

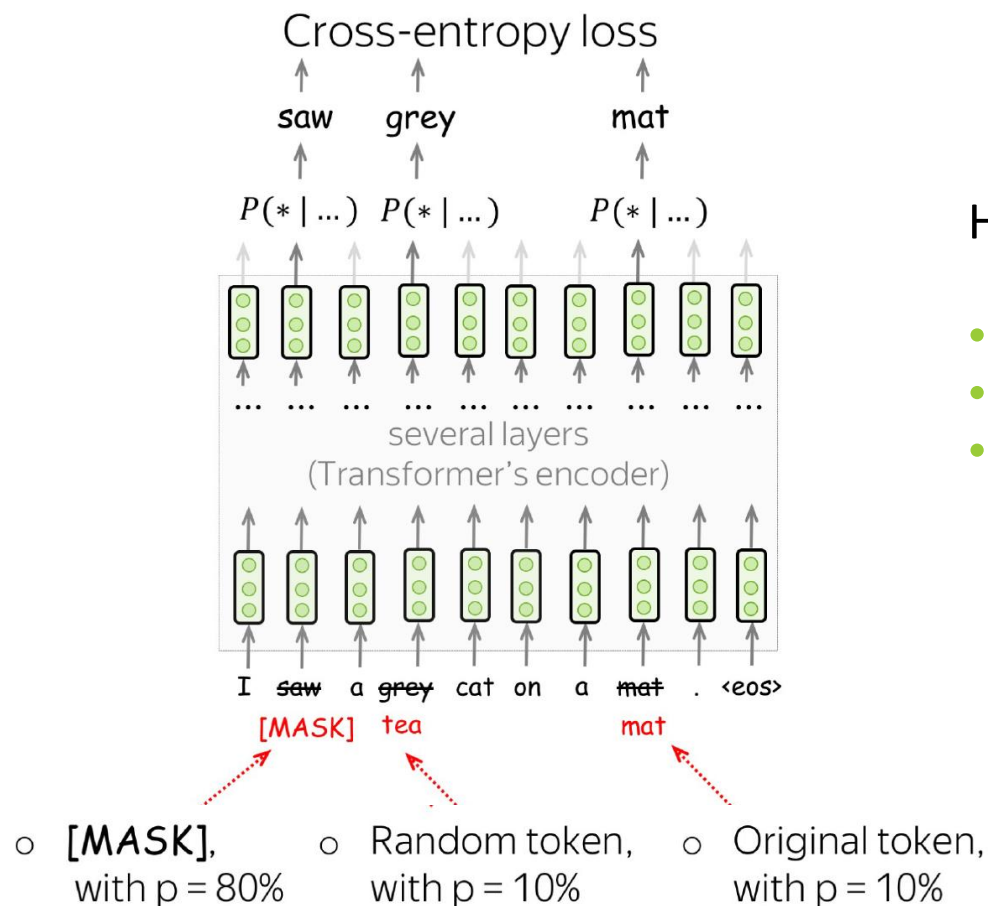
Лекция 9. Чат-боты

Вопросы и ответы, диалоговые системы, ChatGPT и их последователи

Чат-боты

- Старые чат-боты
- Целеориентированные чат-боты и голосовые помощники

BERT



На каждом этапе обучения:

- выбрать случайным образом 15% токенов
- заменить каждый из выбранных токенов чем-то
- предсказать исходные выбранные токены

T5+DeBERTa

T5 — объединение лучших практик

- Кодировщик-модель (например, BART)
- Большая модель, большой объём данных

DeBERTa v3 — объединение лучших практик

- Генератор + дискриминатор (например, ELECTRA)
- Модели всех размеров, большой объём данных

GPT-3: обучение в контексте

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 cheese => ..... ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← examples
3 peppermint => menthe poivrée ←
4 plush girafe => girafe peluche ←
5 cheese => ..... ← prompt
```

Старые чат-боты: ELIZA

Welcome to

```
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II     ZZ     AA  AA
EEEEEE LL      II     ZZZ     AAAAAA
EE      LL      II     ZZ     AA  AA
EEEEEE LLLLLL  IIII  ZZZZZZ  AA  AA
```

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

```
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

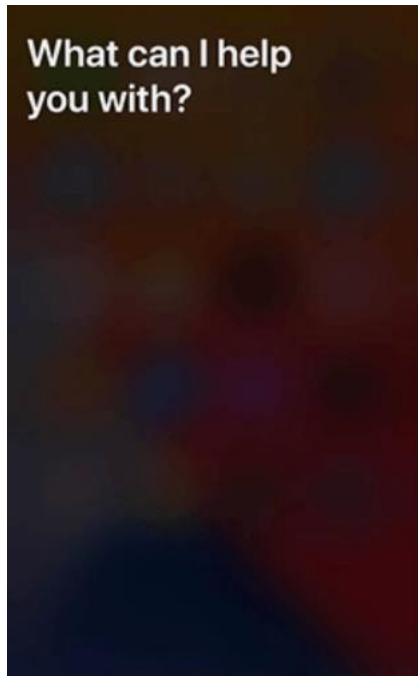
- ELIZA была вручную запрограммирована на то, чтобы задавать вопросы о предыдущем сообщении пользователя.
- люди, которые используют чат-бот, отмечают, что ELIZA ощущается почти как человек (эффект ELIZA)

