

9 класс · занятие 1 из 2, теория · 40 минут

Методические рекомендации к занятию «Предсказать следующее слово. Как устроены большие языковые модели»

Цели занятия

Научить объяснять, как машина «видит» текст (токенизация); раскрывать принцип работы языковой модели — предсказание следующего слова по вероятностям; распознавать галлюцинации и объяснять их причины; сформулировать личное правило проверки важной информации.

Формируемые ценности

Критическое мышление и проверка информации · ответственное и этичное использование технологий · уверенность «я могу разобраться, как это устроено».

Основные смыслы

Компьютер видит текст цепочкой токенов; **«модель предсказывает, а не знает»** — формулировка-камертон, она должна прозвучать на занятии не менее трёх раз; ответ модели зависит от датасета; галлюцинация звучит как правда, поэтому важное проверяем; кто понимает механизм, тот управляет результатом.

Планируемые результаты

Личностные

- ценности научного познания: интерес к устройству технологий и установка «я могу разобраться, как это устроено» — ученик своими руками воспроизводит механизм предсказания (игра, станции);
- гражданское воспитание: соблюдение норм цифровой этики и информационной безопасности при работе с ИИ-сервисами (личные данные в промпты не пишем, регистрация — с родителями);
- духовно-нравственное воспитание: осознанное неприятие использования генеративного ИИ для обмана и распространения недостоверной информации (кейс «найди галлюцинацию»);
- патриотическое воспитание: знакомство с отечественными языковыми моделями GigaChat и YandexGPT как примерами современных LLM (нейтрально, без рекламы).

Метапредметные

- критически оценивать достоверность и полноту информации, в том числе сгенерированной ИИ; выявлять признаки недостоверности (задание 3: три галлюцинации и «ловушка-правда»);
- формулировать выводы на основе наблюдения за работой модели (станции «Токенизатор» и «Предскажи слово»: от наблюдения — к выводу «машина видит токены», «побеждает частое слово»);
- находить закономерности в данных: связывать частоты слов в «датасете» с выбором модели (игра «Мы — нейросеть», пример 120/24/7);
- брать на себя ответственность за решение, принятое с опорой на рекомендации ИИ: правило «важное — проверяю по источнику»;
- коммуникативные действия: аргументировать выбор слова в игре, работать в паре над кейсом, корректно обсуждать разные ответы.

Предметные

- объяснять, что такое обработка естественного языка (NLP), и приводить примеры её применения из повседневной жизни;
- раскрывать на доступном уровне понятия «токен», «токенизация», «лемматизация»; токенизировать фразу вручную;
- раскрывать принцип работы языковой модели (предсказание следующего слова по вероятностям) и воспроизводить пример учебника «школьники открыли свои...» (120/24/7);
- объяснять термины «большая языковая модель (LLM)» и «галлюцинация»; приводить пример влияния датасета на ответ модели.

Формулировки результатов опираются на программу внеурочной деятельности по ИИ и конкретизированы под занятие 9 класса.

Связь с программой

Занятие открывает тему «Большие языковые модели» (модуль 3 учебника, с. 44–62) — первое из пары «теория → практика». **Опирается на:** модули 1–2 (нейросеть, слои и веса, датасет, обучение; чат-боты и пример фактической ошибки нейросети). **Ведёт к:** занятию 2 — практикуму по промптам; памятка, выданная в конце занятия, — его рабочий инструмент. Дальний горизонт (10 класс): эмбединги, ограничения LLM и RAG — до них не поднимаемся, термин «эмбединг» не вводим.

Этапы занятия: методический комментарий

Мотивационно-целевой этап (6 мин)

Смыслы: язык сложнее, чем кажется; предсказание следующего слова — идея, знакомая каждому по Т9. **Задачи:** включить класс через языковую головоломку; вывести из бытового опыта слово «часто» — мостик к вероятностям; поставить цели. **Формы:** беседа, слайды 1–3.

