Урок 7. Промпт-инженер, специалист по дипфейкам и этике ИИ

9 класс · Урок 7 из 34 · 45 минут

Вся программа 9 класса

Профессия промпт-инженера, проверка дипфейков и этика ИИ. Урок для 9 класса о трёх ролях и о том, как их работа помогает безопасно выпускать материалы.

Содержание урока

- 01. Цель урока
- 02. Ход урока
- 03. Этап 1. Введение: Чужой анонс · 4 мин
- 04. Этап 2. Что делает каждый специалист · 11 мин
- 05. Этап 3. Три станции в редакции · 20 мин
- 06. Этап 4. Когда изменённый голос уместен · 5 мин
- 07. Этап 5. Итоги урока · 5 мин

01. Цель урока

Сравнить работу промпт-инженера, специалиста по выявлению дипфейков и специалиста по этике ИИ. Проверить, как результаты их работы дополняют друг друга.

02. Ход урока

Этап	Мин	Учитель	Ученики
Этап 1. Введение: Чужой анонс	4	Открывает вопрос о проверке источника.	Выбирают, что проверить перед публикацией.
Этап 2. Что делает каждый специалист	11	Вводит инструкции, проверку происхождения и правила выпуска.	Сопоставляют работу специалистов с разными доказательствами.
Этап 3. Три станции в редакции	20	Организует смену станций и сбор результатов.	Проверяют три разных материала и выносят решение о выпуске.

Этап	Мин	Учитель	Ученики
Этап 4. Когда изменённый голос уместен	5	Различает создание имитации и открытое изменение записи.	Формулируют условие допустимой публикации.
Этап 5. Итоги урока	5	Подводит итог и сверяет ответы.	Письменно отвечают и сверяют решение.
Итого	45		

03. Этап 1. Введение: Чужой анонс 4 мин

ГОВОРИМ УЧЕНИКАМ

В школьную редакцию приходит голосовое сообщение с неожиданным анонсом. Голос похож на голос ведущего, но запись прислал незнакомый аккаунт. Одновременно помощник редакции подготовил текст объявления и добавил подробность, которой нет в письме организатора. Сегодня разберём, кто в команде отвечает за качество инструкции, подлинность материала и последствия публикации.

Кому передать материал?

Предложите первое действие редакции. Укажите, какая информация после него станет надёжнее.

Для учителя

Подходят сверка текста с письмом и подтверждение голосовой записи по известному контакту. Просите описать результат проверки.

04. Этап 2. Что делает каждый специалист 11 мин

ГОВОРИМ УЧЕНИКАМ

Профессия промпт-инженера связана с разработкой и проверкой инструкций для модели. Такая инструкция называется **промптом**. Если редактор просит «Сделай объявление покороче», модель может заполнить недостающие подробности привычными фразами. Если он добавляет «Используй только письмо, а отсутствующие сведения пометь», проверять результат проще. Специалист прогоняет несколько примеров, в том числе письма с пропусками, и сравнивает, сохранила ли модель факты.

Дипфейк — правдоподобная имитация или подмена внешности либо голоса человека с помощью цифровых технологий, например нейросети. Специалист по выявлению дипфейков проверяет происхождение: кто опубликовал материал первым, есть ли

оригинал, можно ли подтвердить сообщение у человека, которого оно касается. Анализ файла даёт дополнительные признаки. Даже если инструмент отметил подозрительное место, редактору нужны подтверждения из независимых источников.

Специалист по этике ИИ выясняет, кого затронет публикация, было ли согласие и как исправить ошибку. Согласие важно и для настоящих записей, а от имени другого человека нельзя выпускать анонс только потому, что голос похож. Рабочее правило должно объяснить, когда можно публиковать, кто принимает решение и что делать, если поступило возражение.

В текстовой инструкции полезно отделять исходные сведения от того, что можно изменить. Время и место показа сохраняем, длину объявления меняем. С аудио проверка устроена иначе. Слова могли быть подлинными, но из записи вырезали соседнюю реплику и смысл изменился. Поэтому специалист проверяет не только способ создания файла, но и то, из какой ситуации он взят.