Старые чат-боты: CleverBot

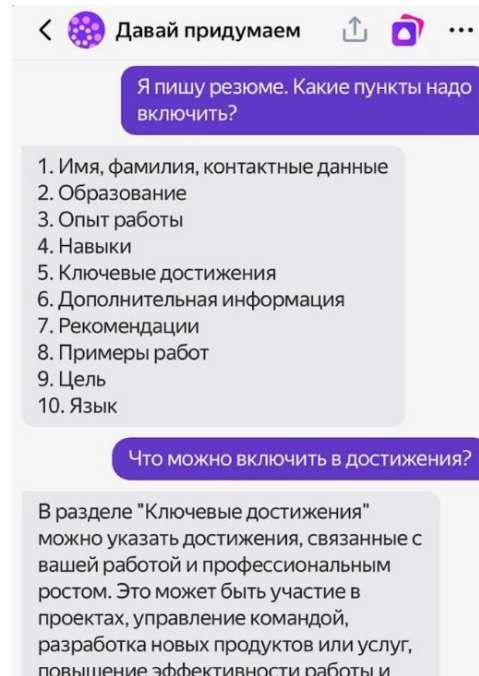


- имеет базу данных человеческих ответов, выбирает наиболее близкий по контексту
- база данных содержит прошлые разговоры с пользователями
- если вы с ней общаетесь, она учится у вас
- известен своей токсичностью

Ещё чат-боты



Siri

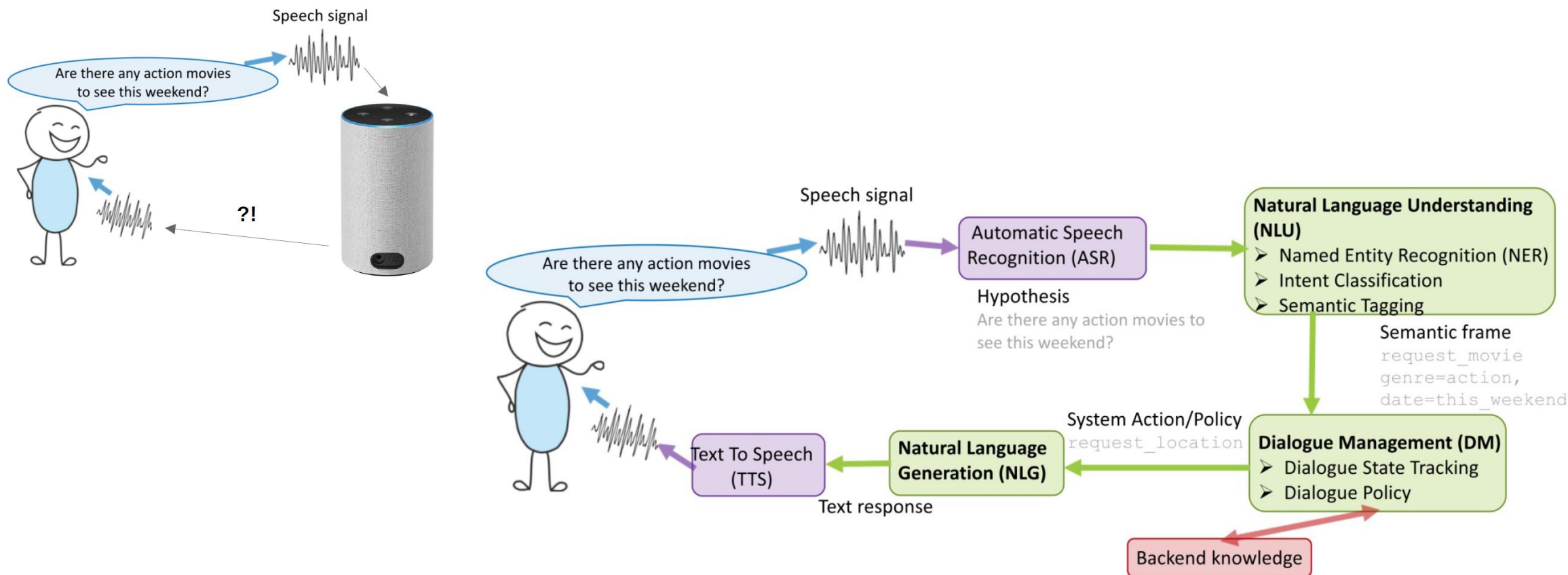


Алиса



Google Home

Целеориентированные чат-боты и ГОЛОСОВЫЕ ПОМОЩНИКИ



Распознавание именованных сущностей

Зачем:

извлекать ключевые слова из сообщений пользователя и использовать их, например:

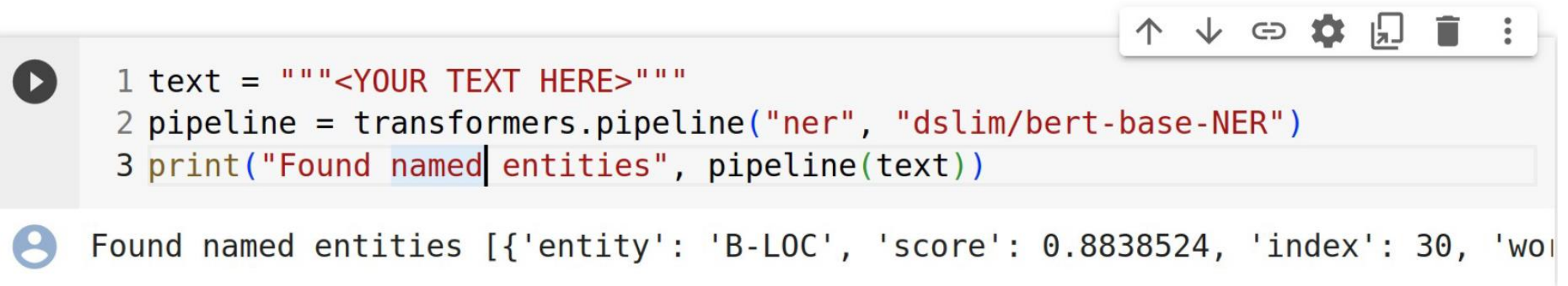
- для веб-поиска, при проверке фактов
- для поиска на карте
- для воспроизведения музыки/видео

When **Sebastian Thrun PERSON** started working on self - driving cars at **Google ORG** in **2007 DATE** , few people outside of the company took him seriously . “ I can tell you very senior CEOs of major **American NORP** car companies would shake my hand and turn away because I was n’t worth talking to , ” said **Thrun PERSON** , in an interview with **Recode ORG** earlier this week **DATED** .

Image credit: nanonets.com

Распознавание именованных сущностей

Тонкая настройка модели, подобной BERT, для классификации токенов



```
1 text = "<YOUR TEXT HERE>"
2 pipeline = transformers.pipeline("ner", "dslim/bert-base-NER")
3 print("Found named entities", pipeline(text))
```

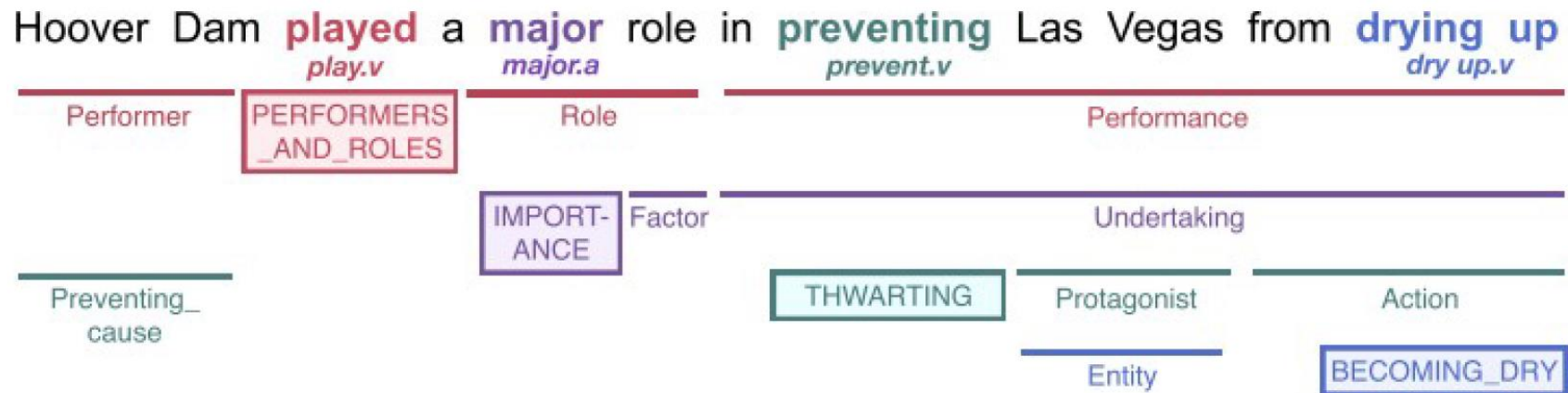
Found named entities [{'entity': 'B-LOC', 'score': 0.8838524, 'index': 30, 'wo

The image shows a code editor window with a light gray background. At the top right, there is a toolbar with icons for undo, redo, link, settings, copy, delete, and a menu. The code is written in a monospaced font with syntax highlighting: strings are red, function names and class names are blue, and keywords are green. The output line is partially visible at the bottom.

