



Financované
Európskou úniou
NextGenerationEU

PLÁN [OBNOVY]

Nitrianske kompetenčné centrum kybernetickej bezpečnosti
Fakulta prírodných vied a informatiky
Univerzita Konštantína Filozofa v Nitre
Tr. A. Hlinku 1, 949 01 Nitra

AI-POWERED KYBERNETICKÉ ÚTOKY

Workshop o rizikách zneužívania umelej inteligencie
v kybernetickej bezpečnosti, so zameraním na deepfake,
AI-generovaný sociálny inžiniering
a overovanie falošného obsahu.

František Forgáč
fforgac@ukf.sk

Obsah

1	Abstrakt	4
1.1	Autori	4
2	Cieľová skupina	4
2.1	Vzdelávacie ciele	4
2.2	Východiskové predpoklady a požiadavky na účastníkov	4
2.3	Použité metódy	4
2.4	Výstupy pre účastníka	4
3	Požiadavky	5
3.1	Prístrojové vybavenie	5
3.2	Priestorové požiadavky	5
3.3	Veľkosť skupiny	5
4	Dĺžka a formát	5
5	Priebeh	5
6	Úvod - Čo je deepfake a prečo je problémom	1
6.1	Ako rozpoznať deepfake	2
6.2	Základné pravidlo	2
7	Deepfake AI incidenty	4
7.1	Incident 1: Deepfake uchádzač na online pracovnom pohovore	5
7.2	Incident 2: Deepfake reklama Ashley James na tabletky na chudnutie . . .	6
7.3	Incident 3: Klonovaný hlas a podvod s urgentnou žiadosťou o peniaze . . .	7
7.4	Incident 4: Falošné AI videá údajnej lúpeže v Louvre	8
7.5	Incident 5: AI-upravená fotografia policajného dôkazu	9
7.6	Incident 6: AI-asistovaný phishing cez Google kontext	10
7.7	Incident 7: Politické deepfaky v indických voľbách	11
7.8	Incident 8: Deepfake investičný pump-and-dump podvod	12
7.9	Incident 9: Deepfake investičné reklamy vo Švédsku	13
7.10	Incident 10: Zneužitie tvárí a hlasov influenceriek vo falošných reklamách .	14
7.11	Incident 11: Deepfake realitného makléra v romantickom podvode	15
7.12	Incident 12: Deepfake Jasona Momou v romantickom podvode	16

7.13 Incident 13: Deepfake Elton John a „push button“ online zárobok	17
7.14 Incident 14: Deepfake reklamy s Tomom Hanksom a Gayle King	18
7.15 Incident 15: Deepfake politické reklamy propagujúce falošné vládne príspevky	19
7.16 Incident 16: Deepfake Elon Musk a investičný podvod v Kanade	20
7.17 Incident 17: Pokus o impersonáciu CEO spoločnosti WPP	21
7.18 Incident 18: Deepfake Taylor Swift a falošný Le Creuset giveaway	22
7.19 Incident 19: Falošný hlas Joea Bidena v robocall kampani	23
7.20 Incident 20: Falošné LinkedIn profily s GAN fotografiami	24

8 Záver	24
----------------	-----------

Abstrakt

Workshop poskytuje prehľad o nových typoch kybernetických útokov, ktoré využívajú umelú inteligenciu. Účastníci sa oboznámia s pojmami ako adversarial AI, temná AI, deepfake či AI-generovaný sociálny inžiniering a ich vplyvom na bezpečnostné systémy.

Autori

Autor1: Mgr. František Forgáč, Phd.

Email1: fforgac@ukf.sk

Univerzita: Fakulta prírodných vied a informatiky, Univerzita Konštantína Filozofa v Nitre

Cieľová skupina

Učitelia ZŠ a SŠ, študenti stredných a vysokých škôl, IT pracovníci samosprávy a bezpečnostní analytici.

Vzdelávacie ciele

- Porozumieť štyrom typom AI-útokov (Adversarial AI, Dark AI, Deepfake, AI sociálny inžiniering)
- Naučiť sa identifikovať falošné multimédiá generované AI
- Získať prehľad o obranných stratégiách proti AI-hrozbám

Východiskové predpoklady a požiadavky na účastníkov

- Základná znalosť práce s počítačom a internetom. Predchádzajúce skúsenosti s kybernetickou bezpečnosťou nie sú potrebné.

