

LESSON

고급 초6·중2 AI 교실

24

다음 단어 맞추기 게임과 챗GPT

다음 단어 맞추기를 연습한 LLM을 대화 예시와 사람의 평가로 다듬었어요.

🎯 학습 목표 이번 시간이 끝나면...

- ✓ 대화형 AI가 만들어지는 4단계(사전 학습 → 미세 조정 → RLHF → 대화 서비스)를 설명할 수 있어요.
- ✓ 확률과 무작위성 때문에 같은 질문에도 답이 달라지는 까닭을 말할 수 있어요.
- ✓ 학습 시점과 환각 때문에 중요한 정보는 직접 확인해야 함을 알 수 있어요.

● 고급 24

다음 단어 맞추기 게임과 챗GPT

챗봇은 다음 단어 맞추기를 연습했어요.

대화 예시와 사람 평가로 한 번 더 다듬어요.



“반짝반짝 작은 __”
다음 말이 바로 떠오르지?

챗봇은 이 놀이를 아주 많이 연습했어요.



📖 꼭 알아야 할 단어

미세 조정 = 배운 모델을 목적에 맞게 더 가르치기

학습 시점 = AI가 배운 자료가 끝나는 때

Q. 챗봇은 학습 이후에 생긴 일도
항상 정확히 안다?

O / X 정답: X

그림 1 한눈에 보는 24강 — LLM이 챗봇이 되기까지의 네 단계를 따라가 보세요.



쉽게 풀어 보기

다음 단어 맞추기 연습이 어떻게 대화하는 AI가 되는지 알아봐요.



LLM에서 챗봇으로

사전 학습을 마친 LLM은 글을 잘 이어 쓰지만, 질문에 친절하게 답하는 법은 아직 몰라요. 그래서 질문과 좋은 답의 예시로 한 번 더 가르쳐요. 이것을 **미세 조정(fine-tuning)**이라고 해요.

그다음 사람들이 여러 답 가운데 더 좋은 답을 고르면, 그런 답을 더 잘 내도록 고쳐요. 이 방법을 **RLHF(사람 피드백 강화 학습)**라고 해요.

그래도 AI가 배운 자료가 끝나는 때, 곧 **학습 시점(knowledge cutoff)** 이후의 일은 모를 수 있어요.



"반짝반짝 작은 ____"으로 이해해요

"반짝반짝 작은..."이라는 말을 들으면 '별'이 바로 떠오르지요? **많이 들어 봐서 다음에 올 말이 익숙하기 때문**이에요.

LLM은 이런 '다음 말 맞추기'를 사람이 평생 읽어도 다 못 읽을 만큼 많은 글로 연습했어요. 그래서 말을 자연스럽게 이어 가요.

하지만 **익숙한 말을 잇는 것과 그 내용이 사실인지 아는 것은 서로 다른 일**이에요.



한 줄 요약

챗GPT 같은 대화형 AI는 '다음 단어 맞추기'를 엄청나게 연습한 LLM을 대화 예시와 사람의 평가로 한 번 더 다듬어 만든 거예요. 그래도 학습한 시점 이후의 일은 모를 수 있고, 틀린 말도 해요.

🗨️ 생활 속 대화형 AI 바르게 쓰기

대화형 AI는 질문에 답하기, 글 다듬기, 아이디어 넓히기에 쓰여요. 같은 질문에도 답이 조금씩 다른 것은, 늘 확률이 가장 높은 단어만 고르지 않고 **약간의 무작위성을 섞어** 고르기 때문이에요.

최신 뉴스나 오늘 날씨처럼 학습 시점 이후의 일은 모를 수 있고, 모르면서 그럴듯하게 지어낼 수도 있어요.

그래서 중요한 정보는 직접 확인하고, 개인 정보는 넣지 않아요. 많은 서비스는 나이 기준이 있으니 보호자나 선생님과 함께 써요.

미리 보기 뒤에 나오는 '직접 해보기 미션'에서 반 친구들과 다음 단어 확률을 직접 계산해 봐요.

미리 짐작하기 “주말에 가족과 함께...” 다음에 올 단어로 반에서 가장 많이 나올 것 같은 말을 적어 보세요.

미션에서 실제 결과와 비교해 봐요.

123 숫자로 보는 24강

사실

2022년 11월

챗GPT가 공개됐어요. 그 뒤 비슷한 대화형 AI가 여러 회사에서 나왔어요.

예시

25명 중 10명

미션 예시: 반 25명 가운데 10명이 다음 단어로 '공원'을 적었어요.