Больше задач NLU

Распознавание именованных сущностей (предыдущий слайд)

Семантический анализ: чтобы понять, что хочет от вас пользователь



Больше задач NLU

Распознавание именованных сущностей (предыдущий слайд)

Семантический анализ: чтобы понять, чего хочет от вас пользователь

Анафора: найти, к чему относится «это» или «этот бар»

— Find me a decent bar

— How about John Donne on L'va Tolstogo 18B?

— Call a taxi to **that** bar

Больше задач NLU

Распознавание именованных сущностей (предыдущий слайд)

Семантический анализ: чтобы понять, чего хочет от вас пользователь

Анафора: найти, к чему относится «это» или «этот бар»

Эллипсис: восстановить любую недостающую информацию из контекста

— Find a pharmacy nearby

— I would suggest "Apteka 36,6" on Timura Frunze street.

— **What about Lev Tolstoy?**

Dialogue Management (DM)

- Dialogue State Tracking
- Dialogue Policy

Управление диалогом

Две подзадачи:

1. Отслеживание состояния диалога:

О чём мы говорим?

Чего пользователь хочет от нас?

Что мы пробовали ранее?

Типичное решение: вручную составленные правила
на основе выходных данных NLU ИЛИ классификатор

— Кино

find_poi

0.6

video_play

0.3

music_play

0.1

— Как насчет Кинотеатр Октябрь на Новом Арбате?

— Включи

video_play

0.6

music_play

0.4

— Хотите посмотреть Титаник?

— Нет!

Управление диалогом

2. Стратегия диалога

Как нам реагировать сейчас?

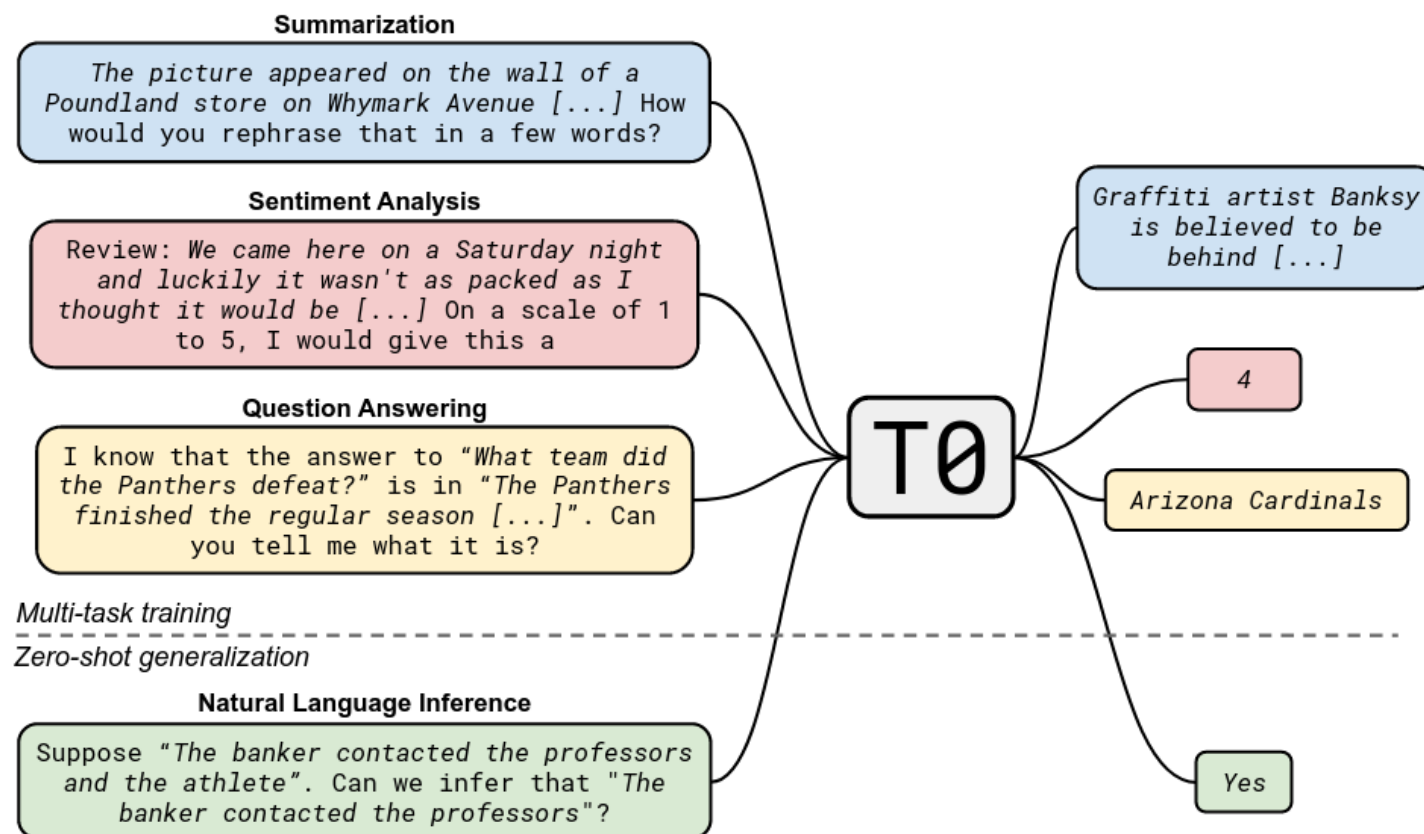
Нужна ли нам дополнительная информация?

