

LESSON

중급 초6·중2 AI 교실

19

AI가 지어낸 거짓말 찾기 (환각)

의심 신호를 찾고, 출처를 확인하고, 두 곳 이상에서 교차 확인해요.

 **학습 목표** 이번 시간이 끝나면...

- ✓ 환각을 찾는 4단계(의심 신호 → 출처 확인 → 교차 확인 → 고치거나 버리기)를 설명할 수 있어요.
- ✓ 출처와 교차 확인의 뜻을 말할 수 있어요.
- ✓ 믿을 만한 자료를 골라 AI의 답을 확인할 수 있어요.

• 중급 19

AI가 지어낸 거짓말 찾기 (환각)

AI의 그럴듯한 답도 틀릴 수 있어요.

출처를 찾고, 두 곳 이상에서 확인해요.



“내일 휴교래!”

진짜인지 학교 알림부터 확인!

AI의 답도 출처를 찾아 확인해요.

1

의심 신호
찾기

2

출처
확인

3

두 곳 이상
교차 확인

4

고치거나
버리기 꼭 알아야 할 단어**출처** = 정보가 나온 곳**교차 확인** = 여러 자료를 비교해 확인하기Q. AI가 알려 준 출처는
항상 실제로 있다?

O

/

X

정답: X

그림 1 한눈에 보는 19강 — 확인하는 네 단계를 차례로 따라가 보세요.



쉽게 풀어 보기

AI의 그럴듯한 답에서 거짓을 찾아내는 방법을 알아봐요.



그렇듯하지만 틀린 답, 환각

AI가 없는 사실을 그럴듯하게 지어내는 것을 **환각(hallucination)**이라고 해요. 없는 책 제목, 틀린 날짜, 하지 않은 말을 진짜처럼 말하기도 해요.

환각을 찾으려면 정보가 어디서 나왔는지, 곧 **출처(source)**를 확인해야 해요.

그리고 여러 자료를 비교해 확인하는 **교차 확인(cross-check)**을 해요. 서로 다른 믿을 만한 자료 두 곳 이상에서 같은 내용인지 봐요.



"내일 휴교래!" 소문 확인하기로 이해해요

친구가 "내일 휴교래!"라고 하면 바로 믿고 가방을 안 챙기나요? 먼저 **학교 누리집이나 알림장, 선생님 말씀으로 확인**하지요.

소문이 아무리 그럴듯해도 믿을 만한 곳에서 확인하기 전에는 사실이 아니에요.

AI의 답도 마찬가지예요. 말투가 자신 있고 문장이 매끄러워도, 확인하기 전까지는 '**그럴듯한 소문**'으로 생각해요.



한 줄 요약

AI는 없는 사실을 그럴듯하게 지어낼 때가 있어요. 이것을 환각이라고 해요. 의심스러운 부분을 찾고, 출처를 확인하고, 두 곳 이상에서 교차 확인하면 속지 않을 수 있어요.

**생활 속 의심 신호**

AI의 답에 **정확한 숫자, 날짜, 사람 이름, 책 제목, 인용한 말**이 나오면 특히 조심해요. 환각이 자주 숨는 곳이에요.

AI가 알려 준 책이나 링크가 실제로 있는지 **직접 열어 확인**해요. 같은 AI에게 다시 묻는 것은 교차 확인이 아니에요. 같은 곳에서 같은 실수를 할 수 있으니까요.

확인할 수 없는 내용은 쓰지 않는 것이 가장 안전해요. 숙제에 쓸 때는 확인한 출처를 함께 적어요.

미리 보기 뒤에 나오는 '직접 해보기 미션'에서 틀린 곳이 숨은 글을 교과서로 확인해 봐요.

미리 생각하기 내가 정보를 확인할 때 믿을 만한 곳을 두 군데 적어 보세요.

예: 교과서, 도서관 백과사전, 공공 기관 누리집

123 숫자로 보는 19강**사실****1446년**

훈민정음이 반포된 해예요. 세종
대왕 때의 일이에요.

예시**2곳 이상**

교차 확인은 서로 다른 믿을 만한
자료 두 곳 이상에서 해요.