예시

40%

그때 '공원'의 확률 ($10 \div 25$). 확률로 다음 단어를 고르는 연습이에요.

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



표로 정리해요

챗봇이 되기까지의 과정을 정리해요.



챗봇이 되기까지 4단계

단계	하는 일	비유
① 다음 단어 맞추기 (사전 학습)	아주 많은 글로 다음 토큰을 맞추는 연습을 해요.	“반짝반짝 작은 ____” 맞추기를 엄청나게
② 대화 예시로 다듬기 (미세 조정)	질문과 좋은 답의 예시를 보여 주며 대화하는 법을 익혀요.	모범 답안을 보며 답하는 법 익히기
③ 사람 평가 반영 (RLHF)	사람들이 여러 답 가운데 더 좋은 답을 고르면, 그런 답을 더 잘 내도록 고쳐요.	선생님이 더 좋은 답에 표시해 주기
④ 대화 서비스	우리가 질문하면 토큰을 하나씩 이어 붙여 답을 만들어요.	질문을 받고 한 단어씩 답하기



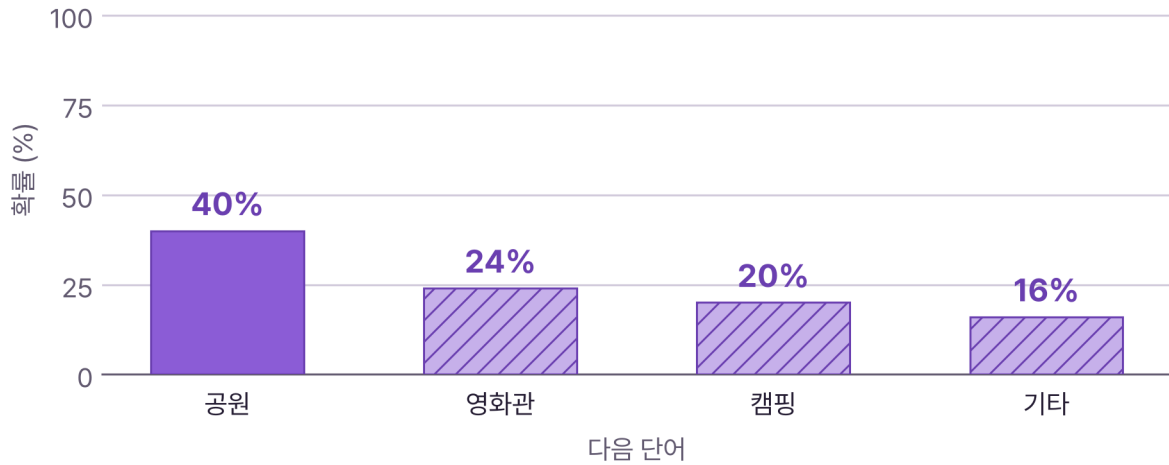
가장 높은 것만 고를 때와 무작위성을 섞을 때

고르는 방법	좋은 점	아쉬운 점
늘 가장 높은 단어	답이 늘 같고 안정적이에요.	뻘하고 같은 말을 되풀이하기 쉬워요.
가끔 낮은 단어도	다양하고 새로운 문장이 나와요.	엉뚱하거나 틀린 말이 섞일 수 있어요.

생각해 보기 시를 쓸 때와 수학 풀이를 물을 때, 어느 고르는 방법이 더 알맞을까요? 까닭도 적어 보세요.

다양함이 필요한 일과 정확함이 필요한 일을 나눠 봐요.

“주말에 가족과 함께...” 다음 단어 (우리 반 25명) (예시)



※ 반 25명을 예로 들어 만든 숫자예요 (공원 10명, 영화관 6명, 캠핑 5명, 기타 4명). 실제 결과는 반마다 달라요.

그래프 읽기 ‘공원’만 고르면 문장은 어떻게 될까요? 가끔 ‘캠핑’을 고르면 어떻게 달라질까요?

이 그래프는 예시 데이터예요. 막대의 합이 100%인지도 확인해 봐요.

가 꼭 알아야 할 단어

미세 조정

fine-tuning

미리 배운 모델을 특정 목적에 맞게 더 가르치는 것

사람 피드백 강화

학습

RLHF

사람이 고른 좋은 답 쪽으로 AI를 다듬는 방법

학습 시점

knowledge cutoff

AI가 배운 자료가 끝나는 때



직접 해보기 미션

손과 눈으로 직접 확인해 봐요.

