



생성형 검색 응답의 시간적 변동: 체계적 문헌고찰 프로토콜

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

수록	제2권 제1호 · pp. 53-78
저자	오현석
투고	2026-03-21
게재 확정	2026-05-18
주제어	체계적 문헌고찰 · 시간적 변동 · 반복 관측 · 재현성 · 연구 프로토콜

초록

같은 질의에 대한 생성형 검색의 응답이 시점마다 달라진다는 관찰은 널리 공유되지만, 그런 관찰이 어떤 조건에서 이루어졌고 무엇이 기록으로 남았는지는 정리된 적이 없다. 이 글은 시간적 변동을 반복 관측으로 다룬 문헌이 관측 조건과 변동의 근거를 어떤 형태로 보고해 왔는지를 확인하기 위한 체계적 문헌고찰 프로토콜이다. 연구 질문과 검색 전략, 선정·배제 기준, 2인 독립 선별 절차, 자료 추출 항목, 보고 투명성을 축으로 삼은 질 평가 등급을 미리 정하고, 양적 통합 대신 기록 항목의 공백을 드러내는 서술 종합을 택한 이유를 밝힌다. 예비 범위 검토에서 관찰된 문헌의 형태를 함께 적되 편수와 수치는 제시하지 않는다.

1. 배경

생성형 검색은 같은 질의에 늘 같은 답을 내놓지 않는다. 오늘 확인한 출처가 다음 주에는 빠져 있고, 한동안 안정적이던 답변이 어느 회차부터 형식을 바꾼다. 현상 자체는 이 분야를 다루는 사람들 사이에서 널리 공유돼 있다.

널리 알려진 것과 정리된 것은 다르다. 변동을 다뤘다는 글은 여러 경로로 나오지만, 그 글들이 각각 언제 무엇을 몇 번 관측했고 무엇을 기록으로 남겼는지는 한자리에 모인 적이 없다. 그래서 두 보고가 어긋날 때 어느 쪽 관측이 더 촘촘했는지 가릴 근거가 없고, 새로 관측을 시작하는 사람은 앞선 관측을 이어받지 못한 채 처음부터 다시 설계한다.

시간이 개입하는 관측에는 다른 주제에 없는 함정이 하나 더 있다. 값이 변했다는 관측은 대상이 변했다는 뜻일 수도 있고 관측 장치가 변했다는 뜻일 수도 있다. 둘을 가르려면 관측자가 자기 환경의 변경 이력을 남겨 두었어야 하는데, 그 기록이 실제로 보고서에 실렸는지는 개별 문헌을 열어 보기 전에는 알 수 없다.

이 글은 그 확인을 한 편씩이 아니라 체계적으로 수행하기 위한 프로토콜이다. 변동이 얼마나 큰지를 묻지 않는다. 변동을 관측했다고 말한 글들이 무엇을 어떻게 기록하고 보고했는지를 묻는다.

2. 연구 질문

주 질문은 하나다. 생성형 검색 응답의 시간적 변동을 반복 관측으로 다룬 문헌은 관측 조건과 변동의 근거를 어떤 형태로 기록하고 보고해 왔는가.

하위 질문은 넷으로 나눈다.

- 1. 시간에 걸친 반복 관측을 수행한 연구는 회차마다 어떤 항목을 남겼다고 밝혔는가
- 2. 대상 서비스의 변경과 관측 환경의 변경을 나누어 다룬 연구가 있는가, 있다면 무엇을 근거로 나누었는가
- 3. 결측 회차와 관측 종단을 어떻게 정의하고 어떻게 세었는가
- 4. 재현을 위해 무엇을 공개했는가 — 질의 목록, 수집 도구, 응답 원문 가운데 어디까지인가

요소	이 고찰에서의 정의
대상(P)	생성형 검색 또는 대화형 응답 시스템의 출력을 둘 이상의 시점에서 관측한 연구
개입·노출(I)	시간의 경과, 질의와 조건을 고정한 채 관측 시점만 달리한 설계
비교(C)	같은 연구 안의 다른 시점, 또는 같은 시점의 다른 조건(대조 질의, 다른 환경)
결과(O)	응답·출처·형식의 변동을 보고한 방식과 그 근거로 제시된 기록 항목
설계(S)	반복 측정 관측 연구. 단일 시점 관측과 모의 환경 연구는 대상에서 뺀다

3. 검색 전략

데이터베이스는 아홉을 쓴다. arXiv, ACL Anthology, ACM Digital Library, IEEE Xplore, Scopus, Web of Science, DBLP, 한국학술지인용색인, 학술연구정보서비스. 영어와 한국어 문헌을 함께 받는다.

