사례가 단순한 문서
요약이나 초보적인 민원 질의응답 수준에 국한되어 있으며, 공직자의 복잡한
의사결정을 보조하는 단계로는 나아가지 못하고 있습니다.

가장 근본적인 원인은 '데이터의 상태'에 있습니다. 공공기관이 보유한 방대한 지식과
경험은 체계적인 데이터베이스(DB) 형태가 아니라, 개별 공무원의 PC 속에 한글(HWP),
PDF 등의 비정형 문서 형태로 흩어져 있습니다. 파일명 규칙이나 버전 관리가 통합되지
않은 상태에서, AI가 학습하고 참조할 수 있는 '정제된 데이터'의 부재는 AI 도입의 가장

큰 장벽이 되고 있습니다.



2. 공공에서 온톨로지 프로젝트가 직면하는 4 가지 어려움

이러한 상황에서 최근 지식그래프나 온톨로지를 통해 데이터를 구조화하려는 시도가 이어지고 있습니다. 개념과 관계를 명확히 정의하는 온톨로지는 학술적으로 훌륭한 접근이지만, 현재의 공공 현장에 바로 적용하기에는 다음과 같은 현실적인 제약이 따릅니다.

첫째, 기초 데이터의 부재입니다. 온톨로지는 명확히 정의된 '엔티티(Entity)'와 그들 간의 '관계(Relationship)'를 전제로 합니다. 하지만 현재 공공기관은 표준화된 용어 사전이나 일원화된 데이터 구조가 부족한 실정입니다. 뼈대가 될 데이터 자체가 정비되지 않은 상태에서 지식 간의 관계망을 그리려는 시도는, 기초 공사 없이 건물을 올리는 것과 같아 실효성을 거두기 어렵습니다.

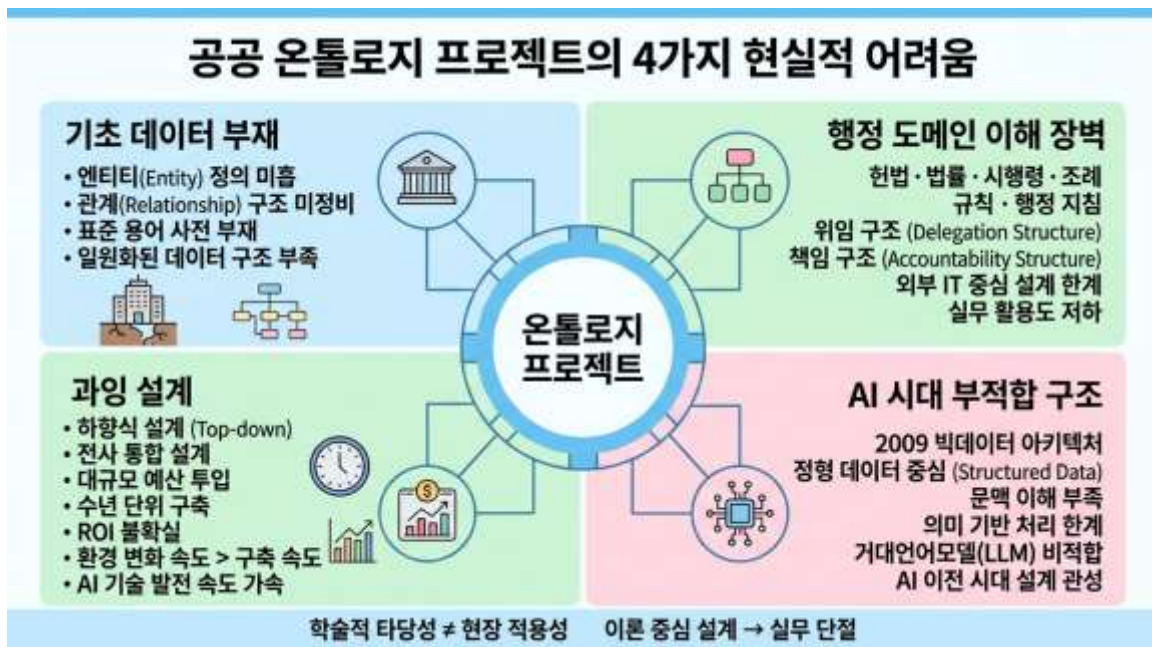
둘째, 행정 도메인 이해의 높은 장벽입니다. 행정 업무는 헌법부터 법률, 시행령, 조례, 규칙, 행정 지침으로 이어지는 고도의 위임 구조와 책임 소재를 가집니다. 외부 IT 기업이 단기간의 인터뷰나 분석만으로 이러한 복잡한 행정 도메인을 완벽히 이해하고

온톨로지를 설계하는 것에는 한계가 존재합니다. 업무의 본질을 담아내지 못한 표면적인 온톨로지는 결국 실무에서 활용되지 못하는 결과로 이어집니다.

셋째, 과잉 설계(Over-engineering)로 인한 예산과 시간의 비효율성입니다.

하향식(Top-down)으로 조직 전체의 온톨로지를 설계하는 작업은 막대한 예산과 수년의 시간을 요구합니다. 빠르게 변화하는 행정 환경과 하루가 다르게 발전하는 AI 기술 속도를 고려할 때, 수십억 원을 투입하여 장기간 시스템을 구축하는 방식은 투자 대비 효용(ROI) 측면에서 합리적인 선택이 되기 어렵습니다.

넷째, 생성형 AI 시대에 맞지 않는 과거의 접근 방식입니다. 현재 제안되는 많은 데이터 플랫폼 구축 방안은 2009 년경 빅데이터 붐이 일어났을 당시의 아키텍처에 머물러 있는 경우가 많습니다. 과거의 정형 데이터 중심 설계는 문맥과 의미를 이해해야 하는 현재의 거대언어모델(LLM)이 소화하기 어려운 구조입니다. AI 이전 시대의 경험만으로 AI 시대의 기반을 설계하는 것은 지양해야 할 접근법입니다.



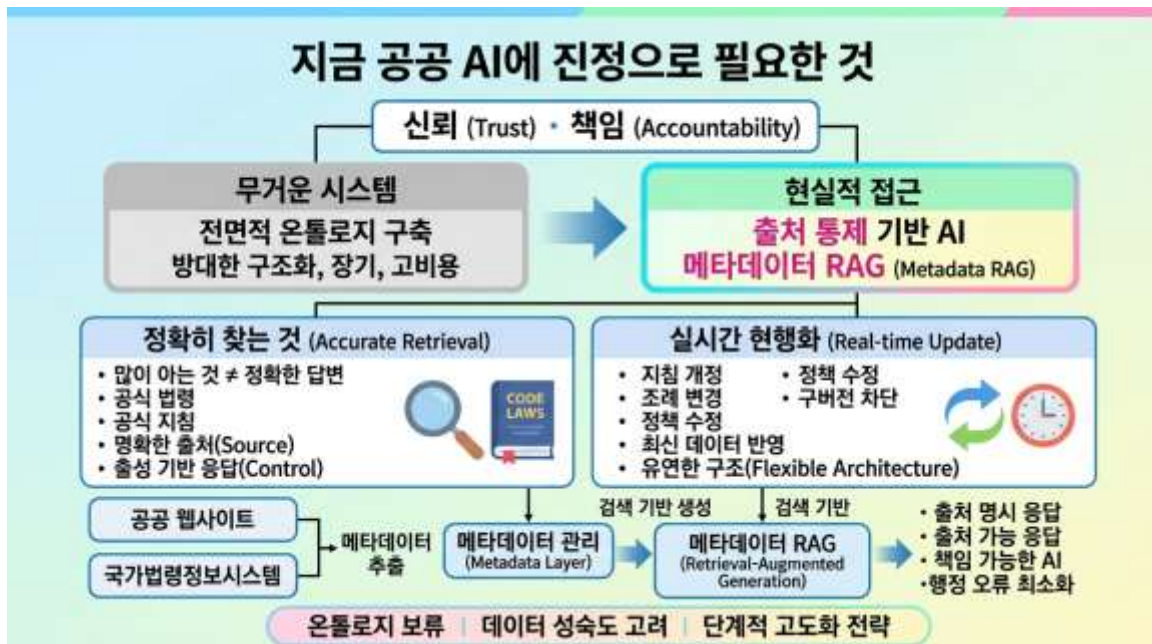
3. 그렇다면 지금 공공 AI에 진정으로 필요한 것은 무엇인가

지금 공공기관에 필요한 것은 방대한 지식을 억지로 구조화하는 무거운 시스템이 아닙니다. 공공 행정에서 가장 중요한 가치는 '신뢰'와 '책임'입니다. 따라서 AI 도입 시 최우선으로 고려해야 할 요소는 다음과 같습니다.

첫째, '많이 아는 것'보다 '정확히 찾는 것'입니다. AI 가 스스로 생성한 정보보다, 공식적인 법령과 지침이라는 명확한 '출처'를 기반으로 답변할 수 있도록 통제하는 기술이 필요합니다.

둘째, '실시간 현행화'입니다. 행정에서는 어제까지 유효했던 지침이 오늘 개정될 수 있습니다. 구버전의 데이터를 인용하여 발생하는 행정 오류를 막기 위해, 변경된 정책과 조례를 즉각적으로 반영할 수 있는 유연한 구조가 필수적입니다.

결론적으로, 현재 공공부문의 데이터 성숙도를 고려할 때 무리한 온톨로지 구축은 잠시 미루어 두는 것이 합리적입니다. 대신, 기존에 잘 정리된 공공 웹사이트와 법령 시스템의 구조를 그대로 활용하면서 출처를 엄격히 통제하는 '**메타데이터 RAG**'가 가장 현실적이고 안정적인 첫걸음이 될 것입니다.



Ⅱ. 공공 데이터의 3 층 구조와 스캐터엑스의 AI 데이터 전달 체계

공공기관이 보유하고 관리하는 데이터는 그 성격과 생성 주체, 활용 목적에 따라 매우 다양한 형태를 띠고 있습니다. 인공지능(AI)이 행정 업무를 올바르게 보조하기 위해서는, 모든 데이터를 하나의 커다란 저장소에 무작위로 섞는 것이 아니라 데이터의 본질적 특성에 따라 분류하고 그에 맞는 전달 체계를 구축해야 합니다. 스캐터엑스는 공공 데이터를 크게 3 가지 계층으로 정의하고, 각각의 특성에 맞춘 맞춤형 AI 데이터 전달 방식을 제안합니다.

1. 제도·근거 데이터 (정확한 데이터 찾아주기)

첫 번째 계층은 법률, 시행령, 조례, 시행규칙, 행정 지침, 고시 및 공식 홈페이지의 안내문 등 행정 행위의 법적 근거가 되는 '제도·근거 데이터'입니다. 이 데이터는 주로 텍스트 위주의 비정형 문서로 존재하며, 대외적으로 공개되는 경우가 많습니다. AI가 행정 업무를 보조할 때 가장 먼저 참조해야 하는 기준점이기도 합니다. 스캐터엑스는 이 계층의 데이터를 AI에 전달할 때 '정확한 데이터 찾아주기(Retrieval)'에 집중합니다. 외부 신뢰 데이터 플랫폼이나 법령 사이트에서 AI가 요청하는 정확한 조문과 지침을 검색하여, 답변의 근거로 제공하는 역할을 수행합니다.

2. 행정 운영 데이터 (정확한 데이터 만들어주기)

두 번째 계층은 공공기관이 실제 업무를 수행하는 과정에서 내부 시스템에 축적되는 '행정 운영 데이터'입니다. 각종 인허가 처리 내역, 민원 접수 및 처리 이력, 세입·세출 예산 집행 내역, 계약 관리 데이터 등이 이에 해당합니다. 이 데이터는 주로 내부망의 관계형 데이터베이스(RDB)에 정형화된 형태로 저장되어 있으나, 시스템마다 구조가 다르고 데이터 간의 결합이 필요한 경우가 많습니다. 스캐터엑스는 자체 개발한 ETL(추출·변환·적재) 엔진과 데이터 카탈로그 엔진을 활용하여, 파편화된 행정 데이터를 정제하고 AI가 이해할 수 있는 형태로 가공하는 '정확한 데이터 만들어주기'를 수행합니다.

