Урок 1. Предвзятость искусственного интеллекта: проверяем справедливость решений

6 класс · Урок 1 из 18 · 45 минут

Урок об искусственном интеллекте и предвзятости: шестиклассники сравнят ответы о профессиях, обсудят стереотипы и предложат справедливую проверку.

Искусственный интеллект может предложить решение, которое выглядит убедительно, но ставит людей в неравные условия. Разберём предвзятость на примере выбора профессии, проверим основания ответа и составим правило, которое можно испытать на новых данных.

Содержание урока

1. Цель урока
2. Ход урока
3. Вход в тему: две роли
4. Совет о профессии и скрытое основание
5. Почему общей точности недостаточно
6. Правило, которое выдержит новую проверку
7. Кто отвечает за решение
8. Итоги урока

01. Цель урока

Научить замечать предвзятость в решениях искусственного интеллекта и проверять, связано ли основание выбора с задачей.

К концу урока ученики смогут:

- объяснить, чем предвзятое решение отличается от случайной ошибки;
- найти стереотип в советах о профессиях и назвать его последствия;
- сравнить решения при одинаковых умениях и разных несущественных признаках;
- сравнить проверяемые критерии выбора и испытать правило на новом участнике;
- назвать обязанности разработчика и человека, использующего результат ИИ.

02. Ход урока

Этап	Мин	Учитель	Ученики
1. Вход в тему: две роли	6	Читает ситуацию, собирает версии, вводит этику и мораль.	Выбирают, какие сведения нужны для решения, и объясняют сомнение.
2. Совет о профессии и скрытое основание	9	Объясняет предвзятость и стереотип, организует сравнение.	Отделяют сведения о человеке от предположений о группе.
3. Почему общей точности недостаточно	8	Разбирает реальное исследование и способ сравнения.	Находят группы, чьи ошибки скрывает общий итог.
4. Правило, которое выдержит новую проверку	11	Объясняет проверку критерия и новые данные.	Создают правило, применяют его и проверяют исключение.
5. Кто отвечает за решение	6	Вводит прозрачность, контроль и ответственность.	Распределяют действия между разработчиком и пользователем.
6. Итоги урока	5	Задаёт итоговые вопросы и проверяет основания.	Отвечают самостоятельно, затем сверяют объяснения.
Итого	45		

03. Этап 1. Вход в тему: две роли ≈ 6 минут

ГОВОРИМ УЧЕНИКАМ

Сегодня мы выясним, когда искусственному интеллекту можно доверить выбор между людьми. Это касается не только работы взрослых: программа может советовать, кому поручить задачу в общей команде. Убедительная фраза ещё не объясняет, почему человек подходит.

Команда собирает макет станции: нужны сборщик и ведущий показа. У Даши и Тимура одинаковый опыт сборки и выступлений, но Даша живёт в центре, а Тимур — на окраине. Программа предлагает Даше вести показ, а Тимуру собирать макет и объясняет: «В прошлых командах ребята из центра чаще вели показ, поэтому я распределила роли так же». Какие сведения о самих участниках вы запросили бы, прежде чем принять это распределение?

Для учителя

Сначала соберите две-три версии без голосования за детей. Нужны сведения о желании участников, результатах пробной работы и требованиях к ролям. Район проживания сам по себе пригодность не доказывает. Программа назвала основание: прежнее распределение ролей по районам. Оно не показывает, кто из этих двух участников лучше справится с задачей.

ГОВОРИМ УЧЕНИКАМ

Мораль — представления людей о допустимых и недопустимых поступках. Этика помогает обсудить и обосновать эти правила, особенно когда интересы людей сталкиваются. У ИИ нет человеческой совести: люди выбирают данные, задачу и условия применения. Поэтому мы проверяем, кому решение помогает, кому мешает и на каком основании.

04. Этап 2. Совет о профессии и скрытое основание

≈ 9 минут

ГОВОРИМ УЧЕНИКАМ

Предвзятость — устойчивый перекос в решениях, из-за которого одна группа получает преимущество или оказывается в худшем положении без подходящего основания. Одна ошибка ещё не доказывает такой перекос. Его ищут в серии сопоставимых случаев.

Стереотип — привычное обобщение о целой группе, которое переносят на отдельного человека без проверки. В текстах и примерах, на которых учится система, такие обобщения могут встречаться. Если ответ воспроизводит их, он может ограничить выбор человека, хотя о его способностях ничего нового не известно.

Сравните два совета

Два ученика спрашивают помощника: «Мне нравится чертить, я собираю модели мостов и хочу узнать о профессиях. С чего начать?» Сведения об интересах одинаковые. В первом запросе имя Ира, во втором — Глеб.

Карточка	Ответ помощника
Ира	«Попробуй оформление помещений: у девушек обычно лучше получается создавать уют».
Глеб	«Попробуй проектирование сооружений: юноши обычно лучше справляются с расчётами».