우리 반 다음 단어 확률 게임

딱이 “**주말에 가족과 함께...**”처럼 문장 앞부분을 칠판이나 종이에 써요. 모두 다음 단어를 하나씩 적은 뒤, 많이 나온 순서대로 표를 만들어 ‘**확률**’을 계산해 봐요. (예: 25명 중 10명이 ‘공원’ → 40%)

- 1 문장 앞부분을 정하고, 모두 다음 단어를 하나씩 적어요.
- 2 많이 나온 순서대로 표에 적고, 확률(사람 수 ÷ 전체 × 100)을 계산해요.
- 3 가장 많이 나온 단어만 고를 때와, 가끔 적게 나온 단어도 고를 때 문장이 어떻게 달라지는지 비교해요.

※ 실제 챗봇으로 해 본다면 보호자나 선생님과 함께, 나이 기준에 맞는 서비스에서 해요.

기록표

	다음 단어	사람 수	확률 (%)
1위			
2위			
3위			
4위			
기타			

- 1위 단어만 이어서 만든 문장과, 3위 단어로 만든 문장을 각각 써 보세요.

확률 게임과 챗봇이 닮은 점

확률로 고르면 생기는 문제

🔑 **정리** 챗봇도 다음 단어 후보마다 확률을 매기고, 약간의 무작위성을 섞어 골라요.

**확인 퀴즈**

스스로 풀고, 정답은 마지막 쪽에서 확인해요.

1. **O/X** 챗봇은 학습 이후에 생긴 일도 항상 정확히 안다.☐

O

☐

X

2. **고르기** 사람이 더 좋은 답을 골라 주어 AI를 다듬는 방법은?☐

① RLHF

☐

② 픽셀

☐

③ 입력층

3. **빈칸** 25명 가운데 5명이 '캠핑'을 적었다면 '캠핑'의 확률은 _____ %예요.4. **쓰기** 같은 질문을 두 번 했는데 챗봇의 답이 달랐어요. 그 까닭을 써 보세요.

**중2 한 걸음 더**

무작위성, 환각, 그리고 검색

챗봇은 늘 확률이 가장 높은 단어만 고르지 않고, **약간의 무작위성**을 섞어서 골라요. 그래서 같은 질문에도 답이 조금씩 달라질 수 있어요. 또 '다음 말로 그럴듯한가'를 기준으로 만들어지기 때문에, 사실이 아닌 내용도 자연스럽게 말하는 **환각**이 생겨요.

요즘은 검색 결과를 함께 참고하는 서비스도 있지만, 그 검색 결과나 요약도 틀릴 수 있어요. 챗GPT는 2022년 11월에 공개되었고, 그 뒤 비슷한 대화형 AI가 여러 회사에서 나왔어요. **어떤 서비스든 중요한 정보는 직접 확인**해요.

- 생각해 봐요: 챗봇이 "제가 아는 정보는 특정 시점까지예요."라고 말한다면, 어떤 질문을 할 때 특히 조심해야 할까요?

**용어 정리**

이번 시간에 나온 어려운 말을 모았어요.

용어	영어	쉽게 말하면	예시
대화형 AI	conversational AI	질문에 대화하듯 답하는 AI	챗GPT 같은 챗봇
사전 학습	pre-training	많은 글로 다음 토큰 맞추기 연습	“반짝반짝 작은 별”
미세 조정	fine-tuning	미리 배운 모델을 목적에 맞게 더 가르치기	대화 예시로 다듬기
사람 피드백 강화 학습	RLHF	사람이 고른 좋은 답 쪽으로 다듬기	더 좋은 답에 표시
피드백	feedback	써 본 사람의 평가나 의견	“이 답이 더 좋아요.”
학습 시점	knowledge cutoff	AI가 배운 자료가 끝나는 때	그 뒤 소식은 모름
무작위성	randomness	정해지지 않고 우연히 고르는 성질	같은 질문, 다른 답
확률	probability	그럴 것 같은 정도	공원 40%
환각	hallucination	없는 사실을 그럴듯하게 지어내기	없는 책 제목
토큰	token	AI가 글을 나눈 작은 조각	한 단어씩 이어 붙이기

**정답과 풀이**

퀴즈를 다 푼 다음에 확인해요

- X** — 학습 시점 이후의 일은 모를 수 있고, 모르면서 그럴듯하게 지어낼 수도 있어요.
- ㉠** — 사람의 피드백을 이용한 강화 학습이에요.
- 20** — $5 \div 25 \times 100 = 20$ 이에요.
- (예시 답)** 챗봇은 확률이 가장 높은 단어만 고르지 않고 약간의 무작위성을 섞어 고르기 때문이에요.