Warsztaty 7: Deepfake i oszustwa AI
Prompty do weryfikacji podejrzanych tresci

ZASADA

AI moze pomoc wypisac czerwone flagi.
AI nie potwierdza, ze podejrzana tresc jest bezpieczna.
Zrodlo sprawdzam samodzielnie.

Nie wklejam:
- kodow,
- hasel,
- danych logowania,
- danych karty,
- PESEL-u,
- pelnych dokumentow,
- prywatnych rozmow innych osob,
- aktywnych linkow, ktorych nie chce klikac.

1. PODSTAWOWY PROMPT

Pomoz mi spokojnie ocenic podejrzana tresc.

Opis sytuacji:
[opisz bez danych prywatnych, bez linkow do klikania i bez kodow]

Wypisz:
1. sygnaly presji,
2. prosby o pieniadze, kod, dane albo klikniecie,
3. miejsca, gdzie ktos moze sie podszywac,
4. informacje, ktorych nie da sie potwierdzic z opisu,
5. pytania, ktore musze sprawdzic drugim kanalem.

Nie rozstrzygaj, czy to na pewno prawda albo falsz.
Na koncu podaj najbezpieczniejszy pierwszy krok bez klikania w link
i bez placenia.

2. GDY KTOS PROSI O KOD ALBO PIENIADZE

Ocen te sytuacje pod katem oszustwa.

Opis:
[opisz sytuacje bez kodow, numerow kont i danych prywatnych]

Sprawdz:
1. czy jest presja czasu,
2. czy ktos prosi o kod, przelew, dane albo aplikacje,
3. czy tozsamosc osoby jest potwierdzona,
4. jaki drugi kanal sprawdzenia jest najbezpieczniejszy,
5. czego nie powinienem/powinnam robic pod presja.

Nie podawaj instrukcji wykonania platnosci.
Podaj tylko bezpieczne kroki zatrzymania i sprawdzenia.

3. GDY WIDZE REKLAME Z OBIECUJACYM ZYSKIEM

Pomoz mi ocenic reklame, ktora obiecuje szybki zysk.

Opis reklamy:
[opisz bez klikania w link]

Wypisz:
1. obietnice zysku,
2. elementy presji,
3. czy reklama korzysta z autorytetu osoby albo instytucji,
4. jakie informacje powinny byc na oficjalnej stronie,
5. gdzie moge sprawdzic ostrzezenia samodzielnie.

Nie oceniaj atrakcyjnosci inwestycji.
Skup sie na ryzyku oszustwa i bezpiecznym pierwszym kroku.

4. GDY DOSTAJE FILM ALBO NAGRANIE

Pomoz mi sprawdzic, czy nagranie wymaga ostroznosci.

Opis nagrania:
[opisz co widzisz i slyszysz]

Wypisz:
1. sygnaly w obrazie,
2. sygnaly w glosie,
3. niespojnosci tresci,
4. brakujacy kontekst: data, miejsce, pelne zrodlo,
5. co sprawdzic przed przeslaniem dalej.

Nie rozstrzygaj pewnie, czy to deepfake.
Podaj pytania do weryfikacji i bezpieczny krok.

5. GDY CHCE OSTRZEC BLISKICH

Pomoz mi napisac spokojna wiadomosc do rodziny albo znajomych.

Cel:
Chce ostrzec przed prosbami o kod, pieniadze albo klikniecie w link,
ale bez straszenia i bez zawstydzania.

Napisz:
1. wersje krotka,
2. wersje ciepla i rodzinna,
3. jedno zdanie-zasade,
4. propozycje hasla albo procedury rodzinnej.

Ton:
spokojny, konkretny, bez paniki.

6. GDY KTOS JUZ KLIKNAL ALBO PODAL DANE

Pomoz mi ulozyc spokojny plan ograniczenia szkody.

Opis sytuacji:
[opisz bez hasel, kodow, numerow kart, PESEL-u i aktywnych linkow]

Wypisz:
1. jakie dowody warto zachowac,
2. ktore konta, karty albo hasla moga wymagac zabezpieczenia,
3. z jaka instytucja skontaktowac sie oficjalnym kanalem,
4. czego nie kasowac przed zgloszeniem,

5. jak opisac sytuacje rodzinie albo zaufanej osobie bez wstydu.

Nie podawaj instrukcji platnosci ani obchodzenia zabezpieczen.
Skup sie na zatrzymaniu szkody, zgloszeniu i bezpiecznym kontakcie.