대상 기간은 2022년 1월부터 검색 시행일까지다. 그 이전의 문헌은 이 고찰이 묻는 관측 조건 자체를 갖지 않는다.

검색어는 세 묶음을 교차한다. 묶음 안은 OR 로, 묶음 사이는 AND 로 잇고 데이터베이스마다 구문을 맞춘다.

묶음	검색어
대상	generative search, AI search, conversational search, answer engine, 생성형 검색, 대화형 검색
현상	temporal change, longitudinal, drift, variability, over time, 시계열, 변동, 재현성
관측	repeated measurement, audit, monitoring, snapshot, 반복 측정, 관측, 감사

실제로 실행한 질의문은 데이터베이스 이름, 실행 일자, 적용한 필터와 함께 그대로 부록에 남긴다.

회색문헌은 받되 따로 센다. 기관 기술 보고서와 사전 인쇄본, 학술대회 발표자료는 본 대상에 넣고, 사업자 공지와 상업 목적의 글은 별도 목록으로만 관리해 본 종합에 섞지 않는다.

포함된 문헌의 참고문헌을 거슬러 확인하고, 그 문헌을 인용한 뒤 글도 확인한다. 양방향 추적은 새 문헌이 더 나오지 않을 때까지 반복한다.

4. 선정·배제 기준

구분	기준
포함	둘 이상의 시점에서 같은 질의 또는 비교 가능한 질의로 응답을 관측한 문헌
포함	관측 시점이나 관측 기간이 본문에 적혀 있는 문헌
포함	응답 내용·인용 출처·응답 형식 가운데 하나 이상을 관측 단위로 삼은 문헌
포함	영어 또는 한국어로 작성된 문헌
배제	단일 시점 관측만 보고한 문헌
배제	관측 없이 변동을 논평하거나 전망한 글
배제	합성 로그나 모의 응답만으로 변동을 다룬 문헌
배제	관측 절차를 확인할 본문이 없는 초록·포스터 요약
배제	전문을 구할 수 없는 문헌. 별도 목록에 사유와 함께 남긴다

5. 선별 절차

제목·초록 선별과 전문 선별의 두 단계로 나눈다. 각 단계는 검토자 두 명이 상대의 판정을 모르는 상태에서 따로 매기며, 이 고찰은 시점 기술을 주로 보므로 1단계에서 초록에 관측 기간이 드러나지 않아도 배제하지 않고 전문까지 넘긴다.

본 선별에 앞서 보정 단계를 둔다. 무작위로 고른 일부를 두 사람이 함께 판정해 기준의 해석 차이를 좁힌 뒤 본 선별로 넘어간다. 보정에서 손본 기준은 개정 이력으로 남긴다.

불일치는 토의로 조정한다. 토의로 좁혀지지 않으면 편집위원회가 지정한 제3의 검토자가 판정하고 그 사유를 기록한다. 조정 전 일치도는 단계별로 계산해 함께 보고한다. 조정 후의 결과만 실는 것은 판정이 갈렸다는 사실을 감추는 일이다.

선별 흐름은 표준 흐름도 형식으로 제시한다. 다만 각 단계의 수는 검색을 실제로 수행한 뒤에만 적을 수 있으므로, 이 프로토콜에서는 틀만 공개하고 칸을 비워 둔다. 배제 사유는 전문 단계에서만 문헌별로 적는다.

6. 자료 추출 항목

추출표는 프로토콜 확정 시점에 서식으로 고정한다. 두 사람이 나누어 추출하고 서로의 표를 교차 확인한다.