Použité metódy

- Prednáška,
- Rozbor prípadových štúdií (rozpoznávanie fake obsahu)
- Diskusia

Výstupy pre účastníka

- Certifikát o účasti
- Prístup k materiálom
- Praktické zručnosti pri rozpoznávaní AI hrozieb a fake obsahu

Požiadavky

Prístrojové vybavenie

- Projektor, internetové pripojenie

Priestorové požiadavky

- Prednášková miestnosť

Veľkosť skupiny

- Maximálne podľa kapacity prednáškovej miestnosti

Dĺžka a formát

- Trvanie: 3 hodiny
- Forma: prezenčne

Priebeh

0:00-0:05 | Otvorenie workshopu

- predstaviť tému,
- vysvetliť, prečo sú deepfaky bezpečnostný problém,
- nastaviť pravidlo: nebudeme sa sústreďiť iba na technické chyby, ale aj na kontext a zámer.

0:05-0:20 | Čo je deepfake

- definícia deepfake,
- rozdiel medzi AI-generovaným obsahom, upraveným obsahom a deepfake obsahom,
- typy deepfake obsahu: obraz, video, hlas, text, live deepfake,
- krátke vysvetlenie, ako deepfake vzniká,
- prečo sú verejne dostupné fotografie, videá a hlasové nahrávky rizikom.

0:20-0:30 | Ako deepfake rozpoznať

- vizuálne znaky: pery, oči, mimika, okraje tváre, svetlo, ruky, objekty,
- zvukové znaky: intonácia, pauzy, emócie, nesúladi s obrazom,
- kontextové znaky: neznámy zdroj, tlak, peniaze, heslá, politická manipulácia,
- postup overovania S.T.O.P.

0:30-0:40 | Inštrukcie k práci s incidentmi Lektor vysvetlí, že účastníci budú analyzovať 20 incidentov. Pri každom incidente majú určiť:

- typ incidentu,
- komu mohol incident uškodiť,
- aké sú technické znaky AI,
- aké sú kontextové znaky podvodu,
- ako by sa obsah dal overiť,
- aké preventívne pravidlo by podobnému incidentu zabránilo.

0:40-1:30 | Analýza incidentov 1-10

- Deepfake uchádzač na online pracovnom pohovore
- Deepfake reklama Ashley James
- Klonovaný hlas a urgentná žiadosť o peniaze
- Falošné AI videá údajnej lúpeže v Louvre
- AI-upravená fotografia policajného dôkazu
- AI-asistovaný phishing cez Google kontext
- Politické deepfaky v indických voľbách
- Deepfake investičný pump-and-dump podvod
- Deepfake investičné reklamy vo Švédsku
- Zneužitie tvárí a hlasov influenceriek vo falošných reklamách

1:30-1:40 | Spoločná diskusia k incidentom 1-10

1:40-1:50 | Prestávka

1:50-2:40 | Analýza incidentov 11-20

- Deepfake realitného makléra v romantickom podvode
- Deepfake Jasona Momou v romantickom podvode
- Deepfake Elton John a „push button“ online zárobok
- Deepfake reklamy s Tomom Hanksom a Gayle King
- Deepfake politické reklamy propagujúce falošné vládne príspevky
- Deepfake Elon Musk a investičný podvod v Kanade
- Pokus o impersonáciu CEO spoločnosti WPP
- Deepfake Taylor Swift a falošný Le Creuset giveaway

- Falošný hlas Joea Bidena v robocall kampani
- Falošné LinkedIn profily s GAN fotografiami

2:40-2:50 | Spoločná diskusia k incidentom 11-20

- Ktorý incident bol najťažšie rozpoznateľný?
- Ktorý incident bol najnebezpečnejší a prečo?
- Kde technické znaky nestačili?
- Aký rozdiel je medzi deepfake zábavou, neoznačenou reklamou a podvodom?
- Ako by sme podobné incidenty vysvetlili žiakom alebo kolegom?

2:50-3:00 | Zhrnutie obranných pravidiel a záver

- Neveriť iba hlasu alebo obrazu.
- Pri peniazoch a heslách vždy overiť druhým kanálom.
- Celebrity a politici v reklamách sú podozrivý signál.
- Neoverené krízové videá nezdieľať.
- Pri investíciách ignorovať garantovaný výnos.
- Pri online pohovoroch a firemných hovoroch používať dodatočné overenie identity.
- Pri podozrivom obsahu používať S.T.O.P

Úvod - Čo je deepfake a prečo je problémom

Deepfake je forma syntetického obsahu vytvoreného pomocou umelej inteligencie. Môže ísť o video, obrázok, hlasovú nahrávku alebo text, ktorý pôsobí dôveryhodne, ale v skutočnosti zobrazuje alebo napodobňuje niečo, čo sa nikdy nestalo. Deepfake môže napríklad vytvoriť dojem, že známa osoba povedala vetu, ktorú nikdy nepovedala, že riaditeľ firmy vydal pokyn na prevod peňazí, alebo že príbuzný v telefonáte žiada urgentnú pomoc.

