

Рабочий лист · занятие 1, теория · 9 класс

Предсказать следующее слово

1 Разрежь фразу на токены

1,5 мин · сам

Ты — токенизатор. Раздели фразу из учебника вертикальными чёрточками | на токены тремя способами — и каждый раз сосчитай, сколько токенов получилось.

Способ 1 · режь на слова

токенов: _____

К о с и л к о с о й к о с о й к о с о й к о с о й**Способ 2 · режь на кусочки слов — и внутри слова тоже**

токенов: _____

К о с и л к о с о й к о с о й к о с о й к о с о й**Способ 3 · режь на символы — каждая буква отдельно**

токенов: _____

К о с и л к о с о й к о с о й к о с о й к о с о й

Сравни с соседом: совпали ли ваши нарезки? Потом проверь себя у станции «Токенизатор». **Токен** — базовая единица текста: слово, часть слова или символ. «Единственно правильной» нарезки нет — главное, резать всегда по одним и тем же правилам.

2 Побудь языковой моделью

1 мин · сам

а) Раунд 3 станции «Предскажи слово»: на экране — фраза с пропуском. До того как откроются варианты, запиши самое **вероятное** продолжение (не самое интересное!). Потом — самое **невероятное**: слово, которое модель не выберет никогда.

вероятное продолжение**невероятное продолжение**

б) Перенеси с доски варианты продолжения и их частоты — сколько голосов набрало каждое слово. **Обведи слово-чемпион.**

слово**частота**

Эта табличка — «внутренности» языковой модели: побеждает не самое умное слово, а самое частое.

3 Найди три галлюцинации

3 мин · пара

Работайте в паре: **подчеркните** всё подозрительное — в ответе спрятаны **три фактические ошибки** (галлюцинации). Подсказка: вся правда была в ролике Искры.

Промпт · Расскажи коротко историю чат-ботов.**Ответ модели**

Первые чат-боты появились в 1990-е годы. Самым известным из них была программа Элиза, созданная Джозефом Вейценбаумом: она переводила тексты с русского языка на английский. Настоящий прорыв случился в 2017 году, когда придумали архитектуру «трансформер», — нейросети научились учитывать контекст всей беседы. На её основе создали GPT: это название расшифровывается как «глобальная поисковая технология». Для обучения первой GPT-нейросети использовали датасет из нескольких миллиардов веб-страниц — если распечатать его, потребуется 400 тысяч тонн бумаги.

Выпиши три ошибки — а если знаешь, то и как на самом деле:

1. _____

2. _____

3. _____

Ловушка: что звучит как наглая выдумка, но это правда?

Как проверить: назови 1–2 источника, где выяснишь истину.

4 Собери свой вывод

2 мин · сам

а) Закончи фразы: Языковая модель не _____, а _____

Больше всего сегодня меня удивило _____

Теперь, прежде чем поверить ответу нейросети, я _____

б) Насколько ты разобрался(ась), как устроена языковая модель? Обведи цифру:

пока туман

1

2

3

4

5

могу объяснить
другу

в) Вопрос, который у меня остался: _____



Искра: теперь ты знаешь мой секрет: я не знаю ответов — я их предсказываю. И ошибаюсь тем же уверенным тоном. Так что проверка фактов — твоя суперсила, не моя.