AI 시대 역사학의 신뢰와 대중적 설득

사료 해석의 한계와 LLM을 활용한 연구 및 공공역사

2026년 11월 6일

역사학의 과제는 객관성을 보증할 수 없는 조건에서도 더 믿을 만한 설명을 만들고, 그 설명이 사회에서 작동하도록 하는 데 있다. 대규모 언어모델인 LLM은 한 연구자가 검토할 수 있는 자료와 대안의 범위를 넓힐 수 있다. 그러나 처리량의 확대가 곧 진실의 확보를 뜻하지는 않는다. 또한 연구의 정교함만으로, 확신과 소속감을 주는 서사를 원하는 대중을 설득할 수 있는 것도 아니다.

이 글은 김율 교수의 부사와 스콜라철학 강연에서 출발해, 박구병 교수의 진실·화해위원회 강연 기록을 구체적 사례로 삼는다. 중심 질문은 두 가지다. LLM의 양적 능력으로 역사가 개인의 해석에 대한 의존을 어떻게 줄일 것인가. 그리고 그렇게 얻은 잠정적 설명을, 권위 있는 정답을 원하는 사회에 어떤 형태로 전달할 것인가.

연구 도구의 확대와 역사학의 변화

로베르토 부사의 Index Thomisticus는 토마스 아퀴나스와 관련 저작의 단어를 변화형과 표제어에 따라 정리하고, 등장 위치와 문맥을 찾아볼 수 있게 한 색인이다. 단어의 빈도를 세는 기능도 있지만, 핵심은 개념이 실제 문장에서 어떻게 사용되었는지 비교할 수 있다는 데 있다. 김율 교수는 이러한 도구를 활용해온 경험과 최근 AI를 통한 옛 문헌의 전사 작업을 소개한다.[1]

여기서 "최종 판단은 인간의 몫"이라는 원칙은 책임의 소재를 밝힌다. 다만 AI와 함께 어떤 새로운 연구를 할 것인가에 대한 답은 더 필요하다. 자료를 찾고 읽는 범위가 달라지면, 연구자가 설정하는 질문과 비교 단위도 달라질 수 있다. 도구가 기존 연구의 속도를 높이는 데서 그칠지, 연구의 구조를 바꾸게 될지가 쟁점이다.

앞으로의 역사학은 AI에게 해석을 맡기는 문제와, AI를 사용해 자신의 해석이 받는 검토를 확대하는 문제를 구별해야 한다. 후자의 방향에서는 연구자가 마지막에 승인만 하는 사람이 아니라, 자료 선택과 분류, 가설 비교의 전 과정에 참여한다.

객관성을 보증할 수 없는 조건

국가기록은 통치의 정당화나 조직의 자기보호를 위해 작성·편집될 수 있다. 민간의 증언과 구전에도 기억의 변형, 전승 과정의 생략, 현재의 이해관계가 개입할 수 있다. 어느 한 종류의 기록을 객관성의 최종 보증으로 삼기는 어렵다. 그렇다고 모든 설명을 같은 정도로 믿어야 하는 것은 아니다.

가령 통치를 정당화하기 위해 작성된 기록은 사건의 충실한 보고로 읽기 어렵더라도, 국가가 무엇을 정당화해야 했고 어떤 분류로 사람을 관리했는지 보여줄 수 있다. 사료의 편향은 제거해야 할 잡음이면서 동시에 연구해야 할 역사적 흔적이기도 하다. 연구자는 기록이 주장하는 내용과, 그 기록이 만들어졌다는 사실에서 얻는 정보를 구별해야 한다.

따라서 객관성의 보증이 없다는 명제와 믿을 정도를 높이는 메커니즘이 없다는 명제는 다르다. 독립적인 경로에서 나온 흔적들을 함께 설명하는가, 불리한 증거마다 예외를 추가해야 유지되는가, 어떤 자료가 나오면 해석을 수정할 것인가를 물을 수 있다. 이런 검토는 진실을 보증하지 않지만 설명 간의 차이를 드러낸다.

