



STACJA 4

Lustro Kłamstw – deepfake i dezinformacja



KARTA PRACY DLA DRUŻYNY

Tematyka: Krytyczne myślenie wobec obrazów i nagrań wideo w sieci, rozpoznawanie fałszywych materiałów stworzonych przy pomocy sztucznej inteligencji.

Deepfake to technologia wykorzystująca sztuczną inteligencję (AI) do tworzenia fałszywych, ale bardzo realistycznych zdjęć, filmów i dźwięków (np. klonowania głosu). Technologia ta może służyć oszustom w działalności przestępczej np. do kradzieży tożsamości, wyłudzenia pieniędzy lub danych osobowych, ale też do cyberprzemocy.

Zadanie 1

Wyobraźcie sobie zdjęcie lub video wygenerowane przez AI, na którym Wasz ulubiony sportowiec rzekomo reklamuje podejrzaną aplikację hazardową do wygrywania pieniędzy. Jego zdjęcie zostało wygenerowane bez jego zgody.

Jakie szczegóły na zdjęciu lub filmie pomogą wam rozpoznać, że to oszustwo (deepfake)? Wymieńcie trzy wskazówki.

Zadanie 2

Jak nazywa się zjawisko polegające na celowym rozpowszechnianiu treści fałszywych lub wprowadzających w błąd? Technologia deepfake może być wykorzystywana do tworzenia takich informacji.

Odpowiedź:
.....
.....



STACJA 4

Lustro Kłamstw – deepfake i dezinformacja



KARTA MENTORA - OPIEKUNA STACJI

Rola mentora – jesteś przewodnikiem, który zadaje pytania, aby pomóc drużynie rozwiązać zadanie. Nie przekazujesz wiedzy, nie podajesz rozwiązań, ale naprowadzasz tylko pytaniami. Mentor nie krytykuje odpowiedzi drużyny, lecz zadaje jedno z pytań. Możesz wybrać je z zaproponowanej listy pytań pomocniczych. Zaliczenie stacji następuje, gdy uczeń własnymi słowami sformułuje prawidłową odpowiedź związaną z ochroną prywatności.

Instrukcja dla mentora

1. Daj drużynie 1 minutę na przeczytanie „Pigułki wiedzy”.
2. Przekaż drużynie „Kartę pracy” i włącz stoper, czas wykonania zadań to 8 minut.
3. Jeśli drużyna nie zna odpowiedzi, NIE podawaj rozwiązania. Zadawaj pytania naprowadzające oraz pytanie finałowe.
4. Po prawidłowym rozwiązaniu wręczasz drużynie fragment „Kodu odporności cyfrowej”.
5. Jeśli drużyna podjęła złą decyzję – zabierz odpowiedni „Żeton decyzji”.

Żetony decyzji: *Prawda/Odpowiedzialność*, które drużyna oddaje, jeśli udzieli złych odpowiedzi.

Fragment hasła „Kodu odporności cyfrowej”, który drużyna otrzymuje po wykonaniu zadania: „*Myślę.*”

Cel mentora: Rozwijanie samodzielnego myślenia i podejrzliwości wobec materiałów audio/wideo w sieci, nauka rozpoznawania deepfake’ów oraz reagowania na oszustwa.

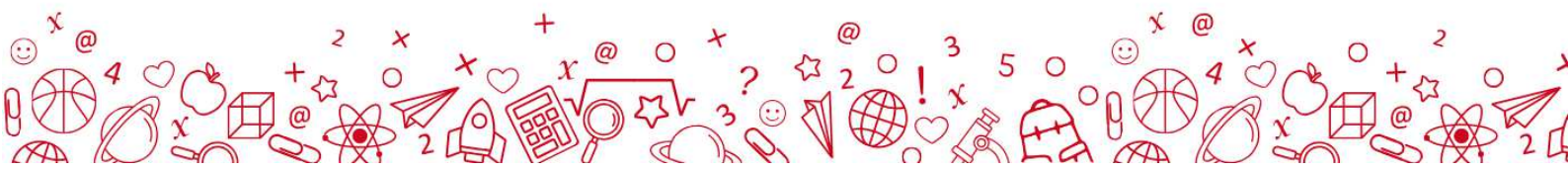
Główne zagadnienia: Krytyczne myślenie, rozpoznawanie sfałszowanych obrazów i nagrań audio/wideo (deepfake), dezinformacja (*fake news*).

Pytania naprowadzające

Czy przysłowie „zobaczyć znaczy uwierzyć” jest jeszcze prawdziwe w dzisiejszym internecie? Dlaczego?

Czym jest technologia deepfake?

Wyobraźcie sobie, że dostajecie nagranie, na którym znany sportowiec reklamuje aplikację do wygrywania pieniędzy. W jakim celu oszuści stworzyli to nagranie?



Przeglądacie się podejrzanemu filmikowi ze znanym sportowcem. Na jakie szczegóły ciała i elementy tła zwróćcie uwagę?

W jakim celu rozprzestrzenia się fałszywe wiadomości (dezinformację) i jak zachowuje się odpowiedzialny cyfrowy detektyw, gdy natrafi na szokujące nagranie?

Gdyby oszuści użyli AI do podrobienia głosu Waszej mamy lub taty i zadzwonili z prośbą o szybki przelew, jak możecie sprawdzić, czy to naprawdę oni?

