

# LLM 기반 고고학 정보 플랫폼의 현재와 미래

잠든 데이터를 깨워 사라지는 역사를 되살리고 세계의 표준을 열다

김홍연 (소프트웨어 엔지니어)

한국방송통신대학교 컴퓨터학과 졸업, 한국전통문화대학교 문화유적학과 졸업

본 연구는 『헤리티지: 역사와 과학』 제58권 제3호에 게재된 논문 「대형 언어 모델(LLM)을 활용한 고고학 정보화 연구 -발굴조사보고서의 메타데이터 자동 추출 파이프라인 개념 검증-」에 근거한다.

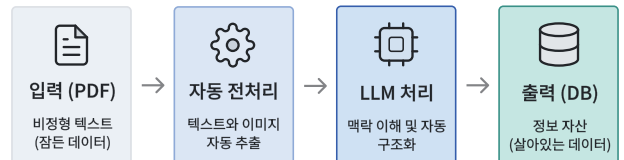
## 제1장 데이터에 갇힌 고고학, 연구의 한계가 교육의 위기로

고고학 연구는 여전히 개별 보고서를 수작업으로 분석하는 단계를 넘어서지 못하고 있다. 대규모 비교나 통계 분석 같은 새로운 연구 방법론을 적용하는 데 한계가 있다. 이러한 ‘연구 환경의 정체’는 학술적 성과가 사회와 교육 현장으로 확산되지 못하는 결과를 낳았으며, 이는 역사 교과서의 선사시대 서술 축소와 같은 교육의 위기를 초래했다.



## 제2장: 학술적 해법: LLM으로 비정형 텍스트의 맥락을 이해하다

『헤리티지: 역사와 과학』에 게재된 새로운 방법론이 문제의 근본적인 해결책은 잠든 데이터를 AI가 자유롭게 활용하는 ‘살아있는 데이터’로 만드는 것이다. 기존의 정보화 방식이 방대한 사전 설계와 표준화를 전제로 했다면, LLM은 사전에 구조를 정의하지 않아도 비정형 텍스트를 직접 이해하고 처리할 수 있다. 즉, LLM으로 정형화를 먼저 진행한 후 이를 바탕으로 역으로 정교한 분류 체계를 도출하는 것도 가능하다.



## 제3장: 기술적 실증: 데이터가 눈을 뜨는 순간

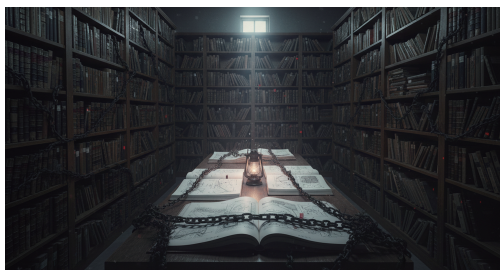
개념 검증(PoC)으로 증명된 보고서의 정보화

이론은 개념 검증(PoC)을 통해 현실이 되었다. 직접 개발한 LLM 파이프라인은 실제 발굴조사보고서 PDF로부터 유적·유구·유물의 핵심 정보를 성공적으로 추출하고, 이를 검색과 분석이 가능한 구조화된 데이터로 자동 변환했다. 잠들어 있던 텍스트가 비로소 상호작용 가능한 ‘살아있는 정보’로 깨어나는 순간을 기술적으로 실증한 것이다.



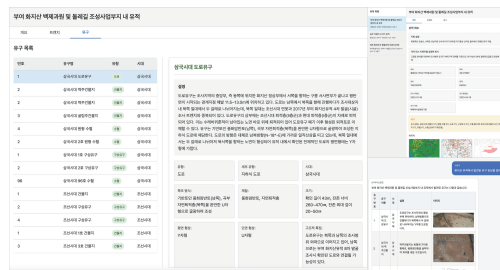
데이터가 깨어나는 전 과정을 영상으로 확인하세요.  
(poc.heripo.com)

Before



\* Gemini 모델로 만든 AI 생성 이미지입니다.

After



## 제4장: 10년의 여정: 한 고고학도의 꿈이 기술이 되기까지

### 기술에서 찾은 새로운 가능성

이러한 시도는 한 고고학도의 10년 이상에 걸친 집념의 산물이다. 제도적 해결을 모색하다 한계를 깨달은 후 “현장을 바꾸는 것은 기술”이라는 결론 아래 IT 엔지니어가 되어 오랜 꿈을 코드로 직접 실현하고 있다.



2013: 문제의식



2015: 제도적 접근



2025: 기술적 실현

## 제5장: 가장 현실적인 길: 왜 개인의 R&D로 시작하는가?

### ‘선(先)증명, 후(後)협력’ 전략

문제의 핵심은 구조적 딜레마에 있다. 이 딜레마의 현실적 해법은 민간이 먼저 기술 가치를 증명하여 국가의 협력을 유도하는 새로운 상향식 접근이며, 본 프로젝트는 그 가능성을 여는 ‘마중물’이다.

#### 전통적 방식 (Top-down)

복잡하고 느린 흐름



#### 본 연구 방식 (Bottom-up)

간결하고 빠른 흐름



## 제6장: 본 연구가 꿈꾸는 미래: 살아 숨 쉬는 데이터 생태계

본 프로젝트는 단기적으로 구조화된 데이터 탐색·분석 플랫폼을 구축, 장기적으로는 데이터 생성부터 출판까지 전 과정을 잇는 ‘통합 워크플로우 솔루션’을 완성하여 살아 숨 쉬는 고고학 생태계를 구축하는 것을 최종 목표로 한다.

#### 0단계 (완료): 데이터 활용

‘리서치 레이더’ 출시

#### 1단계 (현재): 데이터 깨우기

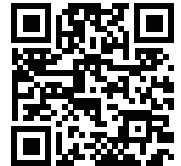
보고서 자동 구조화

#### 3단계 (단기): 데이터 융합

무료 플랫폼 제공

#### 4단계 (장기): 데이터 창조

통합 워크플로우 솔루션



리서치 레이더

## 제7장: 제언: ‘정보고고학’, 연구의 새로운 지평을 열다

### 고고학 연구를 향해: 질문의 한계를 넘어서다

“살아있는 데이터는 연구자들을 단순 반복 작업에서 해방 시킬 것이다. 이제 과거에는 불가능했던 질문을 던질 수 있다. 전국 단위 토기 형식의 시기별 분포 변화를 1시간 만에 시각화하고, 특정 유구의 유형별 출토 유물 조합을 통계적으로 분석하는 등 새로운 차원의 연구가 가능해진다.”

### 사회와 세계를 향해: 새로운 가능성을 열다

“연구의 깊이가 깊어질 때 고고학의 목소리는 더 큰 힘을 갖게 된다. 살아있는 데이터로 우리 역사의 99%를 복원하여 교과서에 그 깊이를 더하고, K-신석기가 넷플릭스로 가는 문화콘텐츠의 길을 열 것이다. 나아가 이 모델은 세계 고고학계의 새로운 표준이 될 잠재력을 품고 있다.”