묶음	추출 항목	비어 있을 때의 표기
서지	저자, 발표 연도, 문헌 유형, 동료 심사 여부	해당 없음
대상	관측 대상의 종류, 접근 표면(화면·API), 본문에 적힌 버전	미기재
설계	관측 시작·종료 시점, 회차 수, 관측 간격, 질의 수	미기재
조건	계정 상태, 지역·언어 설정, 대화 이력 처리, 접속 경로	미기재
기록	회차마다 남겼다고 밝힌 항목, 응답 원문 보관 여부	미기재
변경	대상 쪽 변경과 자체 도구 변경의 언급 유무, 구분 근거	구분 없음
결측	결측과 중단의 정의, 세는 방식, 보고한 위치	미기재
공개	질의 목록·수집 도구·응답 원문·추출표의 공개 범위와 접근 경로	비공개
한계	저자가 스스로 적은 한계와 일반화 범위	미기재

미기재와 해당 없음을 구분한다. 적을 자리가 있었는데 비어 있는 것과 애초에 해당하지 않는 것은 다른 사실이다. 이 구분이 흐려지면 기록 관행의 공백을 세는 일 자체가 성립하지 않는다.

7. 질 평가

이 고찰은 문헌이 내린 결론의 참·거짓을 가리지 않는다. 가리는 것은 보고의 투명성이다. 기존 편향 위험 도구는 개입 연구를 전제하므로 그대로 옮겨 쓰기 어렵고, 대신 네 축을 두어 축마다 세 단계로 매긴다.

축	충족(2점)	부분 충족(1점)	미충족(0점)
시점 명시	관측 시작·종료와 회차 간격을 모두 적음	기간만 적음	시점을 알 수 없음
조건 기록	계정·지역·표면·도구 버전을 모두 밝힘	일부만 밝힘	조건 언급 없음
변경 구분	대상 변경과 자체 변경을 근거와 함께 나눔	언급만 함	구분 없음
재현 자료	질의 목록과 응답 원문을 함께 공개	둘 중 하나만 공개	공개 없음

네 축에 축마다 0~2점을 매기므로 합계의 범위는 0점에서 8점(=4축 × 2점)이다. 이 범위를 셋으로 나누어 6점 이상을 상, 3점에서 5점을 중, 2점 이하를 하로 묶는다. 등급은 배제의 근거가 아니라 종합에서 무게를 다르게 두는 근거로만 쓴다. 등급은 두 사람이 독립으로 매기고 불일치는 5절과 같은 방식으로 조정한다.

8. 종합 방법

양적 통합은 하지 않는다. 대상 시스템과 질의, 관측 간격, 변동의 정의가 문헌마다 다르므로 효과크기를 한 척도로 모을 수 없고, 억지로 모으면 서로 다른 것을 더한 평균이 나온다. 이 고찰이 묻는 것도 변동의 크기가 아니라 보고의 형태다.

대신 서술 종합을 쓰되 그 근거를 표로 드러낸다. 문헌을 행으로, 6절의 추출 묶음을 열로 놓은 행렬을 만들어 어느 칸이 비어 있는지를 그대로 보인다. 빈칸의 분포가 이 고찰의 주된 산출물이 된다.

문헌은 관측 설계에 따라 묶는다. 고정 간격 반복, 사건 전후 비교, 비정기 재확인 of 세 갈래를 잠정 분류로 두고, 추출 과정에서 맞지 않는 사례가 나오면 갈래를 늘린다. 분류를 바꾸면 바꾼 시점과 이유를 남긴다.

세는 대상은 편수가 아니라 관행이다. 무엇을 한 문헌이 얼마나 되는지는 대상 문헌이 모두 확정된 뒤에만 적고, 그때에도 분모를 함께 적는다.

9. 예비 범위 검토에서 본 것

프로토콜을 짜기 위해 소규모 예비 검색을 했다. 아래는 그 과정에서 눈에 띈 문헌의 형태이며 편수나 비율이 아니다. 본 검색에서 달라질 수 있다.

첫째, 이름이 갈린다. 같은 현상을 두고 변동, 표류, 비결정성, 불안정성이 섞여 쓰이고 어느 말도 정의된 채로 쓰이지 않는 경우가 있었다. 검색어 묶음을 넓게 잡은 것은 이 때문이다.

둘째, 시점이 흐리다. 관측을 했다고 적으면서 언제 했는지는 “최근”으로만 밝힌 글이 눈에 띈다. 이런 글은 같은 조건으로 다시 관측해 견줄 방법이 없다.

셋째, 두 종류의 변경을 나눈 서술이 드물다. 변동이 관측됐다는 서술은 있으나 그것이 대상 쪽에서 왔다는 근거를 따로 댄 대목은 예비 검색 범위에서 찾기 어려웠다. 자체 수집 도구를 고친 이력을 밝힌 글도 마찬가지로였다.

