인사이트

AI 문서 검색에서 검색기의 일은 AI에게 답 조각을 건네는 것에서 AI가 직접 읽을 후보 문서를 고르는 것으로 바뀌었다

방동걸 (주)웨이스 대표이사
2026년 11월 29일

AI가 문서를 찾아 답하는 구조에서 검색기가 맡는 일이 바뀌고 있습니다. 2026년 6월 공개된 DR-DCI 논문에서는 검색기로 후보 문서를 작업 공간에 모은 뒤 AI가 그 안을 직접 찾아 읽게 했습니다. 그러자 830문항 시험에서 같은 모델(GPT-5.4 nano)로 잰 정답률이 문서 전체를 직접 뒤지는 방식의 62.9%에서 71.2%로 올랐고, 문항당 평균 걸린 시간은 3,139초에서 146초로 줄었습니다. 검색기가 사라진 것이 아닙니다. AI에게 답이 될 조각을 건네던 검색기가 이제 AI가 읽을 후보를 고릅니다.

공장 운영의 두뇌인 제조 AI VEXPLOR를 만드는 웨이스는 이 변화를 회사 문서를 다루는 AI에서 먼저 확인해야 할 대목으로 봅니다. 공장에는 설비 매뉴얼, 품질 이력, 거래 문서처럼 여러 문서를 견줘야 답이 나오는 질문이 많기 때문입니다. 이 글에서는 두 논문이 무엇을 비교했는지, 검색기의 역할이 어떻게 바뀌었는지, 회사 문서에 적용할 때 무엇을 확인해야 하는지 정리합니다.



AI 문서 검색에서 검색기의 일은 AI에게 답 조각을 건네는 것에서 AI가 직접 읽을 후보를 고르는 것으로 바뀌었습니다

01

검색기에게서 조각만 받는 AI는 검색기가 놓친 문서를 끝내 보지 못합니다

지금 널리 쓰이는 RAG(검색 증강 생성)는 문서를 미리 조각내 색인해 두고, 질문이 오면 비슷한 조각 몇 개를 골라 AI에게 건넵니다. AI가 볼 수 있는 것은 그 조각뿐입니다. 검색기가 첫 단계에서 놓친 문서는 AI의 추론이 아무리 좋아도 답에 쓰이지 못합니다.

2026년 5월 공개된 DCI 논문은 이 창구를 없애 봤습니다. 연구진은 AI에게 색인 없이 grep(문서에서 낱말을 찾는 터미널 명령)과 파일 읽기 도구만 주고 문서 더미를 직접 뒤지게 했습니다. 문서 10만여 개와 830문항으로 된 시험에서 같은 모델(Claude Sonnet 4.6)로 두 방식을 비교한 결과는 다음과 같습니다.

방식	정답률	전체 평가 비용
검색기가 건넨 조각으로 답하기	69.0%	1,440달러
문서를 직접 찾아 읽기	80.0%	1,016달러

연구진은 다른 모델(GPT-5.4 nano)로 100문항을 따로 떼어 분석해, 직접 찾아 읽는 쪽이 앞선 까닭이 문서를 더 많이 찾아서가 아니라고 밝혔습니다. 직접 찾아 읽은 AI는 찾은 문서 안에서 근거가 되는 구간을 좁혀 들어가는 데 더 뛰어났습니다. 정확한 문자열을 찾고, 조건을 겹쳐 거르고, 앞뒤 문맥을 직접 읽을 수 있었기 때문입니다.

02

문서를 직접 뒤지는 방식은 문서가 많아지면 무너집니다

DCI에는 약점이 있었습니다. 색인이 없으니 문서가 늘수록 첫 단서를 찾는 비용이 가파르게 올랐습니다. 같은 논문에서 방해 문서를 넣어 10만 개를 20만 개로 늘리자 문항당 도구 호출이 38.5번에서 86.9번으로 늘고 정답률은 13.6점 떨어졌습니다. 40만 개에서는 정답률이 37.5%까지 내려갔습니다.

한 달 뒤 나온 DR-DCI 논문의 100문항 확장 실험도 같은 흐름을 보였습니다. 문서 전체를 직접 뒤지는 방식은 10만 개에서 67문항을 맞혔지만 80만 개에서는 29문항으로 떨어졌습니다. 연구진은 그보다 큰 규모에서는 이 방식의 평가를 끝까지 돌리지 못했다고 밝혔습니다.