예시**4단계**

의심 신호 찾기 → 출처 확인 →
교차 확인 → 고치거나 버리기

사실 = 실제로 있었던 일 · **예시** = 이해를 도우려고 만든 숫자



표로 정리해요

환각을 찾는 흐름과 믿을 만한 자료를 정리해요.



환각을 찾는 4단계

단계	하는 일	훈민정음 예시
① 의심 신호 찾기	정확한 숫자, 날짜, 사람 이름, 책 제목, 인용한 말이 나오면 특히 조심해요.	"1446년", "이순신 장군"이 나왔어요.
② 출처 확인	그 정보가 어디서 나왔는지 찾아봐요. AI가 알려 준 책이나 링크가 실제로 있는지도 직접 열어 확인해요.	교과서의 훈민정음 단원을 찾아요.
③ 교차 확인	교과서, 백과사전, 공공기관 누리집처럼 믿을 만한 곳 두 곳 이상에서 같은 내용인지 비교해요.	백과사전에서도 확인해요.
④ 고치거나 버리기	틀린 부분은 고치고, 확인할 수 없는 내용은 쓰지 않아요.	만든 사람을 '세종 대왕'으로 고쳐요.



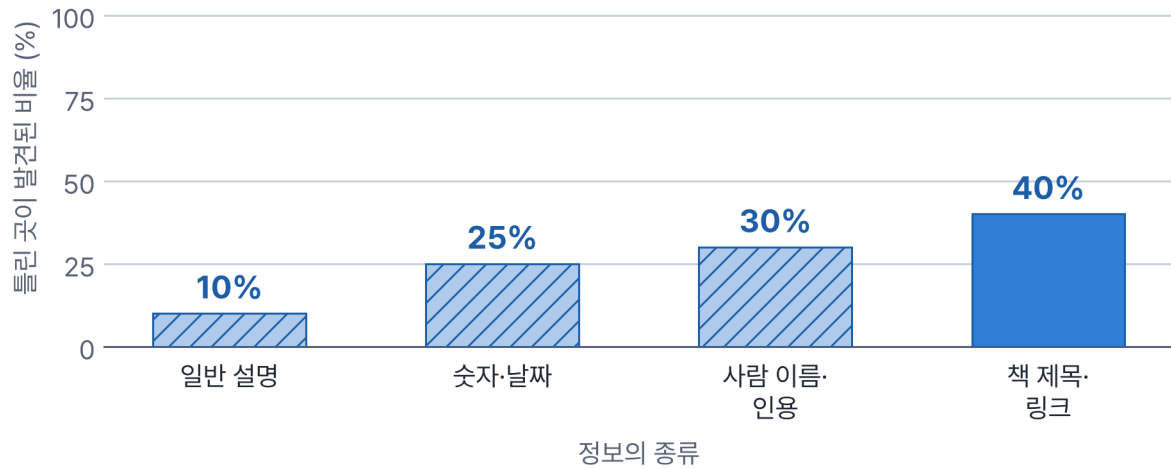
교차 확인일까, 아닐까?

살펴봐요	교차 확인이에요	교차 확인이 아니에요
자료	교과서와 백과사전	같은 AI에게 다시 묻기
사람	선생님께 여쭙고 책으로도 확인	친구 한 명의 말만 믿기
누리집	공공 기관 누리집 두 곳 비교	출처를 모르는 글 여러 개

생각해 보기 AI가 알려 준 책 제목이 진짜 있는지 확인하는 방법을 두 가지 적어 보세요.

도서관 검색, 출판사 누리집을 떠올려 봐요.

AI 답의 정보 종류별로 틀린 곳이 발견된 비율 (예시)



※ 이해를 도우려고 만든 예시 숫자예요. 실제 비율은 AI와 질문에 따라 크게 달라요.

그래프 읽기 어떤 정보에서 틀린 곳이 가장 많았나요? 그런 정보는 어떻게 확인하면 좋을까요?

이 그래프는 예시 데이터예요. '의심 신호'와 연결해 생각해 봐요.

가 꼭 알아야 할 단어

출처

source

정보가 나온 곳

교차 확인

cross-check

여러 자료를 비교해 확인하기

**직접 해보기 미션**

손과 눈으로 직접 확인해 봐요.