역사학이 이념이나 신앙처럼 굳어지는 위험은 주관이 개입한다는 사실만으로 발생하지 않는다. 어떤 증거가 나와도 자기 설명을 보호할 수 있게 되는 순간 그 위험이 커진다. 과학과 종교라는 비유에서 유효한 쟁점도 이 수정 가능성이다. 학문의 신뢰를 연구자의 선의나 권위에만 맡겨두기 어렵다는 뜻이다.

LLM의 양적 능력과 더 진한 회색

"더 진한 회색"은 불확실성을 없애지 못하더라도, 자료와 해석을 더 폭넓게 대조해 자의적 판단의 여지를 줄이려는 지향이다. LLM의 가치는 한 사람이 감당하기 어려운 규모로 반대 사료 탐색, 경쟁 가설 생성, 다른 지역과 언어권의 비교를 반복할 수 있다는 가능성에 있다.

그러나 LLM의 포괄성과 처리량은 실제로 무한하지 않다. 답변을 많이 생성하는 일은 독립적인 증거를 늘리는 일과도 다르다. 같은 보고서를 인용한 글 천 편은 독립된 증거 천 개가 아니다. 서로 다른 AI의 동의 역시 독립된 관찰자의 일치로 곧바로 계산할 수 없다. 여러 LLM 사이에서 오류가 겹친다는 연구는 이 점을 경계하게 한다. 해당 연구가 역사 해석을 직접 검증한 것은 아니라는 한계도 함께 보아야 한다.[4]

양을 늘릴 때 세어야 할 것은 문장의 수보다 새로 확인한 증거의 경로, 실제로 검토한 경쟁 설명, 기존 해석의 취약점이다. 이야기 생성 능력이 커질수록 어떤 자료에도 그럴듯한 설명을 붙여 반증을 회피하는 능력도 커질 수 있다. 자료 탐색과 가설 생성을 평가 절차에 연결해야 하는 이유다.

박구병 강연 기록에서 얻는 연구 문제

박구병 교수의 강연은 라틴아메리카의 과거사 정리와 한국의 과제를 비교한다. 기록의 생산적 대목은 질의응답에서 나타난다. 강연자가 자료 접근과 인력의 제약을 설명한 뒤, 피해 당사자이자 관련 기구 활동 경험이 있는 정지석 동문은 이미 존재하는 명단에도 개인별 신청과 입증을 반복하는 조사 방식이 지연을 만드는 것 아니냐고 묻는다. 박 교수도 사건 전체보다 개인별 조사로 해석하는 관행을 인정한다.[2]

이 반론은 "여러 시각이 존재한다"는 수준을 넘어, 조사 지연의 원인을 비교할 수 있는 가설로 바꾼다. 다음 표는 그 문제 제기에서 도출한 연구 제안이며, 강연자가 입증한 결론은 아니다.

경쟁하는 설명	추가로 확인할 차이
자료와 인력이 부족하다	자료 접근이나 담당 인력이 개선된 뒤 처리 속도가 달라지는가
개인별 입증 절차가 병목이다	자료가 충분한 사건에서도 중복 확인과 대기가 반복되는가
정치적 부담이 진척을 제약한다	비슷한 자료·인력 조건에서 사건 성격에 따라 진척이 달라지는가

여러 원인이 동시에 작용할 수 있다. 자료의 범위와 비교 조건을 밝힌 뒤, 사건 처리 이력이나 보완 요구, 담당자 배정, 공개 회의록 등을 대조해야 한다. 관찰되는 차이를 곧바로 인과관계로 단정해서는 안 되며, 현재 자료로 가설들을 구별할 수 없다는 결과도 남길 수 있어야 한다.

개인별 기록에서 사건의 구조로

접근이 허용된 명단, 학적부, 군 관련 기록, 기존 결정서를 연결하면 같은 시기와 경로로 처리된 사례를 찾을 수 있다. AI는 표기와 서술이 다른 기록 사이의 연결 후보를 제시하는 데 활용할 수 있다. 연구자는 이를 검증해 개별 피해 뒤의 반복되는 행위와 조직적 절차를 재구성한다. 명단상의 연관성만으로 개인의 피해나 책임을 확정하지 않는 구분이 필요하다.

