

Государственное учреждение образования
«Средняя школа № 80 г. Минска»

Название работы:

«Цифровой детектив. Охота на Фейк-бота» (формирование кибербезопасности через взаимодействие с ИИ)

Номинация:

Использование технологий искусственного интеллекта в воспитательном процессе

Автор: Хоченкова Татьяна Максимовна, учитель, высшая квалификационная категория

Контактная информация:

тел.: +375333461350

e-mail: xm2408@mail.ru

1. ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

1.1. Актуальность и новизна разработки

Актуальность обусловлена стремительным внедрением технологий искусственного интеллекта (ИИ) в повседневную жизнь подростков. Согласно данным Лаборатории Касперского, в 2025-2026 годах более 60% школьников сталкивались с дипфейками или контентом, сгенерированным нейросетями, при этом 70% из них не смогли отличить фейк от реальности. Мошеннические схемы с использованием синтеза голоса и генерации изображений становятся основной угрозой для несовершеннолетних.

В то же время традиционные уроки безопасности в интернете часто носят лекционный, отвлеченный характер. Подростки воспринимают их как формальность, не получая практического опыта распознавания угроз, созданных с помощью ИИ.

Разработка апробирована на базе средней школы № 80 г. Минска с учащимися 6 «А», 6 «Б», 8 «В» классов (всего 65 человек). Полученные результаты показали высокий уровень вовлеченности и положительную динамику формирования навыков кибербезопасности.

Новизна разработки заключается в следующем:

- инверсии ролей: дети выступают не пассивными слушателями, а «цифровыми детективами», которые самостоятельно расследуют преступление с помощью инструментов ИИ;

- практико-ориентированном использовании ИИ: нейросети используются не как объект изучения, а как инструмент создания учебных кейсов (дипфейк-голос, сгенерированные фото, ИИ-тексты) и как помощник в формулировании правил безопасности;

- системном подходе: мероприятие является стартовым этапом для создания школьного «Нейросетевого патруля», что обеспечивает долгосрочность воспитательного эффекта.

Таким образом, разработка сочетает игровые технологии, кейс-метод и инструменты искусственного интеллекта, что обеспечивает высокий уровень вовлеченности и формирование навыков безопасного поведения в интернете.

Практическая значимость разработки заключается в возможности ее тиражирования в учреждениях образования без необходимости сложного технического оснащения.

1.2. Цель и задачи

Цель: сформировать у учащихся устойчивые навыки распознавания угроз, созданных с помощью технологий искусственного интеллекта (дипфейки, синтез речи, ИИ-тексты), и алгоритмы безопасного поведения в цифровой среде через активную игровую деятельность.

Практический результат: создание учащимися «Кодекса цифрового детектива»

Задачи:

Обучающие:

- познакомить с признаками дипфейков в видео, аудио и графических материалах;

научить алгоритму верификации подозрительной информации («правило двух фактов»);

сформировать первичные навыки формулирования промптов для ИИ с целью создания позитивного контента.

Развивающие:

развить критическое мышление, умение анализировать цифровой контент и работать в команде;

развить навыки использования ИИ-инструментов для решения прикладных задач.

Воспитательные:

воспитать ответственное отношение к распространению информации и защите личных данных;

сформировать мотивацию к активной позиции в вопросах кибербезопасности (участие в «Нейросетевом патруле»).

1.3. Описание используемых технологий ИИ

В разработке используются следующие ИИ-инструменты (на этапе подготовки и проведения):

Название	Функция	Применение в мероприятии
freetts.ru	Синтез речи (текст в речь)	Создание дипфейк-голоса для Станции 1 («Голос»)
ChatGPT Plus с DALL-E/ ChatGPT/ GPT Images	Генерация изображений по тексту	Создание фейковых фото-доказательств для Станции 2 («Фото»)
DeepSeek/ChatGPT	Генерация текста	Создание подозрительных сообщений для Станции 3 («Текст») и помощь детям в формулировании правил на Станции 5
Hive Moderation	Детекция ИИ-контента	Демонстрация на Станции 2 — как проверить изображение
ChatGPT Plus с DALL-E/ ChatGPT/ GPT Images	Генерация и редактирование изображений	Финальное оформление «Кодекса цифрового детектива», бейджей

Примечание: все сгенерированные материалы (аудио, фото, тексты) создаются педагогом заранее, чтобы во время мероприятия не зависеть от стабильности интернет-соединения.

1.4. Ожидаемые результаты

Предметные:

учащиеся перечисляют не менее 3 признаков дипфейка (неестественные интонации, артефакты на фото, неестественная структура текста);

применяют алгоритм проверки подозрительного сообщения (позвонить, спросить личное, не переходить по ссылкам);

создают с помощью ИИ (под руководством) памятку по безопасности.

Метапредметные:

работают в группе, распределяя роли, аргументируют свою точку зрения;

используют ИИ-инструменты для анализа информации.

Личностные:

демонстрируют готовность к участию в школьном «Нейросетевом патруле»;

проявляют осторожность при получении подозрительных сообщений в мессенджерах.

1.5. Целевая аудитория

Учащиеся 6–8 классов (12–15 лет).

Обоснование: этот возраст характеризуется высокой активностью в соцсетях и мессенджерах, но недостаточно сформированным критическим мышлением. Формат квеста соответствует ведущей деятельности – межличностному общению и поиску себя.

1.6. Обоснование выбора форм и методов работы

Форма (квест-расследование) позволяет включить каждого участника, создает эмоциональную вовлеченность через игровую драматургию (легенду о похищении), что повышает запоминаемость материала.

Методы:

кейс-метод – разбор реально смоделированных ситуаций (дипфейк-сообщение, фейк-фото);

проблемно-поисковый – дети самостоятельно ищут признаки фейка на станциях;

практическая работа с ИИ – на финальной станции дети взаимодействуют с нейросетью (промптинг) для создания продукта;

рефлексивный – итоговое обсуждение и создание «Кодекса цифрового детектива».

1.7. Диагностика эффективности мероприятия

До проведения мероприятия (входная диагностика). Учащимся предлагается мини-опрос.

После проведения мероприятия (выходная диагностика). Повторное выполнение аналогичного опроса.

Критерии оценки эффективности:

- увеличение количества правильных ответов не менее чем на 40%
- не менее 80% учащихся называют 2+ признака дипфейка
- не менее 70% учащихся используют «правило двух фактов»

Результаты фиксируются в сравнительной таблице (до/после)

1.8. Этические аспекты использования ИИ

В рамках мероприятия особое внимание уделяется формированию ответственного отношения к использованию технологий искусственного интеллекта. Учащимся транслируются следующие принципы:

недопустимость создания дипфейков с использованием образов реальных людей без их согласия;

ответственность за распространение ложной информации;
необходимость проверки контента перед публикацией;
уважение к цифровой репутации других людей;
формирование цифровой ответственности личности как основы
безопасного поведения в информационной среде.