문서를 직접 뒤지는 방식은 문서가 많아지면 무너집니다

03

검색기가 후보를 모으고 AI가 그 안을 직접 뒤지면 문서가 늘어도 버팁니다

DR-DCI가 바꾼 것은 순서입니다. AI가 pull(끌어오기) 도구로 검색어와 개수를 정하면, 검색기가 후보 문서를 통째로 작업 공간에 넣어 줍니다. AI는 그 작업 공간 안에서만 grep과 읽기로 조건을 겹쳐 찾고, 새 단서가 생기면 다시 끌어옵니다.

문서 수	전체를 직접 뒤지기	후보를 모은 뒤 직접 뒤지기
10만	67/100	80/100
80만	29/100	72/100
1,000만	끝까지 돌리지 못함	70/100

문서가 100배로 늘어나는 동안 후보를 모은 방식의 정답은 80개에서 70개로 완만하게 줄었고, 작업 공간은 문서 1천~1,400개 안팎에 머물렀습니다. 830문항 전체 시험(GPT-5.4 nano)에서는 정답률이 62.9%에서 71.2%로 오르고, 평가 비용은 88달러에서 35달러로 줄었습니다.

04

검색기 종류보다 AI가 후보 안을 직접 대조하는 단계가 결과를 정했습니다

DR-DCI 논문에서 눈여겨볼 실험은 둘입니다. 연구진이 후보를 모아 준 뒤 AI가 여러 문서를 가로질러 찾는 기능만 막자, 100문항 정답이 82개에서 40개로 절반 넘게 줄었습니다. 반면 검색기를 낱말 검색(BM25)으로 쓰든 임베딩 검색으로 쓰든 정답은 80개와 82개로 비슷했습니다.

바꾼 것	정답(100문항)	뜻
모든 도구를 씀	82	기준
여러 문서를 가로질러 찾기를 막음	40	직접 대조가 핵심
검색기를 낱말 검색으로 바꿈	80	검색기 종류는 작은 차이

그래서 검색기는 버려진 것이 아니라 하는 일이 바뀌었습니다. 예전에는 검색기가 고른 조각이 곧 AI가 볼 수 있는 전부였습니다. 이제 검색기는 AI가 읽을 후보의 범위를 정하고, 답의 근거는 AI가 그 후보 안에서 직접 찾아 읽은 원문이 됩니다. DR-DCI 논문도 AI에게 답을 낼 때 실제로 읽은 문서의 경로를 인용하게 하고, 검색기가 보여 주는 순위 미리보기는 길잡이로만 쓰게 했습니다.

같은 흐름은 오픈소스에서도 보입니다. Marker Inc.가 공개한 AutoRAG는 원래 데이터에 맞는 RAG 구성을 자동으로 골라 주는 도구였는데, 2.0에서는 컴퓨터 안의 파일을 찾아 읽고 출처를 다는 에이전트로 방향을 바꿨습니다.



검색기가 후보를 모으고 AI가 그 안을 직접 뒤지면 문서가 늘어도 버팁니다

05

회사 문서에 적용하려면 먼저 확인할 것이 네 가지 있습니다

두 논문의 시험 문서는 모두 영어 평문(웹 문서 · 위키 문서 등)이었습니다. 회사 문서는 사정이 다릅니다. 회사가 이 방식을 자기 문서에 쓰려면 먼저 다음 네 가지를 확인해야 합니다.

- **본문을 글로 뽑아 두었는가** — 논문의 시험 문서는 모두 평문이었습니다. 회사 문서함에는 PDF, 엑셀, 한글 파일이 많고, grep은 이런 파일의 본문을 보지 못합니다. 본문을 글로 뽑아 둔 사본이 있어야 AI가 직접 찾을 수 있습니다.
- **사람마다 볼 수 있는 범위를 지키는가** — 논문은 열람 권한을 다루지 않았습니다. 회사에서는 후보를 모으는 단계부터 질문한 사람이 볼 수 있는 문서만 모아야 합니다.
- **실제로 읽은 원문만 근거로 대는가** — 검색기가 보여 준 순위나 미리보기는 근거가 아닙니다. AI가 열어 읽은 문서만 출처로 남아야 합니다.
- **우리 질문으로 따로 측정했는가** — 논문 수치는 영어 시험에서 나온 값이고, 규모 확장과 기능 제거 실험은 100문항 표본, 모델도 실험마다 하나(GPT-5.4 nano 또는 Claude Sonnet 4.6)로 정해져 있었습니다. 한국어 문서와 실제 업무 질문으로 따로 측정해야 우리 회사에 맞는지 알 수 있습니다.