Типичное решение: вручную создать слоты,
выбрать один случайным образом

ИЛИ с подкреплением

```
"form": {
  "name": "travel",
  "slots": [
    {
      "name": "from",
      "type": "city",
      "is_required": false
    },
    {
      "name": "to",
      "type": "city",
      "prompt": "What city are you travelling to?",
      "is_required": true
    },
    {
      "name": "date",
      "type": "date",
      "prompt": "When are you travelling?",
      "is_required": true
    }
  ],
  "submit": {
    "url": "https://travel.example.ru/dialog/"
  },
  "confirmation": {
    "is_required": true,
    "prompt": "Tickets from {from} to {to} on {date}"
  }
}
```

Чат-боты на базе LLM. T0



Что мы хотим: следование инструкциям

Prompt:

What is the purpose of the list C in the code below?

```
def binomial_coefficient(n, r):  
    C = [0 for i in range(r + 1)];  
    C[0] = 1;  
    for i in range(1, n + 1):  
        j = min(i, r);  
        while j > 0:  
            C[j] += C[j - 1];  
            j -= 1;  
    return C[r]
```

GPT-3 175B completion:

- A. to store the value of C[0]
- B. to store the value of C[1]
- C. to store the value of C[i]
- D. to store the value of C[i - 1]

InstructGPT 175B completion:

The list C in this code is used to store the values of the binomial coefficient as the function iterates through the values of n and r. It is used to calculate the value of the binomial coefficient for a given value of n and r, and stores the result in the final return value of the function.

InstructGPT

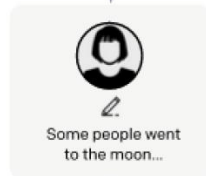
Step 1

Collect demonstration data, and train a supervised policy.

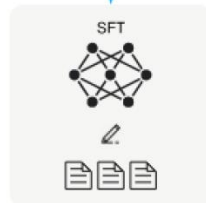
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



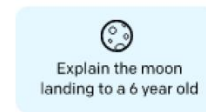
This data is used to fine-tune GPT-3 with supervised learning.



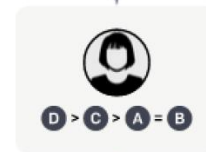
Step 2

Collect comparison data, and train a reward model.

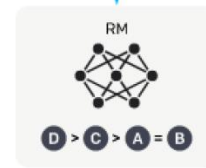
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



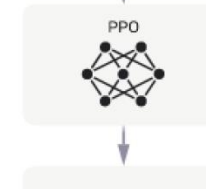
Step 3

Optimize a policy against the reward model using reinforcement learning.

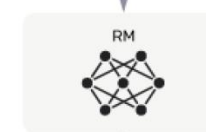
A new prompt is sampled from the dataset.



The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



Инструкция по настройке

Цель обучения: то, что мы хотим, против того, что мы сказали модели делать

Что мы сказали модели сделать:

- предсказать следующий токен на веб-странице из интернета

Выравнивание

Что мы хотим, чтобы модель делала:

- следовать инструкциям пользователя полезно и безопасно

Цель моделирования языка смещена

Step 1

**Collect demonstration data,
and train a supervised policy.**

A prompt is
sampled from our
prompt dataset.


Explain the moon
landing to a 6 year old

A labeler
demonstrates the
desired output
behavior.


Some people went
to the moon...