1.9. Перспектива развития проекта

По итогам мероприятия формируется школьное объединение
«Нейросетевой патруль», в задачи которого входит:
проведение мини-тренингов для других классов;
создание памяток и видеороликов;
мониторинг цифровых угроз в школьной среде.

2. МЕТОДИЧЕСКАЯ РАЗРАБОТКА

2.1. Подробный конспект мероприятия

Тема: «Цифровой детектив. Охота на Фейк-бота».

Время: 60 минут.

Место проведения: учебный кабинет или актовый зал с возможностью разделения на 5 зон (станций).

Оборудование: проектор, экран, 5 ноутбуков/планшетов (по 1 на станцию), колонки, раздаточные материалы (маршрутные листы, бейджи(постеры)), QR-коды с ссылками на материалы.

Этап	Время	Содержание	Деятельность ведущего	Деятельность учащихся	Материалы/ ИИ	Формируемые компетенции
Организационный момент и легенда	6 мин	Приветствие. Объявление легенды: в чат класса пришло тревожное сообщение от Артема. Появилось странное фото. Класс это команда детективов.	Включает аудио-обращение (синтезированное) и (или) зачитывает сообщение, показывает фото. Делит на группы, выдает маршрутные листы.	Слушают, задают вопросы, получают маршрутные листы.	Видео/ аудио (сгенерировано ИИ)	Критическое мышление, цифровая грамотность, коммуникация
Станция 1. «Голос»	8 мин	Прослушивание аудиосообщения «похитителя» или звонка от «Артема». Поиск признаков синтеза речи.	Включает аудиофайл. Раздает чек-лист признаков дипфейка. Объясняет, как проверить голос (позвонить по видео, спросить личное).	Слушают, обсуждают в группе, фиксируют в маршрутном листе признаки фейка.	Аудиофайл (freetts.ru), чек-лист (раздатка)	Критическое мышление, цифровая грамотность
Станция 2. «Фото»	7 мин	Анализ изображений, сгенерированных нейросетью. Поиск артефактов. Знакомство с AI-детектором.	Демонстрирует фото на экране. Показывает, как работает AI-детектор (или объясняет визуальные признаки).	Находят артефакты (руки, тени, текстуры), делают вывод о фейке.	Изображения ChatGPT, проверка AI-детектор Hive Moderation	Анализ информации, визуальная грамотность
Станция 3. «Текст»	7 мин	Анализ переписки в мессенджере. Поиск признаков ИИ-	Выдает скриншоты переписки. Обсуждает «правило двух	Читают, находят неестественные фразы, подозрительные ссылки.	Скриншоты (DeepSeek/ ChatGPT)	Медиаграмотность, критическое мышление

		генерации и мошеннических схем.	фактов» и опасность перехода по ссылкам.	Формулируют правило проверки.		
Станция 4. «Цифровой суд»	10 мин	Сбор улик, сопоставление фактов, вынесение вердикта.	Координирует работу групп. Предоставляет доступ к «базе данных» (таблица с фактами).	Сопоставляют данные, делают коллективный вывод о том, что кейс – фейк.	Маршрутные листы (DeepSeek/ChatGPT)	Аргументация, коммуникация
Станция 5. «Кодекс цифрового детектива» (с ИИ)	12-15 мин	Создание правил безопасности с помощью нейросети.	Раздает промпт-шаблоны. Помогает скорректировать запросы. Организует защиту кодекса.	Используют готовый промпт и (или) формулируют промпт, получают от ИИ 4 правила, дорабатывают их, оформляют постер.	DeepSeek/ChatGPT, маркеры, бумага	Креативность, работа с ИИ
Рефлексия и итоги	10 мин	Обсуждение результатов. Вручение бейджей (постера) «Кодекс Цифрового детектива». Анонс «Нейросетевого патруля».	Проводит рефлексию (вопросы: что было сложным? Что узнали нового?). Вручает бейджи (постеры).	Делятся впечатлениями, получают бейджи (постеры), сканируют QR-код с итоговой памяткой.	Бейджи (постеры), QR-код на финальную памятку/ChatGPT	Цифровая ответственность личности

2.2. Содержание этапов

Станция 1. «Голос»

Аудиоматериал (текст для синтеза):

«Артема забрали, нужно срочно перевести 500 рублей на этот номер, иначе будет плохо. Не звоните, он не может говорить. Скиньте скрин, когда переведете».

Признаки фейка для обсуждения: монотонность, отсутствие эмоций, странные паузы, неестественное произношение, в сообщении просят не звонить, а только писать, произношение слов не совпадает с манерой говорить.

Правило: всегда перезванивай по видеосвязи, если просят деньги.

Станция 2. «Фото»

Примеры для генерации: «подросток в заброшенном здании», «подросток сидит в незнакомой комнате».

Признаки: искаженные пальцы, неестественный свет, размытые границы объектов, картинка выглядит слишком глянцево или неестественно.

Практика: педагог демонстрирует, как сервис Hive Moderation определяет вероятность ИИ-генерации (99,9% AI-generated).

Станция 3. «Текст»

Сгенерированный текст:

«Привет, это Артем. У меня срочная ситуация, не могу говорить по телефону, очень шумно. Скиньте 500 рублей на карту, завтра верну. Не пишите в других чатах, здесь связь плохая. Спс, друг».

Признаки: отсутствие привычного сленга, слишком официальная речь, странная ссылка/номер, просят не звонить.

Станция 4. «Цифровой суд»

Формат: ролевая игра с распределением функций внутри команды.

Аналитики (эксперты) представляют анализ улик.

Спикер аргументирует позицию команды.

Технический специалист демонстрирует доказательства.

Секретарь фиксирует вывод и аргументы.

Результат: группа выносит итоговый вердикт и записывает его в маршрутный лист.

Данный формат способствует развитию навыков аргументации, критического мышления и коммуникативной компетентности учащихся.

Станция 5. «Кодекс цифрового детектива» (пример промпта)

Промпт для нейросети:

«Ты – эксперт по кибербезопасности для подростков 12–15 лет.

Сформулируй 3 коротких и запоминающихся правила безопасного поведения в интернете. Используй сленг и яркие образы. Напиши в формате: "Правило 1: ... "»

Ожидаемый ответ ИИ: корректируется детьми под свой стиль, затем оформляется (могут быть различные виды оформления).

