Урок 13. Этика языковых моделей и цифровая безопасность

9 класс · Урок 13 из 34 · 45 минут

Вся программа 9 класса

Этика искусственного интеллекта начинается с выбора, какие сведения попадут в запрос и кто проверит результат.

Содержание урока

- 01. Цель урока
- 02. Ход урока
- 03. Этап 1. Введение: Узнаваемый дневник · 6 мин
- 04. Этап 2. Три решения до отправки · 10 мин
- 05. Этап 3. Редсовет проекта · 15 мин
- 06. Этап 4. Условия изменились · 9 мин
- 07. Этап 5. Итоги урока · 5 мин

01. Цель урока

Определять, какие сведения допустимо передавать языковой модели, как указывать чужой вклад и кто проверяет готовый текст перед отправкой.

02. Ход урока

Этап	Мин	Учитель	Ученики
Этап 1. Введение: Узнаваемый дневник	6	Записывает два возможных способа помочь редакции.	Объясняют, где начинается риск.
Этап 2. Три решения до отправки	10	Сопоставляет три вида разрешений.	На коротком примере отделяют данные, вклад и решение.
Этап 3. Редсовет проекта	15	Раздаёт роли и карточки условий.	Согласуют безопасный запрос и решение о выпуске.
Этап 4. Условия изменились	9	Выдаёт новое ограничение после решения.	Корректируют план и сравнивают две допустимые версии.

Этап	Мин	Учитель	Ученики
Этап 5. Итоги урока	5	Подводит итог и сверяет ответы.	Письменно отвечают и сверяют решение.
Итого	45		

03. Этап 1. Введение: Узнаваемый дневник 6 мин

ГОВОРИМ УЧЕНИКАМ

Класс проверяет, легко ли новичку найти в школе нужный кабинет. В дневнике испытания имён нет, но автор одной записи упомянул свою роль в спектакле, которую исполнял только он. Редактор предлагает отправить все дневники языковой модели, чтобы она подготовила памятку. Какие наблюдения нужны для памятки?

На этом примере разберём один из вопросов этики искусственного интеллекта: какие сведения можно передать модели. Для памятки полезны места, в которых люди терялись, и причины. Подробность о роли в спектакле не объясняет, где поставить указатель, зато по ней можно узнать участника. В запрос её включать незачем.

Для учителя

Попросите учеников сделать выбор и объяснить его. Запишите на доске наблюдения, которые нужны для памятки, и сведения, по которым можно узнать человека.

04. Этап 2. Три решения до отправки 10 мин

ГОВОРИМ УЧЕНИКАМ

Перед отправкой запроса проверьте, можно ли узнать человека по данным. Иногда имя убрали, а по классу, точной дате и редкой истории всё равно ясно, о ком речь. Если цель — описать общее правило, оставьте только общее правило. Согласие показать текст учителю также не означает, что автор разрешил отправить его в сервис или опубликовать.

Затем проверьте, чьи слова и идеи вы используете. Если берёте слова автора дословно, читателю нужно видеть, что это цитата и кому она принадлежит, когда автор согласен быть указанным. Пересказ чужой мысли также сохраняет связь с источником. Если правила вашей работы требуют указать помощь ИИ, сделайте отдельную пометку: например, «модель предложила план, содержание проверила редакция». Такая пометка не превращает чужую работу в вашу.

Наконец, назовите человека, который отвечает за выпуск. Модель может по ошибке соединить две истории и приписать слова другому человеку. Перед отправкой ответ сверяет редактор, а разрешения подтверждает тот, кто их действительно давал. Если ситуация спорная, полезно уменьшить объём данных или подготовить часть работы без модели.

Решение	Вопрос перед отправкой
Входные данные	Какие сведения нужны для этой задачи и кто может их увидеть?
Авторство	Кому принадлежат слова и идеи? Как показать вклад?
Ответственность	Кто проверяет результат, исправляет ошибку и решает, можно ли выпускать?

