Nowe zagrożenie wykorzystywane przez cyberprzestępców - deepfake!

Czym jest deepfake?

Deepfake to fałszywe wideo lub nagranie głosu, które wygląda bardzo realistycznie – jakby była na nim prawdziwa osoba, choć tak nie jest. Deepfake generowane są przez sztuczną inteligencję.

Przykłady:

- sfałszowany głos (np. pracownika banku, policjanta, bliskiej osoby) która wydaje polecenia dotyczące określonego działania takie jak, np. wykonanie przelewu bankowego;
- sfałszowana (podmieniona) twarz znanej osoby w nagraniu wideo, która namawia do czegoś czego ta osoba nigdy by nie powiedziała;
- fałszywe wiadomości zawarte w reklamach produktów lub usług (np. reklamy z udziałem celebrytów, polityków).

Ze względu na coraz większą dostępność rozwiązań, które wykorzystują AI cyberprzestępcy stosują je do tworzenia przekazu związanego z oszustwami finansowymi, dezinformacją czy nawet szantażem.

Schemat oszustwa z wykorzystaniem deepfake:

1. Oszust wybiera popularną osobę publiczną (np. Elon Musk, Robert Lewandowski, Andrzej Duda), której wizerunek wzbudza zaufanie lub emocje.

2. Przy użyciu sztucznej inteligencji (AI) tworzy się wideo/nagranie głosowe, w którym znana osoba zachęca do zainwestowania np. w akcje Orlen, prosi o wpłatę na „fundację", promuje fałszywy konkurs.

3. Deepfake zamieszczany jest w reklamach sponsorowanych na popularnych portalach, np. Onet, wp.pl, YouTube, Facebook/Instagram.

4. W materiale pojawia się link: „Zainwestuj teraz!", „Kliknij, aby odebrać nagrodę!, „Wspomóż akcję humanitarną!".

5. Ofiara klika i przekazuję np.. dane karty płatniczej, dane do logowania do bankowości elektronicznej

Jak rozpoznać deepfake?

Choć przestępcy coraz bardziej doskonalą się w tworzeniu fałszywych nagrań to często zawierają one zauważalne błędy takie jak:

- błędy generowania obrazu (np. niepoprawna mimika twarzy, brak spójności ruchu);
- dziwny mechaniczny głos (nieprawidłowo wygenerowany głos niepoprawnie intonuje wypowiedzi lub popełnia błędy np. złe akcentowane, nieprawidłowa odmiana liczebników);
- niedoskonałości drugiego planu (rozmazane krawędzie, błędy w zachowaniu osób w tle).

Podstawowym elementem sprawdzenia czy nagranie jest prawdziwe jest sprawdzenie źródła informacji.

Jak się chronić przed atakami?

Zasady bezpiecznego zachowania:

- nie ufaj w ciemno rozmówcy - nawet jeśli głos lub twarz wydają ci się znajome;

- nie wykonuj przelewów na podstawie rozmowy telefonicznej lub wideo - przerwij połączenie i samodzielnie skontaktuj się z rozmówcą;
- nigdy nie podawaj nikomu danych do logowania do bankowości elektronicznej, PIN-ów ani kodów SMS, nie instaluj podejrzanych aplikacji;
- nie działaj pod wpływem emocji, presji czasu czy autorytetu;
- chroń swoje dane w sieci - ograniczając udostępnianie w mediach społecznościowych zdjęć i filmów;
- ostrzeż innych - jeśli za pomocą deepfeke ktoś użył twojego wizerunku, poinformuj znajomych i rodzinę, aby nie stali się ofiarami przestępcy;
- stosuj rozwiązania wieloskładnikowe podczas logowania np. do bankowości internetowej - utrudnia to dostęp do twojego konta.