Урок 4. Цифровая безопасность и надёжные ИИ-сервисы

9 класс · Урок 4 из 34 · 45 минут

Вся программа 9 класса

Безопасность искусственного интеллекта для 9 класса. Учимся проверять ИИ-сервисы, сокращать объём передаваемых данных и распознавать подозрительные запросы.

Содержание урока

- 01. Цель урока
- 02. Ход урока
- 03. Этап 1. Введение: Кнопка импорта · 4 мин
- 04. Этап 2. Куда попадают данные · 10 мин
- 05. Этап 3. Экспертиза трёх сервисов · 13 мин
- 06. Этап 4. Разбор подозрительного приглашения · 13 мин
- 07. Этап 5. Итоги урока · 5 мин

01. Цель урока

Оценить надёжность ИИ-сервиса по его происхождению, запросам доступа и правилам хранения данных. Выбрать безопасный способ выполнить задачу.

02. Ход урока

Этап	Мин	Учитель	Ученики
Этап 1. Введение: Кнопка импорта	4	Предлагает выбрать между быстрым импортом и проверкой.	Формулируют вопросы к экрану загрузки.
Этап 2. Куда попадают данные	10	Объясняет хранение, разрешения и минимизацию данных.	Сопоставляют свойства сервиса с рисками.
Этап 3. Экспертиза трёх сервисов	13	Даёт сравнить разные сочетания признаков надёжности.	Выбирают способ работы и обосновывают допуск данных.

Этап	Мин	Учитель	Ученики
Этап 4. Разбор подозрительного приглашения	13	Разбирает проверку приглашений и действия после ошибки.	Проигрывают три ситуации и сравнивают порядок шагов.
Этап 5. Итоги урока	5	Подводит итог и сверяет ответы.	Письменно отвечают и сверяют решение.
Итого	45		

03. Этап 1. Введение: Кнопка импорта 4 мин

ГОВОРИМ УЧЕНИКАМ

Вы делаете буклет для школьного фестиваля. Редактор обещает сам разложить выступления по времени и предлагает загрузить анкеты участников целиком. В таблице есть телефоны для связи с организатором. Сегодня разберём, что означает безопасность искусственного интеллекта для того, кто пользуется сервисом, и как закончить работу, сохранив данные участников.

Два вопроса до загрузки

Запишите, что нужно узнать о сервисе и о самой таблице, прежде чем нажать «Импорт». Выслушайте два варианта в классе.

Для учителя

Спросите, кому попадёт файл и какие поля реально нужны. К остальным вопросам вернитесь после объяснения.

04. Этап 2. Куда попадают данные 10 мин

ГОВОРИМ УЧЕНИКАМ

При загрузке в онлайн-сервис файл уходит на сервер. Владелец сервиса может хранить его или историю запросов по своим правилам. Посмотрите, указаны ли разработчик, срок хранения и способ удаления, использует ли сервис загрузки для улучшения моделей и кому могут быть доступны данные. Закрытый профиль ограничивает видимость для других пользователей, а правила хранения объясняют, что происходит внутри сервиса. Ссылка с поддельного сайта вообще ведёт к другому владельцу, поэтому открывайте знакомый адрес самостоятельно или через школьный каталог.

Минимизация данных означает, что для работы передают только необходимое. Для буклета достаточно темы, времени и площадки. Телефоны можно убрать. Замена фамилии

инициалами помогает лишь тогда, когда остальная информация тоже не позволяет узнать человека. Посмотрите и на разрешения. Доступ к одному выбранному файлу и разрешение читать всё облачное хранилище дают сервису разные возможности.

На экране входа и на экране выдачи разрешения происходят разные вещи. При входе вы подтверждаете, кто вы. Когда выдаёте разрешение, вы позволяете приложению, например, читать документы или менять их. Знакомый способ входа не означает, что любой запрос доступа безопасен. Проверьте название приложения и перечисленные права, прежде чем согласиться. Одного лишнего разрешения бывает достаточно, чтобы сервис увидел документы, которые вы не собирались открывать.

Реальный случай. Чужие заголовки в истории чатов

20 марта 2023 года в ChatGPT произошёл сбой. Из-за ошибки во вспомогательной программной библиотеке часть пользователей увидела заголовки чатов других людей, которые были активны в это же время. Разработчик сообщил, что исправил ошибку. Этот случай показывает, почему даже при обычном использовании стоит сокращать объём сведений, которые уходят в облачный сервис.

05. Этап 3. Экспертиза трёх сервисов 13 мин

ГОВОРИМ УЧЕНИКАМ

Бывает, что у сервиса есть понятные правила, но он просит слишком широкий доступ. Бывает наоборот: доступ ограничен одним файлом, а про хранение ничего не сказано. Оценка зависит от задачи. Открытое расписание и внутренний файл с контактами требуют разных решений. Если для работы хватит открытого фрагмента, можно подготовить его отдельно или использовать школьную программу для работы с таблицами.