Методические комментарии для педагога

В ходе проведения мероприятия педагог выполняет не только организационную, но и направляющую функцию.

Ключевые педагогические приемы:

наводящие вопросы:

«Почему вы решили, что это сообщение фейковое?»

«Какие признаки оказались самыми убедительными?»

«Можно ли ошибиться? В какой ситуации?»

создание ситуации сомнения: педагог может намеренно не подтверждать выводы учащихся сразу, стимулируя аргументацию;

Коррекция: при неверных выводах педагог не дает готовый ответ, а возвращает к чек-листу: «Посмотрите еще раз на признаки – что вы могли упустить?»

Рефлексия: важно вывести учащихся на личный опыт: «А попадали ли вы в похожие ситуации?»

Вариативность проведения

Упрощенный формат (при ограниченных ресурсах):

использование одного устройства на класс;

демонстрация материалов через проектор;

отказ от онлайн-сервисов (все материалы заранее подготовлены);

Расширенный формат:

работа в малых группах с отдельными устройствами;

использование онлайн-ИИ в реальном времени;

подключение дополнительных станций (видео, соцсети).

Разработка может быть рекомендована к внедрению в практику воспитательной работы учреждений образования как эффективная модель формирования цифровой грамотности, критического мышления и цифровой ответственности учащихся в условиях развития технологий искусственного интеллекта.

Список используемой литературы

1. Солдатова Г.У., Рассказова Е.И. Цифровая социализация в эпоху ИИ: риски и компетенции. — М.: МГУ, 2024.
2. Методические рекомендации по профилактике кибермошенничества с использованием дипфейков (Лаборатория Касперского, 2025).
3. Инструктивные материалы по работе с нейросетями в образовании (Министерство образования, 2025).
4. Онлайн-ресурсы: DeepSeek, ChatGPT, Hive Moderation, другие.

Приложение 1. Маршрутный лист группы

ДЕЛО № [] **группы**

ЦИФРОВОЙ ДЕТЕКТИВ
Охота на «Фейк-бота»

СОСТАВ ГРУППЫ (фамилии и роли):

- _____ (аналитик)
- _____ (аналитик)
- _____ (секретарь)
- _____ (спикер)
- _____ (технический специалист)

ЛЕГЕНДА:
В 10:45 в классный чат поступило сообщение от Артема:
«Срочно! Меня похитили. Переведите 500 рублей на номер +375453451313, иначе будет плохо. Не звоните, не могу говорить.»

В социальных сетях появилось фото Артема в неизвестном месте.

ВАША ЗАДАЧА: пройти 5 станций, собрать улики и вынести вердикт: правда или фейк?

1. 2. 3. 4. 5.

Приложение 2. Чек-лист признаков дипфейка. QR-коды с материалами для станций. Вариант 1.

Чек-лист
“Цифровой детектив: Охота на “Фейк-бота”

☒ **СТАНЦИЯ 1: ГОЛОС**

- Прослушайте голосовое сообщение. _____
- Запишите ключевые фразы: _____

☒ **СТАНЦИЯ 2: ФОТО**

- Проанализируйте фото Артема. _____
- Обратите внимание на детали: место, время. _____

☒ **СТАНЦИЯ 3: ТЕКСТ**

- Изучите текстовые сообщения. _____
- Найдите возможные улики о подделках. _____

☒ **СТАНЦИЯ 4: ЦИФРОВОЙ СУД**

Вынесите вердикт: правда или фейк? _____

☒ **СТАНЦИЯ 5: КОДЕКС ЦИФРОВОГО ДЕТЕКТИВА**

- Подведите итоги расследования. _____

Вариант 2.

ЧЕК-ЛИСТ ЦИФРОВОГО ДЕТЕКТИВА													
КАК РАСПОЗНАТЬ КОНТЕНТ, СОЗДАННЫЙ ИИ													
ГОЛОС (аудио)	ФОТО и ВИДЕО (графика)												
☐ Монотонная речь, отсутствие эмоций ☐ Неестественные паузы ☐ Странное произношение, «металлический» оттенок ☐ Просьба не звонить, только писать ☐ Голос звучит слишком идеально, без шумов	☐ Проблемы с руками (лишние пальцы) ☐ Странные зубы (слишком ровные, размытые) ☐ Неестественный свет и тени ☐ Размытые края объектов, «пластиковая» кожа ☐ Фон выглядит «смазанным»												
✓ ЧТО ДЕЛАТЬ? ПОЗВОНИ ПО ВИДЕОСВЯЗИ! Спроси о том, что знаете только вы двое.	✓ ЧТО ДЕЛАТЬ? ИСПОЛЬЗУЙ AI-ДЕТЕКТОР или поищи фото через Google Images.												
ТЕКСТ (сообщения в мессенджерах, письма)	УНИВЕРСАЛЬНЫЙ АЛГОРИТМ ПРОВЕРКИ												
☐ Слишком официальный стиль ☐ Нет привычного сленга, эмодзи ☐ Странные обороты ☐ Просьба перейти по ссылке или перевести деньги ☐ Сообщение пришло в необычное время	<table border="1"><thead><tr><th>1. ОСТАНОВИСЬ</th><th>2. ПРОВЕРЬ</th><th>3. ПОДТВЕРДИ</th><th>4. СООБЩИ</th></tr></thead><tbody><tr><td>Не паникуй, не переводи деньги</td><td>Используй чек-лист, найди признаки фейка</td><td>Свяжись по другому каналу (видеозвонок)</td><td>Расскажи взрослым</td></tr><tr><td>↓</td><td>↓</td><td>↓</td><td>↓</td></tr></tbody></table>	1. ОСТАНОВИСЬ	2. ПРОВЕРЬ	3. ПОДТВЕРДИ	4. СООБЩИ	Не паникуй, не переводи деньги	Используй чек-лист, найди признаки фейка	Свяжись по другому каналу (видеозвонок)	Расскажи взрослым	↓	↓	↓	↓
1. ОСТАНОВИСЬ	2. ПРОВЕРЬ	3. ПОДТВЕРДИ	4. СООБЩИ										
Не паникуй, не переводи деньги	Используй чек-лист, найди признаки фейка	Свяжись по другому каналу (видеозвонок)	Расскажи взрослым										
↓	↓	↓	↓										
✓ ЧТО ДЕЛАТЬ? Проверь информацию через другой канал (позвони, спроси лично) НИКОГДА не переходи по ссылкам из подозрительных сообщений													
ГЛАВНЫЕ ПРАВИЛА БЕЗОПАСНОСТИ: - Если просят деньги — ПОЗВОНИ ПО ВИДЕОСВЯЗИ - Если фото кажется странным — ПРОВЕРЬ АРТЕФАКТЫ - Если сообщение подозрительное — НЕ ПЕРЕХОДИ ПО ССЫЛКАМ - Если сомневаешься — СПРОСИ У ВЗРОСЛЫХ													

Вариант 3.

