



출처 표시와 이용자 검증 행동: 체계적 문헌고찰 프로토콜

각주로 붙은 출처를 사람들은 실제로 눌러 보는가. 이 물음이 어떻게 연구돼 왔는지를 모으기 위한 프로토콜입니다.

윤소연 · 발행 2026-06-30 · 최종 수정 2026-06-30

수록	제2권 제1호 · pp. 79-104
저자	윤소연
투고	2026-03-27
게재 확정	2026-05-29
주제어	체계적 문헌고찰 · 출처 표시 · 검증 행동 · 인간 대상 연구 · 비몰림 위험

초록

생성형 검색은 답변과 함께 출처를 제시하지만, 그 표시가 이용자의 이해와 확인 행동에 어떤 영향을 주는지는 정리돼 있지 않다. 이 글은 출처 표시 형식을 조작 변인으로 다룬 인간 대상 연구를 체계적으로 모으기 위한 프로토콜이다. 연구 질문과 검색 전략, 선정·배제 기준, 자료 추출 항목, 비몰림 위험 평가 계획을 수행 전에 공개한다. 고찰 결과는 담고 있지 않으며, 예비 범위 검토에서 드러난 문헌의 형태만 서술한다.

1. 배경

생성형 검색은 완성된 문장을 내놓으면서 그 아래에 출처를 붙인다. 각주 번호로 붙기도 하고, 문장 안에 링크로 박히기도 하고, 도메인 이름만 적히기도 한다. 링크가 아예 없는 경우도 있다.

이 표시가 무엇을 위한 것인지는 분명하다. 이용자가 원문을 대조해 답의 진위를 가릴 수 있게 하려는 것이다. 그런데 **그 장치가 실제로 그렇게 쓰이는지**는 다른 문제다. 출처가 붙어 있다는 사실만으로 답을 더 믿게 되고 오히려 확인을 덜 하게 된다면, 표시는 의도와 반대로 작동한 셈이 된다.

이 물음은 시스템 쪽 관측으로는 답할 수 없다. 사람이 무엇을 눌렀고 무엇을 읽었고 무엇을 믿었는지를 재야 하므로 인간 대상 연구가 된다. 그래서 이 고찰은 앞의 두 편과 방법론적 성격이 다르다. 표본은 문서가 아니라 사람이고, 통제해야 할 것은 수집 조건이 아니라 과제 설계와 참가자 특성이다.

2. 연구 질문

출처 표시 방식은 이용자의 이해와 검증 행동에 어떤 영향을 주는가. 아래 넷으로 나눈다.

- 출처 표시 형식을 조작 변인으로 다룬 연구가 있는가. 어떤 형식을 비교했는가
- 검증 행동을 무엇으로 측정했는가
- 참가자 표본과 과제는 어떻게 설계됐는가
- 사전 등록과 윤리 절차를 보고한 연구는 얼마나 되는가

축	이 고찰에서의 정의
대상	생성형 검색 또는 대화형 답변 인터페이스를 이용하는 성인
개입	출처 표시 형식의 변경 (각주형 · 인라인 · 링크 없는 도메인 표기 · 불확실성 문구 동반)
비교	출처 표시가 없는 조건, 또는 다른 표시 형식
결과	원문 확인 행동, 답변 정확도 판단, 주관적 신뢰, 과제 수행 정확도

3. 검색 전략

학술 데이터베이스는 ACM Digital Library, IEEE Xplore, Scopus, Web of Science, PsycINFO 를 쓴다. 국내 문헌은 KCI 와 RISS 를 함께 검색한다. 사전 출판본을 놓치지 않기 위해 arXiv 의 cs.HC·cs.IR 범주를 포함한다.

검색어는 세 묶음을 교차한다.

- 시스템: generative search, conversational search, AI assistant, LLM, 생성형 검색, 대화형 검색
- 개입: citation display, source attribution, provenance, reference presentation, 출처 표시, 인용 표기
- 결과: trust, verification behavior, click-through, credibility judgment, 신뢰, 검증 행동

대상 기간은 2019년 1월부터 검색 시행일까지로 한다. 대화형 답변 인터페이스가 일반 이용자에게 닿기 전의 연구도 설계 참고가 되므로 하한을 넉넉히 둔다. 언어는 한국어와 영어로 제한하고, 그 밖의 언어는 제외 사유를 기록한다.