위원회 목록에서 피해 인정의 과정으로

강연의 칠레 사례는 위원회 종료 뒤에도 배상 기구, 추가 조사, 사법 절차가 이어지는 과정을 보여준다.[2] 이를 국가별 후속 기록과 연결하면, 위원회의 존속 기간보다 피해자가 인정과 배상에 도달하는 경로와 시간을 비교할 수 있다. 국가마다 조사 범위와 인정 기준이 다르므로, 공통 이름만으로 수치를 합쳐서는 안 된다.

공개 기록과 AI의 역사 이해

강연을 공개 웹에 올리는 작업은 AI가 역사에 답할 때 참조할 자료를 늘릴 수 있다. GEO는 Generative Engine Optimization, AEO는 Answer Engine Optimization의 약자다. 둘 다 AI의 검색과 답변에서 콘텐츠가 발견되고 활용되도록 하는 활동을 가리키며, 지식의 정확성이나 균형 자체를 뜻하지는 않는다.[3]

웹에 공개된 뒤에도 검색에 수집되고, 해당 질문에서 선택되고, 답변에 정확하게 반영되는 단계가 남는다. AI가 검색한 글을 참고하는 것과 모델 학습에 반영되어 지속적으로 바뀌는 것도 별개다. 공개 글의 의의는 우선 맥락과 이견을 갖춘 참조 자료를 만들었다는 데 있으며, 실제 답변 개선은 별도로 확인해야 한다.

박구병 교수 강연 기록은 발표 슬라이드와 발화, 질문자와 답변을 구분한다. 이를 활용하면 강연자의 주장뿐 아니라 그 주장에 대한 현장의 저항까지 전달할 수 있다.[2] 더 나아가 반론을 후속 자료와 연결하고, 무엇이 해결되었으며 어떤 해석이 바뀌었는지 갱신한다면 공개 기록은 지속적인 공동 연구의 기반이 된다.

출처의 보존과 출처 비판

출처를 달았다는 사실만으로 좋은 역사 설명이 완성되지는 않는다. 출처들이 같은 원문을 되풀이할 수 있고, 국가기관이 사용한 분류 자체가 연구 대상을 왜곡할 수도 있다. 누가 어떤 목적으로 기록했는지, 다른 자료와 독립적인지, 누구를 빠뜨렸는지에 대한 검토를 함께 보존해야 한다.

강연록에서도 발화의 보존과 사실 확인은 구별해야 한다. 예컨대 게시글의 남아공 약 7천 명 사면이라는 설명은 신청과 허가를 나누어 확인할 대목이다. 남아공 법무부 공식 표에서 7,112는 신청 총수이며, 허가와 거절은 별도로 집계된다.[2][9] 원음을 대조해 녹취 오류인지 발화상의 착오인지 확인하고, 원래 발언과 편집자의 보충을 분리해둘 수 있다.

이를 위해 각 대목에 화자, 원문 위치, 작성 시점, 증거의 성격, 편집·정정 이력을 연결할 필요가 있다. 증언, 위원회 판단, 학자의 해석, 편집자의 보충을 하나의 평평한 사실 목록으로 합치지 않는 것이다. 생성된 주장마다 원문으로 돌아갈 수 있게 하고, 연결이 확인되지 않은 설명은 가설로 남겨야 한다.

Hughes-Warrington의 역사학과 AI 윤리 논의도 새로운 증거와 해석에 대한 개방성, 질문의 역할을 다룬다.[7] 이런 개방성을 실제 연구에 적용하려면 출처 표시를 넘어, 반론이 후속 조사와 해석 수정에 연결되는 절차를 만들어야 한다.