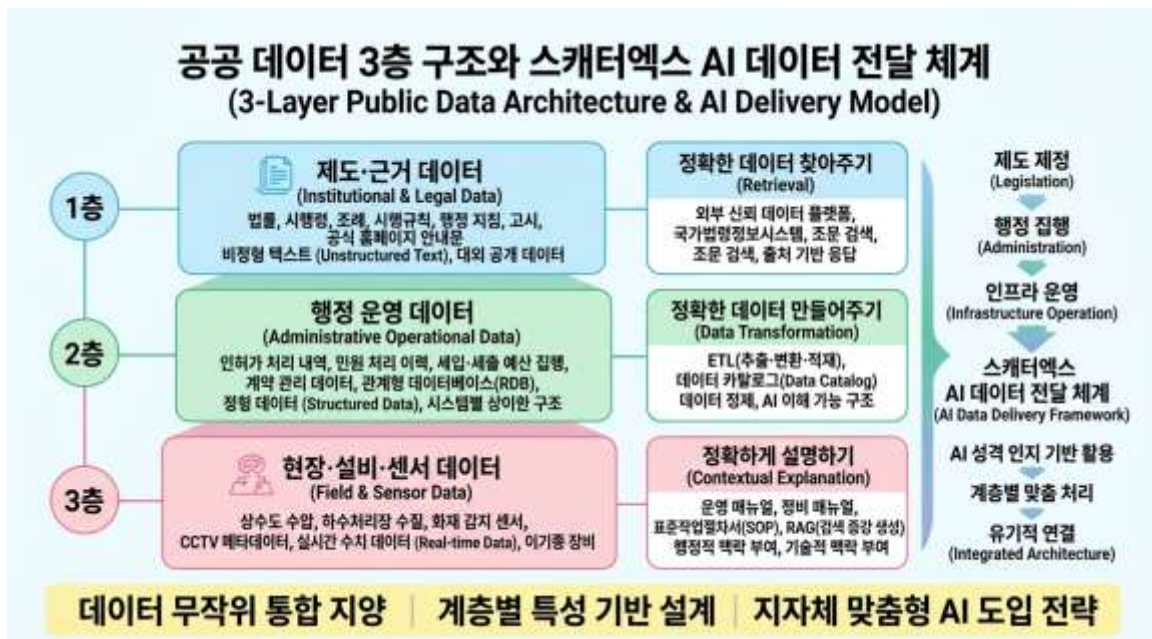
3. 현장·설비·센서 데이터 (정확하게 설명하기)

세 번째 계층은 스마트시티 인프라나 재난 안전 관리 등 물리적 현장에서 발생하는 '현장·설비·센서 데이터'입니다. 상수도 수압, 하수처리장 수질, 화재 감지 센서, CCTV 메타데이터 등 실시간으로 발생하는 이기종 장비의 수치 데이터가 주를 이룹니다.

이러한 단순 수치 데이터는 그 자체로는 AI가 의미를 파악하기 어렵습니다.

스캐터엑스는 장비의 운영 및 정비 매뉴얼, 비상 대응 표준작업절차서(SOP) 등을 기반으로 RAG(검색 증강 생성) 기술을 적용하여 센서 값에 행정적, 기술적 맥락을 부여하는 '정확하게 설명하기' 방식을 채택하고 있습니다.

특히 광역 및 기초자치단체는 법규를 제정(1 계층)하고, 행정을 집행(2 계층)하며, 도시 인프라를 운영(3 계층)하는 복합적인 역할을 수행합니다. 따라서 지자체의 AI 도입은 이 3층 구조의 데이터를 유기적으로 연결하고, AI가 각 데이터의 성격을 정확히 인지하여 활용할 수 있도록 돕는 데이터 전달 체계의 확립에서부터 시작되어야 합니다.



제3장: 메타데이터 RAG 와 '순도 높은 데이터' 가공

인공지능이 공공 행정의 신뢰할 수 있는 보조자로 자리 잡기 위한 핵심은 '어떤 기술을 사용하는가'를 넘어 '어떤 데이터를 어떻게 전달하는가'에 달려 있습니다. 스캐터엑스는 기존의 범용적인 AI 검색 기술이 가진 구조적 한계를 극복하고, 공공기관이 요구하는 엄격한 신뢰성과 감사 추적성을 확보하기 위해 '메타데이터 RAG(Metadata-Retrieval-Augmented Generation)' 아키텍처와 4 단계 '순도 높은 데이터' 가공 방법론을 제안합니다. 이 장에서는 백서의 기술적 핵심인 메타데이터 RAG의 원리와 데이터 전처리 기술, 그리고 공공 웹 구조를 활용하는 실용적 전략을 상세히 서술합니다.

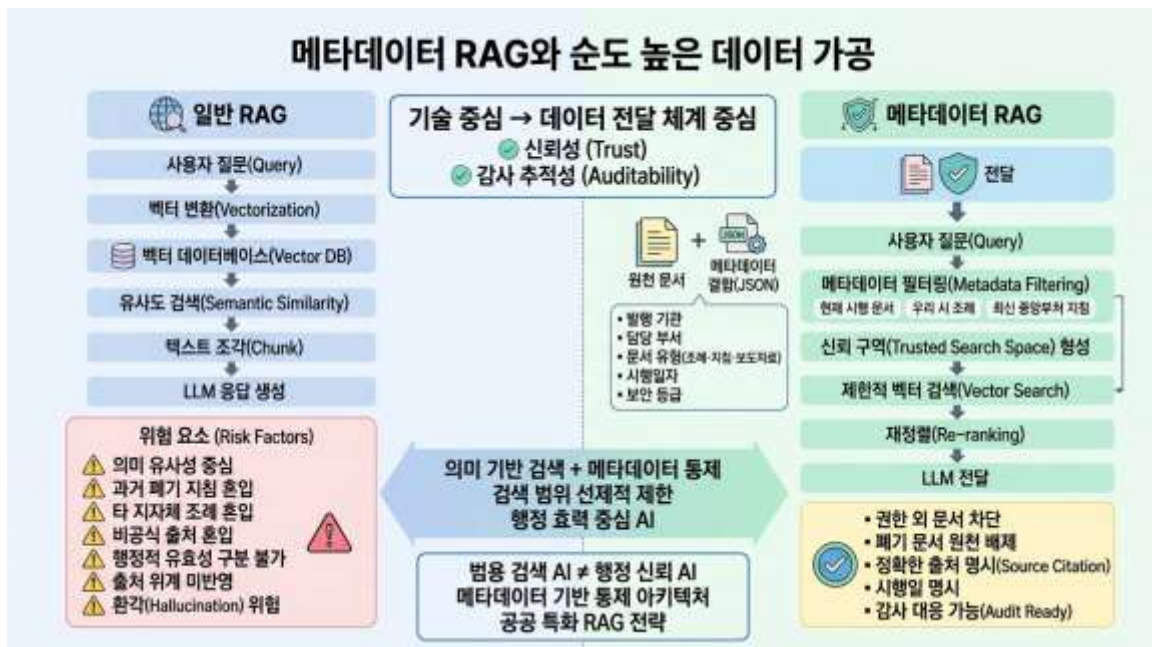
1. 일반 RAG의 구조적 한계와 메타데이터 RAG의 차별성

대규모 언어 모델(LLM)의 환각(Hallucination) 현상을 줄이기 위해 도입된 RAG(검색 증강 생성) 기술은 현재 AI 산업의 표준으로 자리 잡고 있습니다. 일반적인 RAG는 사용자의 질문을 고차원 공간의 벡터(Vector) 형태의 숫자로 변환한 뒤, 벡터 데이터베이스에 저장된 방대한 문서 중 의미적 거리(유사도)가 가장 가까운 텍스트 조각(Chunk)을 찾아 LLM에 전달하는 방식으로 작동합니다. 즉, '의미의 유사성(Semantic Similarity)'이 검색의 유일한 기준이 됩니다.

이러한 방식은 일상적인 대화나 범용 정보 검색에는 유용하지만, 정확한 근거가 생명인 공공 행정에서는 치명적인 약점을 노출합니다. 예를 들어, 실무자가 "침수 피해 가구에 대한 재난지원금 지급 기준을 알려줘"라고 질문했을 때, 일반 RAG는 의미가 유사하다는 이유만으로 5년 전에 폐기된 과거의 지침, 우리 지자체가 아닌 타 지자체의 조례, 심지어 신뢰할 수 없는 개인 블로그의 글을 답변의 근거로 혼합하여 제시할 위험이 있습니다. 행정에서는 과거의 지침과 현재의 지침이 '의미'는 유사할지라도 '법적 효력'은 완전히 다릅니다. 일반 RAG는 이러한 행정적 유효성과 출처의 위계를 구분하지 못합니다.

스캐터엑스가 설계한 ‘메타데이터 RAG’는 벡터 유사도 검색에 의존하기 전에, 문서에 부여된 구조화된 메타정보를 활용하여 검색의 범위를 엄격하게 통제하는 기술입니다. 원천 문서를 데이터베이스에 저장할 때 단순한 텍스트뿐만 아니라 ‘발행 기관’, ‘담당 부서’, ‘문서 유형(조례, 지침, 보도자료 등)’, ‘시행일자’, ‘보안 등급’ 등의 메타데이터를 JSON 등의 형태로 단단히 결합하여 저장합니다.

사용자의 쿼리가 입력되면 시스템은 즉시 벡터 검색을 수행하지 않습니다. 먼저 메타데이터 필터링(Metadata Filtering)을 통해 ‘현재 시행 중인 우리 시의 조례’와 ‘행정안전부의 최신 지침’으로 검색 대상(Search Space)을 좁힙니다. 이후 엄격히 제한된 신뢰 구역 안에서만 의미 기반 벡터 검색을 수행하고, 검색된 결과물의 가중치를 재정렬(Re-ranking)하여 LLM 에 전달합니다. 이 프로세스는 AI 가 권한 밖의 문서나 폐기된 문서를 참조하는 것을 원천적으로 차단하며, 최종 생성된 답변에 정확한 출처와 시행일을 명시할 수 있어 향후 감사 시 완벽한 소명 자료를 제공하는 기반이 됩니다.



2. AI 의 문맥 이해도를 높이는 4 단계 '순도 높은 데이터' 가공 기술

메타데이터 RAG 가 의도한 대로 정확하게 작동하기 위해서는, 벡터 데이터베이스에 입력되는 데이터의 품질이 보장되어야 합니다. 아무리 좋은 검색 엔진이라도 정제되지

많은 쓰레기 데이터(Garbage Data)가 입력되면 잘못된 결과를 출력할 수밖에 없습니다. 스캐터엑스는 수집된 원천 데이터를 AI가 가장 잘 이해할 수 있는 형태로 정제하는 4 단계의 '순도 높은 데이터' 가공 프로세스를 거칩니다.

1 단계: 보안 등급에 따른 데이터 물리적·논리적 분류

공공기관의 데이터는 대국민 공개가 가능한 데이터와 내부망에서만 다루어야 할 민감 데이터로 나뉩니다. 첫 번째 단계에서는 문서의 메타데이터와 내용을 분석하여 보안 등급을 부여하고, 접근 권한을 분리합니다. 이를 통해 민원인을 위한 대민 챗봇과 내부 직원을 위한 업무용 에이전트가 참조하는 데이터 풀(Pool)을 엄격하게 분리하여 정보 유출의 위험을 방지합니다.

2 단계: 부서별 권한 및 도메인 분리

행정 기관 내에서는 동일한 용어라도 부서의 업무 성격에 따라 의미가 달라질 수 있습니다. 예를 들어 '지원금'이라는 단어는 복지 부서와 기업 지원 부서에서 각각 다른 조례와 지침을 따릅니다. 두 번째 단계에서는 데이터의 소유 부서와 업무 도메인을 메타데이터에 태깅하여, AI가 질문자의 소속 부서나 질문의 맥락에 맞추어 정확한 도메인의 문서를 우선 참조하도록 맥락(Context)의 혼선을 방지합니다.