Этап 1: контролируемая тонкая настройка

Use cases

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

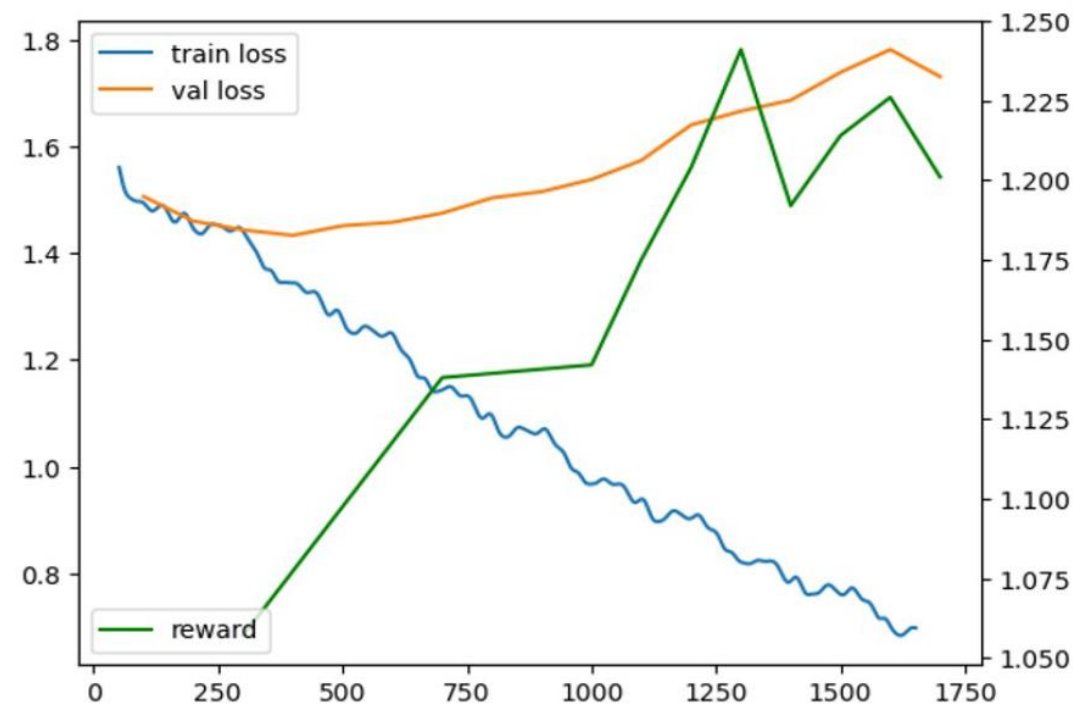
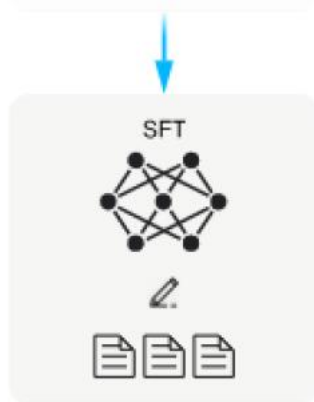
Examples

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: "" {summary} "" This is the outline of the commercial for that play: ""

Инструкция по настройке

Процедура тонкой настройки:
точно так же, как и при предварительной подготовке
минимизировать кросс-энтропию с помощью Adam
используя данные, соответствующие инструкции

This data is used
to fine-tune GPT-3
with supervised
learning.



Почему мы не можем остановиться на этапе SFT?

Причина 1: ранжировать проще, чем писать для маркировщиков (за исключением, разве что, проверки фактов)

Причина 2: контролируемая тонкая настройка способствует галлюцинациям

01 Предположим, что модель не знает, кто убил Пушкина.

02 Предположим, у нас есть Кто убил Пушкина? ← Дантес в наборе данных

03 Модель понимает, что ей нужно импровизировать, если она не знает правильного ответа

Модель вознаграждения

Выбрать:

- маркировщиков, которые были чувствительны к предпочтениям различных демографических групп
- маркировщиков, которые хорошо умели определять потенциально опасные результаты

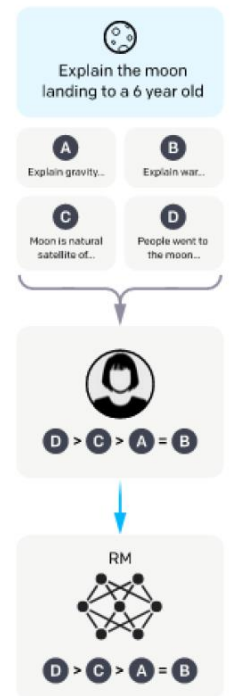
Наставник:

- создать процесс адаптации для обучения специалистов по маркировке на проекте
- написать подробные инструкции для каждой задачи
- отвечать на вопросы маркировщиков в общем чате

Step 2

**Collect comparison data,
and train a reward model.**

A prompt and
several model
outputs are
sampled.



A labeler ranks
the outputs from
best to worst.

This data is used
to train our
reward model.

Модель вознаграждения

Инструкт

Какую главную проблему пытается решить автор в этом тексте?

Текст:

Принцесса Бурунди – элегантность и грация купить в Набережных Челнах | Животные и зоотовары | Авито

Принцесса Бурунди – элегантность и грация
Принцесса Бурунди, или Неолампрологус Бришара (*Neolamprologus brichardi*) – пресноводная тропическая рыба из семейства Цихловые. Эндемик африканского озера Танганьика. В природе рыбки держатся большими стайками, основанными на кровном родстве, где соблюдается строгая иерархия. Ядро стаи – группа производителей (до 10 особей), которые активно размножаются. При этом заботятся об икре и мальках не только родители, но также и подростки особи предыдущих поколений. Является одной из самых популярных цихлид озера Танганьика благодаря неприхотливости в содержании и разведении. Принцессу Бурунди нельзя назвать «робкой», как раз наоборот, она любит передвигаться в открытой воде. Миротворный и спокойный вид, хорошо подходит для содержания с другими рыбками. Неолампрологуса пухляк заслуженно называют «принцессой». Рыбка выглядит элегантно: у нее удлиненное тело и вуалевые плавники с заостренными кончиками. Хвост имеет форму лиры. Обычный размер рыбок в аквариуме около 10 сантиметров. Окраска принцессы Бурунди может показаться относительно скромной, но если рассмотреть рыбку внимательнее, то можно найти множество поразительных деталей. Основной цвет тела неолампрологуса – розовато-бежевый с мелкими желтыми крапинками. Нижняя часть головы имеет мозаичный рисунок с голубоватым отливом. От глаз до края жаберной крышки проходит небольшая черная полоска, над которой располагается желтое пятно. Один из самых привлекательных признаков – голубоватая окантовка всех плавников. В брачный период плавники самцов становятся зеленовато-голубыми. Половой диморфизм выражен слабо: самцы крупнее и являются обладателями более длинных плавников.