대중이 원하는 확실성과 서사의 역할

더 정교한 설명을 만드는 일과 사람들이 그 설명을 받아들이게 하는 일은 다르다. 대중이 역사에서 원하는 것은 사실만이 아니다. 자신의 고통에 대한 인정, 공동체의 존엄, 현재의 혼란을 이해할 설명, 행동의 방향을 찾기도 한다. 역사학이 이런 요구에 복잡성과 판단 유보만 제공한다면 확실한 적과 구원의 약속을 제시하는 서사가 설득력을 얻을 수 있다.

페라리를 "레이싱 본능"으로 설명하는 대중서사를 생각해볼 수 있다. 이 표현은 창업자와 자동차, 경주와 승리를 기억하기 쉬운 성격으로 묶는다. 독자는 사건을 이해하면서 감탄의 방향도 얻는다. 역사학자가 그 서사를 단순화라고 지적하는 데서 멈추면 설명은 정밀해져도 독자가 얻던 의미는 비어버릴 수 있다.

역사학은 그 매력에서 출발해, 실제로 그런 성취가 어떤 노동과 기술적 선택, 자금과 조직, 실패를 통해 이루어졌는지 이야기할 수 있다. 이것은 특정 페라리 저작에 대한 사실 판정이 아니라 서사 구성의 예다. 인물의 성격으로 압축된 설명을 구체적인 역사 과정으로 펼치면서도 이야기의 응집력을 유지하자는 제안이다.

독일의 사례에서 보아야 할 선택과 제도

제1차 세계대전 패전과 전후 질서에 대한 분노는 나치 부상의 중요한 조건이었다. 여기에 대공황, 반유대주의, 정치적 폭력, 민주주의의 약화와 보수 엘리트들의 선택이 얽혔다. 히틀러는 1933년 힌덴부르크의 임명으로 총리가 되었다. 따라서 독일 전체가 상처 때문에 일제히 권위주의에 매혹되었다는 하나의 심리로 설명하기는 어렵다.[5]

높은 학문과 문화가 민주주의를 자동으로 보호하지는 않는다. 그렇기에 역사학은 분노가 어떤 조직과 선택을 통해 권력이 되었는지 보여주어야 한다. 어느 지점에서 무엇을 달리할 수 있었는지를 밝히면, "상처받은 민족은 독재자를 선택한다"는 운명론을 벗어나 현재의 제도와 선택을 생각할 수 있다.

비판을 견딜 수 있는 소속감

공동체의 존엄을 세우는 데 과거의 무결함이 반드시 필요한 것은 아니다. 잘못을 인정한 사람, 기록을 지킨 사람, 부당한 명령을 거부한 사람, 제도를 바꾼 사람에게서도 존엄을 찾을 수 있다. 과거의 잘못 하나가 드러날 때마다 공동체 전체가 무너진다고 느끼게 만드는 서사는 사실 확인을 자기방어의 문제로 바꾼다. 역사학은 공동체를 이해하는 자원을 풍부하게 만들어 이 압박을 줄이는 데 기여할 수 있다.

분명하게 설명하고 수정할 수 있는 권위

모든 독자에게 매번 사료 검증부터 시작하라고 요구하기는 어렵다. 사람들은 일상을 살아가기 위해 판단을 전문가와 기관에 위임한다. 역사학의 공공적 역할에는 그 위임을 받을 만한 설명과 실적을 만드는 일이 포함된다. 권위의 존재보다 중요한 것은 그 권위가 자신의 오류와 이의 제기를 어떻게 다루는가이다.

역사학자는 현재 자료로 가장 설득력 있는 설명을 분명하게 제시할 수 있어야 한다. 동시에 상당히 확실한 사실, 논쟁 중인 원인의 비중, 아직 밝혀지지 않은 경로를 구체적으로 나누어야 한다. "사건은 확인되지만 명령 전달 경로는 규명되지 않았다"는 설명은 전부 불확실하다는 말과 다르다. 독자에게 무엇을 믿어도 되는지와 어디까지 판단을 유보해야 하는지를 함께 알려준다.

