★ ПРИКЛЮЧЕНИЯ КАЯ В СТРАНЕ ИИ ★

Часть 2

Как базовая модель становится помощником

Герда, в прошлый раз мы узнали, как большая языковая модель учится продолжать тексты. Но это же не конец обучения?

Конечно, это только предобучение

Базовая модель

Модель после предобучения. Она умеет работать с текстом и предсказывать вероятные продолжения, но ещё не всегда хорошо следует инструкциям пользователя.

1. Fine-tuning: специализация

Дальше модель дообучается на более специализированных данных. Это называется fine-tuning

Это как университет после школы

медицинские документы | юридические тексты | программный код | финансовые документы | техническая документация | данные компании

Fine-tuning

Дополнительное обучение уже предобученной модели на более специализированных данных или задачах.

2. Instruction tuning

Важно научить модель следовать инструкциям, а не просто продолжать текст.

Примеры: инструкция → ответ

1. Переведи на английский: Доброе утро. → Good morning.
2. Сократи этот текст до трёх предложений. → Краткое резюме.
3. Объясни школьнику, что такое токен. → Токен — кусочек текста, с которым работает модель.

Instruction tuning

Вид fine-tuning, при котором модель обучают на примерах вида: инструкция → хороший ответ.

3. Но что такое хороший ответ?

Объясни, что такое чёрная дыра.

A коротко и понятно

B очень длинно и научно

C понятно, с аналогией

Понятно. Важно обучить модель не только давать правильные ответы, но и выбирать те, которые люди считают понятными и полезными

Хм, а я думала, тюнинг это что-то про тонированную Ладу

Черная дыра — это мой бюджет на вкусняшки.

Хороший ответ

У одного запроса может быть несколько правильных ответов. Люди могут предпочитать один из них из-за формы, ясности или полезности.

4. Human feedback

Объясни ребёнку, что такое электричество.

A слишком сложно — $I = \frac{U}{R}$

B коротко и понятно

C слишком абстрактно

Поэтому людям показывают несколько вариантов ответов! Чтобы оценить удачность ответов

Верно

Human feedback

Обратная связь от людей о том, какие ответы модели они считают более удачными.

Продолжение на следующей странице...

ПРИКЛЮЧЕНИЯ КАЯ В СТРАНЕ ИИ

А зачем нужна Reward model?

Человек не может проверять все ответы, поэтому оценки пользователей используются для обучения еще одной модели

Нужно оценить 1 000 000 ответов

И тут оценки! Это обязательно?

Зачем ещё одна модель?

Люди не могут постоянно оценивать каждый ответ. Поэтому их оценки используют, чтобы обучить отдельную модель-оценщика.

6. Reward model

Ее задача оценивать ответы, а не отвечать

B 0,89

C 0,67

Модель-оценщик (reward model)

Понятно, кто тут двоечница

На самом деле в современных подходах используются и другие алгоритмы без оценщика, но я училась по классике

Reward model

Модель, которая учится оценивать ответы на основе человеческих предпочтений.

7. RLHF

Далее идет стадия RLHF (Reinforcement Learning from Human Feedback).

Цикл RLHF:

вопрос пользователя → ответ модели → оценка (reward model) → корректировка модели → повторяем много раз

Пользователь задает вопрос
Модель генерирует ответ.
Reward модель дает оценку.
Основная модель немного корректируется

Ну окээй.. Это все?

RLHF

Способ дообучения, при котором модель подстраивается под оценки, основанные на человеческой обратной связи.

8. Что меняется внутри модели?

до обучения RLHF → веса изменяются → после обучения RLHF

А что меняется?

То, о чем говорили в прошлой части: корректируются веса

Все помешались на весах

Веса

Во время обучения меняются числовые параметры модели — веса. Они влияют на то, какие ответы становятся более вероятными.

9. Повторим всё по шагам

1 Pretraining
Учусь языку и закономерностям в тексте.

2 Fine-tuning
Дополнительно обучаюсь на специальных данных и задачах.
медицина · право · код · данные компании

Instruction tuning: учусь выполнять инструкции пользователя.

3 Human feedback / RLHF
Учусь чаще давать ответы, которые люди оценивают выше.

Давайте повторим стадии обучения

Схема

Общие знания → специализация и инструкции → человеческие предпочтения.

10. Финал

Вот это какой огромный труд!

Да, мы учимся постоянно

А потом вам скажут: «Все хорошо, но нужно переделать»

БАЗОВАЯ МОДЕЛЬ → СПЕЦИАЛИЗАЦИЯ И ИНСТРУКЦИИ → ЧЕЛОВЕЧЕСКИЕ ПРЕДПОЧТЕНИЯ

Финал

За качественным ответом ИИ стоит несколько этапов обучения: общие знания, специализация и подстройка под человеческие предпочтения.

Продолжение следует...