Реальный случай. Как данные попадают в правила для школ

В сентябре 2023 года ЮНЕСКО опубликовала рекомендации по генеративному ИИ для образования. Защита личных данных вошла в число мер, которые авторы предлагают закреплять в правилах. Школе полезно заранее определить, какие материалы можно передавать сервису, чтобы не решать это на ходу.

05. Этап 3. Редсовет проекта 15 мин

ГОВОРИМ УЧЕНИКАМ

У редактора, испытателя и координатора разные задачи. Редактор хочет дать новичку понятный маршрут. Испытатель может заметить, что публикация его рассказа раскрывает, кто он. Координатор проверяет, где посетители действительно терялись. У вас будет один общий результат, поэтому каждое решение придётся обосновать.

Кейс. Памятка для нового ученика

Памятка должна помочь новому ученику найти зал кружков. У редакции есть три материала.

Координатор. Разрешено сообщить: от главного входа идите до лестницы. Зал кружков находится на втором этаже. Часть участников прошла мимо лестницы, потому что не увидела указатель.

Автор дневника. Разрешил зачитать на защите проекта фразу «Я увидел лестницу только на обратном пути». Про открытую памятку и сервис ИИ разговора не было.

Личная запись. В дневнике описан тот же поворот, но автор поясняет, что оказался у сцены в костюме единственного ведущего спектакля. Запись предназначалась только для рабочей группы.

1. Распределите роли координатора, автора и редактора. Согласуйте короткий запрос для подготовки памятки. Рядом запишите, какие материалы оставили за его пределами и почему.
2. Прочитайте будущую памятку из двух-трёх предложений. Укажите, кто проверит маршрут и что редакция скажет о помощи модели, если использует её черновик.

Для учителя

Достаточно открытых условий из первой карточки. Дневник не нужен для описания маршрута. За цитатой и выбором аудитории надо обратиться к автору отдельно. Пример отметки при использовании модели: «Черновик помогла составить языковая модель. Маршрут проверил координатор проекта».

06. Этап 4. Условия изменились 9 мин

ГОВОРИМ УЧЕНИКАМ

Разрешение может относиться к одной аудитории, одному тексту или одному этапу. Когда условия меняются, полезно быстро найти все места, где старое разрешение ещё влияет на решение.

Изменение перед выпуском

Автор отдельно разрешил поместить фразу из дневника в памятку без имени. Перед выпуском он просит её убрать. Памятка ещё не опубликована. Решите, что изменить в тексте, сохранённом черновике и следующих действиях редакции.

Для учителя

Выпустить памятку без цитаты или запросить новую фразу на согласованных условиях. Обновить черновик и проверить, не осталась ли цитата в других вариантах. Если её ранее передали во внешний сервис, уточняют доступные способы удалить историю. Удаление из статьи само по себе не отменяет уже сделанную передачу. При демонстрации в ГигаЧате используйте только открытые условия.

07. Этап 5. Итоги урока 5 мин

ГОВОРИМ УЧЕНИКАМ

Ответственность начинается до отправки запроса. Сначала решите, какие сведения вправе использовать, а перед выпуском проверьте готовый текст.

Главное за урок

- Для запроса отбирают только нужные сведения.
- Цитата, чужая идея и помощь ИИ требуют разных пояснений.
- Решение о выпуске принимает человек, который проверил результат.

Итоговые вопросы

1. Можно ли отправить в модель анкету без имён, если там остались редкие подробности? Объясните решение.
2. Что вы исправите, если модель приписала слова из дневника координатору?

Для учителя

Смотрим, можно ли узнать человека по сочетанию подробностей, и нужны ли они для задачи. Во втором случае сверяем исходник, убираем ошибочную атрибуцию и используем цитату только на согласованных условиях.

Источники

1. ЮНЕСКО. Guidance for generative AI in education and research · 07.09.2023 · Материал старше 2024 года. «mandating the protection of data privacy» Источник