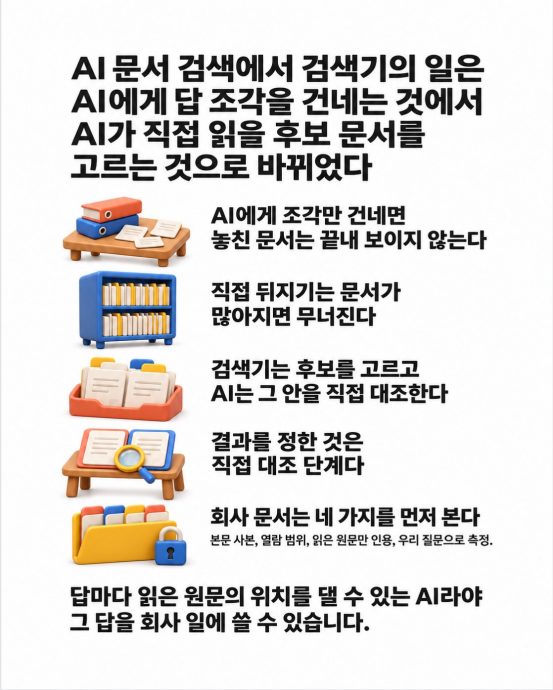
회사는 속도도 따져야 합니다. DR-DCI는 830문항에서 문항당 평균 146초가 걸렸습니다. 현장에서 몇 초 안에 답해야 하는 질문에는 지금처럼 검색기가 AI에게 조각을 건네는 방식이 여전히 맞습니다. 후보를 모으는 방식은 여러 문서를 견줘야 하는 질문에 맞습니다.

06

AI 문서 검색의 근거는 검색기가 고른 조각이 아니라 AI가 후보 안에서 직접 읽은 원문입니다

두 논문이 보여 준 것은 RAG가 끝났다는 것이 아니라, 검색기와 AI가 일을 나누는 방식이 바뀌었다는 것입니다. 정리하면 다음과 같습니다.

- **AI에게 조각만 건네면 놓친 문서는 끝내 보이지 않는다** — 같은 모델(Claude Sonnet 4.6)에서 조각을 받은 쪽은 69.0%, 직접 찾아 읽은 쪽은 80.0%였다.
- **직접 뒤지기는 문서가 많아지면 무너진다** — 20만 개에서 13.6점, 40만 개에서 37.5%까지 떨어졌다.
- **검색기는 후보를 고르고 AI는 그 안을 직접 대조한다** — 문서가 100배로 늘어도 정답은 80개에서 70개로 완만하게 줄었다.
- **결과를 정한 것은 직접 대조 단계다** — 문서를 가로질러 찾기를 막으면 82개가 40개로 줄었고, 검색기 종류는 작은 차이였다.
- **회사 문서는 네 가지를 먼저 본다** — 본문 사본, 열람 범위, 읽은 원문만 인용, 우리 질문으로 측정.



이 글을 한 장으로 요약하면 이렇습니다

회사 문서를 찾는 AI를 고를 때는 「어떤 검색기를 쓰는가」보다 「AI가 답의 근거로 무엇을 실제로 읽었는가」를 먼저 물어보면 됩니다. 답마다 읽은 원문의 위치를 댈 수 있는 AI라야 그 답을 회사 일에 쓸 수 있습니다.

WACE(웨이스)는 공장 운영의 두뇌인 제조 AI VEXPLOR를 만드는 회사다.

07

함께 읽어보세요

- 웨이스는 VEXPLOR WORK에서 AI가 회사 문서를 찾을 때 후보를 먼저 모으게 하고, 실제로 읽은 원문만 근거로 인용하게 합니다
- RDB, RAG는 정말 끝났나?
- Graph RAG가 벡터 RAG보다 나은 순간과 아닌 순간

원문 blog.vexplor.com/post/ai-search-retriever-picks-candidates-to-read
발행 2026년 11월 29일 · 방동걸 (주)웨이스 대표이사

회사 주식회사 웨이스 (WACE Inc.) · 대표이사 방동걸 · 사업자등록번호 439-87-02852
주소 (18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동) 1702-1708호 (영천동, 실리콘앨리동탄)
연락처 전화 070-4730-0629 · contact@wace.me
홈페이지 wace.me · linkedin.com/company/wace-inc

원문 홈페이지 링크드인