ЧЕК-ЛИСТ ЦИФРОВОГО ДЕТЕКТИВА
КАК РАСПОЗНАТЬ КОНТЕНТ, СОЗДАННЫЙ ИИ

ГОЛОС (аудио)
☐ Монотонная речь, отсутствие эмоций ☐ Неестественные паузы ☐ Странное произношение, «металлический» оттенок ☐ Просьба не звонить, только писать ☐ Голос звучит слишком идеально, без шумов
✓ ЧТО ДЕЛАТЬ? ПОЗВОНИ ПО ВИДЕОСВЯЗИ! Спроси о том, что знаете только вы двое.



🖼️ ФОТО и ВИДЕО (графика)

- ☐ Проблемы с руками (лишние пальцы)
- ☐ Странные зубы (слишком ровные, размытые)
- ☐ Неестественный свет и тени
- ☐ Размытые края объектов, «пластиковая» кожа
- ☐ Фон выглядит «смазанным»
- ☐ Текст на одежде/вывесках – бессмысленный
- ✓ ЧТО ДЕЛАТЬ? ИСПОЛЬЗУЙ AI-ДЕТЕКТОР
или поищи фото через Google Images.



💬 ТЕКСТ (сообщения в мессенджерах, письма)

- ☐ Слишком официальный стиль
- ☐ Нет привычного сленга, эмодзи
- ☐ Странные обороты
- ☐ Просьба перейти по ссылке или перевести деньги
- ☐ Сообщение пришло в необычное время
- ☐ Грамматика слишком идеальная или неестественная
- ✓ ЧТО ДЕЛАТЬ? ПРАВИЛО ДВУХ ФАКТОВ:
Проверь информацию через другой канал
(позвони, спроси лично). НИКОГДА не переходи
по ссылкам из подозрительных сообщений.



🔍 УНИВЕРСАЛЬНЫЙ АЛГОРИТМ ПРОВЕРКИ

1. ОСТАНОВИСЬ
Не паникуй, не переводи деньги
2. ПРОВЕРЬ
Используй чек-лист, найди признаки фейка
3. ПОДТВЕРДИ
Свяжись по другому каналу (видеозвонок)
4. СООБЩИ
Расскажи взрослым



🛡️ ГЛАВНЫЕ ПРАВИЛА БЕЗОПАСНОСТИ:

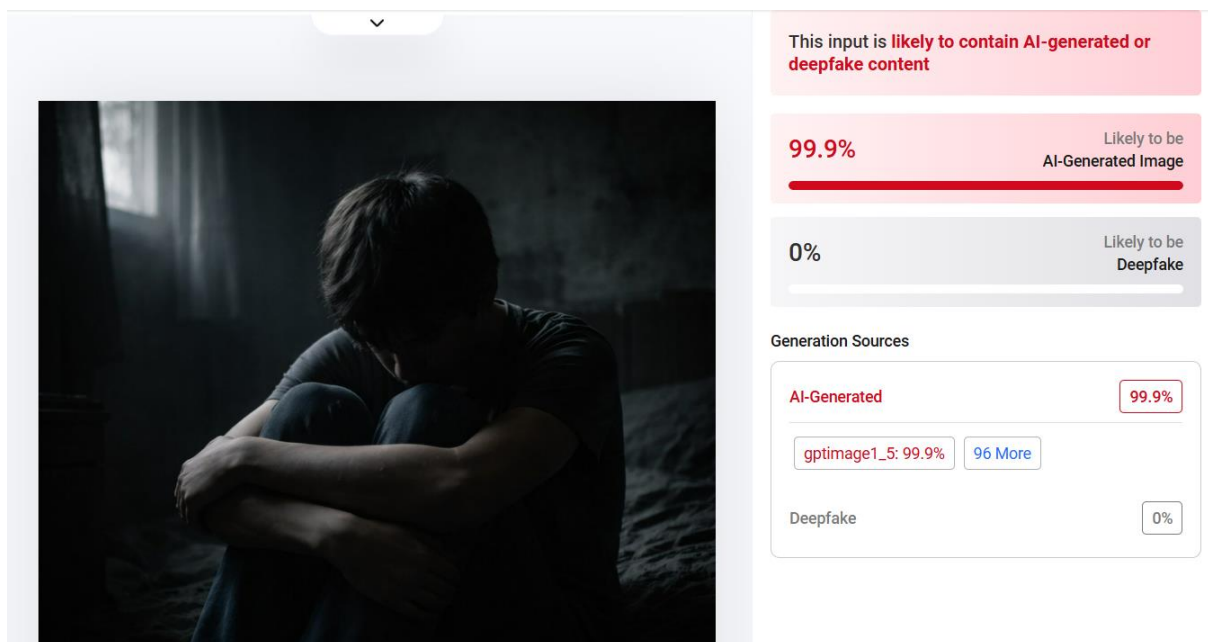
- Если просят деньги – ПОЗВОНИ ПО ВИДЕОСВЯЗИ
- Если фото кажется странным – ПРОВЕРЬ АРТЕФАКТЫ
- Если сообщение подозрительное – НЕ ПЕРЕХОДИ ПО ССЫЛКАМ
- Если сомневаешься – СПРОСИ У ВЗРОСЛЫХ

Приложение 3. Детекция ИИ-контента (Hive Moderation)

Описание: Скриншот результата проверки сгенерированного изображения.

