DAY 06 · 1부 AI가 뭔지

AI는 왜 없는 사실을 태연하게 지어낼까? — 환각

하루 5분 AI 수업 · 2026년 9월 12일

자신 있게 말한다고 사실인 건 아니다

한 줄로 말하면

1. AI는 사실을 찾는 게 아니라 **확률이 높은 다음 말**을 고를 뿐이라, 틀린 문장도 자신 있게 만들어낼 수 있어.

2. 이걸 **환각(hallucination)**이라 불러 — 거짓말할 의도가 아니라, 확률적으로 그럴싸한 답을 만드는 구조의 한계야.

3. 학습 데이터에 없거나 드문 사실, 최신 정보, 정확한 숫자·날짜를 물을 때 특히 잘 틀려 — **항상 다른 곳에서 확인**하자.

그림 한 장

정답이 없어도, 확률은 어떻게든 고른다

질문:
“이 주장을 뒷받침하는
논문이 뭐야?”

→

AI: 그럴듯한 이름·제목을
확률적으로 조합

→

**결과: 존재하지 않는
논문과 저자**

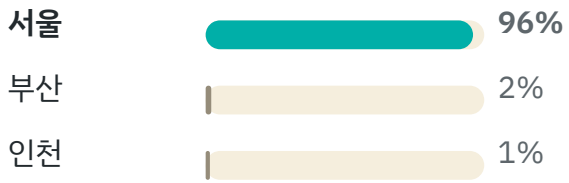
확인해보면: 그런 이름의 저자도, 그런 제목의 논문도 실제로는 없었다.

핵심: 중요한 사실은 항상 다른 출처로 다시 확인하자.

질문에 따라 확률이 얼마나 다르게 퍼질까

96%

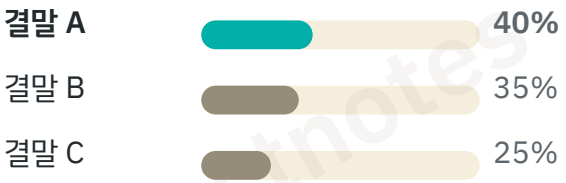
쉬운 질문 · '대한민국의 수도는?'



학습 데이터에 아주 많이 등장 → 확률이 한쪽에 몰린다

40%

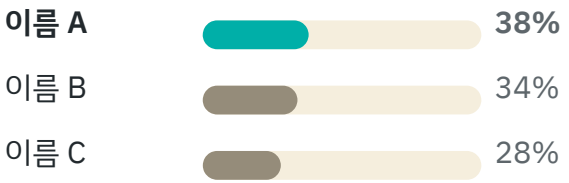
애매한 질문 · '이 소설의 숨은 결말은?'



확실한 정답이 부족 → 확률이 고르게 퍼진다

38%

없는 질문 · '그 논문의 저자는?'



정답 자체가 없어도, 확률은 어떻게든 고른다

확률이 고르더라도, AI는 그중 하나를 '자신 있게' 골라 말한다.

※ 수치는 이해를 돕기 위한 예시다.

확인해보기

세 문제만 풀어 보자

Q1 환각(hallucination)이 생기는 근본적인 이유는?

① AI가 거짓말을 하려는 의도가 있어서
② AI가 확률적으로 그럴듯한 말을 고르는 구조라서
③ 서버 오류 때문에
④ 사용자가 질문을 잘못해서

Q2 **(O / X)** 환각은 학습 데이터에 없거나 드문 내용을 물을 때 더 자주 나타난다.

Q3 **(O / X)** AI가 자신 있는 말투로 대답하면 그 내용은 항상 사실이다.

왜 그런지까지 알고 가자

Q1

② AI는 사실 검색이 아니라 확률이 높은 다음 말을 고르는 구조라서, 정답이 없어도 그럴듯한 말을 만들어낸다.

Q2

O 드문 사실일수록 확률 분포가 흐릿해지고, 그 빈틈을 그럴듯한 말로 메운다.

Q3

X 자신감 있는 말투와 사실 여부는 별개다. 중요한 정보는 꼭 따로 확인하자.

자신 있게 말한다고
사실인 건 아니다.

내일은 '기억의 한계' — AI는 왜 방금 한 말도 잊어버릴까?