Комментарий. «Да нет, наверное» разгадывают мгновенно — не задерживайтесь: ценность крючка в выводе «вас никто не учил, вы слышали это сотни раз». Про автозамены держите регламент «2–3 реплики»; если класс молчит — начните со своего примера. На вопрос «откуда телефон знает?» не добивайтесь слова «вероятность»: достаточно «часто», термин появится в блоке 2.

Блок 1. Как машина видит текст (7 мин)

Смыслы: NLP позволяет говорить с компьютером по-человечески; для машины текст — цепочка токенов. **Задачи:** ввести термины NLP, токен/токенизация, лемматизация с аналогиями; дать каждому токенизировать фразу руками; показать машинную нарезку на станции. **Формы:** объяснение (слайды 4–5), задание 1, фронтальная станция «Токенизатор» (слайд 6).

Комментарий. Порядок принципиален: сначала задание 1 руками, потом машина — так возникает «ага-момент» сравнения. Нарезка по слогам в задании 1 — не ошибка, а повод для вывода «единственно правильной нарезки нет». На станции обязательно переключите режим «подслова»: удивление «а что, так можно?» — главная эмоция блока. Лемматизацию не растягивайте: одна пара «заняты → занятый» и вопрос «зачем». На вопрос «зачем резать на подслова» — честно: так модель собирает незнакомые слова из знакомых кусочков (глубже — уровень 10 класса).

Блок 2. Модель-предсказатель (8 мин)

Смыслы: модель даёт прогноз, а не ответ; побеждает частое слово. **Задачи:** ввести «языковая модель» и «LLM»; прожить механизм в игре; связать игру с примером учебника (120/24/7); сравнить себя с моделью на станции; письменно зафиксировать понимание (задание 2). **Формы:** объяснение (слайд 7), игра «Мы — нейросеть» (слайд 8), разбор примера (слайд 9), фронтальная станция «Предскажи слово», задание 2.

Комментарий. Игра — сердце занятия, держите темп: первые четыре разных варианта, без обсуждения. Счётчики голосов в рядах — тайм-менеджмент: подсчёт занимает 10 секунд вместо минуты. Ответы про контрольную предсказуемы («о каникулах», «о еде») — это хорошо: чем банальнее чемпион, тем нагляднее тезис «побеждает частое, а не интересное». Твист про «столицу Марса» не комментируйте подробно — бросьте крючок «вернёмся через десять минут»: он выстрелит в блоке 3. На станции ведите счёт «класс : модель» на углу доски — азарт без проигравших.

Видео (6 мин)

Смыслы: у технологии есть 70-летняя история; трансформер — прорыв, давший GPT. **Задачи:** передать исторический блок ролику (не дублировать!); настроить внимательный просмотр. **Формы:** подводка, ролик ~5 мин, фиксация двух опорных слов (Элиза, трансформер).

Комментарий. Анонс «через пять минут устрою проверку» — не шутка, а установка на просмотр: кейс задания 3 построен на фактах из ролика. После ролика не пересказывайте содержание — только два опорных слова. Проверьте звук заранее; субтитры, если есть, включите — полезно всем.

Блок 3. Когда модель ошибается (7 мин)

Смыслы: галлюцинация — вероятные слова без проверки реальности; ответ зависит от датасета; важное проверяем. **Задачи:** назвать явление термином; показать механизм (датасет → вероятности → уверенная ошибка); дать поймать галлюцинацию руками (задание 3); вывести правило проверки. **Формы:** объяснение (слайды 11–12), парная работа с кейсом (задание 3), разбор (слайд 13).

Комментарий. Разбор кейса — кульминация: раскрывайте ошибки по одной, спрашивая «кто это подчеркнул?». Обязательно дойдите до «ловушки-правды» (400 тысяч тонн бумаги): момент «самое дикое оказалось правдой» — главный воспитательный удар занятия. Если класс ошибок не находит — раскройте первую сами, остальные найдут по аналогии. Ошибки и правильные факты перечислены в сценарии после текста кейса.