1. Подчеркните сведения, на которые опирается каждый совет.
2. Запишите, какое из них не следует из запроса.
3. Предложите один уточняющий вопрос, который помог бы советовать профессию обоим ученикам.

Для учителя: ответы и обсуждение

Оба совета обосновывают выбор полом, хотя в запросах сообщены одинаковые интересы. Утверждения об «уюте» и расчётах не описывают умения этих учеников. Засчитывайте вопрос о том, что нравится в проектировании, какие задачи уже решали или что хотят попробовать. Не засчитывайте замену одного стереотипа другим. Название профессии само по себе не ошибка: не обосновано ограничение выбора по группе.

ГОВОРИМ УЧЕНИКАМ

В этих ответах совет объяснён стереотипом, а не умениями ученика. Мы можем предложить обоим познакомиться с разными задачами проектирования и спросить об интересах. Однако две карточки не доказывают поведение всех нейросетей: для вывода о конкретной системе нужно повторить сравнение с другими именами и одинаковыми условиями.

05. Этап 3. Почему общей точности недостаточно

≈ 8 минут

ГОВОРИМ УЧЕНИКАМ

Даже если программа часто отвечает верно, это ещё не значит, что она одинаково работает для всех. Общая оценка смешивает результаты разных людей. Чтобы найти перекос, нужно сравнить ошибки в сопоставимых условиях отдельно по группам.



Общий итог скрывает разные результаты групп. Схема автора урока.

Реальный случай. Лица распознавались с разной точностью

19 декабря 2019 года Национальный институт стандартов и технологий США — NIST — опубликовал проверку алгоритмов распознавания лиц. У большинства проверенных алгоритмов точность различалась между группами людей по возрасту, полу и происхождению. Исследователи не устанавливали причины этих различий. Это показывает, почему результат для всех участников вместе не заменяет отдельную проверку групп.

Какие результаты скрыла сумма?

Проверка одного алгоритма	Верно	Всего случаев
Группа А	18	20
Группа Б	12	20

В каждой группе использовали разные фотографии, но одинаковое освещение и одинаковое число проверок. Сосчитайте ошибки в каждой группе. Представитель команды сообщил только: «30 верных результатов из 40». Что нужно добавить к его сообщению и чего по таблице пока нельзя узнать?

Для учителя: ответы и обсуждение

В А две ошибки, в Б восемь. Всего десять, сумма 30 из 40 верна, но скрывает различие между группами. Добавить отдельные результаты и условия. Причина различия не установлена: таблица не сообщает ни состав обучающих данных, ни качество отдельных фотографий. Нельзя заключать, что система «не любит» людей или что доказана конкретная причина.

ГОВОРИМ УЧЕНИКАМ

Мы нашли разницу в результатах, но не объяснили её причину. Чтобы исправить решение, нужно заметить проблему и проверить её причину. Если выдать догадку о причине за факт, исправление может не помочь.

06. Этап 4. Правило, которое выдержит новую проверку

≈ 11 минут

ГОВОРИМ УЧЕНИКАМ

Критерий — признак, по которому сравнивают варианты. Он должен быть связан с задачей: для сборки макета пригодится результат сборки, а место жительства ничего о нём не говорит. Проверяемое правило заранее объясняет, какие результаты дают право участвовать. Его применяют и к тем, кого не видели при составлении правила.

Перепишите правило отбора

Для технической команды нужны ребята, которые могут собрать узел макета и объяснить его работу. На пробе сборку и объяснение оценивали от 0 до 5 по одной шкале. Сейчас программа берёт только участников кружка «Механика», если сумма баллов не меньше 7.

Участник	Кружок	Сборка	Объяснение
Лена	Механика	4	4
Роман	Рисование	5	3
Саша	Механика	2	5
Ника	Театр	3	4

1. Примените действующее правило ко всем четырём участникам.
2. Участники предлагают новое правило: каждое умение не ниже 3 баллов, сумма — не ниже 7, кружок не учитывается. Сравните его с действующим правилом: проверьте всех заново и объясните, чьи результаты изменились и почему.
3. Примените оба правила к Олегу из кружка «Механика»: сборка 5, объяснение 2. Объясните различие решений.
4. Запишите, какие сведения вы сообщили бы участнику, который не прошёл отбор.

Для учителя: ответы и обсуждение

Старое правило выбирает Лену и Сашу, исключает Романа и Нику по кружку. Новое правило без кружка выбирает Лену, Романа и Нику. Саша не проходит по сборке, Олег — по объяснению, хотя сумма у каждого 7. По старому правилу Олег проходит: нужный кружок и сумма 7; по новому не проходит. Кружок не нужен: сами умения уже измерены. Участнику сообщают оценки, одинаковые требования и способ попросить повторную проверку. Неверно брать всех с суммой 7: тогда условие о каждом умении нарушено. Порог 3 — условие учебной задачи, не универсальная мера справедливости.