사실과 수치의 불확실성을 전달한 실험에서는 신뢰 하락이 작거나 제한적이었다. 다만 정체성과 정치적 갈등이 걸린 역사 문제에 같은 결과를 일반화할 수는 없다.[6] "대중은 불확실성을 싫어한다"는 전제도 전달 방식과 상황에 따라 검토해야 한다. 현재 설명의 한계를 이해 가능한 형태로 제시하는 방법을 시험할 필요가 있다.

LLM을 통한 설명과 검토의 연결

AI는 독자의 기존 지식과 관심에 맞춰 설명의 출발점과 길이를 조절할 수 있다. 그러나 독자가 원하는 방향으로 사실과 책임의 기준까지 바꾸어서는 안 된다. 같은 근거 위에서 서로 다른 접근 경로를 제공하는 것이 설계 목표가 되어야 한다.

첫 화면에서는 짧고 분명한 설명을 제공한다. 이어 그 설명을 지탱하는 결정적인 증거와 원문 위치를 보여준다. 독자는 필요에 따라 주요 반론과 미해결 쟁점으로 들어갈 수 있다. 이후 오류가 확인되면 기존 설명과 수정 이유를 연결해 남긴다. 열린 검토의 부담을 처음부터 독자에게 전부 넘기지 않으면서, 검토를 원하는 사람의 접근은 보장하는 방식이다.

이때 AI가 생성한 역할 토론만으로 견해의 다양성이 확보되었다고 볼 수 없다. 실제 당사자의 질문, 다른 연구자의 반론, 서로 다른 경로의 자료가 들어와야 한다. 문체가 다른 여러 답변보다 중요한 것은 설명을 바꾸게 만드는 실제 저항이다.

역사 지식만으로 불평등과 정치적 배제, 정보 유통 권력의 문제를 해결할 수는 없다. 학계와 언론, 박물관, 교육기관이 오류를 고치고 반론에 응답하며 설명을 개선하는 실적을 쌓아야 한다. LLM은 이 과정을 연결할 수 있지만, 기록의 비공개나 제도의 무응답을 기술만으로 없애지는 못한다.

강연 기록에서 시작하는 검증 가능한 연구

박구병 교수 강연 기록을 출발점으로 삼으면 거대한 AI 역사학 체계를 먼저 세우지 않아도 작은 연구를 시작할 수 있다. 아래 절차는 실제로 수행한 실험 결과가 아니라, 공개 기록을 연구 가능한 대상으로 발전시키기 위한 제안이다.

질의응답이 역사적 설명에 미치는 영향

동일한 질문, 예컨대 한국의 과거사 조사가 지연되는 이유를 놓고 강연 본문만 제공한 조건과 질의응답까지 제공한 조건을 비교한다. 가능하면 주제와 분량을 맞춘 비교 자료도 사용해, 단순히 텍스트가 늘어난 효과와 반론이 추가된 효과를 구별한다. 답변은 반복 수집하고 사용 모델과 설정, 자료 버전을 기록한다.

평가할 항목은 조사 방식에 관한 설명이 추가되는지, 질문자의 주장을 확정 사실과 구별하는지, 여러 원인을 검토하는지, 후속 확인에 필요한 자료를 제시하는지 등이다. 인용 여부만 세지 말고, 인용한 대목이 주장을 실제로 뒷받침하는지도 확인한다. 같은 질문에서 출처를 바꾸거나 반례를 추가했을 때 설명이 어떻게 변하는지도 살펴볼 수 있다.

자료와 해석을 바꾸어도 남는 주장

자료군을 달리하고, 국가안보·관료조직·피해 경험 등 서로 다른 분석 틀을 적용해본다. 그 결과를 평균내기보다, 상당한 변경에도 유지되는 주장과 특정 가정에서만 유지되는 주장을 나눈다. 여러 설명이 모두 같은 원자료에 의존한다면 그 공통 결론의 취약성도 기록한다. 모든 역사 주장에 하나의 신뢰도 숫자를 붙일 필요는 없다.