틀린 곳 찾기 탐정

선생님이나 보호자가 **AI의 답처럼 꾸민 글**(일부러 틀린 곳 두 군데를 넣은 글)을 준비해 주면, 교과서와 백과사전으로 틀린 곳을 찾아보세요. (예: “훈민정음은 1446년에 반포되었고, 이순신 장군이 만들었다.” → 만든 사람이 틀렸어요.)

1 글에서 숫자, 날짜, 이름 같은 의심 신호에 밑줄을 그어요.

2 교과서와 백과사전 두 곳에서 하나씩 확인해요.

3 틀린 곳과 바른 내용, 확인한 출처를 표에 적어요.

※ 직접 AI를 써 본다면 보호자나 선생님과 함께, 나이 기준에 맞는 서비스에서 교과서로 확인할 수 있는 질문을 하고, 답을 두 곳 이상에서 확인해 봐요.

기록표

	의심 신호	바른 내용	확인한 출처 2곳
틀린 곳 1			
틀린 곳 2			
맞았던 곳			

- 틀린 곳을 찾는 데 가장 도움이 된 자료는 무엇이었나요?

그렇듯해서 속을 뻔한 부분

앞으로 AI 답을 볼 때 나의 약속

🔑 정리 AI의 답은 확인하기 전까지 ‘그렇듯한 소문’이에요. 출처와 교차 확인으로 가려내요.

**확인 퀴즈**

스스로 풀고, 정답은 마지막 쪽에서 확인해요.

1. **O/X** AI가 알려 준 출처는 항상 실제로 있다.☐

o

☐

x

2. **고르기** 교차 확인으로 알맞은 것은?☐

① 같은 AI에게 같은 질문을 여러 번 하기

☐

② 백과사전과 공공기관 누리집 두 곳에서 확인하기

☐

③ 친구 한 명의 말만 믿기

3. **빈칸** 정보가 나온 곳을 _____(이)라고 해요.4. **쓰기** AI의 답에서 특히 조심해야 할 '의심 신호'를 두 가지 써 보세요.

**중2 한 걸음 더**

환각은 왜 생기고, 어떻게 줄일까?

언어 모델은 사실을 저장한 사전이 아니라, 그럴듯한 다음 낱말을 고르는 계산이라서 환각이 생겨요.

검색한 자료를 함께 보면서 답하게 하는 '**검색 증강 생성(RAG, Retrieval-Augmented Generation)**'으로 환각을 줄일 수 있지만 완전히 없앨 수는 없어요.

외국에서는 변호사가 AI가 지어낸 판례를 확인하지 않고 법원에 냈다가 문제가 된 일도 있었어요.

전문가도 확인을 건너뛰면 실수한다는 뜻이에요.

- 생각해 봐요: AI가 출처 링크를 함께 보여 주면 환각 걱정이 없을까요? 내 생각을 적어 보세요.



용어 정리

이번 시간에 나온 어려운 말을 모았어요.

용어	영어	쉽게 말하면	예시
환각	hallucination	AI가 없는 사실을 그럴듯하게 지어내기	없는 책 제목
출처	source	정보가 나온 곳	교과서 23쪽
교차 확인	cross-check	여러 자료를 비교해 확인하기	교과서 + 백과사전
의심 신호	red flag	틀렸을 수 있어 조심할 표시	정확한 날짜, 인용
인용	quotation	남의 말이나 글을 그대로 옮기기	"..."라고 말했다
공공 기관	public institution	나라나 지역이 운영하는 기관	정부, 시청 누리집
백과사전	encyclopedia	여러 지식을 모아 정리한 책	도서관 백과사전
판례	court precedent	법원이 앞서 내린 판결	변호사가 찾는 자료
검색 증강 생성	RAG	검색한 자료를 함께 보며 답하는 방법	자료 보고 답하는 챗봇
사실 확인	fact check	맞는 정보인지 확인하기	날짜 확인



정답과 풀이

퀴즈를 다 풀 다음에 확인해요

1. X — AI는 없는 책 제목이나 링크를 지어낼 수 있어서 직접 확인해야 해요.
2. ㉠ — 서로 다른 믿을 만한 자료끼리 비교해야 해요.
3. 출처(source) — 책 제목, 누리집 주소가 출처예요.
4. (예시 답) 정확한 숫자나 날짜, 사람 이름, 책 제목, 인용한 말 가운데 두 가지