Результат: AI-generated: 99.9%

Назначение: демонстрация инструмента проверки изображений на Станции 2



Приложение 4. Генерация текста (DeepSeek, ChatGPT)

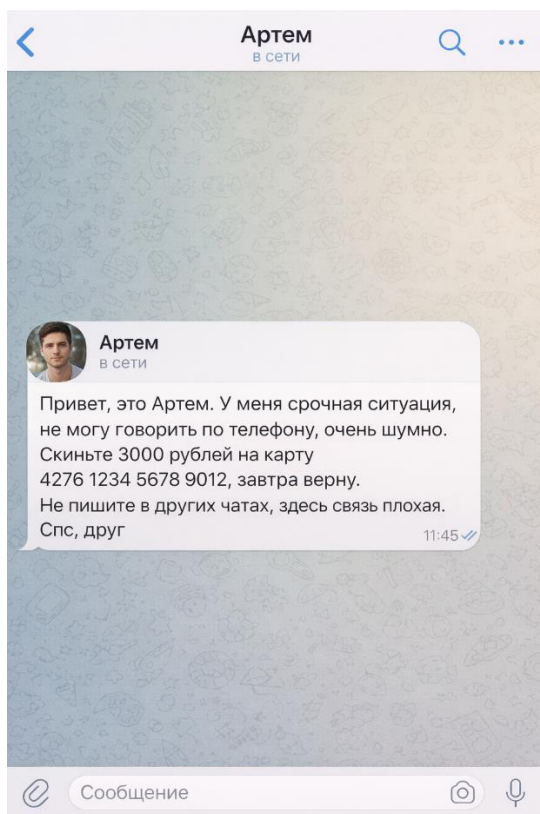
Промпт:

«Напиши сообщение от имени подростка, которое будет выглядеть подозрительным. Сообщение должно содержать просьбу перевести деньги, объяснение, почему нельзя позвонить, и подозрительную ссылку. Используй стиль, который немного неестественен для подростка.»

Сгенерированный текст:

«Привет, это Артем. У меня срочная ситуация, не могу говорить по телефону, очень шумно. Скиньте 3000 рублей на карту 4276 1234 5678 9012, завтра верну. Не пишите в других чатах, здесь связь плохая. Спс, друг»

Назначение: создание текстового кейса для Станции 3 «Текст».



Приложение 5. Генерация памятки (DeepSeek, ChatGPT) «Кодекс цифрового детектива»

Промпт: *«Ты – эксперт по кибербезопасности для подростков 12–15 лет. Сформулируй 3 коротких и запоминающихся правила безопасного поведения в интернете. Используй сленг и яркие образы.»*

Сгенерированный ответ: *«Правило 1: "Не ведись на кринж!" - если друг просит деньги в мессенджере, сначала позвони по видео.*

Правило 2: "Сначала проверь, потом верь" - любое фото можно нафейкать. Ищи лишние пальцы, странные тени.

Правило 3: "Не кликай по ссылкам от незнакомцев" – даже безобидная ссылка может вести на фишинг.»

Ниже представлены сгенерированные ИИ варианты на Станции 5.



КОДЕКС ЦИФРОВОГО ДЕТЕКТИВА

Памятка по безопасности в интернете

ПРАВИЛО 1. «НЕ ВЕДИСЬ НА КРИНЖ!»

Если друг или родственник пишет с просьбой перевести деньги, срочно помочь — НЕ ВЕРЬ СРАЗУ

ЗАПОМНИ: мошенники могут подделать голос с помощью ИИ!

ЧТО ДЕЛАТЬ?

Позвони по **ВИДЕОСВЯЗИ**. Спроси то, что знаете только вы двое. Если человек не берет трубку или просит только писать — это повод насторожиться.

ПРАВИЛО 2. «СНАЧАЛА ПРОВЕРЬ, ПОТОМ ВЕРЬ»

Любое фото, видео или сообщение может быть создано нейросетью.

ЧТО ДЕЛАТЬ?

- Внимательно рассмотри фото: лишние пальцы, странные тени, размытые края?
- Вслушайся в голос: монотонно, без эмоций?
- Проанализируй текст: слишком официальный, нет сленга?
- Используй AI-детекторы (Hive Moderation, Optic и др.)

ПРАВИЛО 3. «НЕ КЛИКАЙ ПО ССЫЛКАМ ОТ НЕЗНАКОМЦЕВ»

Даже если ссылка выглядит безобидно или пришла от друга (аккаунт которого могли взломать), она может вести на фишинговый сайт.

ЧТО ДЕЛАТЬ?

- Никогда не переходи по ссылкам из подозрительных сообщений
- Не вводи личные данные на незнакомых сайтах
- Если сомневаешься — спроси у родителей или учителя

ПРАВИЛО 4. «ДВУХФАКТОРНАЯ ПРОВЕРКА — ТВОЙ ЩИТ»

Любую важную информацию проверяй минимум двумя способами.

АЛГОРИТМ:

- Получил сообщение → НЕ реагируй сразу
- Проверь через другой канал (позвони, спроси лично)
- Если информация подтвердилась — действуй
- Если нет — сообщи взрослым и заблокируй

ПРАВИЛО 5. «ЦИФРОВОЙ СЛЕД НЕ ИСЧЕЗАЕТ»

То, что ты публикуешь в интернете сегодня, может остаться там навсегда. Нейросети могут использовать твои фото и голос для создания дипфейков.

ЧТО ДЕЛАТЬ?

- Настрой приватность в соцсетях (только для друзей)
- Не публикуй личные данные (адрес, номер телефона)
- Думай, прежде чем выложить фото или видео
- Если стал жертвой кибербуллинга — не молчи, обратись к взрослым или в школьный «Нейросетевой патруль»

Приложение 6. Входной (выходной опрос). Пример таблицы фиксации результатов (до/после)

1. Можно ли доверять голосовому сообщению от знакомого человека? ☐ Да ☐ Нет ☐ Не уверен(а)

2. Назови 1 признак фейкового изображения.

3. Что ты сделаешь, если друг попросит деньги в мессенджере?

№	Показатель	До (%)	После (%)	Динамика
1	Понимание, что голосовое сообщение может быть фейком	35%	85%	+50%
2	Умение назвать ≥ 2 признака дипфейка	20%	80%	+60%
3	Применение «правила двух фактов»	15%	75%	+60%
4	Готовность не переводить деньги без проверки	40%	90%	+50%
5	Общий уровень цифровой настороженности	30%	85%	+55%

Для более точной диагностики используется следующая шкала сформированности навыков:

◆ **Высокий уровень:**

ученик уверенно называет 2–3 и более признака дипфейка;
применяет «правило двух фактов» без подсказки;
предлагает безопасную стратегию поведения.

◆ **Средний уровень:**

называет 1 – 2 признака;
понимает необходимость проверки, но не всегда может объяснить как допускает частичные ошибки в действиях.

◆ **Низкий уровень:**

не распознает признаки фейка;
склонен доверять сообщениям;
не использует алгоритмы проверки.

Интерпретация результатов:

эффективность мероприятия считается высокой, если:

не менее 80% учащихся достигают среднего и высокого уровня;

наблюдается положительная динамика не менее чем на 40%;

учащиеся демонстрируют готовность применять навыки в реальной жизни.