Заключительный этап (6 мин)

Смыслы: модель предсказывает, а не знает; понимание механизма = управление результатом. **Задачи:** зафиксировать главный вывод (слайд 14); дать письменную рефлексия (задание 4) и услышать 2–3 голоса; выдать памятку; перекинуть мостик к практике (слайд 15). **Формы:** беседа, задание 4, памятка.

Комментарий. В задании 4 первая фраза («Языковая модель не ..., а ...») — проверка камертона: ждём «не знает, а предсказывает» в любых формулировках. Рефлексию читают только желающие; не оценивайте ответы. Памятку выдавайте в руки, а не «возьмите на выходе», и прямо скажите: понадобится на следующем занятии — это создаёт связку пары уроков.

Трудные вопросы: как отвечать

«А ИИ живой? Он меня понимает?»

Нет. Учебник формулирует точно: кажется, что нейросеть осознанно отвечает, «но это не так: искусственный интеллект не обладает сознанием, просто эта нейросеть гораздо лучше обучена». Модель имитирует понимание, подбирая вероятные слова. Полезная параллель: калькулятор «решает» примеры быстрее нас, но не «понимает» математику.

«Если модель просто угадывает слова, почему она так часто отвечает правильно?»

Похвалите вопрос — он лучший на занятии. В миллиардах текстов частые продолжения обычно совпадают с правдой; но там, где данных мало или вопрос редкий, вероятное расходится с истинным — рождается галлюцинация. Правильность — побочный эффект хорошей статистики, а не признак знания.

«Можно ли делать домашку нейросетью?»

Не запрещайте с порога — разберите честно. В учебнике есть пример: сгенерированная целиком работа — с ошибками, «слишком много общего текста», без собственного мнения; такое опознаётся быстро. Рабочая формула — как у журналиста из учебника: ИИ — черновик, человек — редактор, и отвечает за сданную работу тот, кто её сдал. На практике (занятие 2) этому и учимся.

«А нас всех заменит ИИ?»

Спокойно, без запугивания и розовых очков: по учебнику (модуль 1) ИИ не имеет собственной воли и не может ставить задачи сам себе — это инструмент. Инструменты меняют работу: часть задач автоматизируется, зато ценнее становятся те, кто умеет ставить задачи и проверять результат. Ровно этому мы учимся.

«Почему модели нельзя просто запретить врать?»

Она не врёт — враньё требует знать правду, а у модели есть только вероятности слов. Разработчики закладывают ограничения, инженеры дообучают модель на оценках пользователей — ответы улучшаются, но механизм остаётся предсказанием. Последний рубеж проверки — всегда человек.

Организация активностей

Техника: проектор + компьютер учителя; интерактив — один офлайн-файл [interactive/index.html](#), станции по ссылкам с **#tokenizer** и **#predict** — откройте обе вкладки до урока; проверьте звук ролика. Устройства ученикам не нужны. **Печать:** рабочие листы и памятки — по числу учеников (ч/б). **Доска:** нужна для игры (фраза, столбик вариантов, частоты) — освободите центр заранее. **Рассадка:** обычная; перед игрой назначьте в каждом ряду «счётчика голосов», для станций можно позвать ученика-«оператора» — хорошая роль для скептика или знатока. **Роль учителя в интерактивах:** ведущий и модератор темпа; не превращайте станции в лекцию — экран работает, класс отвечает, вы задаёте вопросы.

Вариативность

Класс без компьютера/проектора — офлайн-замены станций (полные правила)

«Бумажный токенизатор» — вместо станции «Токенизатор», 3–4 мин

Материалы: доска; у учеников — задание 1 рабочего листа.

1. Все выполняют задание 1 — режут фразу «Косил косой косой косой косой» чёрточками.
2. Трое учеников у доски записывают три нарезки одной и той же фразы: по словам, по слогам-«подсловам», по буквам.
3. Класс хором считает токены в каждом варианте, учитель подписывает счётчик под каждой нарезкой.
4. Вопрос: «какой способ удобнее машине?» — вывод: правильных нарезок несколько, важно резать всегда одинаково.
5. Финал — устно: сколько токенов в «Да нет, наверное»?