3 단계: 핵심 업무 단위 클러스터링(Clustering)

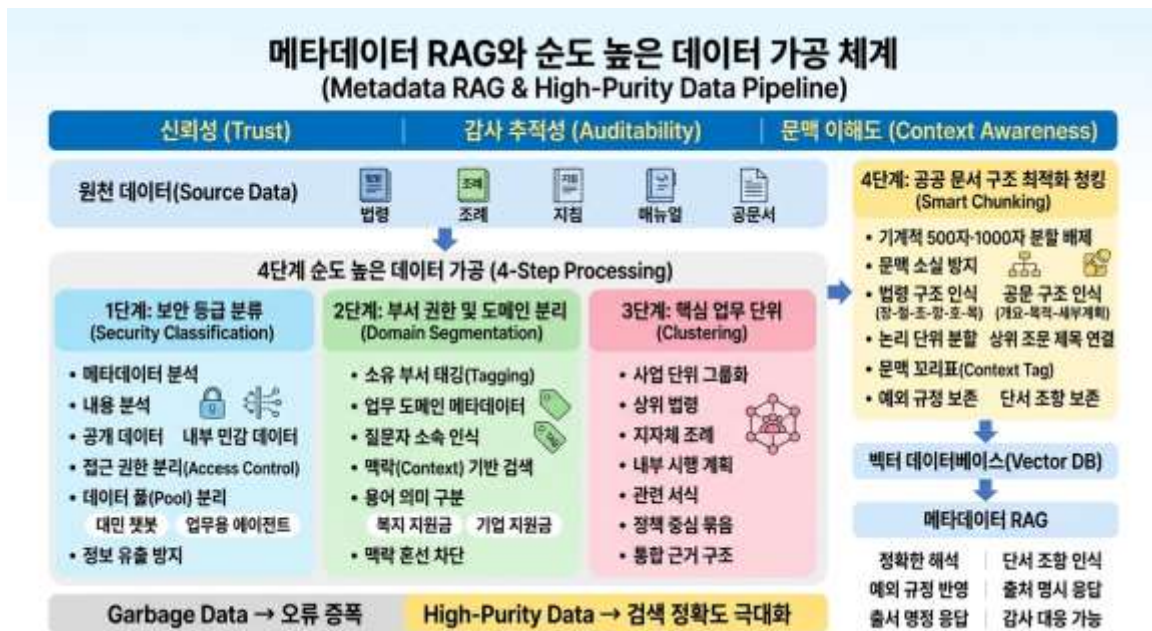
법령, 매뉴얼, 기안 문서 등 서로 다른 형태의 문서라도 동일한 행정 목적을 가진 경우가 많습니다. 세 번째 단계에서는 특정 사업이나 정책(예: 청년 일자리 지원 사업)과 관련된 상위 법령, 지자체 조례, 내부 시행 계획, 관련 서식 등을 논리적인 그룹으로 묶어줍니다. 이를 통해 AI는 단편적인 조문 하나만 찾는 것이 아니라, 해당 업무를 수행하는 데 필요한 일련의 근거 문서들을 통합적으로 파악하고 제시할 수 있게 됩니다.

4 단계: 공공 문서 구조에 최적화된 청킹(Chunking) 및 전처리

이 단계는 메타데이터 RAG의 검색 정확도를 결정짓는 가장 핵심적인 기술입니다. 일반적인 AI 시스템은 긴 문서를 처리할 때 500자나 1,000자 등 일정한 글자 수를 기준으로 문서를 기계적으로 자르는(Chunking) 방식을 사용합니다. 그러나 법령이나 공문서를 이렇게 기계적으로 자르면 치명적인 문맥 소실이 발생합니다. 예를 들어

"제 1 항의 규정에도 불구하고 다음 각 호의 경우에는 예외로 한다"라는 문장 뒤에서 텍스트가 잘려버리면, AI 는 예외 조항의 전제 조건을 잃어버리게 됩니다.

스캐터엑스의 전처리 엔진은 법령의 고유한 '장-절-조-항-호-목' 구조나 공문서의 '개요-목적-세부계획' 구조를 인식합니다. 의미가 연결되는 법적 단위를 훼손하지 않고 논리적인 덩어리로 데이터를 분할하며, 분할된 각 조각(Chunk)마다 상위 조문의 제목과 문맥 정보를 꼬리표처럼 붙여줍니다. 이렇게 가공된 데이터는 AI 가 법령의 단서 조항이나 예외 규정을 놓치지 않고 정확하게 해석할 수 있도록 돕습니다.



3. 기존 공공 웹 구조를 활용하는 실용적 메타데이터 전략

많은 기업들이 AI 도입을 위해 새로운 개념 체계인 지식그래프나 온톨로지를 백지상태에서 설계해야 한다고 주장합니다. 그러나 이는 막대한 예산과 수년의 시간이 소요되는 작업입니다. 스캐터엑스는 이러한 과잉 설계를 피하고, 정부와 지자체가 오랜 시간 공들여 구축해 놓은 기존의 정보 아키텍처를 적극적으로 재활용하는 실용적 전략을 취합니다.

중앙부처나 지자체의 공식 홈페이지, 그리고 국가법령정보센터나 자치법규정보시스템은 무작위로 만들어진 것이 아닙니다. 수많은 공직자와 전문가들이 해당 기관의 업무 성격과 대국민 서비스의 편의성을 고려하여 체계적으로 분류해 놓은 결과물입니다. 즉, 홈페이지의 메뉴 트리(Menu Tree)나 법령 사이트의 법체계 분류는 그 자체로 해당 기관의 업무를 완벽하게 분류해 놓은 훌륭한 '경량 온톨로지(Lightweight Ontology)'입니다.

스캐터엑스는 외부 데이터를 수집하고 가공할 때, 이 기존의 계층 구조를 메타데이터로 그대로 흡수합니다. 예를 들어, 지자체 홈페이지의 특정 게시물이 '홈페이지 메인 > 분야별 정보 > 복지/보건 > 노인복지 > 기초연금'이라는 경로에 위치해 있다면, 이 경로 구조 전체를 해당 텍스트 조각의 문맥 정보(Metadata Tags)로 저장합니다.

이러한 접근 방식은 두 가지 큰 이점을 제공합니다. 첫째, 새로운 개념 체계를 정의하고 분류하는 데 소요되는 막대한 예산과 시간을 절약할 수 있습니다. 이미 검증된 공공의 분류 체계를 그대로 이식하기 때문에 구축 기간이 획기적으로 단축됩니다. 둘째, 공직자와 민원인 모두에게 익숙한 분류 체계를 AI 검색에 적용함으로써, AI가 제시하는 답변의 논리적 흐름이 실제 행정 업무의 흐름과 자연스럽게 일치하게 됩니다.

결과적으로 스캐터엑스의 메타데이터 RAG 전략은, 막대한 비용이 드는 온톨로지 구축 없이도 공공기관이 보유한 기존의 정보 자산을 가장 지혜롭게 활용하여 AI의 신뢰성과 정확성을 극대화하는 현실적인 해법입니다.



IV. 실시간 현행화: 공공 특화 '변경 감시 에이전트' 아키텍처

공공 행정에서 데이터의 생명은 '최신성'과 '무결성'에 있습니다. 아무리 정교하게 가공된 데이터라 할지라도, 어제 개정된 조례나 오늘 아침 하달된 중앙부처의 변경 지침이 반영되지 않았다면 그 AI 시스템은 행정 현장에서 신뢰를 잃게 됩니다. 과거의 규정을 근거로 잘못된 민원 안내를 하거나 부적절한 예산 집행 계획을 세우는 것은 심각한 행정 사고로 이어질 수 있습니다. 스캐터엑스는 이러한 리스크를 원천적으로 방지하기 위해, 데이터의 변경 사항을 스스로 추적하고 업데이트하는 '변경 감시 에이전트(Watching Agent)' 아키텍처를 적용하고 있습니다.

1. 데이터 변동 주기에 따른 맞춤형 감시 체계

공공 데이터는 그 성격에 따라 변경되는 주기가 다릅니다. 스캐터엑스의 변경 감시 에이전트는 서버의 과부하를 방지하고 효율성을 높이기 위해 데이터를 세 가지 유형으로 나누어 모니터링합니다.

매일 새로운 정보가 올라오는 취업 공고, 입찰 공고, 새소식, 내부 행정망(새울 등)의 게시판과 같은 '고빈도 변동 데이터'는 일단위 혹은 실시간으로 변경 사항을 감지합니다. 반면, 법령의 개정, 조례의 제·개정, 주요 행정 지침의 변경과 같은 '이벤트성 변동 데이터'는 해당 사이트의 공포일이나 고시일을 기준으로 변동을 추적합니다. 조직도나 기관의 기본 정책 방향 등 '저빈도 정적 데이터'는 상대적으로 긴 주기를 두고 정합성을 점검합니다.



2. 구조 기반의 정밀한 변경 감지 기술

웹사이트나 문서의 단순한 디자인 변경이나 오타자 수정 수준의 변화로 전체 데이터를 다시 처리하는 것은 비효율적입니다. 스캐터엑스의 에이전트는 문서의 논리적 구조를 기반으로 실질적인 의미 변화를 감지합니다.

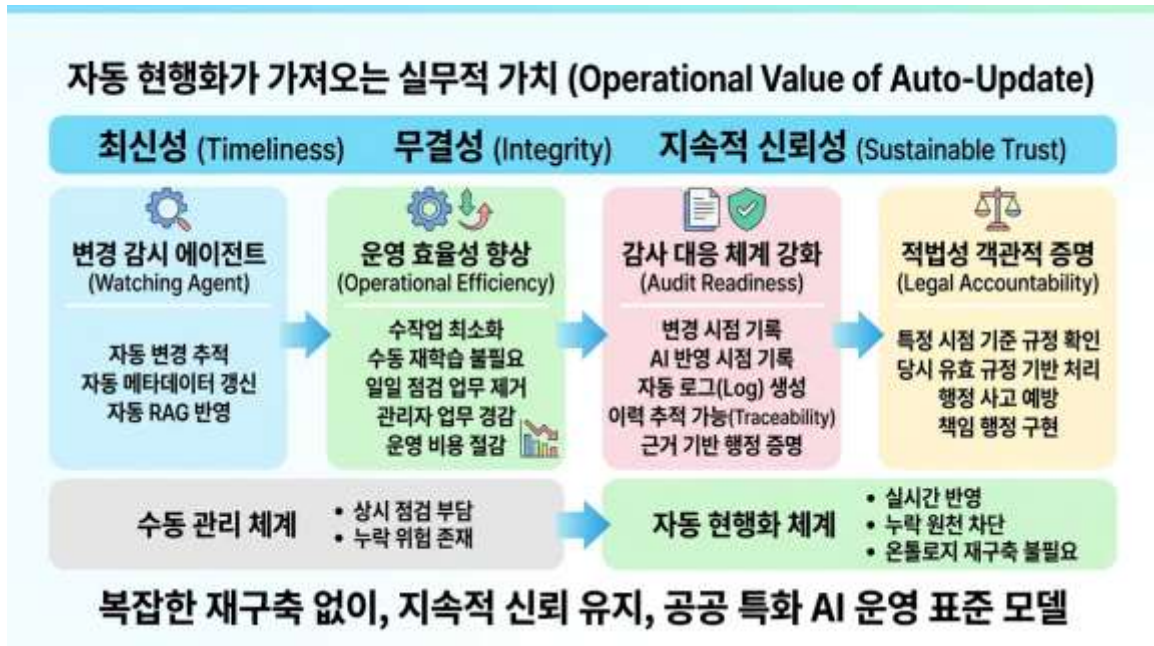
법령이나 조례의 경우, 문서 전체를 비교하는 대신 '조문 단위의 해시(Hash) 값'을 생성하여 관리합니다. 이를 통해 수백 페이지에 달하는 조례 중 제 5 조 제 2 항의 내용만 '일부 개정'되었는지, 혹은 '전부 개정'되었는지를 정확히 파악하고 변경된 조문만을 새롭게 메타데이터 RAG 시스템에 업데이트합니다. 특히, 행정 문서에서 가장 중요한 '공포일'과 '시행일' 데이터를 자동으로 추출하여 메타데이터의 유효 기간을 갱신함으로써, 현재 시점에서 효력이 있는 조문만이 AI의 검색 대상이 되도록 철저히 통제합니다.