넷째, 결측이 보이지 않는다. 반복 관측에서 회차가 빠지는 일은 흔한데, 빠진 회차를 어떻게 처리했는지 적은 대목은 눈에 띄지 않았다.

다섯째, 공개의 폭이 좁다. 질의 목록까지 공개한 사례는 있었으나 응답 원문까지 공개한 사례는 확인하지 못했다. 원문이 없으면 뒤에 오는 사람은 같은 관측을 다시 할 수 있을 뿐 이어서 할 수는 없다.

이상은 모두 본 고찰이 확인해야 할 가설이지 결론이 아니다.

10. 한계

- 표면에서 이루어진 관측을 다룬 문헌만 모은다. 내부에서 무엇이 바뀌었는지를 담은 자료는 공개되지 않으므로 이 고찰의 시야 밖에 있다
- 대상도 문헌도 계속 늘어난다. 검색 시행일 이후의 문헌은 갱신 전까지 반영되지 않으며, 이 주제에서 그 지연은 다른 주제보다 무겁게 작용한다
- 언어를 영어와 한국어로 제한한다. 다른 언어권의 관측 관행은 빠지고, 그 빠짐의 크기는 추정하지 못한다
- 회색문헌의 상당수는 색인되지 않아 검색으로 닿지 않는다. 인용 추적으로 일부를 메우지만 빠짐을 없앨 수는 없다
- 보고된 기록만 볼 수 있다. 기록은 했으나 지면 제약으로 실지 않은 경우와 애초에 기록하지 않은 경우를 이 고찰은 구분하지 못한다
- 판정에 사람이 개입한다. 무엇을 구분 근거로 볼지에는 해석의 여지가 있고, 일치도 보고로 그 폭을 드러낼 뿐 없애지는 못한다

11. 공개와 갱신

프로토콜 본문은 게재와 동시에 학회 홈페이지에 둔다. 이후 고치는 경우 고친 내용과 날짜, 사유를 개정 이력에 덧붙이고 이전 판을 함께 남긴다. 확정 뒤에 바뀐 기준은 결과 해석에 영향을 주므로 감추지 않는다.

검색 기록은 데이터베이스별 질의문과 실행 일자, 적용한 필터를 그대로 공개한다. 추출표 서식과 질 평가 기준표는 빈 서식 상태로 먼저 공개해, 다른 연구자가 같은 서식으로 자신의 문헌을 정리하고 우리 결과와 나란히 놓을 수 있게 한다.

갱신 주기는 열두 달로 둔다. 그 사이에도 새 문헌은 쌓이므로 포함 문헌 목록에는 확인하는 대로 추가분을 반영하고, 종합은 정기 갱신 때 다시 낸다. 갱신할 때마다 이전 판과 무엇이 달라졌는지를 따로 적는다.

맺음

이 글은 프로토콜이며 고찰 결과를 담고 있지 않다. 어떤 시스템의 응답이 얼마나 변했는지도, 어느 연구가 더 옳은지도 말하지 않는다. 그런 말을 하기 전에 무엇을 어떤 순서로 확인할 것인지를 먼저 적어 공개하는 것이 이 글의 전부다.

검색을 실제로 수행한 뒤의 결과는 다음 호에 따로 신는다. 그때의 보고가 이 프로토콜과 어긋난다면 어긋난 자리와 이유를 밝히는 것까지가 이 글에 딸린 약속이다.

참고문헌

- Aggarwal, P. et al. (2023). *GEO: Generative Engine Optimization*. arXiv:2311.09735. — 생형 검색에 대한 문서 쪽 개입을 다룬 초기 연구다. 시점을 달리한 재확인이 왜 필요한지를 생각하게 한 출발점이며, 포함 기준의 경계 사례로도 검토했다.
- 한국AI검색학회 측정·평가 분과. 반복 측정 설계 초안. 내부 검토 중. — 결측을 값으로 채우지 않는 원칙과 조정 전 일치도를 함께 보고하는 방침을 이 초안에서 가져왔다.
- 한국AI검색학회 검색엔진 분과. 수집기 운영 기록 서식. 내부 검토 중. — 6절 추출 항목의 조건·기록 묶음은 이 서식을 문헌 추출용으로 옮긴 것이다.