설명의 품질과 공공적 수용을 따로 평가하기

연구의 품질은 원문에 근거한 주장 비율, 출처 연결의 정확성, 독립 자료와 경쟁 가설의 검토, 반증에 따른 수정으로 살펴볼 수 있다. 공공적 수용은 독자가 핵심 설명과 미해결 부분을 구별하는지, 수정 이유를 이해하는지, 반대 입장의 자료도 열어볼 수 있는지 등으로 별도 평가할 수 있다. 많이 인용되거나 많이 동의받은 답변을 곧바로 좋은 역사라고 부르지 않는 것이다.

이런 연구의 기술적 요소가 전부 가정에 머무는 것은 아니다. 필라델피아의 56년치 부동산 증서를 다룬 연구는 OCR, 검색, LLM을 결합해 역사 문서에서 정보를 추출하고 검증하는 방법을 제시한다.[8] 다만 특정 문서의 추출 성능이 역사적 인과 설명이나 다른 자료의 신뢰성을 보증하지는 않는다.

"더 진한 회색"은 모든 주장을 흐릿하게 만드는 상태일 필요가 없다. 확인된 부분은 더 선명하게, 해석에 의존하는 부분의 경계도 더 선명하게 만들 수 있다. 역사학이 제공할 것은 판단을 끝없이 미루라는 요구만이 아니라, 현재의 설명을 믿고 활용하면서도 새로운 증거 앞에서 수정할 수 있는 근거와 경험이다.

참고 자료

본문의 번호는 아래 자료를 가리킨다. 연구 설계와 공공역사에 관한 제안은 각 자료에서 직접 입증된 결과와 구별된다.

[1] 김율
로베르토 부사 키드의 생애와 AI 및 스콜라철학 강연자료. 첨부 PDF 5쪽, 7~9쪽, 14~15쪽. 부사의 색인과 문헌 연구 경험, AI 활용의 역할을 설명하는 자료.

[2] 박구병 강연 및 이수화 정리
라틴아메리카 진실·화해위원회의 과거사 정리와 한국에 제공하는 시사점. 서울대 사학인 포럼, 2026년 10월 29일. 발표자료와 녹취를 매칭한 공개 기록. 특히 제5~6절과 제9절 질의응답.
원문 열기

[3] Google Search Central
Optimizing your website for generative AI features on Google Search. AEO와 GEO의 정의, 검색 결과를 활용한 생성형 답변의 작동을 설명하는 공식 문서.
원문 열기

[4] Elliot Kim 외
Correlated Errors in Large Language Models. 2025년 공개 연구. 여러 LLM 사이의 오류 상관을 다루며, 역사 해석 자체를 시험한 연구는 아니다.
원문 열기

[5] United States Holocaust Memorial Museum
Adolf Hitler and the Nazi Rise to Power 1918–1933. 나치의 부상과 히틀러의 총리 임명, 정치·경제적 조건에 대한 역사 해설.
원문 열기

[6] Anne Marthe van der Bles 외
The effects of communicating uncertainty on public trust in facts and numbers. PNAS 117권 14호, 2020, 7672–7683쪽. 불확실성 전달에 관한 실험이며 정치적 역사 논쟁에 대한 직접 검증은 아니다.
원문 열기

[7] Marnie Hughes-Warrington
Ethics for Artificial Historians. History and Theory 64권 2호, 2025, 159–177쪽. 역사 서술의 논리와 새로운 증거·해석에 대한 개방성을 논의한다.
원문 열기

[8] Federal Reserve Bank of Philadelphia
Combining AI and Established Methods for Historical Document Analysis. Consumer Finance Institute Discussion Paper 25-2, 2025. 역사 문서 분석에서 OCR·검색·LLM을 결합한 방법과 검증을 다룬다.
원문 열기

[9] South African Department of Justice
TRC Amnesty Hearings and Decisions. 2000년 11월 1일 기준 표는 총 7,112건의 신청을 허가·거절 및 기타 상태로 구분한다. 이 표를 최종 허가자 수와 혼동해서는 안 된다.
원문 열기