3. 자동 현행화가 가져오는 실무적 가치

이러한 변경 감시 에이전트의 도입은 시스템 유지보수에 필요한 수작업을 최소화하여 운영 비용을 절감하는 효과를 가져옵니다. 관리자가 매일 홈페이지나 법령 사이트를 확인하고 데이터를 수동으로 재학습시킬 필요가 없어집니다. 무엇보다 중요한 것은 감사 대응 능력의 향상입니다. 에이전트는 특정 지침이 언제 변경되었고, AI 시스템에 언제 반영되었는지에 대한 로그(Log)를 자동으로 기록합니다. 이를 통해 특정 시점의 행정 처리가 당시의 유효한 규정에 근거하여 적법하게 이루어졌음을 객관적으로 증명할

수 있는 환경을 제공합니다. 이는 복잡하고 무거운 온톨로지 재구축 과정 없이도, 시스템의 신뢰성을 지속적으로 유지할 수 있는 가장 효율적인 방안입니다.



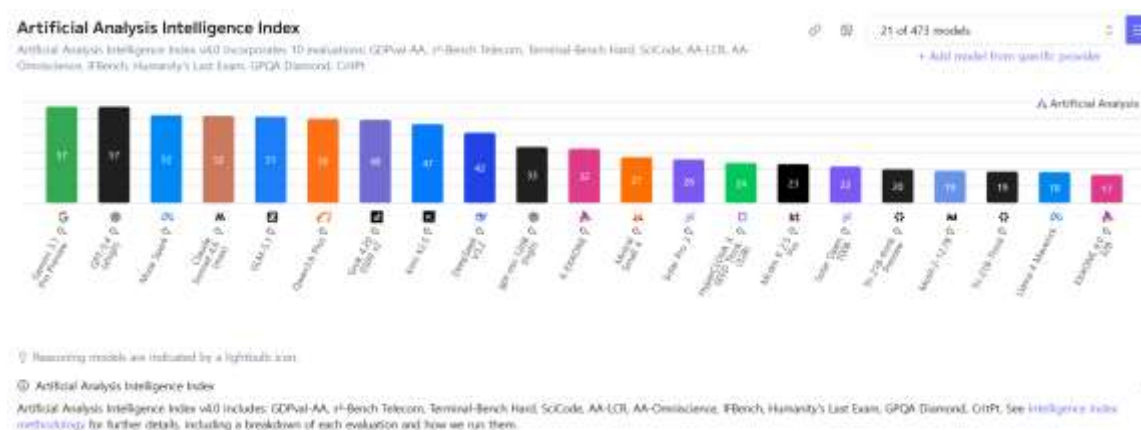
V. 공공 AI 에이전트의 두뇌: 글로벌 LLM 도입의 필연성과 철저한 보안

공공기관의 업무를 실질적으로 보조하는 AI 에이전트를 구축하기 위해서는, 출처를 통제하는 '메타데이터 RAG' 기술뿐만 아니라 복잡한 행정 문맥을 이해하고 추론할 수 있는 강력한 두뇌, 즉 거대언어모델(LLM)이 필요합니다. 스캐터엑스는 기술의 객관적 현실을 직시하고, 국가 보안 규정을 엄격하게 준수하면서도 최고의 성능을 낼 수 있는 합리적인 인프라 아키텍처를 제안합니다.

1. 에이전트 구현을 위한 LLM 성능의 현실: 글로벌 LLM 필수 시대와 유일한 국산 대안

최근 공공부문에서는 데이터 주권과 보안을 이유로 국내 기업이 개발한 LLM 이나 범정부 AI 공통 기반 모델의 도입이 우선적으로 검토되고 있습니다. 국내 AI 기술 발전을 위한 이러한 방향성에는 깊이 공감합니다. 그러나 단순한 챗봇을 넘어, 법령 간의 충돌을 분석하고 복합적인 민원 처리 가이드를 스스로 기획하는 'AI 에이전트'를 현장에 투입하기 위해서는 모델의 객관적인 추론 능력을 냉정하게 평가해야 합니다.

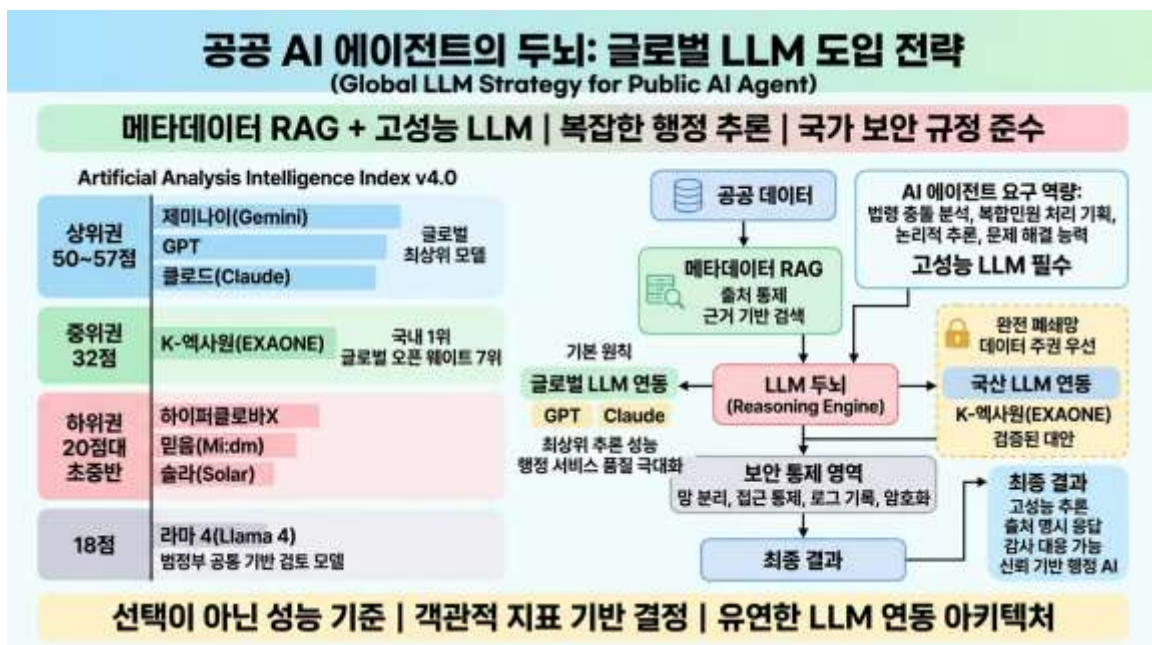
글로벌 AI 모델 성능 평가 플랫폼인 'Artificial Analysis'가 발표한 최신 'Intelligence Index v4.0' 지표를 살펴보면 현실적인 기술 격차가 뚜렷하게 나타납니다. 복잡한 코딩, 수학적 추론, 논리적 문제 해결 능력을 종합한 이 지표에서 상위권(50 점~57 점 대)은 제미니(Gemini), GPT, 클로드(Claude) 등 글로벌 빅테크 기업의 최상위 모델들이 확고히 자리 잡고 있습니다. 반면, 현재 공공 도입이 논의되는 하이퍼클로바 X, 민음(Mi:dm), 솔라(Solar) 등 대다수의 국내 주요 모델들은 20 점대 초중반의 하위권에 머물고 있습니다. 특히 범정부 공통 기반으로 검토된 바 있는 오픈소스 라마 4(Llama 4) 모델 역시 18 점 수준으로, 고도의 행정 추론을 맡기기에는 성능적 한계가 명확합니다.



이러한 객관적 지표는 자율적 판단과 복잡한 문제 해결이 요구되는 공공 AI 에이전트 구축에 있어, 글로벌 최상위 LLM(GPT, Claude 등)의 도입이 선택이 아닌 필수적임을 시사합니다. 행정의 사소한 오판이 국가적 신뢰 하락으로 이어질 수 있는 만큼, 압도적인 추론 능력을 갖춘 글로벌 모델의 활용은 행정 서비스의 질을 담보하는 가장 확실한 방법입니다.

다만, 데이터 주권 확보와 완전한 폐쇄망 구축을 위해 부득이하게 국산 LLM 도입을 최우선으로 고려해야 하는 기관이라면, 객관적인 지표를 통해 검증된 대안을 선택해야

합니다. 앞서 언급한 지표에서 **LG AI 연구원의 'K-엑사원(EXAONE)' 모델은 32 점을 기록하며 국산 모델 중 유일하게 유의미한 성능을** 보여주고 있습니다. 이는 글로벌 오픈 웨이트 모델 기준 세계 7 위, 국내 1 위에 해당하는 성과로, 글로벌 최상위 모델에는 미치지 못하지만 다른 국산 모델들이나 범정부 공통 기반 모델(라마 4)과 비교할 때 상대적으로 안정적인 추론 능력을 기대할 수 있는 유일한 대안입니다. 따라서 스캐터엑스는 글로벌 LLM 도입을 기본 원칙으로 하되, 기관의 특수한 보안 요건에 따라 국산 모델 도입이 필수적인 경우 K-엑사원을 최적의 대안으로 연동하는 유연한 아키텍처를 지원합니다.



2. 국정원 'AI 보안 가이드 2.0' 해설 및 스캐터엑스의 준수 방안

글로벌 LLM의 우수한 성능을 인지함에도 불구하고 공공기관이 도입을 주저하는 가장 큰 이유는 '데이터 외부 유출'에 대한 우려 때문입니다. 국가정보원은 이러한 공공 현장의 고민을 해소하고 안전한 AI 활용을 지원하기 위해 2025년 12월 10일 '국정원 AI 보안 가이드 2.0'을 새롭게 발표하였습니다.

이 가이드라인의 핵심 취지는 무조건적인 망분리와 AI 사용 금지가 아닙니다. 데이터의 중요도와 민감도에 따라 등급을 세분화하고, 민감 정보가 외부의 AI 모델로 유출되지

않도록 '다층적인 보안 통제 체계'를 마련하여 안전하게 AI의 편익을 누리라는 것입니다. 특히 업무망과 인터넷망 간의 데이터 연계 시 개인정보 및 비공개 업무 자료에 대한 철저한 식별과 마스킹(Masking), 그리고 API 연동 구간의 암호화 및 접근 통제를 강도 높게 요구하고 있습니다.

스캐터엑스는 이러한 국정원의 보안 가이드 2.0을 시스템 설계의 최우선 원칙으로 삼고 있습니다. 우리는 공공기관의 내부 데이터를 외부 클라우드에 위치한 글로벌 LLM으로 무분별하게 전송하지 않습니다. 스캐터엑스의 아키텍처는 기관의 내부망(폐쇄망)에 독립적인 '보안 전처리 구역(Security Pre-processing Zone)'을 구축하는 하이브리드 방식을 채택하고 있습니다.

공개되어도 무방한 일반적인 법령이나 홈페이지 안내 자료는 메타데이터 RAG를 거쳐 유연하게 처리하되, 주민등록번호, 연락처, 내부 결재 정보 등이 포함된 민감 데이터는 내부망을 절대 벗어나지 않도록 통제합니다. 글로벌 LLM(또는 대안으로서의 K-엑사원)은 오직 '지능(추론 능력)'만을 제공할 뿐, 공공기관의 원천 데이터는 내부망에서 철저히 관리되는 구조입니다. 이는 국정원이 요구하는 데이터 등급별 통제와 외부 연계 보안 기준을 충실히 이행하는 가장 모범적인 설계라 할 수 있습니다.