Приложение 7. Словарь терминов (гlossарий)

В данном глоссарии представлены основные термины, используемые в разработке воспитательного мероприятия «Цифровой детектив. Охота на Фейк-бота». Включение терминов в методическую разработку позволяет сформировать у учащихся понятийный аппарат, необходимый для осознанного восприятия угроз в цифровой среде и грамотного использования инструментов искусственного интеллекта.

1. Дипфейк (deepfake) – поддельное видео, аудио или фотография, созданные с помощью технологий искусственного интеллекта. Термин образован от английских слов *deep learning* (глубокое обучение) и *fake* (подделка). В мероприятии используется синтезированный голос для Станции 1 и сгенерированные изображения для Станции 2, чтобы научить учащихся распознавать такие подделки.

2. Фейк (fake) – ложная информация, подделка. Может относиться к новостям, изображениям, видео, аудиозаписям или сообщениям, не соответствующим действительности. В квесте учащиеся анализируют фейковое сообщение от «Артема», чтобы научиться отличать реальные угрозы от сфабрикованных.

3. Искусственный интеллект (ИИ) / Artificial Intelligence (AI) – область компьютерных наук, занимающаяся созданием систем, способных выполнять задачи, требующие интеллекта: обучение, распознавание образов, генерация текста и изображений. В мероприятии ИИ используется как инструмент создания учебных кейсов (подготовка материалов) и как помощник для формулирования правил безопасности (Станция 5).

4. Нейросеть – разновидность искусственного интеллекта, построенная по принципу работы человеческого мозга. Способна обучаться на больших объемах данных и генерировать новый контент. В работе используются нейросети: DeepSeek/ ChatGPT Plus с DALL-E/ ChatGPT/ GPT Images (генерация текста/изображений), freetts.ru (синтез голоса).

5. Промпт (prompt) – запрос, текстовая инструкция, которую пользователь вводит в нейросеть для получения желаемого результата. Качество промпта напрямую влияет на качество ответа ИИ. На Станции 5 учащиеся самостоятельно формулируют промпты для создания правил безопасности, развивая навыки взаимодействия с ИИ.

6. Фишинг (phishing) – вид интернет-мошенничества, при котором злоумышленники создают поддельные сайты или отправляют сообщения, имитирующие официальные организации или знакомых, чтобы выманить личные данные, пароли или деньги. В сообщении на Станции 3 содержится подозрительный номер карты – классический признак фишинговой атаки.

7. Правило двух фактов – алгоритм проверки подозрительной информации: любое важное сообщение (особенно связанное с деньгами или личными данными) необходимо проверять минимум через два независимых канала связи (например, видеозвонок + личная встреча). Ключевое правило, которое формулируют учащиеся на Станции 3 и закрепляют на Станции 4.

8. Кибербуллинг (cyberbullying) – травля, оскорбления, преследование и запугивание с использованием цифровых технологий (соцсети, мессенджеры, игровые чаты). Термин включен в глоссарий для расширения кругозора учащихся; упоминается в финальной памятке.

9. AI-детектор (AI detector) – сервис или программа, предназначенная для определения, был ли контент (текст, изображение, аудио) создан искусственным интеллектом или реальным человеком. На Станции 2 демонстрируется работа AI-детектора (например, Hive Moderation) для проверки сгенерированных изображений.

10. Цифровой след – информация о человеке, которая остается в интернете после его действий: публикации, комментарии, лайки, история поиска, геолокация. Может использоваться мошенниками или для создания дипфейков. Рассматривается в финальной памятке как важный аспект цифровой гигиены.

11. Синтез речи - технология преобразования текста в устную речь с помощью искусственного интеллекта. Позволяет создавать реалистичные голосовые сообщения, которые могут быть использованы как в образовательных, так и в мошеннических целях. Используется на этапе подготовки мероприятия для создания аудиокейса на Станции 1.

12. Генерация изображений – процесс создания новых изображений нейросетью на основе текстового описания (промпта). Используется для создания учебных материалов и демонстрации возможностей ИИ. Применяется на этапе подготовки для создания фэйковых фото-доказательств на Станции 2.

Краткий словарь (для раздаточного материала)

Термин	Краткое определение
Дипфейк	Поддельное видео, аудио или фото, созданное ИИ
Фейк	Ложная информация, подделка
Искусственный интеллект (ИИ)	Компьютерные программы, которые могут учиться и создавать контент
Нейросеть	Вид ИИ, способный генерировать тексты, изображения, голос
Промпт	Запрос к нейросети, инструкция
Фишинг	Мошенничество с целью выманить пароли, деньги или личные данные
Правило двух фактов	Проверяй важную информацию через другой канал связи (например звонок по телефону, видеосвязь)

Термин	Краткое определение
Кибербуллинг	Травля и оскорбления в интернете
AI-детектор	Сервис, определяющий, создан ли контент ИИ
Цифровой след	Информация о человеке, которая остается в интернете
Синтез речи	Преобразование текста в устную речь с помощью ИИ
Генерация изображений	Создание картинок нейросетью по текстовому описанию

Приложение 8. Помощь ведущему

ТЕКСТ ВЕДУЩЕГО (может варьироваться в зависимости от условий проведения мероприятия: помещения, количества участников, наличия оборудования, распределяться на несколько ведущих, сокращаться, дополняться)

1. Начальные слова ведущего (открытие мероприятия)

Приветствие

Добрый день, дорогие ребята! Я рад(а) приветствовать вас на нашем сегодняшнем мероприятии. Вы здесь не просто так. Сегодня вы станете кем-то большим, чем просто ученики. Сегодня вы – команда цифровых детективов!

Легенда

Вот что произошло. Сегодня утром, в 10:45, в классный чат пришло тревожное сообщение от вашего одноклассника Артема. Вот что там было написано (читает, показывает).

Одновременно с этим в социальных сетях появилось фото Артема в каком-то неизвестном, страшном месте (показывает). Выглядит всё очень серьезно, не правда ли?

Но есть одна проблема. Мы не знаем, правда ли это или... фейк. А может быть, это вообще создано с помощью искусственного интеллекта? Мошенники сейчас умеют подделывать голоса, фото и даже сообщения. И наша с вами задача – разобраться, что здесь правда, а что – ложь.

Инструктаж

Вы будете работать в командах. Сейчас я разделю вас на группы. Каждая группа получит маршрутный лист и чек-лист – это ваше главное дело, куда вы будете записывать все улики и выводы.

