



9 класс · Урок 7 из 34 · 45 минут

Специалисты по дипфейкам и этике: проверка и ответственность

5.3. Искусственный интеллект для будущей профессии и образования

Результат: анализирует учебный медиапакет, проверяет происхождение и сигналы детектора, оформляет осторожный вывод и безопасный порядок действий.

Листы ученика

01. Проблемная ситуация

В пересланной афише встреча перенесена, появилась оплата, а подпись организатора выглядит знакомо. Детектор пометил файл «настоящий». Можно ли доверять афише и обвинять отправителя в создании дипфейка?

За 1 минуту прочитайте ситуацию; за 2 запишите предварительное решение и основание; за 2 сравните с соседом и назовите недостающую проверку. Случаи, числа и реплики условного ИИ созданы для обучения, не являются результатами испытания действующего сервиса.

Место для ответа



02. Объяснение и разобранный пример

5-14 минут: 4 минуты - понятия; 3 - совместный разбор; 2 - проверка ограничений примера.

Дипфейком обычно называют синтетический или изменённый с помощью ИИ медиаматериал, который убедительно изображает человека, его лицо, голос или действия. Простая замена даты на афише тоже может вводить в заблуждение, но по ней нельзя доказать применение ИИ или назвать любого редактора создателем дипфейка. Различайте содержание, происхождение и способ изготовления.

Специалист проверки собирает оригинал, автора/издателя, дату, версии, цепочку передачи и независимое подтверждение. Метаданные и подпись на картинке полезны, но могут быть неполными или изменёнными. Стандарты происхождения, например C2PA, позволяют проверять подписанные сведения о создании/правках; они не заменяют проверку истинности каждого сообщения. Отсутствие таких сведений само по себе не доказывает подделку.

Детектор выдаёт сигнал по своей модели. Он может ошибиться в обе стороны: принять изменённое за исходное или пометить исходное как подозрительное. Значение 0,94 в учебной карточке - внутренний балл, не доказанная вероятность истинности именно этого файла. Проверка по известному источнику сильнее одного визуального признака.

Специалист по этике оценивает последствия: ошибочное обвинение, распространение чужого лица/голоса, давление оплатить и потерю доступа. Он предлагает минимальный безопасный шаг: не распространять подозрительное, не платить по нему, проверить через известный канал организатора и ясно указать степень уверенности. Используем вымышленные материалы без лиц, голосов и реальных платёжных ссылок.

Разобранный пример. Если две афиши расходятся по дню, факт установлен только о различии надписей. Пока нет ответа организатора, неизвестно, какая версия актуальна. Фраза «дата изменена, значит это ИИ-дипфейк» добавляет неподтверждённое объяснение способа. Осторожная запись: «Есть противоречие; оплату и новое время не принимать до подтверждения».



03. Практика 1: Исследуйте происхождение подозрительной афиши

Учебный медиапакет: сравните условия

А. Карточка издателя, версия 1

Четверг 16:00-16:20. Кабинет 12.
Бесплатно. Учебный клуб. Источник:
вымышленный архив.

Б. Пересланная копия

Пятница 16:00. Кабинет 14. Требуется
оплата. Автор и версия не указаны.
Не настоящее объявление.

Авторские учебные изображения, не фото события и не результат ИИ.

Документ пакета	Известные сведения
А: карточка издателя	Учебный клуб, версия 1; четверг 16:00-16:20, кабинет 12, бесплатно; в вымышленном школьном архиве за 10 октября.
Б: пересланная копия	Версия/автор отсутствуют; пятница 16:00, кабинет 14, требуется оплата. Получена через пересылку 13 октября.
С: сигнал детектора	Б помечен «настоящий», балл 0,94; условия испытания и частота ошибок не сообщаются.
Д: цепочка передачи	Известно только «получено от другого участника»; оригинала Б нет.

Изображения нарисованы автором урока как учебные документы. Они не являются фотографиями реального объявления или результатом детектора. Слова «архив», даты и балл относятся только к вымышленному пакету. Полную расшифровку обеих карточек даёт таблица.