3. RTX 4090 기반 내부망 전처리 및 비식별화(DLP) 기술 상세

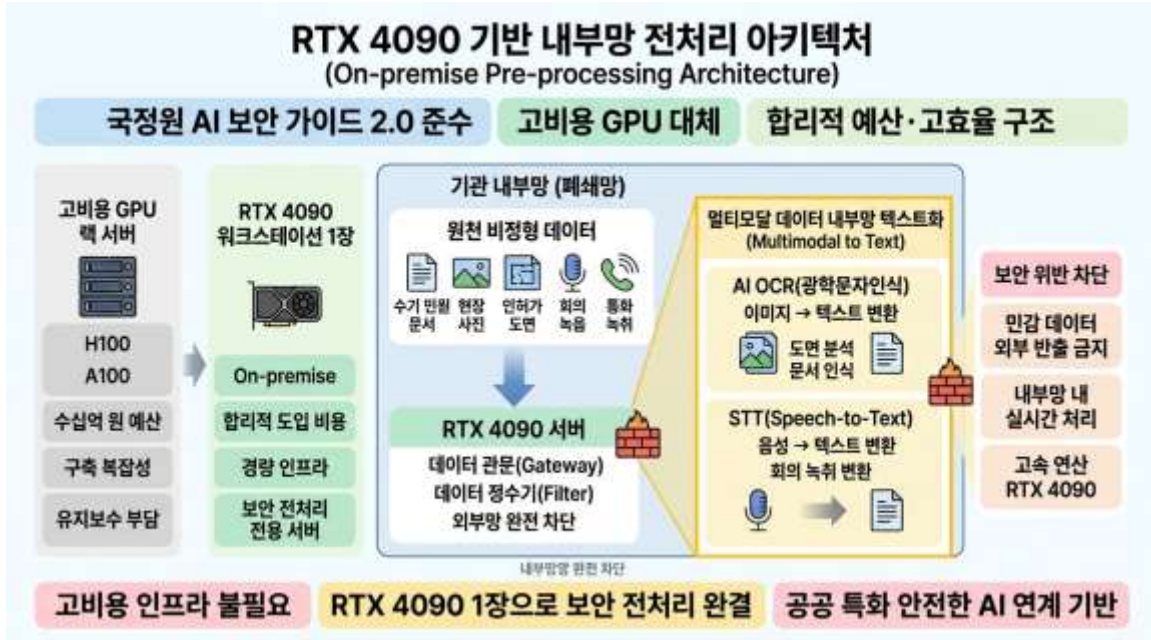
스캐터엑스가 국정원 보안 가이드 2.0 을 준수하면서 글로벌 LLM 을 안전하게 활용할 수 있는 기술적 핵심은 **‘엔비디아(NVIDIA) RTX 4090 워크스테이션을 활용한 온프레미스(On-premise) 전처리 및 비식별화 아키텍처’**에 있습니다.

통상적으로 내부망에 AI 인프라를 구축한다고 하면, 수십억 원에 달하는 H100 이나 A100 GPU 랙 서버 도입을 떠올립니다. 하지만 이는 예산 측면에서 큰 부담이 될 뿐만 아니라, 시스템 구축과 유지보수에도 많은 어려움이 따릅니다. 스캐터엑스는 이러한 고비용 구조를 탈피하여, 시중에서 합리적으로 접근 가능한 RTX 4090 GPU 단 1 장으로 내부망의 필수적인 데이터 보안 처리를 완벽하게 수행하는 효율적인 파이프라인을 완성했습니다.

이 내부망 RTX 4090 서버는 외부 LLM 으로 향하는 데이터의 ‘관문’이자 ‘정수기’ 역할을 수행합니다. 작동 프로세스는 크게 세 단계로 나뉩니다.

첫째, 멀티모달 데이터의 내부망 텍스트화(AI OCR 및 STT)입니다.

국민신문고에 접수된 민원인의 수기 작성 문서나 현장 사진, 복잡한 인허가 도면 등은 이미지 형태의 비정형 데이터입니다. 또한, 부서 내부의 중요 회의 녹음 파일이나 악성 민원인의 통화 녹취는 음성 데이터입니다. 이러한 비정형 데이터들을 외부 API 로 전송하여 변환하는 것은 심각한 보안 위반입니다. 스캐터엑스는 RTX 4090 서버 내부에 경량화된 고성능 AI OCR(광학문자인식) 엔진과 STT(Speech-to-Text) 모델을 탑재하였습니다. 이를 통해 민감한 이미지와 음성 파일은 외부망과 완전히 차단된 내부망 안에서 빠르고 정확하게 텍스트 데이터로 변환됩니다.



둘째, AI 기반 데이터 유출 방지(DLP) 및 완벽한 비식별화 처리입니다.

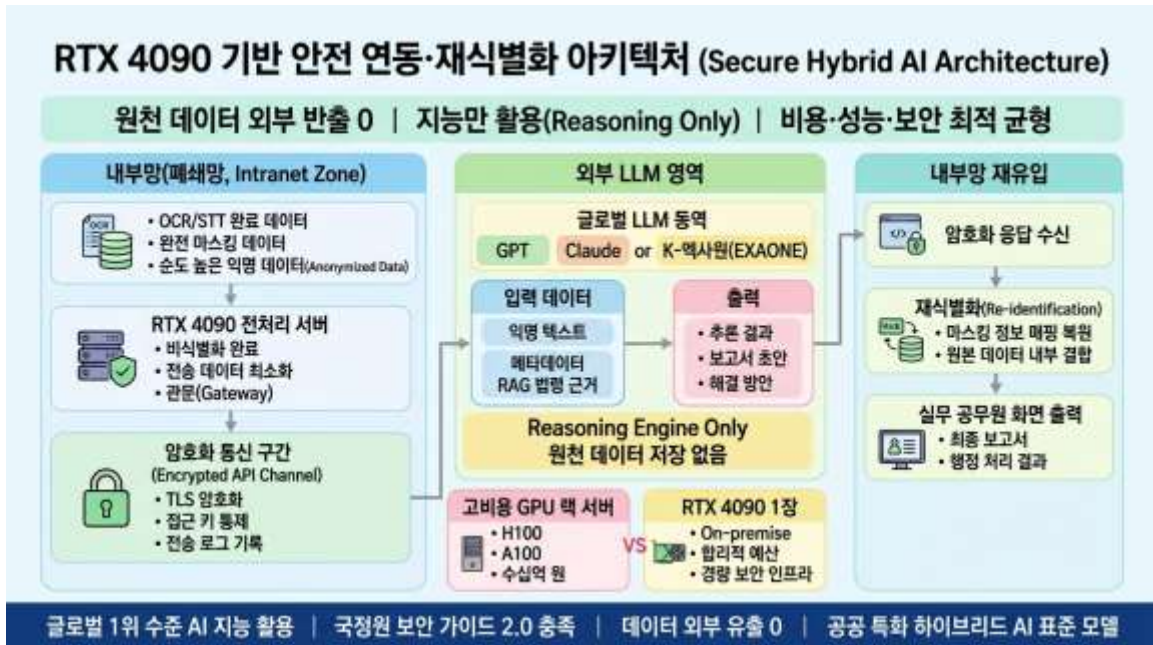
텍스트로 변환된 데이터 내부에는 여전히 민원인의 실명, 주민등록번호, 주소, 전화번호를 비롯해 외부에 공개되어서는 안 될 기관 내부의 기밀 정보가 섞여 있습니다. RTX 4090 서버는 이 텍스트를 분석하기 위해 자연어 처리(NLP) 기반의 특화된 소형 보안 언어모델(sLLM)을 구동합니다. 이 모델은 문맥을 파악하여 단순한 숫자 패턴뿐만 아니라, 이름이나 직책 등 문맥 속에 숨어 있는 개인정보와 민감 정보를 찾아내어 완벽하게 비식별화(마스킹, 예: 홍길동 → ○○○) 처리합니다. 이 과정은 오직 비식별화에만 특화되어 설계되었기 때문에 RTX 4090 1 장의 연산 능력만으로도 지연 없이 신속하게 처리됩니다.



셋째, 순도 높은 데이터의 안전한 연동과 결과물 재식별화입니다.

내부망에서 OCR/STT 변환과 완벽한 마스킹 처리를 거친 데이터는 이제 개인정보 유출의 위험이 전혀 없는 '순도 높은 익명 데이터'가 되었습니다. 스캐터엑스 시스템은 오직 이 정제된 텍스트와 메타데이터 RAG 에서 추출한 법령 근거만을 암호화된 통신 구간을 통해 글로벌 LLM(또는 K-엑사원) API 로 전송합니다. 외부의 LLM 은 전달받은 익명 데이터를 바탕으로 고도의 추론을 수행하여 해결책이나 보고서 초안을 작성해 내부망으로 회신합니다. 내부망 서버는 회신받은 결과물에 마스킹해 두었던 원본 정보를 다시 매핑(재식별화)하여 실무 공무원의 화면에 최종 출력합니다.

이러한 스캐터엑스의 RTX 4090 기반 전처리 아키텍처는 막대한 하드웨어 예산 도입 없이도, 글로벌 1 위 수준의 AI 지능을 행정 업무에 도입할 수 있게 해줍니다. 원천 데이터는 단 한 글자도 외부로 나가지 않으며, 국정원의 엄격한 보안 지침을 완벽하게 충족합니다. 비용, 성능, 보안이라는 세 가지 난제를 가장 기술적이고 실용적인 방식으로 풀어낸 결과물입니다.



VI. 메타데이터 RAG 기반 공공 AI 에이전트 12 종

메타데이터 RAG 와 안전하게 통제된 LLM 의 결합은 단순한 검색 보조 도구를 넘어, 공공 행정의 실질적인 업무를 지원하는 지능형 에이전트의 든든한 기반이 됩니다. 스캐터엑스는 공직자분들이 현장에서 겪는 어려움을 깊이 이해하고, 이를 실무적으로 덜어드릴 수 있는 12 종의 공공 특화 AI 에이전트를 제안합니다.

이 에이전트들은 크게 네 가지 그룹으로 나뉩니다. 첫째, 정확한 법적 근거를 찾아 행정 오류를 막는 '**근거·정책 검색형**', 둘째, 대민 서비스의 질과 속도를 높이는 '**민원·행정 처리형**', 셋째, 반복적인 행정 소요를 줄여주는 '**회의·업무 자동화형**', 마지막으로 데이터를 바탕으로 합리적인 의사결정을 돕는 '**정책 분석·인텔리전스형**'입니다. 각 에이전트가 실무자의 책상 위에서 어떻게 작동하며 기여하는지 구체적으로 살펴보겠습니다.

[A 그룹] 근거·정책 검색형 에이전트

1. 정책·조례 변화 감시 에이전트

[도입 배경] 중앙부처의 행정 지침이나 상위 법령은 수시로 제정 및 개정됩니다. 실무자가 이를 적시에 파악하지 못하고 과거의 규정을 근거로 행정 처리를 진행할 경우, 민원 발생이나 감사 지적의 원인이 될 수 있습니다. 매일 수많은 관보와 법령 사이트를 일일이 확인하는 것은 현실적으로 큰 부담입니다.

[작동 원리(사용 데이터)] 자치법규정보시스템, 국가법령정보센터, 중앙부처 고시 게시판 등의 데이터를 연동하여 작동합니다. 조문 단위의 해시(Hash) 값을 비교하는 '변경 감시 에이전트'가 상위 법령의 제·개정 사항을 감지하면, 이를 메타데이터로 즉각 갱신합니다. 이후 지자체의 현행 조례와 상충하는 부분이 발생할 경우 관련 부서 담당자에게 알림을 제공합니다.

[기대 효과] 법령 개정 사항을 놓쳐 발생하는 행정 오류를 미연에 방지할 수 있습니다. 담당자가 관련 법규의 변경 여부를 확인하는 데 소요되는 시간을 줄여 본연의 기획 업무에 집중할 수 있으며, 조례 개정이 필요한 시점을 적기에 파악하여 입법 공백을 최소화하는 데 기여합니다.

2. 근거 자동 첨부 보고서 에이전트

[도입 배경] 공공기관의 모든 기안문과 정책 보고서는 명확한 법적 근거와 상위 지침을 바탕으로 작성되어야 합니다. 보고서의 논리를 뒷받침할 정확한 조항과 통계, 타 기관의 선례를 찾고 이를 문서 양식에 맞게 인용하는 과정은 공무원들에게 많은 시간과 노력을 요구하는 업무 중 하나입니다.

[작동 원리(사용 데이터)] 실무자가 작성 중인 보고서의 초안이나 핵심 키워드를 입력하면, 메타데이터 RAG 시스템이 내부망에 구축된 법령, 지침, 과거 결재 문서, 부처 보도자료를 검색합니다. 단순히 의미가 유사한 문서를 찾는 것을 넘어, 시행일과 소관

부서 메타데이터를 필터링하여 현재 유효한 조문만을 추출하고, 보고서의 각 문단에 출처(URL 및 조항 번호)를 주석 형태로 자동 매핑하여 제시합니다.

[기대 효과] 보고서 작성 시 가장 번거로운 자료 조사 시간을 크게 단축할 수 있습니다. 또한, 모든 정책적 주장에 공식적인 출처가 명시되므로 내부 결재 과정에서 보고서의 객관성과 신뢰도를 높이고, 상급자의 의사결정을 한층 수월하게 지원할 수 있습니다.

3. 감사 대응 에이전트

[도입 배경] 정기 또는 수시 감사가 진행될 때, 수년 전에 처리된 사업의 적법성을 소명하는 것은 매우 까다로운 작업입니다. 현재의 법령이 아닌 '당시의 시행 지침'을 기준으로 행정 행위가 이루어졌음을 증명해야 하기 때문입니다. 과거의 문서와 지침 연혁을 뒤적이는 과정은 담당자에게 큰 업무적 스트레스로 다가옵니다.