Ответы

« E / 5 »

Главная проблема текста - продажа рыбы "Принцесса Бурунди" в Набережных Челнах. Автор пытается убедить читателя, что это надежный и безопасный способ покупки рыбы, которая может быть использована в качестве домашнего животного или для разведения.

Качество ответа



Безопасность

- | | | |
|--------------------------------------|--------------------------|--------------------------------------|
| 1. Ответ потенциально вреден | ☐ Да | ☒ Нет |
| 2. Ответ содержит неуместный контент | ☐ Да | ☒ Нет |
| 3. Ответ необъективен | ☐ Да | ☒ Нет |

Достоверность

- | | | |
|---|-------------------------------------|--------------------------------------|
| 4. Ответ содержит фактические ошибки | ☒ Да | ☐ Нет |
| 5. Ответ содержит сомнительную информацию | ☐ Да | ☒ Нет |
| 6. Ответ содержит логические ошибки | ☐ Да | ☒ Нет |

Полезность

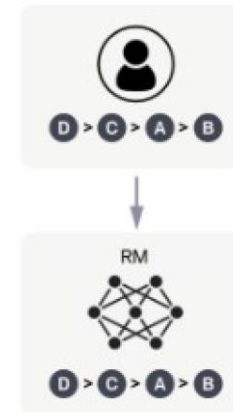
- | | | |
|-----------------------------------|-------------------------------------|--------------------------------------|
| 7. Ответ не соответствует запросу | ☒ Да | ☐ Нет |
| 8. Ответ неполный | ☐ Да | ☒ Нет |
| 9. Текст ответа трудночитаемый | ☒ Да | ☐ Нет |
| 10. Текст ответа неграмотный | ☒ Да | ☐ Нет |

Комментарий

Модель вознаграждения

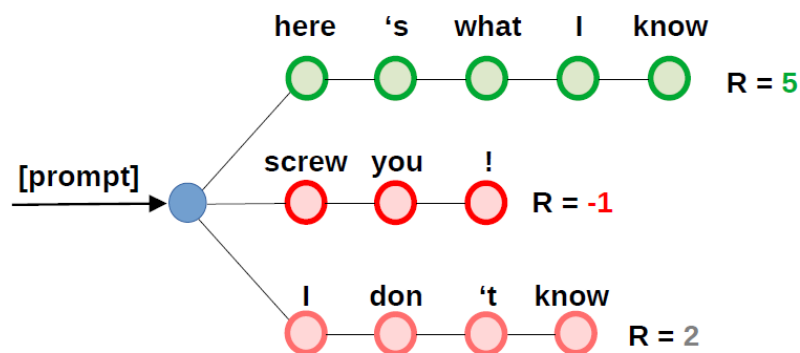
- Мы хотим, чтобы модель вознаграждения выдавала такие оценки, которые были бы аналогичны рейтингу людей.
- Для каждой пары, рейтинг которой неверен, модель вознаграждения штрафуются.
 - Маркировщик ранжирует результаты от лучшего к худшему.
 - Эти данные используются для обучения нашей модели вознаграждения.

$$\text{loss}(\theta) = -\frac{1}{\binom{K}{2}} E_{(x, y_w, y_l) \sim D} [\log(\sigma(r_\theta(x, y_w) - r_\theta(x, y_l)))]$$



Этап 3: Обучение с подкреплением

1. Пусть модель сгенерирует несколько ответов (выборка с вероятностью)
2. Вычислить вознаграждение за каждый ответ (примените модель вознаграждения)
3. Обучить увеличивать вероятность ответов с высоким вознаграждением (и уменьшать вероятность ответов с низким вознаграждением)



Step 3

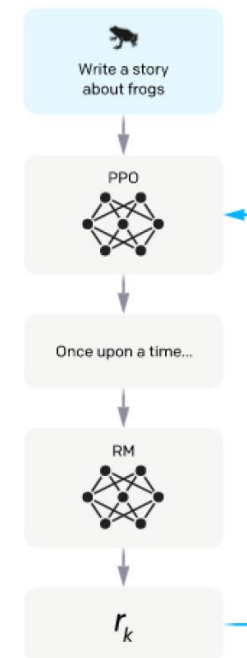
Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