회색문헌은 **포함하되 따로 표시한다**. 업계 블로그와 백서에도 이용자 조사가 실리지만 방법 기술이 부실한 경우가 많으므로, 별도 층으로 분류해 종합할 때 가중을 달리한다. 선별된 문헌의 참고문헌 목록을 한 단계 추적하고, 학술대회 초록집도 손으로 훑는다.

4. 선정·배제 기준

구분	기준
포함	사람을 대상으로 자료를 수집했을 것
포함	출처 표시의 형식·유무가 조작되었거나 비교되었을 것
포함	행동·인지·태도 가운데 하나 이상을 측정했을 것
배제	시스템 성능만 보고하고 이용자 자료가 없는 연구
배제	출처 표시를 다루지 않고 답변 품질만 평가한 연구
배제	원문 전문을 구할 수 없는 문헌
배제	같은 자료를 중복 보고한 문헌 (가장 완전한 판을 남긴다)

5. 선별 절차

제목·초록 단계와 전문 단계를 나누고, 각 단계를 **두 사람이 독립으로** 수행한다. 불일치는 세 번째 검토자가 조정하며, 조정 전 일치도를 카파 계수로 보고한다.

수행 전이므로 검색 건수와 선별 편수는 적지 않는다. PRISMA 흐름도의 빈 틀을 프로토콜에 함께 두고, 검색을 마친 뒤 그 자리를 채워 공개한다.

6. 자료 추출 항목

항목	기록 내용
서지	저자, 연도, 게재처, 문헌 층(학술/회색)
인터페이스	실제 서비스인가 모의 화면인가, 어떤 시스템을 모사했는가
조작 조건	비교한 출처 표시 형식의 목록과 조작 방식
과제	참가자에게 준 과제의 종류와 개수, 정답 유무
표본	인원, 모집 경로, 연령·배경 분포, 보상
설계	참가자 간/내 설계, 순서 통제 방법
결과 지표	측정한 변수와 측정 도구, 주 지표의 사전 지정 여부
보고	사전 등록 여부, 자료·분석 코드 공개 여부
윤리	심의 기구, 사전 동의, 사후 설명

7. 질 평가

인간 대상 연구이므로 효과 크기보다 **비뚤림 위험**을 먼저 본다. 네 축으로 나눈다.

축	살피는 것
선택 비뚤림	모집 경로가 특정 집단에 치우쳤는가. 배정이 무작위였는가
수행 비뚤림	참가자가 연구 목적을 짐작할 수 있었는가. 실험자가 조건을 알았는가
측정 비뚤림	결과 지표가 조작 조건과 독립으로 측정됐는가. 자기보고에만 의존했는가
탈락 비뚤림	중도 이탈률이 조건 간에 달랐는가. 결측을 어떻게 처리했는가

축마다 낮음·불명확·높음으로 판정하고, 판정 근거를 원문 인용과 함께 남긴다. 종합 등급을 하나의 점수로 합산하지 않는다. 합치면 어느 축에서 문제가 있었는지가 사라진다.

8. 종합 방법

양적 통합은 처음부터 계획하지 않는다. 출처 표시 형식의 조작 방식과 결과 지표가 연구마다 다를 것으로 보이며, 서로 다른 것을 켜 값을 하나의 효과 크기로 묶으면 그 숫자가 무엇을 의미하는지 말할 수 없게 된다.

대신 서술적 종합을 한다. 조작 조건을 공통 축으로 재분류하고, 결과 지표를 행동·인지·태도 세 층으로 나눈 뒤 층마다 방향과 근거 수준을 표로 제시한다. 같은 방향을 가리키는 연구가 여럿이어도 **비뿔림 위험이 높은 연구만으로 이루어진 일치는 일치로 세지 않는다.**

9. 예비 범위 검토에서 본 것

수행 전 범위 검토에서 확인한 것은 문헌의 **형태**이며 결과가 아니다. 네 가지가 눈에 띄었다.

첫째, 두 갈래가 따로 자란다. 정보검색 쪽은 출처의 정확성을 시스템 지표로 다루고, 인간컴퓨터상호작용 쪽은 신뢰와 인지 부하를 다룬다. 서로를 인용하는 경우가 드물어 같은 대상을 다르게 부르고 있을 가능성이 있다.