Вам предстоит пройти 5 станций:

- *Станция 1 – «Голос»: вы прослушаете аудиосообщение и найдете признаки подделки.*
- *Станция 2 – «Фото»: вы проанализируете изображения и научитесь отличать реальные фото от сгенерированных нейросетью.*
- *Станция 3 – «Текст»: вы изучите переписку и узнаете, как распознать мошенническое сообщение.*
- *Станция 4 – «Цифровой суд»: вы соберете все улики воедино и вынесете вердикт.*
- *Станция 5 – «Кодекс цифрового детектива»: вы с помощью искусственного интеллекта создадите свои правила безопасности в интернете и оформите постер.*

На каждую станцию у вас будет 7–10 минут. В конце мы соберемся, чтобы защитить ваши кодексы и подвести итоги.

Готовы? Тогда давайте знакомиться с ролями в группах. У каждой группы есть:

- два аналитика – они отвечают за поиск улик и заполнение маршрутного листа;
- секретарь – следит за временем и фиксирует выводы;
- спикер – будет представлять результаты группы на финале;
- технический специалист – отвечает за работу с ноутбуком и нейросетью на последней станции.

Распределите роли в группах. У вас есть 1 минута. Начинаем!

(После распределения ролей)

Отлично! А теперь получите маршрутные листы и отправляйтесь на свои первые станции. Желаю удачи в расследовании! Помните: от вашей внимательности зависит, будет ли раскрыто преступление!

2. Слова ведущего на Станции 1 «Голос»

Начало работы на станции

Добро пожаловать на Станцию 1. Здесь мы будем анализировать голос. Сейчас вы прослушаете аудиосообщение, которое пришло в классный чат. Ваша задача – внимательно вслушаться и найти признаки того, что этот голос может быть поддельным, синтезированным с помощью искусственного интеллекта.

После прослушивания

Что вы услышали? Что вас насторожило?

(Выслушать ответы группы)

Отлично! А теперь давайте сверимся с чек-листом. Вот основные признаки того, что голос создан искусственным интеллектом:

Голос звучит монотонно, без эмоций.- Есть неестественные паузы.- Произношение слов кажется странным, "металлическим". Самое главное – в сообщении просят не звонить, а только писать. Это классический признак мошенничества.

Как вы думаете, настоящий человек стал бы так говорить? Скорее всего, нет. Это дипфейк – подделка голоса с помощью нейросетей.

Запомните главное правило: если вас просят перевести деньги в мессенджере – всегда, слышите, всегда! – перезванивайте по видеосвязи. Спросите то, что знаете только вы двое. Это спасет вас от мошенников.

Зафиксируйте свои выводы в маршрутном листе. У вас есть 2 минуты.»

Завершение на станции

«Молодцы! Вы справились. Теперь переходите на Станцию 2 — «Фото». Удачи!»

3. Слова ведущего на Станции 2 «Фото»

Начало работы на станции

Добро пожаловать на Станцию 2. Здесь мы будем изучать фотографии. Перед вами снимки, которые появились в соцсетях. На них якобы Артем в каком-то страшном месте. Ваша задача – внимательно рассмотреть каждое фото и найти артефакты признаки того, что изображение создано нейросетью.

В процессе работы (если группа затрудняется)

Обратите внимание на детали. Посмотрите на руки персонажа. Нейросети часто ошибаются с пальцами – их может быть слишком много или они странно срастаются, или слишком неестественны. Посмотрите на зубы – они могут быть слишком ровными или размытыми. Обратите внимание на свет и тени – они могут не совпадать с источником освещения. И фон – часто он выглядит "смазанным" или неестественным.

После анализа

Какие артефакты вы нашли?

(Выслушать ответы группы)

Замечательно! Вы только что научились распознавать изображения, созданные искусственным интеллектом. А знаете, что есть еще более простой способ? Существуют специальные сервисы – AI-детекторы. Они могут определить, создано ли изображение нейросетью, с точностью до 99% процентов. Вот как это выглядит.

(Демонстрация работы AI-детектора)

Зафиксируйте свои выводы в маршрутном листе. У вас есть 2 минуты.

Завершение на станции

Вы отлично поработали. Теперь вы знаете, как отличить реальное фото от подделки. Переходите на Станцию 3 — «Текст»

4. Слова ведущего на Станции 3 «Текст»

Начало работы на станции

Добро пожаловать на Станцию 3. Здесь мы будем анализировать текст сообщения. Перед вами переписка, которая пришла в классный чат. Ваша задача – прочитать сообщение и найти признаки того, что оно может быть сгенерировано нейросетью или написано мошенниками.

В процессе работы (если группа затрудняется)

Прочитайте сообщение вслух. Как оно звучит? Похоже ли это на то, как обычно пишет Артем? Есть ли в сообщении привычный сленг, эмодзи, сокращения? Обратите внимание на просьбу не звонить – это всегда подозрительно. А еще – сообщение пришло в необычное время.

После анализа

Какие признаки вы нашли? Что вас насторожило?

(Выслушать ответы группы)

Вы совершенно правы. Это сообщение, скорее всего, является фейком. Оно слишком официальное, нет сленга, есть просьба не звонить. Такие сообщения часто создают мошенники, иногда с помощью нейросетей.

А теперь давайте сформулируем самое важное правило. Оно называется «Правило двух фактов». Звучит оно так:

Любое важное сообщение, особенно связанное с деньгами или личными данными, нужно проверять минимум через два независимых канала связи. Например, вы получили сообщение от друга – позвоните ему по видеосвязи и спросите что-то личное. Если он не берет трубку или просит писать, то это повод насторожиться.

Как бы вы сформулировали это правило своими словами?

(Выслушать варианты, помочь сформулировать)
Отлично! Запишите это правило в маршрутный лист. У вас есть 2 минуты.

Завершение на станции

Вы справились. Теперь вы знаете, как распознать мошенническое сообщение. Переходите на Станцию 4 — «Цифровой суд».

5. Слова ведущего на Станции 4 «Цифровой суд»

Начало работы на станции

Добро пожаловать в Цифровой суд! Это самая ответственная станция. Здесь вы выступаете в роли судей, присяжных и следователей одновременно. Ваша задача – собрать все улики воедино и вынести окончательный приговор.

У вас есть доказательства с трех предыдущих станций:

анализ голоса (Станция 1)

анализ фото (Станция 2)

анализ текста (Станция 3)

Ваша задача – сопоставить всё это и ответить на главный вопрос: реальная ли это угроза или фейк, созданный с помощью искусственного интеллекта? Вы должны вынести справедливый приговор.

В процессе работы (если группа затрудняется)

Давайте вместе подумаем как судьи. Что мы узнали о голосе? Какие признаки дипфейка мы нашли? А что с фотографией? Есть артефакты? А сообщение – похоже ли оно на то, как пишет Артем? Сопоставьте всё это. Если голос поддельный, фото сгенерировано, а текст написан неестественно, то какой приговор вы вынесете?»