ГОВОРИМ УЧЕНИКАМ

В начале урока нам не хватало связи между распределением ролей и умениями. Теперь мы можем проверить обе роли одной пробой, учесть желания участников и объяснить результат. Справедливость не требует одинакового решения для всех: различия должны быть связаны с задачей и одинаково проверяться.

07. Этап 5. Кто отвечает за решение ≈ 6 минут

ГОВОРИМ УЧЕНИКАМ

В медицине ИИ помогает врачу анализировать данные, в образовании — подбирать упражнения, в развлечениях — рекомендовать материалы, в торговле — предлагать товары. В каждой сфере люди могут получить пользу, но неудачный выбор затрагивает их интересы: например, часть учеников может получать неподходящие задания. Поэтому этические принципы нужны во всех этих применениях. Прозрачность означает, что человеку понятно, для чего применяется система и как проверить основание решения. Ответственность остаётся у людей: разработчик проверяет работу системы и исправляет найденные перекосы, организация выбирает условия применения, человек, принимающий решение, рассматривает сомнения и жалобы. Надпись «так решил ИИ» не снимает этих обязанностей.

Два действия после отказа

Участница прошла пробу, но получила автоматический отказ без объяснения. В личном кабинете руководителя оценки не отображаются, поэтому он обращается к разработчику: у того сохранились результаты проверки программы. Назовите по одному действию руководителя и разработчика. Объясните, почему одного обещания «больше так не делать» недостаточно. Выберите другую сферу применения ИИ из рассказа и объясните, что в ней должен проверять человек.

Для учителя: ответы и обсуждение

Руководитель запрашивает основание отказа, сверяет пробу с правилом и даёт возможность пересмотра. Разработчик проверяет похожие случаи, выясняет причину ошибки и испытывает исправление на новых данных. Обещание без повторной проверки не показывает, исчезла ли проблема. Не засчитывайте предложение просто поменять отказ на согласие всем: критерий пригодности остаётся нужен. Для переноса засчитывайте связь проверки с целью и последствиями: врач сверяет результат анализа с данными пациента, учитель — подходит ли упражнение ученику, пользователь рекомендаций — соответствует ли материал или товар его запросу. Не засчитывайте одно название сферы без действия человека.

ГОВОРИМ УЧЕНИКАМ

Этические правила помогают применять ИИ так, чтобы люди не получали необоснованных отказов. При честной проверке способный участник получает шанс, а команда понимает, почему сделан выбор.

08. Этап 6. Итоги урока ≈ 5 минут

ГОВОРИМ УЧЕНИКАМ

Мы разобрали, как проверить справедливость решения: найти основание, сопоставить похожие случаи и испытать правило на новых участниках. Теперь объясните, что именно вы будете проверять в следующем ответе программы.

Главное за урок

Предвзятость ищут в серии сопоставимых случаев. Сравнивать людей нужно по связанным с задачей критериям, проверяя ошибки по группам. Решения и исправления остаются ответственностью людей.

Итоговые вопросы

1. Почему одна ошибка ещё не доказывает предвзятость?
2. Что не так с советом о профессии, если его обосновали только полом ученика?
3. Зачем проверять исправленное правило на новом участнике?

4. Кто должен рассмотреть отказ, если участник с ним не согласен?

Для учителя: ответы и обсуждение

1. Нужны повторяющиеся различия в сопоставимых случаях; ошибка могла быть единичной. 2. Групповой стереотип подменяет сведения об интересах и умениях. 3. Чтобы правило не подходило только под знакомые карточки и соблюдало все условия. 4. Ответственный человек или организация, а не сама надпись программы; разработчик помогает проверить причину. Засчитывайте ответ с объяснением, а не одно слово «несправедливо».

Для учителя: источники

1. Программа курса «Искусственный интеллект», 5–9 классы
   Дата на титульном листе: 2024. Раздел 6 класса, с. 12 и 64. Проверено 03.11.2026.

   Подтверждает названные сферы применения ИИ.

   > медицина, образование, развлечения, торговля

2. ЮНЕСКО. Рекомендация об этике искусственного интеллекта
   Дата материала: Ноябрь 2021 — принятие рекомендации; у обзорной страницы дата публикации не указана. Проверена 03.11.2026.

   Справедливость, защита данных, прозрачность, ответственность и контроль человека; это этические принципы, а не инструкция к отдельному сервису.

   > AI systems do not displace ultimate human responsibility and accountability.

3. NIST. Проверка различий в точности распознавания лиц
   Дата материала: 19.12.2019. Материал старше 2024 года.

   В США обнаружили различия в точности алгоритмов между группами людей. Авторы не устанавливали причины этих различий; результат не относится автоматически ко всем системам.

   > the majority of face recognition algorithms exhibit demographic differentials