Pytanie na wniosek końcowy

Co robicie przed podaniem dalej informacji, zdjęcia lub dodaniem „kciuka w górę” do sensacyjnego nagrania w sieci?



STACJA 4

Lustro Kłamstw – deepfake i dezinformacja



KLUCZ ODPOWIEDZI DLA MENTORA – OPIEKUNA STACJI

Tematyka: Krytyczne myślenie wobec obrazów i nagrań wideo w sieci, rozpoznawanie fałszywych materiałów stworzonych przy pomocy sztucznej inteligencji.

➤ Opis stacji

Deepfake to technologia wykorzystująca sztuczną inteligencję (AI) do tworzenia fałszywych, ale bardzo realistycznych zdjęć, filmów i dźwięków (np. klonowania głosu). Technologia ta może służyć oszustom w działalności przestępczej np. do kradzieży tożsamości, wyłudzenia pieniędzy lub danych osobowych, ale też do cyberprzemocy.

Zadanie 1

Wyobraźcie sobie zdjęcie lub video wygenerowane przez AI, na którym Wasz ulubiony sportowiec rzekomo reklamuje podejrzaną aplikację hazardową do wygrywania pieniędzy. Jego zdjęcie zostało wygenerowane bez jego zgody. Jakie szczegóły na zdjęciu lub filmie pomogą wam rozpoznać, że to oszustwo (deepfake)? Wymieńcie trzy wskazówki.

Klucz odpowiedzi

Należy zwrócić uwagę na: 1) nienaturalne rysy twarzy czy mimikę 2) ruchy ust, które mogą nie pokrywać się z dźwiękiem na nagraniu 3) nienaturalne ruchy ciała lub dziwne wyglądające dłonie 4) treść z błędami albo dziwnymi znakami 5) niespójne czy dziwne elementy tła oraz rozmyte litery w tle. Warto dodać, że programy AI są coraz doskonalsze i coraz trudniej rozpoznać wygenerowane sztucznie obrazy czy nagrania. Trzeba również zwracać uwagę na mało wiarygodny kontekst nagrania, czy obrazu oraz sprawdzać z wiarygodnym autorytetem czy rzetelnym źródłem wiedzy.

Zadanie 2

Jak nazywa się zjawisko polegające na celowym rozpowszechnianiu treści fałszywych lub wprowadzających w błąd? Technologia deepfake może być wykorzystywana do tworzenia takich informacji.

Klucz odpowiedzi:

Dezinformacja.

STACJA 4

Lustro Kłamstw – deepfake i dezinformacja



PIGUŁKA WIEDZY

Główne zagadnienia

Chroń swój wizerunek oraz głos i nie daj się oszukać przez narzędzia AI!

KONIEC EPOKI "ZOBACZYĆ LUB USŁYSZEĆ ZNACZY UWIERZYĆ"

- **Sztuczna inteligencja (AI)**

Narzędzie, które uczy się na podstawie danych i przykładów oraz wykonuje różne zadania np. może odpowiadać na pytania, tłumaczyć lub tworzyć teksty i obrazy. Sztuczna inteligencja nie jest człowiekiem, nie ma uczuć ani ludzkiego rozumu.

- **Deepfake**

Bardzo realistyczny, ale fałszywy obraz, film (deepfake video) lub nagranie dźwiękowe (deepfake audio) stworzone przez sztuczną inteligencję.

- **Dezinformacja**

Celowe tworzenie i rozpowszechnianie fałszywych wiadomości/informacji (*fake news*) w celu wprowadzenia w błąd, oszustwa finansowego lub nękania.

JAK ROZPOZNAĆ DEEPPFAKE?

➤ **Obrazy i nagrania wideo**

Zwróć uwagę na nienaturalną mimikę twarzy, zniekształcenia dłoni, brak naturalnego ruchu ciała, zniekształcone lub dziwne elementy tła czy ruchy ust/oczu niepasujące dokładnie do wypowiadanych słów oraz dziwne znaki i litery.

➤ **Dźwięk (klonowanie głosu)**

Do podrobienia czyjegoś głosu oszustom wystarczy zaledwie kilka sekund nagrania zdobytego z sieci.

➤ **Cyberprzemoc**

Przerabianie zdjęć lub głosu innej osoby, bez jej zgody jest łamaniem prawa. Oszuści potrafią podrobić głos bliskich i zadzwonić z prośbą o pieniądze.



ZŁOTA ZASADA STACJI 4 SUPERMOC DETEKTYWA PRYWATNOŚCI

POWIEDZENIE „ZOBACZYĆ LUB USŁYSZEĆ NIE ZNACZY UWIERZYĆ”. ZAWSZE SPRAWDZAM, CZY INFORMACJA JEST PRAWDZIWA. USTAL Z RODZICAMI TAJNE „HASŁO RODZINNE” (NP. „ZIELONY OGÓREK”), ABY W RAZIE PODEJRZANEGO TELEFONU NATYCHMIAST ZWERYFIKOWAĆ, CZY NAPRAWDĘ PO DRUGIEJ STRONIE JEST KTOŚ BLISKI.

ZATRZYMUJĘ SIĘ. MYŚLĘ. SPRAWDZAM.