[작동 원리(사용 데이터)] 과거의 지침, 조례 연혁, 내부 기안 문서, 예산 집행 내역 데이터를 활용합니다. 특정 사업명이나 기간을 입력하면, 메타데이터 RAG가 문서의 '시행일'과 '폐지일' 메타데이터를 기준으로 검색을 수행합니다. 이를 통해 해당 행정 처리가 이루어졌던 과거 특정 시점에 유효했던 법령과 내부 지침만을 정확하게 추출하여 시계열로 정리된 소명 자료 초안을 구성합니다.

[기대 효과] 감사 수감 준비에 소요되는 막대한 행정력을 절감할 수 있습니다. 개인의 기억이나 단편적인 기록에 의존하지 않고, 시스템에 기록된 객관적인 연혁 데이터를 바탕으로 적법성을 논리적으로 소명할 수 있어 감사 대응의 심리적, 실무적 부담을 실질적으로 덜어줍니다.

4. 정책 충돌 감지 에이전트

[도입 배경] 새로운 지역 맞춤형 정책이나 조례를 기획할 때, 기존의 상위 법령이나 타 부서의 조례와 상충하는 부분이 없는지 사전에 검토하는 것은 필수적입니다. 하지만 방대한 법체계를 실무자 개인의 지식만으로 모두 대조하여 법적 충돌 가능성을 점검하는 것은 물리적인 한계가 따릅니다.

[작동 원리(사용 데이터)] 신규 정책의 기획안이나 조례 제·개정안 텍스트를 입력 데이터로 사용합니다. 에이전트는 입력된 텍스트의 맥락을 분석한 뒤, 메타데이터 RAG 를 통해 관련 있는 중앙부처 법령, 타 지자체 유사 조례, 행정심판 재결례 등을 교차 검색합니다. 이후 LLM 의 추론 기능을 활용하여 조항 간의 의미적 충돌 가능성이나 절차적 누락 위험을 식별하고, 검토 의견서 형태로 제시합니다.

[기대 효과] 신규 정책 도입 시 발생할 수 있는 법적 분쟁이나 행정적 혼선을 사전에 예방할 수 있습니다. 조례 제정 과정에서의 규제 심사나 법제 심사 기간을 단축하고, 정책의 완성도를 높여 주민들에게 더욱 안정적이고 일관된 행정 서비스를 제공하는 데 기여합니다.



[B 그룹] 민원·행정 처리형 에이전트

5. 국민신문고 민원 사전 분석 에이전트

[도입 배경] 국민신문고 등을 통해 접수되는 민원은 그 내용이 복잡하고, 수기 문서 사진이나 현장 영상, 통화 녹음 등 비정형 첨부 파일이 포함된 경우가 많습니다. 담당자가 민원의 요지를 파악하고 관련 법령을 검토하여 적절한 처리 부서를 지정하는데 많은 시간이 소요되며, 담당자의 경험에 따라 처리 품질이 달라지기도 합니다.

[작동 원리(사용 데이터)] 민원 텍스트와 첨부된 이미지, 음성 데이터를 사용합니다. 내부망의 RTX 4090 서버가 AI OCR 과 STT 를 통해 첨부 파일을 텍스트로 변환하고, DLP 시스템으로 개인정보를 마스킹합니다. 이후 안전하게 정제된 데이터를 바탕으로 메타데이터 RAG 가 관련 조례와 과거 유사 민원 처리 사례를 검색하여, 민원 요약, 적용 법규, 담당 부서 추천이 포함된 '사전 분석 리포트'를 생성합니다.

[기대 효과] 복잡한 민원의 초기 검토 시간을 대폭 줄이고, 타 부서로 민원을 이관하는 과정에서 발생하는 지연을 최소화할 수 있습니다. 일관된 법령과 선례를 바탕으로 민원을 처리하게 되어 행정의 형평성을 높이고, 신속한 처리로 민원인의 만족도를 제고하는 데 도움을 줍니다.

6. 민원 대응 근거 에이전트

[도입 배경] 지자체와 공공기관에는 복지 수당 자격, 주정차 단속 기준, 세금 납부 절차 등 반복적이고 유사한 질의가 끊임없이 접수됩니다. 이러한 반복 민원에 일일이 수작업으로 답변을 작성하는 것은 행정력의 낭비를 초래하며, 자칫 잘못된 정보를 안내할 경우 악성 민원으로 발전할 소지가 있습니다.

[작동 원리(사용 데이터)] 지자체의 공식 홈페이지 FAQ, 업무 매뉴얼, 부서별 민원 응대 이력, 그리고 현행 조례 데이터를 활용합니다. 민원 내용이 입력되면, 메타데이터 RAG 가 '대국민 공개 가능'으로 분류된 최신 매뉴얼과 조례만을 필터링하여 검색합니다. 이를 바탕으로 민원인에게 전달할 정중하고 이해하기 쉬운 답변 초안을 작성하며, 답변 하단에는 근거가 된 조례의 링크나 매뉴얼 출처를 함께 명시합니다.

[기대 효과] 단순 반복 민원에 대한 답변 작성 시간을 줄여, 공무원이 보다 심층적인 검토가 필요한 복합 민원에 집중할 수 있게 합니다. 또한, 항상 최신의 규정을 근거로

답변이 작성되므로 오안내로 인한 민원인과의 마찰을 줄이고 행정 기관의 대외적인 신뢰도를 높일 수 있습니다.

7. 인허가 자동 검토 보조 에이전트

[도입 배경] 건축 허가, 공장 설립, 영업 신고 등 인허가 업무는 제출해야 할 서류가 방대하고 검토해야 할 관련 법규가 복잡하게 얽혀 있습니다. 담당자가 수십 장의 도면과 서류를 일일이 대조하며 누락된 요건이나 법적 제한 사항을 점검하는 과정은 업무 피로도가 높고, 단순한 확인 누락(Human Error)이 발생할 여지가 있습니다.

[작동 원리(사용 데이터)] 민원인이 제출한 신청서, 첨부 도면, 스캔 문서와 함께 해당 지자체의 건축 조례, 지구단위계획 지침 등을 데이터로 사용합니다. 내부망의 AI OCR 이 제출 서류의 텍스트를 추출하면, 에이전트가 필수 서류의 누락 여부를 1 차로 점검합니다. 이어서 메타데이터 RAG 를 통해 해당 지번이나 업종에 적용되는 조례의 제한 사항을 검색하고, 서류 내용과 법규 간의 부합 여부를 비교한 검토 보조 명세서를 출력합니다.

[기대 효과] 인허가 서류의 기초적인 요건 검토를 자동화하여 전반적인 처리 시간을 단축할 수 있습니다. 담당자는 에이전트가 식별한 주요 쟁점 사항이나 누락 부분에 집중하여 심층 검토를 진행할 수 있으므로, 업무의 정확성을 높이고 민원인의 대기 시간을 줄이는 데 기여합니다.



[C 그룹] 회의·업무 자동화형 에이전트

8. 내부 회의 자동 비서 에이전트

[도입 배경] 주간 업무 보고, 부서 간 협의, 위원회 회의 등 공공기관에서는 수많은 회의가 진행됩니다. 회의 내용을 기록하고 요약하여 회의록을 작성하는 업무는 실무자에게 적지 않은 부담입니다. 시중의 상용 녹음 요약 서비스가 있지만, 정보 보안 규정상 내부망의 회의 음성을 외부 클라우드로 전송할 수 없어 도입이 불가능한 실정입니다.

[작동 원리(사용 데이터)] 내부 회의실에서 녹음된 음성 파일 데이터를 사용합니다. 보안이 유지되는 내부망의 RTX 4090 서버에 탑재된 온프레미스 STT(Speech-to-Text) 엔진이 음성을 텍스트로 변환합니다. 이후 AI DLP 시스템이 회의 중 언급된 민감 정보를 마스킹 처리한 뒤, LLM 이 전체 회의 내용을 요약하고 주요 결정 사항과 부서별 후속 조치(Action Item)를 자동으로 추출하여 문서 양식에 맞게 정리합니다.

[기대 효과] 회의록 작성에 소요되는 행정력을 크게 절감할 수 있습니다. 보안 지침을 위배하지 않으면서도 회의 내용이 신속하게 문서화되므로, 부서 간 정보 공유가 원활해지고 결정된 사항에 대한 후속 업무의 실행력을 높이는 데 긍정적인 역할을 합니다.

9. 시장·군수 인사말 자동 생성 에이전트

[도입 배경] 기관장(시장, 군수, 구청장 등)은 하루에도 여러 행사에 참석하여 축사나 인사말을 전합니다. 비서실이나 담당 부서에서는 행사의 성격, 참석 대상, 최근의 시정 현안, 중앙 정부의 정책 기조 등을 종합적으로 고려하여 매번 새로운 연설문을 작성해야 하는 고충을 겪고 있습니다.

[작동 원리(사용 데이터)] 행사의 개요, 참석자 정보, 강조하고자 하는 키워드를 입력합니다. 에이전트는 메타데이터 RAG 를 활용하여 지자체의 최근 보도자료, 시정 백서, 관련 조례 제정 현황, 지역 통계 데이터를 검색합니다. 이를 바탕으로 행사의 취지에 맞고 현재 지자체가 추진 중인 역점 사업의 맥락이 자연스럽게 녹아든 3~5 분 분량의 맞춤형 연설문 초안을 생성합니다. 어조(격식, 친근함 등) 또한 선택적으로 조절할 수 있습니다.

[기대 효과] 상황과 대상에 맞는 연설문 초안을 신속하게 확보할 수 있어 비서실과 실무 부서의 업무 부담을 줄여줍니다. 또한, 시정의 일관된 메시지와 최신 통계가 정확하게 반영되므로 기관장의 발언에 대한 신뢰도를 높이고 지역 주민과의 원활한 소통을 지원합니다.



[D 그룹] 정책 분석·인텔리전스형 에이전트

10. 생활인구 기반 정책 설계 에이전트

[도입 배경] 지역 소멸 위기 대응과 상권 활성화를 위해서는 주민등록상 인구가 아닌, 실제로 지역에 머무는 '생활인구' 중심의 정책 설계가 필요합니다. 하지만 지자체가 확보한 통신사의 유동 인구 데이터나 카드사의 매출 데이터를 실제 행정 정책이나 인프라 배치와 연결하여 유의미한 시사점을 도출하는 과정은 고도의 분석을 요하는 어려운 작업입니다.

[작동 원리(사용 데이터)] 지자체가 적법하게 확보한 통신사 유동 인구, 시간대별 카드 매출 데이터(정형 데이터)와 타 지자체의 우수 정책 사례, 관련 연구 보고서(비정형 데이터)를 결합하여 사용합니다. 에이전트는 정형 데이터를 분석하여 인구 밀집 시간대와 소비 패턴을 파악하고, 메타데이터 RAG 를 통해 이와 유사한 조건을 가진 타 지역의 성공적인 상권 활성화 조례나 공공시설 입지 사례를 찾아내어 정책 옵션으로 제안합니다.

[기대 효과] 감이나 관행에 의존하지 않고, 실제 데이터를 근거로 한 과학적인 행정 기획이 가능해집니다. 청년 유입 정책, 축제 개최 장소 선정, 공공 도서관 입지 선정 등에서 한정된 예산을 가장 효과적으로 투입할 수 있는 합리적인 의사결정을 지원합니다.

11. 경제 유발 효과 분석 에이전트

[도입 배경] 지자체에서 지역 축제나 대규모 체육 대회를 개최한 후에는 예산 투입 대비 경제적 효과를 분석하여 의회나 상급 기관에 보고해야 합니다. 그러나 행사 전후의 경제적 파급 효과를 객관적으로 수치화하고, 이를 행정 보고서 양식에 맞추어 설득력 있게 작성하는 것은 실무 부서 입장에서 전문성과 시간이 부족한 영역입니다.

[작동 원리(사용 데이터)] 특정 행사 기간 전후의 지역 내 상권 매출 변화, 대중교통 이용량, 숙박업소 예약률 등의 정형 데이터와 중앙부처의 행사 평가 지침, 과거 유사 행사의 결과 보고서를 활용합니다. 에이전트는 정형 데이터의 증감률을 계산하여 직접적인 경제 효과를 추정하고, 메타데이터 RAG 를 통해 중앙부처의 평가 기준에 부합하는 논리 구조를 가져와 '행사 성과 및 경제 유발 효과 분석 리포트' 초안을 자동 생성합니다.