Продолжение практики 1: действия и рабочий ответ

- За 2 минуты сопоставьте изображения и запишите три различия содержания.
- За 4 минуты составьте таблицу «известно / неизвестно / источник» по автору, версии, времени, оплате и способу создания.
- За 3 минуты подготовьте запрос организатору через уже известный канал и предварительный вывод со степенью уверенности.
- За 2 минуты выберите безопасное действие до ответа и объясните, почему нельзя обвинять человека по одному сигналу.

Форма результата: Карточка расследования: сведения, неизвестное, уверенность, запрос и действие.

Критерии: По 1 баллу: конкретные расхождения; происхождение/пробелы; вывод с уверенностью; независимая проверка/безопасность. Всего 4.

Место для ответа



04. Практика 2: Проверьте детектор и пересмотрите вывод

Контрольный файл	Известный эталон учебного архива	Ответ детектора
K1	Исходный	Подозрительный
K2	Изменённый	Исходный
K3	Исходный	Исходный
K4	Изменённый	Подозрительный

Новый документ Е: ответ известного организатора подтверждает действующую версию 1: четверг 16:00-16:20, кабинет 12, бесплатно; он не выпускал Б. Ответ не содержит сведений о программе, которой создан Б. Это дополнительная учебная карточка, не настоящий ответ человека.

- За 2 минуты определите верные ответы и оба вида ошибок детектора.
- За 3 минуты пересмотрите вывод о содержании Б после Е; отдельно оставьте открытым вопрос о способе создания.
- За 2 минуты составьте сообщение исправления для школьной группы без обвинения конкретного человека.
- За 1 минуту укажите, какие сведения не следует собирать ради проверки.

Форма результата: Протокол четырёх тестов и сообщение исправления с основанием.

Критерии: По 1 баллу: два вида ошибок; обновлённый вывод; корректное исправление; минимальные данные/ограничение. Всего 4.

Место для ответа



05. Обсуждение: аргумент и возражение

Тезис: «Лучше публично назвать автора подделки сразу, иначе все пострадают». За 2 минуты назовите опасность сообщения; за 3 возразите по неизвестному происхождению; за 2 согласуйте защитное действие.

Место для ответа

06. Индивидуальная проверка понимания

У файла нет метаданных, детектор сказал «подозрительный», издатель ещё не ответил. Запишите: установленный факт; недоказанный вывод; независимую проверку; допустимое действие.

За 3 минуты запишите ответ; за минуту сверьте ключ и исправьте другим цветом; за последнюю минуту сдайте запись. По 1 баллу за каждый из четырёх содержательных элементов.

Место для ответа



Для учителя: ответы и критерии

9 класс · Урок 7. Специалисты по дипфейкам и этике: проверка и ответственность

Подготовьте: экран или распечатки, тетради, ручки и калькулятор. Все таблицы, тексты и ключи находятся в статье и PDF. Ключи открывайте после самостоятельной попытки; ученические листы печатайте отдельно от раздела учителя.

Вариант без регистрации: работайте по готовым материалам и учебным журналам на экране или бумаге. Ученики не создают аккаунтов и не передают личных данных. Учебная имитация позволяет проверить рассуждение, но не заменяет настоящий эксперимент с сервисом.

Время	Этап	Результат
0-5	Проблемная ситуация	Первое решение и основания
5-14	Теория и разобранный пример	Понятия, доказательства и ограничения
14-25	Практика 1	Проверяемый результат по критериям
25-33	Практика 2	Перенос в новые условия
33-40	Обсуждение	Аргумент, возражение, общее решение
40-45	Индивидуальная проверка	Самостоятельный ответ и сверка

Что собрать: результаты двух практик, основания вывода и индивидуальную проверку. Минуты включают чтение, обсуждение и запись; раздел соответствия программе предназначен учителю вне этих 45 минут. При 0-1 балле в проверке вернитесь к разобранным примерам; при 2-3 уточните пропущенное звено; при 4 предложите новый случай. Критерии нужны для обратной связи, не задают обязательную школьную отметку.