둘째, 조작 조건의 이름이 제각각이다. 같은 각주형 표시를 두고 citation, reference, attribution, provenance 가 뒤섞여 쓰인다. 검색어를 넓게 잡아야 하는 이유다.

셋째, 실제 서비스를 대상으로 한 연구보다 **모의 화면을 만든 연구**가 많아 보인다. 통제에는 유리하지만 외적 타당도를 따로 평가해야 한다.

넷째, 사전 등록을 밝힌 문헌이 눈에 잘 띄지 않았다. 그래서 7절의 비뿔림 평가에 사전 등록 여부를 별도 항목으로 두었다.

다섯째, 결과 지표의 층이 섞여 보고되는 경우가 있었다. 링크를 눌렀는가(행동), 답이 맞다고 보았는가(인지), 얼마나 믿었는가(태도)는 서로 다른 것인데 한 문단에서 함께 서술되곤 한다. 8절에서 세 층을 나누어 종합하기로 한 것은 이 때문이다.

여섯째, 모집 경로가 특정 집단에 몰릴 가능성이 있다. 온라인 패널이나 대학 구성원을 쓴 연구가 많아 보이며, 이 경우 참가자가 검색 결과를 평소보다 꼼꼼히 확인할 개연성이 있다. 6절의 자료 추출 항목에 모집 경로와 보상을 따로 둔 이유다.

10. 한계

- **수행 전 프로토콜이다.** 실제 검색에서 문헌이 매우 적으면 고찰이 성립하지 않을 수 있고, 그때는 범위 검토로 형식을 바꾸어 그 사실을 보고한다
- **언어 제한이 있다.** 한국어와 영어로 한정해 다른 언어권의 연구를 놓친다
- **회색문헌의 층을 나누는 기준이 주관적이다.** 판정 근거를 모두 기록해 공개하는 것으로 대신한다

- **빠르게 낡는다.** 인터페이스가 바뀌면 조작 조건의 의미도 바뀌므로, 오래된 연구의 결과를 현재 화면에 그대로 옮길 수 없다
- **학회에 연구윤리 심의 기구가 없다.** 이 고찰 자체는 사람을 대상으로 자료를 수집하지 않아 심의 대상이 아니지만, 후속 실험을 학회 이름으로 수행하려면 외부 심의 절차를 먼저 마련해야 한다

11. 공개와 갱신

프로토콜 전문과 검색어 조합, 데이터베이스별 검색 기록, 자료 추출 양식을 학회 홈페이지에 공개한다. 검색을 마치면 PRISMA 흐름도와 선별 사유 목록을 같은 자리에 올린다.

프로토콜을 수행 도중에 고치는 경우 **원문을 지우지 않고 개정 이력을 덧붙인다.** 무엇을 언제 왜 바꿨는지가 남아야 결과를 읽는 쪽이 판단할 수 있다.

고찰은 1년마다 검색을 다시 돌려 갱신한다. 갱신에서 결론이 바뀌면 바뀐 사실과 그 이유를 함께 적는다.

12. 맺음

이 글은 프로토콜이며 고찰 결과를 담고 있지 않다. 출처 표시가 검증 행동을 늘린다거나 줄인다고 말하지 않는다. 다만 그런 주장을 하려면 무엇을 먼저 모아야 하고 무엇으로 걸러야 하는지를 수행 전에 적어 두었다.

같은 절차를 다른 연구자가 밟아 다른 결론에 이른다면, 그 차이가 어디서 왔는지를 이 문서로 따져 볼 수 있다. 그것이 프로토콜을 먼저 공개하는 이유다.

참고문헌

- Aggarwal, P. et al. (2023). *GEO: Generative Engine Optimization*. arXiv:2311.09735. — 답변에 출처가 실리는 방식을 다룬 초기 연구로, 2절의 개입 정의를 세우는 출발점으로 삼았다.
- 한국AI검색학회 인용·출처 분과. 인용 표시 유형 정리 초안. 내부 검토 중. — 조작 조건을 네 형식으로 나눈 분류가 이 초안을 따른다.
- 한국AI검색학회 측정·평가 분과. 반복 측정 설계 초안. 내부 검토 중. — 주 지표를 사전 지정하는 절차를 이 초안에서 가져왔다.