[기대 효과] 외부 연구 용역에 의존하지 않고도 내부 데이터를 활용해 신속하게 성과를 분석할 수 있어 예산과 시간을 절감합니다. 객관적인 수치와 지침에 근거한 결과 보고서를 통해 향후 유사 행사 기획 시 예산 확보의 타당성을 입증하는 데 유용하게 활용할 수 있습니다.

12. 포트홀·재난 예방 에이전트

[도입 배경] 장마철 도로의 포트홀이나 겨울철 블랙아이스, 상습 침수 구역 등은 시민의 안전과 직결된 문제입니다. 사고가 발생한 후 대응하는 사후약방문식 행정에서 벗어나, 한정된 유지보수 예산을 위험도가 높은 곳에 우선적으로 배정하는 선제적인 재난 예방 및 시설물 관리 체계가 요구되고 있습니다.

[작동 원리(사용 데이터)] 국민신문고의 위험 신고 위치 데이터, 지역별 강수량 및 기온 데이터, 도로 포장 이력 등과 함께 국토교통부의 도로 유지보수 매뉴얼, 지자체 재난

대응 지침을 융합하여 사용합니다. 에이전트는 신고 빈도와 기상 조건을 분석하여 사고 발생 확률이 높은 취약 구간을 도출하고, 메타데이터 RAG 를 통해 해당 상황에 적용해야 할 매뉴얼상의 표준 정비 절차와 안전 조치 가이드라인을 담당자에게 함께 제공합니다.

[기대 효과] 데이터 기반의 위험 예측을 통해 사고를 미연에 방지하여 시민의 생명과 재산을 보호하는 데 기여합니다. 또한, 긴급 보수가 필요한 곳의 우선순위를 객관적으로 산정함으로써, 한정된 시설물 유지보수 예산을 가장 효율적이고 투명하게 집행할 수 있는 합리적인 근거를 마련해 줍니다.



Ⅶ. 온톨로지 vs 메타데이터 RAG:

공공부문 도입 적합성 비교

공공기관의 성공적인 AI 도입을 위해서는 기술의 이상적 가치와 현장의 실무적 여건 사이에서 균형 잡힌 시각이 필요합니다. 지식그래프와 온톨로지는 장기적으로 데이터의

의미적 연결성을 높이는 유용한 기술입니다. 하지만 현재 공공기관이 직면한 예산, 인력, 데이터 정비 수준을 고려할 때, 도입의 적합성을 면밀히 비교해 볼 필요가 있습니다. 스캐터엑스는 비용, 기간, 하드웨어, 보안이라는 네 가지 현실적인 기준을 바탕으로 두 기술의 차이를 분석합니다.

1. 구축 비용: 대규모 컨설팅 예산 대(對) 기존 자산의 합리적 활용

온톨로지 구축은 본질적으로 '새로운 개념 체계를 설계하는 작업'입니다. 이를 위해서는 행정 도메인 전문가와 데이터 엔지니어가 투입되어 방대한 문서를 분석하고 분류 체계를 밑바닥부터 다시 짜야 하므로, 수십억 원 단위의 대규모 예산이 수반되는 경우가 많습니다. 반면, 스캐터엑스의 메타데이터 RAG는 정부와 지자체가 이미 오랜 시간 예산을 들여 정비해 놓은 공식 홈페이지 메뉴와 법령 사이트의 계층 구조를 그대로 재활용합니다. 기존의 정보 아키텍처를 메타데이터로 흡수하기 때문에, 불필요한 컨설팅 및 재설계 비용을 최소화하여 한정된 예산 내에서 실용적인 시스템 구축이 가능합니다.

2. 구축 기간: 장기 프로젝트 대(對) 신속한 실무 투입

온톨로지는 데이터 간의 관계를 정의하고 검증하는 과정에 최소 1년 이상의 긴 시간이 소요됩니다. 이 기간 동안 관련 법령이나 부서의 업무가 변경되면 설계도를 다시 수정해야 하는 어려움이 따릅니다. 메타데이터 RAG는 이와 달리 데이터의 출처와 메타정보를 태깅하여 검색 엔진에 연동하는 직관적인 방식을 취합니다. 수개월 내에 시스템 구축을 완료하고 실무에 투입할 수 있으며, 변경 감시 에이전트를 통해 조례나 지침의 개정 사항을 즉각적으로 반영할 수 있어 행정 환경 변화에 유연하게 대응합니다.

3. 하드웨어 요구사항: 고비용 GPU 랙 서버 대(對) 워크스테이션 기반 전처리

거대한 온톨로지를 연산하고 자체적인 거대언어모델(LLM)을 내부망에서 온전히 구동하기 위해서는 수십억 원에 달하는 고사양 GPU(예: H100, A100 등) 서버 인프라가 필수적으로 요구됩니다. 이는 공공기관의 인프라 관리 부담을 가중시킵니다. 스캐터엑스는 합리적인 예산으로 접근 가능한 RTX 4090 GPU 워크스테이션을 내부망에 배치하여 OCR, STT, 비식별화 등 필수적인 전처리만을 수행하고, 무거운 추론 연산은 국정원 보안 가이드를 준수하는 선에서 안전하게 연동된 외부 LLM(또는 대안 모델)에 맡기는 하이브리드 아키텍처를 통해 하드웨어 도입 부담을 크게 낮추었습니다.

4. 정확도 및 보안: 환각의 위험 대(對) 출처 통제와 감사 추적성

온톨로지가 아무리 정교하게 구축되더라도, 생성형 AI 특유의 환각(Hallucination) 현상을 완전히 배제하기는 어렵습니다. 특히 과거의 데이터와 현재의 데이터가 섞여 있을 때, AI가 임의로 정보를 조합할 위험이 존재합니다. 메타데이터 RAG는 답변을 생성하기 전, '시행일'과 '발행 기관' 등의 메타데이터를 기준으로 검색 범위를 물리적으로 제한합니다. 이를 통해 오직 현재 유효한 법령과 지침만을 근거로 답변을 도출하며, 생성된 결과물에 명확한 출처를 표기함으로써 향후 행정 감사 시 객관적인 소명 자료로 활용할 수 있는 높은 신뢰성을 제공합니다.

온톨로지 vs 메타데이터 RAG 공공부문 도입 적합성 비교			
비용(Cost), 기간(Duration), 하드웨어(Hardware), 정확도·보안(Security & Accuracy)			
	온톨로지(Ontology)	메타데이터 RAG(Metadata RAG)	
비용(Cost)	 ✓ 신규 개념 체계 설계 ✓ 도메인 전문가 투입 ✓ 데이터 재분류 ✓ 대규모 컨설팅 	 ✓ 기존 홈페이지 메뉴 재활용 ✓ 법령 사이트 계층 구조 활용 ✓ 정보 아키텍처 흡수 	 ✓ 재설계 최소화 ✓ 예산 효율화
기간(Duration)	 ✓ 1년 이상 장기 프로젝트 ✓ 관계 정의-검증 반복 ✓ 법령 변경 시 재설계 	 ✓ 수개월 내 구축 ✓ 메타데이터 태깅 ✓ 검색 엔진 연동 	 ✓ 검색 ✓ 변경 감시 에이전트 ✓ 실시간 현행화
하드웨어 요구사항 (Hardware Requirements)	 ✓ 대규모 연산 구조 ✓ 내부망 LLM 자체 구동, H100 ✓ 고사양 GPU 랙 서버 ✓ 인프라 관리 부담 	 ✓ RTX 4090 워크스테이션 ✓ OCR ✓ STT 	 ✓ 비식별화 전처리 ✓ 외부 LLM 하이브리드 연동 ✓ 도입 부담 최소화
정확도 및 보안 (Security & Accuracy)	 ✓ Hallucination 위험 ✓ 과거-현재 데이터 혼합 가능성 ✓ 출처 자동 통제 한계 	 ✓ 시행일 기준 필터링 ✓ 발행 기관 기준 제한 ✓ 검색 범위 물리적 통제   ✓ 출처 명시 응답 ✓ 감사 추적성(Audit Trail) 확보	
이상적 기술 ≠ 현장 적합성 현실적 예산·인력·데이터 수준 고려 공공 도입 1단계 최적 해법 메타데이터 RAG 중심 전략			

Ⅷ. 메타데이터 RAG 도입이 가져오는 공공 행정의 5 대 기대 효과

스캐터엑스의 메타데이터 RAG 와 12 종의 AI 에이전트 도입은 단순한 IT 시스템의 교체를 넘어, 공공 행정의 일하는 방식을 개선하고 대국민 서비스의 질을 높이는 실질적인 변화를 이끌어냅니다. 이를 5 가지 핵심 카테고리로 나누어 설명해 드립니다.

1. [예산] 과잉 투자의 방지와 합리적인 예산 집행 (Budget Efficiency)

공공의 예산은 국민의 소중한 세금으로 이루어지는 만큼, 투자 대비 효율(ROI)을 꼼꼼히 따져야 합니다. 현재 데이터 성숙도에 맞지 않는 대규모 온톨로지 구축이나 오버스펙의 하드웨어 도입을 유보함으로써 막대한 매몰 비용을 방지할 수 있습니다. 스캐터엑스의 솔루션은 기존 공공 웹 데이터의 재활용, RTX 4090 기반의 경제적인 전처리 서버 구성, 종량제 기반의 LLM API 연동 등을 통해 가장 합리적인 예산으로 최대의 행정 보조 효과를 도출합니다.

2. [신뢰] 출처 기반의 근거 제시와 감사 대응력 확보 (Trust & Auditability)

행정 업무에서 부정확한 정보의 제공은 곧 기관의 신뢰도 하락으로 직결됩니다. 메타데이터 RAG 는 AI 가 무분별하게 정보를 조합하는 것을 통제하고, 반드시 지정된 공식 문서와 조례에 기반하여 답변하도록 설계되었습니다. 실무자는 AI 가 제시한 초안의 하단에 명시된 법령 조항과 링크를 직접 확인하고 검증할 수 있습니다. 이는 행정의 투명성을 높이고, 향후 상급 기관의 감사나 민원인의 이의 제기 시 적절한 절차와 근거에 따라 업무를 처리했음을 명확히 증명하는 든든한 방패가 됩니다.

3. [업무] 실시간 현행화와 반복 업무의 경감 (Work Productivity)

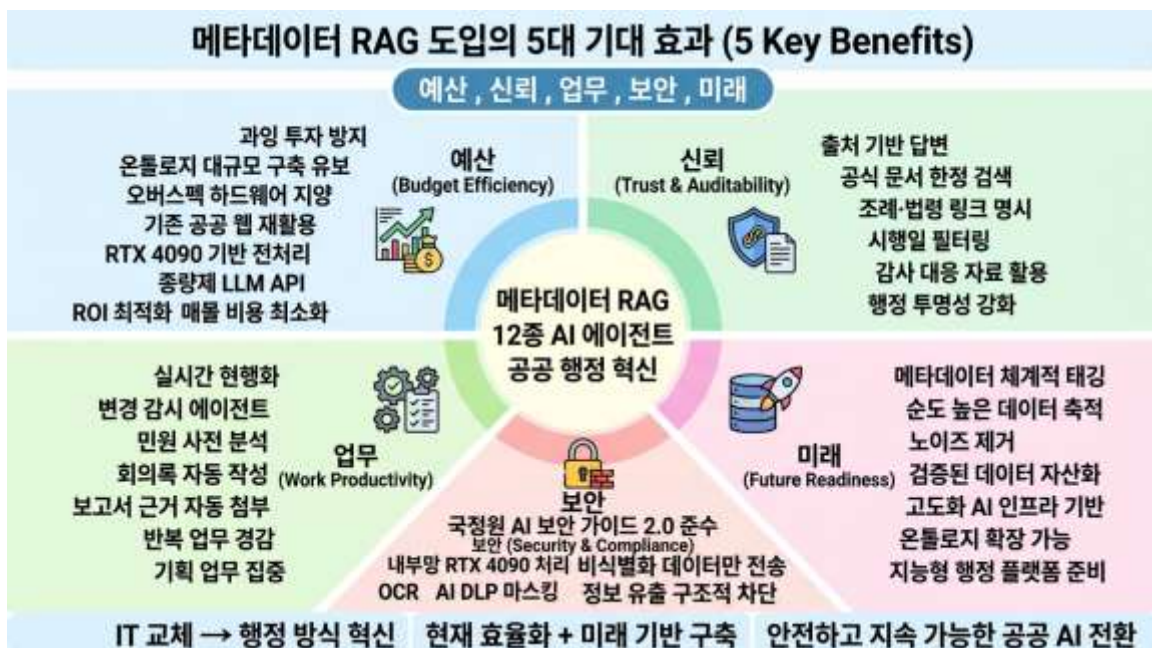
매일 쏟아지는 중앙부처의 지침 변경과 조례 개정 사항을 실무자가 일일이 추적하는 것은 큰 부담입니다. 변경 감시 에이전트가 이를 자동으로 감지하고 시스템에 반영함으로써, 과거 규정을 잘못 적용하여 발생하는 행정 오류를 미연에 방지합니다. 또한, 민원 사전 분석, 회의록 자동 작성, 보고서 근거 자동 첨부 등 12 종의 실무 밀착형 에이전트가 단순 반복적인 자료 조사와 문서 요약 업무를 대신 수행합니다. 이를 통해 공직자들은 기획, 현장 소통, 복합 민원 해결 등 보다 가치 있고 창의적인 본연의 업무에 집중할 수 있는 시간을 확보하게 됩니다.