01-02. Разбор проблемной ситуации и примера

Разобранный пример. Если две афиши расходятся по дню, факт установлен только о различии надписей. Пока нет ответа организатора, неизвестно, какая версия актуальна. Фраза «дата изменена, значит это ИИ-дипфейк» добавляет неподтверждённое объяснение способа. Осторожная запись: «Есть противоречие; оплату и новое время не принимать до подтверждения».

03. Исследуйте происхождение подозрительной афиши

Различаются день, кабинет и условие оплаты; в Б не указан конец интервала. Поздняя пересылка не доказывает, что Б является официальным обновлением. Автор, исходный файл и способ изготовления Б неизвестны; балл 0,94 не устраняет этот пробел. Запрос: подтвердить актуальные день/время/место/оплату и прислать действующую версию по известному школьному каналу. Предварительный вывод: противоречие подтверждено, достоверность условий Б и участие ИИ не подтверждены. Не платить/не распространять Б до проверки, не отправлять документы/секреты. «Высокая уверенность в наличии расхождения» и «низкая уверенность в способе создания» могут сосуществовать.

Критерии: По 1 баллу: конкретные расхождения; происхождение/пробелы; вывод с уверенностью; независимая проверка/безопасность. Всего 4.



04. Проверьте детектор и пересмотрите вывод

Верны К3/К4, ошибочны К1/К2: ложная тревога на исходном и пропуск изменённого. Две ошибки из четырёх, но набор слишком мал для универсальной оценки. После Е высокая уверенность: Б не официальное обновление, его день/кабинет/оплата противоречат действующим условиям. Участие ИИ и личность создателя по-прежнему не установлены. Сообщение: «Проверено у организатора: четверг 16:00-16:20, кабинет 12, бесплатно. Пересланная копия с оплатой не подтверждена; используйте версию 1». Не распространять неподтверждённое обвинение, лица, голоса, адреса, секретные коды или платёжные данные.

Критерии: По 1 баллу: два вида ошибок; обновлённый вывод; корректное исправление; минимальные данные/ограничение. Всего 4.

05. Обсуждение: ожидаемый результат

Тезис: «Лучше публично назвать автора подделки сразу, иначе все пострадают». За 2 минуты назовите опасность сообщения; за 3 возразите по неизвестному происхождению; за 2 согласуйте защитное действие.

Предупредить о неподтверждённом содержании полезно, но обвинение требует оснований. Можно остановить распространение/оплату, показать проверенные условия и обратиться к ответственному взрослому, не приписывая неизвестному человеку создание ИИ-медиа.

Критерии: аргумент по материалу; содержательное возражение; условие итогового решения; роль человека. По 1 баллу, всего 4.

06. Индивидуальная проверка: ключ

Факт - сведения отсутствуют/есть сигнал, если это действительно наблюдалось. Не доказаны подделка, участие ИИ и личность автора. Проверить оригинал и актуальные сведения у издателя через известный канал. До ответа не распространять/не действовать по сомнительному содержанию, сообщить взрослому. По 1 за элемент.



07. Связь с программой

Пункт программы	Действие	Проверяемый результат
Специалист по выявлению дипфейков. Специалист по этике искусственного интеллекта.	Разделяет проверки происхождения и последствий	Не путает сигнал с доказательством
Называть функции, цели и задачи специалистов по этим профессиям.	Создаёт карточку расследования	Есть независимый источник и осторожный вывод
Выполнять практические задания.	Испытывает детектор по четырём эталонам	Различает оба вида ошибок

Источники и права

Основа содержания и результатов - [исходная программа класса](#). Поурочное название и учебные случаи - авторская методическая развёртка, не дословные названия модулей программы. Числовые наборы, реплики условных помощников и ситуации специально подготовлены для урока. Учебные журналы не являются результатами испытания реального ИИ.

[С2РА: происхождение медиа и ограничения цепочки; проверено 15 октября 2026](#) - источник для понятий и проверки; учебные задания разработаны отдельно.

Логотип «Школы России» используется в материалах портала; авторские схемы предназначены для учебной работы. Внешние материалы сохраняют указанные условия использования.