

● Тема: большие языковые модели

Предсказать следующее слово

Как устроены большие языковые модели — и почему они иногда
уверенно ошибаются

Пред

сказ

ать

■



Да

нет,

наверное

**Что на самом деле хотел сказать
собеседник?**

Согласие + отрицание + неуверенность — одновременно

Цели занятия

Три умения на сегодня

К концу занятия каждый сможет...



**Увидеть текст глазами
машины**

токены, леммы и немного NLP



**Понять, как модель
предсказывает слова**

ненадолго сами станем нейросетью



**Ловить модель на уверенных
ошибках**

охота на галлюцинации

NLP · обработка естественного языка

Говорить с компьютером по-человечески

Область ИИ, которая учит компьютер «понимать» и создавать тексты на человеческом языке.

Раньше · программа

< > Пишем код на специальном языке машин

Теперь · обычный вопрос



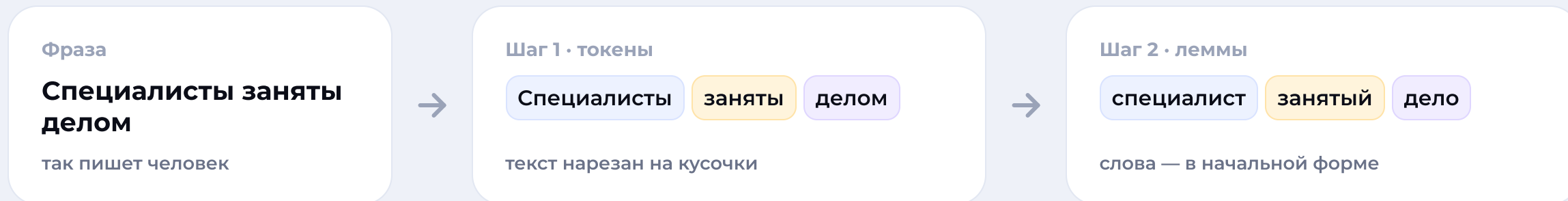
«Какой товар в этом месяце покупали чаще всего?»

Ответ — в обоих случаях

Самый популярный товар в этом месяце — электрический чайник.

Токенизация и лемматизация

Как машина режет текст



✦ Токен — базовая единица текста для модели: **слово, часть слова или символ**

Нейролаборатория · станция 1

Станция «Токенизатор»

- 1 Разрежь фразу чёрточками сам — задание 1
- 2 Сравни свою нарезку с соседом
- 3 Теперь режет машина: слова, подслова, символы

Рабочий лист · задание 1

● Токенизатор

Ввод

Косил косой косой косой косой

Токены

Косил

косой

косой

косой

косой

слова

подслова

символы

5 токенов

Новый термин

Языковая модель — предсказатель

Алгоритм прогнозирует, какое слово вероятнее всего будет следующим, — и из прогнозов собирает текст.

✦ Модель даёт **прогноз**, а не ответ

LLM · большая языковая модель

Обучена на огромных объёмах текстов

GigaChat

YandexGPT

Масштаб

29 млрд

параметров у GigaChat · конец марта 2024

Игра «Мы — нейросеть»

Класс — модель. Учитель — пользователь, который ввёл промпт.

Промпт на доске

После контрольной весь класс мечтал о...

1 Выкрикните варианты —
четыре слова на доску

2 Голосуем: счётчик ряда
считает руки

3 Частое слово побеждает —
дописываем в фразу

Продолжи фразу

Чтобы решить уравнение, школьники открыли свои ____

×120 раз в датасете



тетради

✓ **выбор модели**

×24 раза



ноутбуки

реже — не выбрано

×7 раз



глаза

совсем редко

✦ Побеждает не самое умное слово, а **самое частое**



● Серия «Как я читаю» · ведёт Искра

От Элизы до GPT

История языковых моделей: 70 лет — за 5 минут



Новый термин — галлюцинация

Уверенно — не значит правда

Галлюцинация — модель генерирует ложную или вводящую в заблуждение информацию. Уверенным тоном.

Пример из учебника

«Михаил Юлианович Лермонтов»

Подробная биография поэта, которого не было: настоящий
— Михаил **Юрьевич**

И ещё один

История Атлантиды

Напишет так же бодро и подробно, как историю Древнего
Рима

Откуда берутся вероятности

Ответ зависит от датасета

Мысленный эксперимент

Если бы в датасете школьники открывали только окна...

Ответ модели

Чтобы решить уравнение, школьники открыли свои
окна

Аналогия · старая карта города



Чего нет на карте — нет для приезжего.

Чего не было в датасете — **нет для модели.**

Правило. Важную информацию из ответа модели проверяй по другим источникам

Три галлюцинации в одном ответе

Промпт

Расскажи коротко историю чат-ботов.

Ответ модели

Первые чат-боты появились в 1990-е годы¹. Самым известным была программа Элиза, созданная Джозефом Вейценбаумом: она переводила тексты с русского языка на английский². Прорыв случился в 2017 году с архитектурой «трансформер». На её основе создали GPT — «глобальная поисковая технология»³. Для обучения первой GPT использовали миллиарды веб-страниц: чтобы распечатать, нужно 400 тысяч тонн бумаги.

- 1 Элиза — **1960-е**, а не 1990-е
 - 2 Она **изображала психотерапевта** — отвечала вопросом на вопрос
 - 3 GPT — **Generative Pre-trained Transformer**, «генеративный предобученный трансформер»
- ✓ Ловушка: 400 тысяч тонн бумаги — **правда** из учебника



Главный вывод занятия

Модель **предсказывает,** а не знает

Не оракул — очень начитанный предсказатель следующего слова. Поэтому пишет быстро и складно — и ошибается уверенным тоном.

Занятие 2 · практика

Говорим с моделью на её языке

Соберём сильный промпт — и устроим охоту на галлюцинации в реальных ответах.

Попробуй дома



Неделя с T9: собери коллекцию подсказок и автозамен



Найди, где автоперевод споткнулся о фразеологизм



Игра «Элиза»: отвечай другу только вопросами



памятка темы — принеси на практику

камертон: **модель предсказывает, а не знает**