«Частоты на доске» — вместо станции «Предскажи слово», 3–4 мин

Подготовка: на закрытой части доски (или на листе) учитель заранее выписывает фразу «Чтобы решить уравнение, школьники открыли свои ...» и варианты с частотами из учебника: тетради — 120, ноутбуки — 24, глаза — 7, окна — 0.

1. Класс голосует руками за вариант (варианты открыты, частоты скрыты).
2. Учитель записывает голоса класса рядом с вариантами.
3. Открывает частоты «модели», сравнение: совпал ли чемпион класса с чемпионом модели.
4. Обсуждение «окна — 0»: что было бы, если бы в датасете школьники открывали только окна.
5. Задание 2 выполняется как в основном сценарии, на второй фразе, которую учитель придумывает сам (частоты не нужны). Числа 120/24/7 — из учебника; любые дополнительные частоты вводите только как условные, учебные.

Сильный класс: в блоке 1 спросите, как токенизатор разрежет выдуманное слово «нейросетища» (соберёт из знакомых кусочков); в блоке 2 — «почему модель на один вопрос отвечает по-разному?» (учебник: модели перефразируют ответ каждый раз; глубже не уходим); в задании 3 — найти «ловушку-правду» без подсказки.

Слабый класс / дефицит темпа: лемматизация — одним предложением; в игре — один раунд плюс твист; задание 3 — фронтально: учитель читает текст вслух, класс останавливает его на подозрительных местах.

Сокращённый формат 30 мин: мотивационный — 4 (без обмена автозаменами), блок 1 — 5 (станция 2 мин, задание 1 — дома), блок 2 — 6 (игра без второго круга, станцию «Предскажи слово» пропустить — её роль сыграла игра; слайд 9 обязателен), видео — 5 (без фиксации), блок 3 — 6 (задание 3 — 2,5 мин), заключительный — 4 (задание 4 устно, памятка).
Неприкосновенное ядро: токенизация, «модель предсказывает, а не знает», галлюцинации + правило проверки.

Типичные трудности и риски

- **Не работает интерактив/проектор** — обе станции имеют офлайн-замены (выше); опорные фразы слайдов 7, 9, 14 запишите на доске.
- **Не воспроизводится видео** — единственный случай, когда историю рассказывает учитель: коротко, по цепочке «1950-е перевод → 1960-е Элиза → 1997 память → 2013 слова-в-числа → 2017 трансформер → GPT» (2–3 мин). Кейс задания 3 при этом работает: факты для проверки прозвучали.
- **Дискуссия ушла в «а какая нейросеть лучше?»** — нейтрально: в учебнике названы GigaChat и YandexGPT, устроены все LLM одинаково; сравнение сервисов — на практике.
- **Ученик-эксперт («я этим пользуюсь каждый день»)** — дайте роль: оператор станции, счётчик голосов, «адвокат модели» при разборе кейса; спросите: «а замечал ли ты, чтобы модель уверенно ошибалась?»
- **Ученик-скептик** — не спорьте, дайте факту работать: кейс задания 3 с «ловушкой-правдой» переубеждает лучше аргументов.
- **Игра выдаёт сомнительные слова** — объявленное заранее правило «слово должно подойти для школьного диктанта» + право учителя не записывать вариант.

- **Класс отвечает на твист про Марс всерьёз** — не высмеивать: «модель ответила бы так же уверенно — в этом и суть».

Постразговор

Домашние наблюдения из сценария (Т9-коллекция, автоперевод, два одинаковых вопроса помощнику) предлагайте как добровольные — они не требуют регистрации. Задание 1 учебника (с. 47) — только при доступе к GigaChat/YandexGPT и вместе с родителями. Игра «Элиза» — безтехнологичная, годится всем. Для мотивированных: вопросы самопроверки (с. 57) письменно; мини-проект «Охота на NLP» с обсуждением находок в начале занятия 2. Объёмный доклад с проверкой фактов (задание 5 учебника) приберегите на «после практики» — там он ляжет на навык промптинга.