Deepfake technológia sa opiera o strojové učenie, generatívnu AI a neurónové siete. Systém sa učí z veľkého množstva fotografií, videí alebo hlasových nahrávok cieľovej osoby. Na základe týchto dát sa následne pokúša vytvoriť nový obsah, ktorý napodobňuje jej tvár, mimiku, pohyb, hlas alebo štýl vyjadrovania. Čím viac kvalitných verejných dát o osobe existuje, tým ľahšie môže byť vytvoriť presvedčivý podvrh.

Jednou z techník, ktoré sa pri deepfake obsahu používajú, sú generatívne adversariálne siete, známe ako GAN. Zjednodušene fungujú ako súťaž dvoch systémov. Jeden systém vytvára falošný obsah a druhý sa ho snaží odhaliť. Tým, že sa oba systémy opakovane zlepšujú, výsledný falošný obsah môže byť čoraz realistickejší. Preto sú moderné deepfaky často ťažko rozpoznatelné iba pohľadom.

Deepfake nie je iba video. Môže mať viacero foriem:

- **Obrazový deepfake:** AI-generovaná alebo upravená fotografia, napríklad falošný profil na sociálnej sieti.
- **Video deepfake:** zmanipulované video, v ktorom osoba hovorí alebo robí niečo, čo sa nestalo.
- **Hlasový deepfake:** klonovaný hlas použitý v telefonáte, hlasovej správe alebo robocalle.
- **Textový deepfake:** AI-generovaný text, ktorý napodobňuje štýl konkrétnej osoby, napríklad e-mail od nadriadeného.
- **Live deepfake:** manipulácia tváre alebo hlasu v reálnom čase počas videohovoru.

Deepfake technológia môže mať aj legitímne využitie, napríklad vo filme, vzdelávaní, dabingu, hrách alebo pri rekonštrukcii historických scén. Problém vzniká vtedy, keď sa používa bez súhlasu osoby alebo s cieľom klamať, manipulovať, vydierať, poškodiť reputáciu alebo získať peniaze.

V kybernetickej bezpečnosti sú deepfaky nebezpečné najmä preto, že zneužívajú dôveru. Ľudia sú zvyknutí veriť tomu, čo vidia a počujú. Ak útočník dokáže napodobniť tvár alebo hlas známej osoby, môže obeť ľahšie presvedčiť, aby vykonala rizikovú akciu. Môže ísť o prevod peňazí, otvorenie škodlivého odkazu, poskytnutie hesla, zdieľanie citlivých údajov alebo šírenie nepravdivej informácie.

Typické oblasti zneužitia deepfake technológií:

- finančné podvody,
- phishing a sociálny inžiniering,
- falošné investičné reklamy,
- politická manipulácia,
- poškodzovanie reputácie,

- romantické podvody,
- falošné dôkazy alebo krízové videá,
- vydieranie a nelegálny explicitný obsah,
- podvody vo videohovoroch.

Ako rozpoznať deepfake

Nie každý deepfake sa dá spoľahlivo rozpoznať voľným okom. Staršie alebo menej kvalitné deepfaky však často obsahujú viditeľné chyby. Pri videu alebo obrázku si treba všímať najmä:

Vizuálne znaky

neprirodzený pohyb pier; nesúlad medzi hlasom a pohybom úst; strnulú alebo neprirodzenú mimiku; príliš hladkú alebo rozmazanú pokožku; zvláštne okraje tváre, vlasov alebo krku; neprirodzené žmurkanie alebo pohľad; zlé osvetlenie alebo tieň; deformované ruky, prsty, zuby, okuliare alebo uši; rozdielnu kvalitu tváre a zvyšku videa; objekty, ktoré sa počas videa menia alebo miznú.

Hlasové znaky

Pri hlasovom deepfake treba sledovať:

neprirodzenú rovnomernú intonáciu; zvláštne pauzy; chýbajúce emócie; opakovanie rovnakých fráz; vyhýbanie sa konkrétnym otázkam; tlak na rýchlu reakciu; odmietanie spätného overenia cez iný kanál.