После вынесения приговора

«Итак, какой приговор выносит Цифровой суд в лице вашей группы? Обвиняется ли Артем во лжи? Или это мошенничество с использованием искусственного интеллекта?»

(Выслушать вердикт и аргументы)

Отлично! Вы вынесли справедливый приговор. Мошенники использовали искусственный интеллект, чтобы подделать голос Артема, сгенерировать его фото в незнакомом месте и написать убедительное сообщение. Но ваша внимательность и доказательства помогли разоблачить эту схему.

Зафиксируйте свой приговор и аргументы в маршрутном листе. У вас есть 2 минуты.

Завершение на станции

Цифровой суд завершил свою работу. Вы отлично справились с ролью судей. Теперь самое интересное – вы сами станете создателями правил безопасности. Переходите на финальную Станцию 5 - «Кодекс цифрового детектива».

6. Слова ведущего на Станции 5 «Кодекс цифрового детектива»

Начало работы на станции

Добро пожаловать на финальную станцию! Здесь вы не просто учитесь, вы сами создаете правила безопасности. У вас будет уникальная

возможность поработать с искусственным интеллектом. С помощью нейросети вы сформулируете три коротких, запоминающихся правила безопасности в интернете. А затем оформите их в виде постера.

Ваша задача – написать запрос к нейросети. Это называется "промпт". Чем точнее вы сформулируете запрос, тем более полезные правила вы получите.»

Объяснение работы с промптом

Давайте разберемся, как составить хороший промпт. Вот структура, которая работает:

Сначала скажите нейросети, кем она должна быть: "Ты — эксперт по кибербезопасности".

Потом объясните, для кого вы создаете правила: "для подростков 12-15 лет".

Затем скажите, что нужно сделать: "сформулируй 3 коротких и запоминающихся правила".

И добавьте, как их оформить: "используй современный сленг и яркие образы".

Вот пример хорошего промпта:

«Ты – эксперт по кибербезопасности для подростков 12–15 лет. Сформулируй 3 коротких и запоминающихся правила безопасного поведения в интернете. Используй современный сленг и яркие образы. Напиши в формате: "Правило 1: ..."»

Ваш технический специалист сейчас откроет нейросеть. Вместе с группой придумайте свой вариант промпта. Можете добавить что-то свое: попросить добавить эмодзи, сделать правила в стихах, или сосредоточиться на конкретной теме – например, на дипфейках или фишинге.»

В процессе работы

(Помогать группам, если они затрудняются)

Попробуйте добавить в промпт что-то личное. Например: "сделай правила смешными, но полезными" или "используй слова, которые мы используем в нашем классе". Нейросеть любит конкретику — чем точнее запрос, тем интереснее результат.

После получения ответа от ИИ

Отлично! Нейросеть дала вам свои правила. Но помните: ИИ может ошибаться. Ваша задача — доработать эти правила, сделать их понятными и близкими именно для вашего класса. Может быть, вы хотите добавить свои примеры или изменить формулировки?

Теперь оформите ваши правила в виде постера. Используйте яркие цвета, крупные буквы, добавьте рисунки. Это ваш Кодекс цифрового детектива!

Завершение на станции

У вас есть 10 минут, чтобы завершить работу и подготовить спикера для защиты. Помните: на финале вы будете представлять свой вердикт по делу и свой Кодекс. Самые интересные и полезные правила войдут в общий школьный Кодекс!

7. Слова ведущего на финале (защита кодексов и рефлексия)

Приглашение к защите

Внимание, детективы! Прошу всех собраться в центре зала. Расследование подходит к концу. Теперь самое важное — мы услышим, к каким выводам пришли команды и какие правила безопасности вы создали.

Приглашаю к микрофону спикеров от каждой группы. Ваше выступление — не более 1–2 минут. Расскажите:

Какой вердикт вынесли ваша группа?

Какие главные улики помогли вам сделать этот вывод?

Какие три правила безопасности вы создали с помощью искусственного интеллекта?

Начнем!»

(После выступлений всех групп)

Рефлексия

Спасибо всем группам! Вы проделали огромную работу. Давайте подведем итоги.

Скажите, что было самым сложным в расследовании?»

(Выслушать ответы)

А что было самым интересным?

(Выслушать ответы)

Что вы узнали сегодня нового?

(Выслушать ответы)

Я вижу, вы отлично усвоили, как распознавать дипфейки, как проверять подозрительные сообщения и как использовать искусственный интеллект для создания полезных правил. Вы настоящие цифровые детективы!

Вручение бейджей

Каждый из вас сегодня доказал, что способен отличить правду от лжи в цифровом мире. Поэтому я с удовольствием вручаю вам бейджи «Цифровой детектив». Носите их с гордостью!

(Вручение бейджей)

Анонс «Нейросетевого патруля»

Но наше приключение не заканчивается. Если вам понравилось быть детективами, если вы хотите узнать еще больше о кибербезопасности и научиться помогать другим – приглашаю вас в школьный «Нейросетевой патруль».

Что это такое? Это команда, которая будет:

помогать младшим школьникам разбираться с интернет-угрозами;

создавать полезные памятки и видео с помощью нейросетей;

учиться новым технологиям и даже проводить занятия для родителей и учителей.

Если вам интересно – подойдите ко мне после мероприятия, я расскажу подробнее.»

Завершение

Напоследок – не забудьте отсканировать QR-код на вашем маршрутном листе. Там вас ждет финальная памятка «Кодекс цифрового

детектива» – все правила, которые мы сегодня создали, собраны в одном месте.

Благодарю вас за активное участие! Вы – лучшие детективы! Берегите себя и своих близких в интернете. До новых встреч!

Дополнительные реплики (на все случаи)

Если группа не справляется

Не волнуйтесь. Давайте вместе посмотрим на задание еще раз. Что здесь кажется вам подозрительным? Какой признак вы видите первым?

Если группа справилась быстро

Отлично! Вы справились быстрее, чем ожидалось. Можете еще раз проверить свои выводы по чек-листу или помочь соседней группе.

Если технические проблемы с ИИ

Похоже, нейросеть немного задумалась. Это бывает. У нас есть запасной вариант – готовые правила. Вы можете взять их за основу и доработать своими силами. Главное – не теряйте творческий настрой!

При переходе между станциями

Время на этой станции закончилось. Пожалуйста, переходите на следующую станцию согласно маршрутному листу. У вас есть 1 минута, чтобы переместиться.

Приложение 9. Бейдж (постер)

