

LLM 기반 고고학 정보 플랫폼의 현재와 미래

김흥연(한국방송통신대학교)

디지털과 AI 기술이 발전하는 시대적 흐름 속에서도 고고학 데이터는 텍스트 기반의 발굴조사보고서 속에 ‘잠들어 있어’ 체계적인 활용이 어렵다. 결국 수십 년간 축적된 방대한 발굴 성과는 대중적 지식이나 교육 콘텐츠로 가공·확산되지 못하는 구조적 한계에 부딪혔다. 2022년 개정 역사 교과서에서 선사시대 서술이 사실상 삭제된 것¹⁾도 이와 무관하지 않다고 본다.

고고학계에서도 디지털과 AI를 활용한 연구가 활발해지는 흐름이다. 하지만 대부분의 기반 데이터가 발굴조사보고서 속에 ‘잠들어’ 있어, 분석에 앞서 막대한 수작업 전처리와 선행되어야 한다는 공통된 한계에 부딪힌다. 결국 현재의 연구들은 특정 유적이나 유물에 국한된 소규모 분석에 머물 뿐, 거시적 문화 연구로 확장되기 어려운 구조적 제약을 안고 있다.

이러한 구조적 제약을 극복하기 위해서는 ‘잠들어 있는’ 데이터를 연구자, 교육자, AI가 자유롭게 활용하는 ‘살아있는 데이터’로 전환해야 한다. 이에 발표자는 선행 연구를 통해, 대형 언어 모델(LLM)이 발굴조사보고서의 맥락을 이해하고 유적·유구·유물 정보를 자동으로 추출·구조화하는 파이프라인을 제안하고 그 가능성을 검증했다²⁾. 그 결과 수개월이 걸리던 데이터베이스 구축 작업을 단 몇 분 만에 완료할 수 있는 혁신적인 가능성을 확인했다.

이러한 시도는 10년 이상에 걸친 집념의 산물이다. 발표자는 고고학 학부 시절의 문제 의식에서 출발하여 제도적 해결을 위해 정책학 석사과정에 진학하는 등 다각도의 노력을 기울였으나, 현장의 근본적인 변화를 이끌기에는 명확한 한계가 있음을 깨달았다.

1) 미래엔 웹사이트에서 고등학교 교과서(<https://22txbook.m-teacher.co.kr/book/view.mrn?id=81>)와 중학교 교과서(<https://22txbook.m-teacher.co.kr/book/view.mrn?id=29>)를 웹 기반 e-book으로 바로 살펴볼 수 있다.

2) 김흥연, 2025, 「대형 언어 모델(LLM)을 활용한 고고학 정보화 연구-발굴조사보고서의 메타데이터 자동 추출 파이프라인 개념 검증-」, 『헤리티지: 역사와 과학』 58(3), 국립문화유산연구원.

문제의 핵심은 구조적 딜레마에 있었다. 국가 주도의 하향식(Top-down) 사업은 급변하는 기술 환경에 대응하기엔 경직적이고, 매장유산 분야의 낮은 정책 우선순위로 인해 대규모 예산 확보도 어렵다. 이 딜레마의 현실적 해법은 민간이 먼저 기술의 가치를 실체로 증명하고 국가의 협력을 유도하는 새로운 상향식(Bottom-up) 접근에 있다. 본 프로젝트는 그 첫걸음이자, 한국고고학회가 제시한 ‘디지털 시대의 적응’과 ‘고고학의 외연 확장’이라는 과제³⁾에 대한 한 구성원의 실천적 응답이다.

본 프로젝트는 구체적인 로드맵에 따라 단계적으로 실현되고 있다. 이미 ‘리서치 레이더’(0단계) 출시⁴⁾를 통해 기술적 실현 가능성을 입증했으며, 현재는 LLM 기반 보고서 자동 구조화 파이프라인(1단계)을 통해 ‘잠들어 있는 데이터’를 ‘깨우는’ 작업에 집중하고 있다. 단기적으로는 이렇게 구조화된 데이터를 누구나 쉽게 탐색·분석하는 ‘데이터 융합 플랫폼’(2단계)을 구축하여, 전문 연구의 기반을 다지는 동시에 고고학 지식의 대중적 외연을 확장하고자 한다. 장기적으로는 데이터가 생성되는 현장부터 출판까지 전 과정이 유기적으로 연결되는 ‘통합 워크플로우 솔루션’(3단계)을 완성하여, 데이터가 ‘태어날 때부터 살아 숨 쉬는’ 고고학 연구 생태계를 구축하는 것을 최종 목표로 한다.

궁극적으로 본 발표는 고고학 빅데이터가 살아 숨 쉬는 ‘정보고고학’ 패러다임의 시작과 그 구체적인 로드맵을 제시한다. 살아있는 데이터는 교과서에 우리 역사의 깊이를 더하고 문화적 상상력의 원천이 될 것이며, 나아가 LLM 기반의 차세대 데이터 표준을 선도하여 ‘한국 고고학의 세계화’라는 한국고고학회의 비전⁵⁾을 실현하는 열쇠가 될 것이다.

3) “제30대 한국고고학회 이성주 회장 새해 인사(한국고고학회 공지사항)”, 한국고고학회, <https://www.kras.or.kr/?c=5/37&uid=202160>.

4) “리서치 레이더 구독”, heripo, <https://heripo.com/research-radar/subscribe>.

5) 앞의 한국고고학회 학회장 새해 인사 웹페이지.