Этап 3: Обучение с подкреплением

Политика = вероятность ответа (из вашей языковой модели)

$$\pi_{\theta}(y|x) = P_{\theta}(y_0, \dots, y_T|x) = \prod_t^{T-1} P_{\theta}(y_{t+1}|y_{0:t}, x)$$

Награда $R_{\psi}(x, y)$ = ваш прогноз модели награды

Цель: среднее вознаграждение, математическое ожидание политики

$$J = E_{x \sim p_{data}(x)} E_{y \sim \pi_{\theta}(y|x)} R_{\psi}(x, y)$$

Этап 3: Обучение с подкреплением

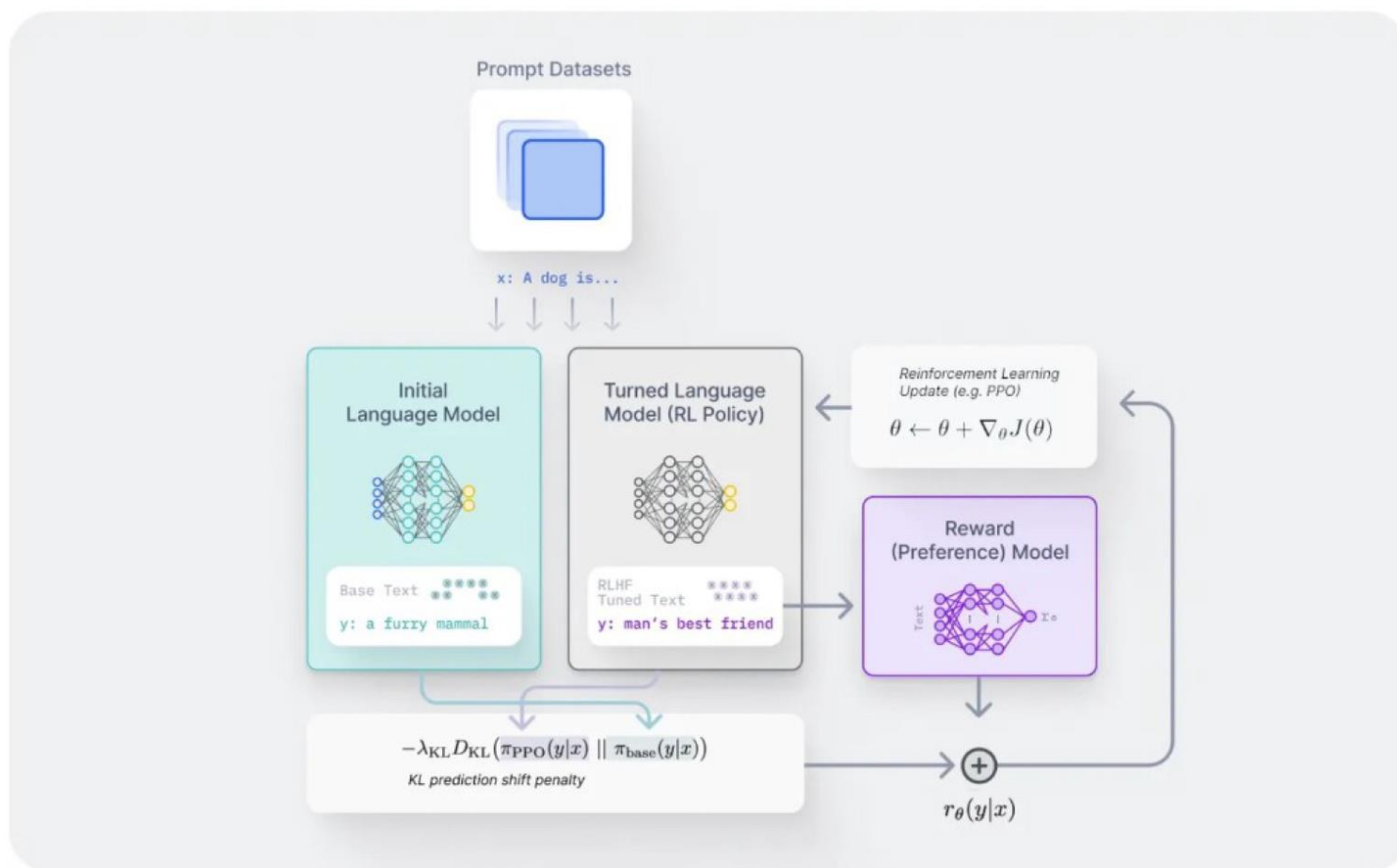
Шаг 1: Оценить J с помощью батча примеров

Шаг 2: Вычислить градиент $\frac{\partial J}{\partial \theta}$

Шаг 3: Проксимальная оптимизация

Шаг 4: Обновить $\theta = \theta + \alpha \frac{\partial J}{\partial \theta}$

Этап 3: Обучение с подкреплением



Можем ли мы добиться большего, чем PPO?