Заявка на школьный каталог

Организатору надо сделать 30 диаграмм по открытому расписанию к завтрашнему утру. Из внутреннего файла с контактами и пометками ему позже понадобится одна диаграмма. А и Б строят серию автоматически, В позволяет строить графики по одному. Изучите три макета, предложите способ для каждой задачи и укажите, что вы изменили бы в доступе или настройках выбранного сервиса.

Макет	Кто отвечает и что обещано	Что запрашивает
А	Назван разработчик. Загрузки хранятся 7 дней, затем удаляются. Указан контакт для удаления раньше.	Чтение папки облачного диска целиком
Б	Названа компания. В справке объяснён экспорт. О сроке хранения и использовании файлов сведений нет.	Только выбранный файл

Макет	Кто отвечает и что обещано	Что запрашивает
В	Редактор работает на школьном компьютере. Инструкция указывает, что файлы остаются на компьютере, функции ИИ нет.	Только выбранный файл

Для учителя

Внутренний файл оставляют в одобренной школьной среде, и для диаграммы подходит В. Для открытой копии допустим Б, если документ действительно публичный и школе подходит такой режим. Для А можно создать отдельную папку с открытой копией и выдать доступ только к этой папке, если это технически возможно. Попросите связать каждое решение с конкретным разрешением или правилом хранения.

06. Этап 4. Разбор подозрительного приглашения 13 мин

ГОВОРИМ УЧЕНИКАМ

Срочность, обещание бонуса или знакомый логотип могут убедить нажать ссылку. Так работает **фишинг**: человека пытаются заставить выдать данные или открыть доступ. Сверьте приглашение через организатора по уже известному каналу. Пароли делайте разными для разных сервисов и включайте второй способ подтверждения входа. Код подтверждения вводят только там, где вы сами начали вход и уверены в адресе.

Если пароль уже ввели на подозрительной странице, перейдите на настоящий сайт самостоятельно, смените пароль и завершите другие сеансы, если сервис это позволяет. Доступ к папке или приложению отмените отдельно. О школьной учётной записи сразу сообщите ответственному взрослому или администратору.

Посмотрим, как это меняет порядок действий. Человеку, который только прочитал сообщение, нужно проверить отправителя и адрес. Если вы уже вошли в настоящий аккаунт, но дали доступ постороннему приложению, проверьте список приложений с доступом. Новый пароль и удалённое разрешение решают разные задачи.

Один случай, три момента

Организатор районного кинопоказа разослал приглашения через школьный сайт. Почти одновременно вам пришло личное сообщение со ссылкой, которая обещает «собрать программу показа за минуту» и предлагает войти школьной учётной записью.

1. В парах разыграйте по одному моменту: ссылку ещё не открывали, пароль уже ввели или приложению дали доступ к папке с работами.
2. Запишите первое действие и два следующих. Назовите, что вы хотите остановить каждым шагом.

3. Найдите пару с другим моментом и сравните планы.

Для учителя

До входа — проверить приглашение у организатора по известному адресу и зайти с официальной страницы. После ввода — смена пароля на настоящем сайте, выход из сеансов, уведомление администратора. После выдачи разрешения — отозвать его, сообщить об инциденте и проверить, к чему был открыт доступ. В реальной ситуации действуют правила школы.

07. Этап 5. Итоги урока 5 мин

ГОВОРИМ УЧЕНИКАМ

Безопасный вариант полезно продумывать заранее, пока ещё можно выбрать, какие сведения и куда отправить.

Главное за урок

- Для работы сервиса обычно достаточно части исходного файла.
- Надёжность проверяют по адресу, разработчику, доступам и правилам обращения с данными.
- Смену пароля и отзыв доступа к файлам проверяют отдельно.

Итоговые вопросы

1. Какие поля вы оставите, если надо отсортировать расписание по времени, а в таблице есть телефоны участников?
2. Какие два признака сервиса вы проверите до загрузки файла?
3. Что предпримете, если уже открыли приложению доступ ко всему школьному диску?

Для учителя

Оставляют поля, необходимые для расписания.

Подходят срок хранения и способ удаления, адрес и разработчик, область запрашиваемого доступа.

Отзовите доступ и сообщите человеку, который отвечает за общий диск.

Источники

1. OpenAI. March 20 ChatGPT outage · 24.03.2023 · Материал старше 2024 года. «allowed some users to see titles from another active user’s chat history» Источник
2. Национальный центр кибербезопасности Великобритании. ChatGPT and LLMs: what’s the risk · 14.03.2023 · Материал старше 2024 года. «not to include sensitive information in queries to public LLMs» Источник
3. CISA. Secure Our World · август 2024. «Recognize and report phishing. Use strong passwords. Turn on multifactor authentication» Источник