Реальный случай. Звонки с подменой голоса и номера

21 января 2024 года перед голосованием в штате Нью-Гэмпшир жители США получили автоматические звонки с голосом, имитировавшим Джо Байдена. В мае американская Федеральная комиссия по связи описала применение генерации голоса и подмены номера звонящего. Сходство голоса и знакомый номер усиливали впечатление, что звонок настоящий. Для проверки пригодился бы независимый источник, который получатель открыл сам.

05. Этап 3. Три станции в редакции 20 мин

ГОВОРИМ УЧЕНИКАМ

Самая полезная проверка на каждой станции своя. Текст сравнивают с источником, аудио — с подтверждениями происхождения, а решение о публикации — с правилами, учитывающими интересы участников. Ни одна проверка в одиночку не решает все задачи редакции. Запишите результат так, чтобы следующая команда поняла, на чём он основан.

Редакционный выпуск

Работайте тройками. На каждой станции оставьте одну короткую запись для выпускающего редактора. После сигнала переходите к следующей. В конце соберите единое решение: что выпустить сейчас, что исправить и что отложить до проверки.

Текст. Письмо сообщает только «Показ 15:40 в малом зале». Помощник написал «Показ 15:40, вход свободный». Перепишите инструкцию так, чтобы она подходила и для следующего письма с другими пропусками. Придумайте, как испытать её.

Аудио. Файл переслали без исходного сообщения, детектор выдал «подозрительно». Есть прежняя подтверждённая запись ведущего и его контакт из редакции. Выберите

порядок проверки и запишите, какого подтверждения ждёте.

Выпуск. Два участника разрешили публиковать общий разговор, третий согласился только на использование внутри кружка. Из одного фрагмента редакция хочет сделать короткий ролик. Запишите условие выпуска и способ отозвать ошибочную публикацию.

Для учителя

На первой станции ждите инструкции с опорой на письмо и теста хотя бы на одном другом пропуске. На второй — подтверждения по известному контакту или поиска оригинала, с пониманием, что детектор даёт сигнал. На третьей подходят монтаж фрагмента без реплик и идентифицирующих деталей третьего участника или новое согласие. Правило исправления указывает ответственного и канал для обращения.

06. Этап 4. Когда изменённый голос уместен 5 мин

ГОВОРИМ УЧЕНИКАМ

В редакции могут изменить голос собеседника, чтобы защитить его от узнавания. Тогда цель, согласие и пометка помогают аудитории понять, что она слышит. Сравните это с анонсом от имени человека, который его не записывал. Технически в обоих случаях звук меняется, но последствия зависят от того, как его представили слушателю.

Подумайте, как редактор передаст это правило новому сотруднику. Нужны простые условия, по которым можно принять решение в реальном выпуске: кто согласовал изменение, где это записано и какую пометку увидит слушатель. Если условия нельзя проверить, они останутся пожеланием и не помогут, когда редакция торопится.

Пометка и ответственность

Запишите короткую пометку к интервью с изменённым голосом. Решите, кто перед публикацией проверит, что по остальным деталям человека тоже трудно узнать.

Для учителя

Подходит «Голос собеседника изменён по его просьбе». Ответственность может взять редактор, который сверит другие детали записи с собеседником.

07. Этап 5. Итоги урока 5 мин

> ГОВОРИМ УЧЕНИКАМ
>
> Объясните работу каждой роли по её результату. По результату будет понятно, какие ошибки замечает каждый специалист.

Главное за урок

- Инструкцию для модели проверяют на разных входных данных.
- Подлинность записи проверяют по нескольким независимым признакам.
- Этика ИИ помогает решить, кого затронет выпуск и кто исправит ошибку.

Итоговые вопросы

1. Как испытать инструкцию, которая просит модель сохранять факты из письма?
2. Что сделаете, если знакомый голос и результат детектора дают разные подсказки?
3. Когда понадобится специалист по этике, даже если запись подлинная?

Для учителя

Возьмите несколько писем, в которых пропущены разные сведения. Сравните ответы модели с письмами.

При спорных сигналах проверяйте источник и подтверждайте сообщение независимо.

Этическая оценка нужна, если в записи есть частные сведения или участие согласовано только для ограниченной аудитории.

Источники

1. Федеральная комиссия по связи США. Решение FCC 24-59 · 23.05.2024. «using AI-generated voice cloning technology and caller ID spoofing» Источник