Warsztaty 7: Deepfake i oszustwa AI
Przyklady fejków i czerwone flagi

Material do cwiczen. Wszystkie sytuacje sa fikcyjne.
Nie zawieraja prawdziwych nazwisk, marek, numerow telefonow ani linkow.

Cel:
Nie uczymy sie tworzyc deepfake'ow ani oszustw.
Uczymy sie zatrzymywac akcje, nazywac sygnaly ostrzegawcze i sprawdzac zrodlo.

Rytual przy kazdym przykladzie:
STOP -> zrodlo -> drugi kanal -> kod/link/pieniadze -> decyzja.

JAK PROWADZIC TO CWICZENIE

1. Pokaz albo przeczytaj przyklad.
2. Popros uczestnikow, zeby nie oceniali od razu: prawda czy falsz.
3. Zadaj trzy pytania:
   - Czego ta tresc chce ode mnie?
   - Jaka presje buduje?
   - Co sprawdzam drugim kanalem?
4. Dopiero potem pokaz czerwone flagi.
5. Na koncu wybierzcie pierwszy bezpieczny krok.

PRZYKLAD 1: GLOS BLISKIEJ OSOBY

Co uczestnik dostaje:
Krotka wiadomosc glosowa z nieznanego numeru.

Tresci nie odtwarzamy jako prawdziwego nagrania. Czytamy opis:
"Hej, to ja. Stoję przy kasie, karta mi nie przechodzi. Telefon mi pada.
Wyslij mi kod BLIK, oddam za chwile. Nie oddzwaniaj, bo nie mam czasu."

Na co zwrocic uwage:
[ ] glos brzmi znajomo, ale numer jest nieznany
[ ] prosba dotyczy kodu BLIK, czyli pieniedzy
[ ] pojawia sie zakaz oddzwaniania
[ ] jest presja czasu i zawstydzenie
[ ] brakuje potwierdzenia drugim kanalem
[ ] glos moze brzmiec plasko, zbyt rowno albo bez naturalnych pauz

Pytania do sali:
- Czy znajomy glos wystarcza do podania kodu?
- Jaki numer wybieramy do oddzwonienia?
- Co mowimy, jesli ktos naciska?

Bezpieczny krok:
Nie podaje kodu. Rozlaczam sie albo nie odpowiadam w tym kanale.
Oddzwaniam na znany wczesniej numer tej osoby.

PRZYKLAD 2: TELEFON Z INSTYTUCJI

Co uczestnik dostaje:
Telefon od osoby podajacej sie za pracownika waznej instytucji.

Opis:
Rozmowca mowi, ze konto albo sprawa urzedowa jest zagrozona. Kaze zostac na lini
i,

wejsc w przeslany link, zainstalowac aplikacje i potwierdzic dzialanie kodem.
Mowi: "Jesli sie rozlaczymy, system zamknie sprawe".

Na co zwrocic uwage:
[ ] rozmowca nie pozwala zakonczyc rozmowy
[ ] link i aplikacja przychodza od rozmowcy, nie z oficjalnego zrodla
[ ] prosba dotyczy kodu, danych albo potwierdzenia operacji
[ ] autorytet instytucji ma zmniejszyc czujnosc
[ ] nie ma czasu na samodzielne sprawdzenie numeru
[ ] glos moze brzmiec jak odczytywany skrypt

Pytania do sali:
- Dlaczego "nie rozlaczaj sie" jest czerwona flaga?
- Skad bierzemy oficjalny numer?
- Co robimy z linkiem wyslanym podczas rozmowy?

Bezpieczny krok:
Koncze rozmowe. Nie klikam linku i nie instaluje aplikacji.
Samodzielnie wpisuje adres oficjalnej strony albo dzwonie na oficjalny numer.

PRZYKLAD 3: REKLAMA INWESTYCYJNA Z "EKSPERTEM"

Co uczestnik widzi:
Reklame w mediach spolecznosciowych z osoba wygladajaca jak ekspert albo znana t
warz.

Opis:
Film obiecuje prosty zarobek i "dostep tylko dzis". Pod filmem jest przycisk
prowadzacy do formularza. Formularz prosi o numer telefonu, e-mail i kwote
pierwszej wplaty.

Na co zwrocic uwage:
[ ] obietnica szybkiego i pewnego zysku
[ ] znana twarz albo autorytet ma zastapic dowody
[ ] presja: tylko dzis, ostatnie miejsca, tajna metoda
[ ] link prowadzi poza oficjalna strone instytucji
[ ] formularz zbiera dane kontaktowe i informacje o pieniadzach
[ ] w filmie usta, glos, gesty albo rytm wypowiedzi moga nie pasowac

Pytania do sali:
- Czy znana twarz jest dowodem, ze oferta jest prawdziwa?
- Co sprawdzamy na stronie KNF albo oficjalnej stronie instytucji?
- Dlaczego podanie samego telefonu tez moze byc ryzykowne?

Bezpieczny krok:
Nie klikam reklamy. Nie zostawiam danych.
Sprawdzam ostrzezenia i oficjalne informacje samodzielnie znalezionym kanalem.

PRZYKLAD 4: SZOKUJACE WIDEO DO UDOSTEPNIENIA

Co uczestnik dostaje:
Film przeslany w komunikatorze z dopiskiem:
"Zobacz, zanim usuna. Przekaz dalej, bo media milcza."