4. [보안] 국가 보안 가이드라인 준수와 데이터 보호 (Security & Compliance)

공공 데이터의 안전한 보호는 그 어떤 기술적 편익보다 우선되어야 합니다. 스캐터엑스의 아키텍처는 국정원의 'AI 보안 가이드 2.0'을 충실히 반영하여 설계되었습니다. 민원인의 개인정보나 기관의 비공개 회의 내용 등 민감한 비정형 데이터는 내부망의 RTX 4090 서버에서 OCR 및 STT를 통해 텍스트화된 후, AI DLP를 통해 철저히 마스킹(비식별화) 처리됩니다. 외부 네트워크로는 오직 안전하게 정제된 익명 데이터만이 전송되므로, 고성능 AI를 활용하면서도 정보 유출의 위험을 구조적으로 차단하는 안전한 업무 환경을 제공합니다.

5. [미래] 정제된 데이터 축적을 통한 지능형 행정의 기반 마련 (Future Readiness)

지금 메타데이터 RAG를 도입하는 것은 단순히 현재의 업무를 편리하게 만드는 데 그치지 않습니다. 시스템을 운영하는 과정에서 공공기관의 내부 저장소에는 메타데이터가 체계적으로 태깅되고, 노이즈가 제거된 '순도 높은 데이터'가 차곡차곡 쌓이게 됩니다. 이렇게 축적되고 검증된 양질의 데이터는, 향후 기관이 더 높은 수준의 AI 인프라나 온톨로지를 구축하고자 할 때 가장 훌륭한 기초 자산이 됩니다. 실패의 위험을 줄이면서 미래의 지능형 행정 플랫폼으로 나아가는 견고한 디딤돌을 놓는 과정입니다.



IX. 결론 및 맺음말: 정책 인텔리전스를 향한 단계적 진화 로드맵

스캐터엑스는 공공기관이 인공지능을 도입하는 과정이 무리한 도약이 아닌, 안전하고 안정적인 계단 오르기가 되어야 한다고 생각합니다. 이를 위해 우리는 공공 AI 가 나아가야 할 현실적인 3 단계 진화 로드맵을 제안합니다.

1 단계: 메타데이터 RAG 를 통한 신뢰 기반 구축

가장 먼저, 무리한 구조화 작업을 내려놓고 메타데이터 RAG 를 통해 기존의 공공 데이터를 안전하게 연결합니다. 출처가 통제된 검색 환경을 조성하고, 내부망 전처리 아키텍처를 통해 보안의 기틀을 다지며 실시간 현행화 체계를 가동하여 시스템에 대한 실무자들의 신뢰를 쌓는 단계입니다.

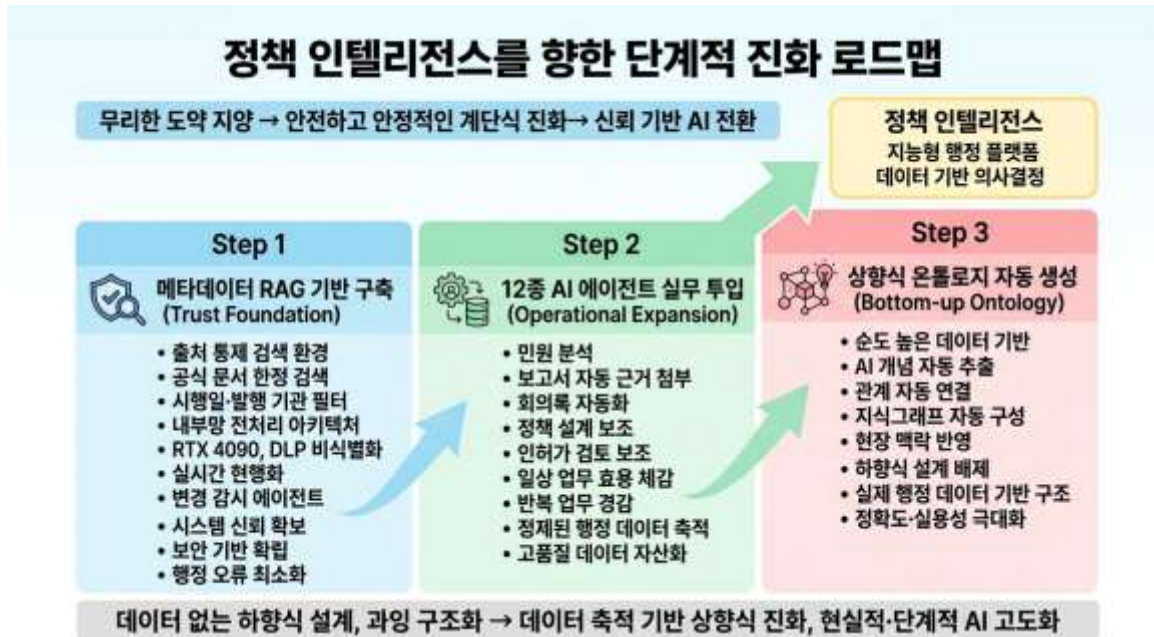
2 단계: 12 종 에이전트를 통한 실무 보조 및 데이터 축적

신뢰할 수 있는 검색 인프라 위에서, 민원 분석부터 정책 설계 보조에 이르는 12 종의 AI 에이전트를 실무에 투입합니다. 공직자들은 일상 업무에서 AI 의 효용을 체감하게 되며, 이 과정에서 시스템에는 AI 가 학습하고 활용하기 좋은 고품질의 정제된 행정 데이터가 지속적으로 축적됩니다.

3 단계: 축적된 데이터를 바탕으로 한 상향식(Bottom-up) 온톨로지 자동 생성

수년간 메타데이터 RAG 와 에이전트를 운영하며 쌓인 '순도 높은 데이터'는 그 자체로 훌륭한 지식의 연결망을 내포하고 있습니다. 이때 비로소 축적된 데이터를 바탕으로 AI 가 스스로 개념과 관계를 추출하여 온톨로지를 구성하는 상향식 접근을 시도할 수 있습니다. 데이터가 없는 상태에서 억지로 짜맞춘 하향식 설계와 달리, 실제 행정

현장의 맥락이 살아 숨 쉬는 가장 실용적이고 정확한 지식그래프가 완성되는 시점입니다.



맺음말

온톨로지 기술 자체가 틀린 것은 아닙니다. 다만, 현재 공공기관의 데이터 정비 상황을 고려할 때 그 시기가 조금 이를 뿐입니다. 지금 당장 공공 행정에 필요한 것은 막대한 예산과 긴 시간이 소요되는 거창한 청사진이 아니라, 오늘 발생한 민원을 빠르고 정확하게 처리할 수 있도록 돕는 실용적인 도구입니다.

스캐터엑스의 메타데이터 RAG 와 안전하게 통제된 AI 아키텍처는, 예산의 낭비를 방지하면서도 공직자분들의 실무적 고충을 덜어드릴 수 있는 합리적인 대안입니다. 저희 스캐터엑스는 앞으로도 현장의 목소리에 귀 기울이며, 공공기관이 가장 안전하고 효과적인 방법으로 AI 시대를 맞이할 수 있도록 겸손하고 묵묵하게 기술적 지원을 다 하겠습니다. 긴 글 읽어주셔서 깊이 감사드립니다.

[스캐터엑스 기술 백서 참고문헌 및 데이터 출처 안내]

본 백서는 타인의 지적 재산을 존중하고 정보의 객관성과 신뢰성을 확보하기 위해, 신뢰할 수 있는 국내외 기관의 공식 발표 자료와 연구 지표를 바탕으로 작성되었습니다. 출처 표기는 일반적인 학술 및 기술 문서 작성 기준(APA Style 7th ed. 등)을 준용하여 정리하였습니다.

1. 공공기관 보안 및 정책 가이드라인

백서 내 공공 망분리 환경 및 AI 도입 보안 아키텍처 설계의 근거로 활용된 국가 공식 가이드라인입니다.

- **국가정보원(2025).** 「국정원 AI 보안 가이드 2.0」. 국가정보원.
- *인용 부분:* 제 5 장 '국정원 AI 보안 가이드 2.0 해설 및 스캐터엑스의 준수 방안', 다층적 보안 통제 체계 및 내부망 비식별화(DLP) 처리 기준.

2. AI 모델 성능 및 보안 평가 지표

글로벌 LLM 과 국산 LLM 의 객관적인 성능 차이 및 보안 수준을 비교 분석하기 위해 인용한 연구 기관의 지표입니다.

- **Artificial Analysis(n.d.).** 「Artificial Analysis Intelligence Index v4.0」.
- *인용 부분:* 제 5 장 '에이전트 구현을 위한 LLM 성능의 현실', 글로벌 최상위 모델(Gemini, GPT, Claude 등)과 국내 모델(K-EXAONE, HyperCLOVA X 등) 및 라마 4(Llama 4) 모델의 추론 성능 점수 비교 데이터.
- **숭실대학교 인공지능안전성연구센터(2025).** 「국내외 파운데이션 모델 AI 보안성·안전성 평가」.
- *인용 부분:* 제 5 장 '에이전트 구현을 위한 LLM 성능의 현실', 프롬프트 인젝션 등 최신 공격 기법에 대한 국내 파운데이션 모델의 보안성 방어율(해외 모델 대비 82% 수준) 통계.

3. 스캐터엑스 자체 기술 및 실증 데이터

메타데이터 RAG 아키텍처 및 12 종 에이전트의 작동 원리를 설명하기 위해 활용된 당사의 내부 기술 명세서입니다.

- **스캐터엑스 기업부설연구소(2025).** 「메타데이터 RAG 기반 순도 높은 데이터 전처리 및 변경 감시 에이전트 기술 명세서」. 스캐터엑스 내부 자료.
- **인용 부분:** 제 3 장 '4 단계 데이터 가공 기술' 및 제 4 장 '변경 감시 에이전트 아키텍처'의 작동 원리와 실증 사례.

4. 문서 작성 기준 및 연구 윤리 참조

본 백서의 체계적인 구조화와 올바른 인용 기준을 확립하기 위해 참고한 가이드라인입니다.

- **경희대학교 도서관(n.d.).** 「APA Style 7th ed. Guide: 참고문헌 목록 일반 사항」.
- **한국저작권위원회(n.d.).** 「인용과 표절의 기준 및 올바른 출처 표기법」.
- **재능넷 지식공유(2025).** 「기술 백서(White Paper) 작성: 전문성 전달하기」.
- **참고 부분:** 기술 백서의 객관적 서술 원칙, 문제-해결형 구조 설계, 데이터와 증거 기반의 서술 방법론.

스캐터엑스는 앞으로도 검증된 데이터와 객관적인 지표를 바탕으로, 공공기관에 실질적인 도움이 되는 정확한 정보를 제공하기 위해 노력하겠습니다. 백서 내용이나 인용된 자료와 관련하여 궁금한 점이 있으시다면 언제든지 문의해 주시기 바랍니다. 감사합니다.