Opis:
Na filmie osoba publiczna mowi cos bardzo ostrego i zaskakujacego.
Nagranie nie ma daty, miejsca, linku do pelnego wystapienia ani kontekstu.

Na co zwrocic uwage:
[ ] silna emocja: gniew, szok, strach albo oburzenie

[ ] presja na szybkie udostepnienie
[ ] brak daty, miejsca i pelnego nagrania
[ ] brak zrodla pierwotnego
[ ] gesty, mimika, usta i glos moga nie pasowac do siebie
[ ] krotki wycinek moze byc wyrwany z kontekstu nawet bez deepfake'u

Pytania do sali:
- Co jest tutaj silniejsze: dowod czy emocja?
- Gdzie szukamy pelnego nagrania?
- Czy dopisek "nie wiem, czy prawdziwe" wystarcza przed udostepnieniem?

Bezpieczny krok:
Nie przesylam dalej. Szukam pelnego zrodla, daty, miejsca i niezaleznego potwier
dzenia.

PRZYKLAD 5: PILNA ZBIORKA Z PODEJRZANYM LINKIEM

Co uczestnik widzi:
Poruszajacy post z obrazem, krotkim filmem i linkiem do wplaty.

Opis:
Post mowi, ze pomoc jest potrzebna natychmiast. Link prowadzi do strony,
ktora przypomina znana organizacje, ale adres jest inny. Na stronie brakuje
pelnych danych organizatora, regulaminu i niezaleznego potwierdzenia.

Na co zwrocic uwage:
[ ] bardzo silna emocja i poczucie winy
[ ] "wplac teraz" zamiast spokojnej informacji
[ ] podobny, ale nieoficjalny adres strony
[ ] brak danych organizatora i regulaminu
[ ] komentarze moga sztucznie wzmacniac presje
[ ] obraz moze byc prawdziwy, ale uzyty w falszywym kontekście

Pytania do sali:
- Jak pomagac bez klikania w link z posta?
- Co musi byc na wiarygodnej stronie zbiorki?
- Czy udostepnienie bez sprawdzenia moze komus zaszkodzic?

Bezpieczny krok:
Nie klikam linku z posta. Jesli chce pomoc, wchodze samodzielnie na oficjalna
strone organizacji albo sprawdzam zbiorke w znanym serwisie.

PRZYKLAD 6: SMS Z LINKIEM I DOPLATA

Co uczestnik dostaje:
SMS z informacja o niedoplatej kwocie, blokadzie paczki albo wygasajacej sprawie
.

Opis:
"Doplata 1,49 zl. Brak platnosci dzis spowoduje anulowanie dostawy.
Oplac tutaj: [podejrzany link usuniety z cwiczenia]"

Na co zwrocic uwage:
[ ] mala kwota ma obnizyc czujnosc
[ ] link przychodzi w wiadomosci, a nie z aplikacji lub oficjalnej strony
[ ] jest presja terminu
[ ] tekst moze miec dziwny szyk, odmiane albo interpunkcje
[ ] prosba prowadzi do platnosci albo danych karty
[ ] nadawca moze udawac kuriera, urzad albo operatora

Pytania do sali:
- Dlaczego mala kwota nadal jest ryzykiem?
- Gdzie sprawdzamy status paczki albo sprawy?
- Co robimy z podejrzanym SMS-em z linkiem?

Bezpieczny krok:
Nie klikam linku. Sprawdzam status w oficjalnej aplikacji albo na stronie wpisanej
samodzielnie. Podejrzany SMS z linkiem mozna przekazac na numer 8080.

SZYBKA LISTA: NA CO PATRZEC

1. Presja
   Teraz, szybko, ostatnia szansa, nie rozlaczaj sie, nikomu nie mow.

2. Prosba o dzialanie
   Kod, przelew, link, aplikacja, haslo, dane karty, dokument, szybka wplata.

3. Tozsamosc
   Kto mowi? Czy moge to potwierdzic poza tym kanalem?

4. Zrodlo
   Czy jest oficjalny adres, data, miejsce, pelny kontekst i regulamin?

5. Technika
   Usta, glos, gesty, tlo, rytm, pauzy, akcent, nienaturalna rownosc wypowiedzi.

6. Logika
   Czy tresc ma sens? Czy obraz, glos, data, miejsce i intencja pasuja do siebie?

7. Emocja
   Czy tresc chce mnie przestraszyc, zawstydzic, wzruszyc albo rozzloscic,
   zanim zdaze sprawdzic?

ZDANIE DO POWTARZANIA

Nie musze od razu wiedziec, czy to deepfake.
Mam zatrzymac akcje, nie podawac kodu i sprawdzic zrodlo drugim kanalem.

ZRODLA DO OMOWIENIA Z GRUPA

- CERT Polska: zglaszanie incydentow: https://incydent.cert.pl/
- CERT Polska: podejrzany SMS z linkiem na numer 8080: https://cert.pl/kontakt/
- NASK: rozpoznawanie deepfake po bledach logicznych, glosowych i jezykowych
- Policja: oszustwa BLIK i wykorzystanie AI do falszywych glosow lub wideo
- CEBRF KNF: falszywe inwestycje i reklamy wykorzystujace deepfake