Kontextové znaky

Dôležité je, že technické chyby nemusia byť vždy prítomné. Moderný deepfake môže vyzeráť veľmi dobre. Preto treba hodnotiť aj kontext. Podozrivé je najmä to, keď obsah:

pochádza z neznámeho alebo neovereného zdroja; vyvoláva strach, hnev alebo urgentnosť; žiada peniaze, heslá alebo osobné údaje; tvrdí niečo mimoriadne bez potvrdenia z dôveryhodných zdrojov; používa známu osobu na propagáciu investície, lieku alebo darčkovej akcie; tlačí používateľa k okamžitému rozhodnutiu; odmieta overenie cez oficiálny kanál.

Základné pravidlo

Pri podozrivom obsahu sa nepýtame iba: „Vyzerá to reálne?“

Správna otázka znie: „Odkiaľ to pochádza, kto z toho profituje a čo odo mňa tento obsah chce?“

Ak video, hlas alebo správa vyvoláva silnú emóciu a zároveň vás tlačí k rýchlej akcii, treba spomaliť a overiť ho.

Postup overovania S.T.O.P.

S - Spomaliť

Nereaguj okamžite. Útočníci často vytvárajú tlak, aby obeť nemala čas premýšľať.

T - Testuj zdroj

Over, kto obsah zverejnil, či ide o oficiálny profil, dôveryhodné médium alebo neznámy účet.

O - Over inde

Hľadaj potvrdenie v ďalších zdrojoch. Pri obrázkoch použi spätné vyhľadávanie, pri videách hľadaj oficiálne vyjadrenia.

P - Potvrď druhým kanálom

Ak ide o peniaze, heslá, osobné údaje alebo interné firemné pokyny, over požiadavku cez iný komunikačný kanál.

Deepfake AI incidenty

Cieľ aktivity: Účastníci budú analyzovať reálne incidenty z AI Incident Database. Pri každom incidente určia, či ide o AI-generovaný obsah, AI-upravený obsah, deepfake, klonovaný hlas alebo AI-asistovaný sociálny inžiniering. Dôležité je, aby nehodnotili iba vzhľad obrázka alebo videa, ale aj kontext, zdroj, spôsob šírenia a požiadavky útočníka.

Incident 1: Deepfake uchádzač na online pracovnom pohovore

Odkazy

<https://www.asiatoday.co.kr/kn/view.php?key=20260319010005830>

<https://www.businessinsider.jp/article/2603-deepfake-job-interview-risk-japan/>

https://www.upi.com/Top_News/World-News/2026/03/20/nkorea-deepfake-online-hiring-interview/1741773979133/

Popis

Japonská IT firma mala počas online pracovného pohovoru komunikovať s uchádzačom, ktorý údajne použil AI-generované video na napodobnenie reálneho IT manažéra Kenbuna Yoshiiho. Podľa AIID boli pri analýze zaznamenané vizuálne a zvukové nezrovnalosti. Verejne dostupné fotografie a kariérne údaje skutočnej osoby mali byť zneužitú na vytvorenie dôveryhodnej identity.

Otázky

- Čo by ste si všímali počas online pohovoru?
- Stačí zapnutá kamera ako dôkaz identity?
- Ako by personalista overil, že osoba vo videohovore je skutočná?

Príznaky

- nesúlady pohybu pier a hlasu,
- neprirodzený pohyb tváre alebo hlavy,
- rozmazané alebo „plávajúce“ okraje tváre,
- zvláštne prechody medzi tvárou, krkom a vlasmi,
- oneskorené reakcie na otázky,
- uchádzač odmieta overenie cez iný kanál,
- identita je postavená hlavne na verejne dostupných fotkách a profile.

Odpoveď

Ide o podozrivý deepfake/identity fraud scenár. Pri online náboe nestačí vizuálna kontrola. Potrebné je overenie identity cez nezávislý kanál, kontrola referencií, prípadne živá nepredvídateľná úloha.

Incident 2: Deepfake reklama Ashley James na tabletky na chudnutie

Odkazy

<https://metro.co.uk/2026/03/14/this-morning-star-hits-back-disgraceful-fake-weight-loss-pill-claims-27429981/>

Popis

Online reklama mala použiť AI-generovanú podobu britskej moderátorky Ashley James na propagáciu tabletiiek na chudnutie. Išlo o falošné odporúčanie celebrity, pričom video malo napodobniť jej tvár aj hlas.

Otázky

- Prečo sú celebrity vhodným cieľom pre deepfake reklamy?
- Ako by ste overili, či daná osoba produkt naozaj odporúča?
- Prečo sú deepfake reklamy na zdravotné produkty obzvlášť nebezpečné?

Príznaky

- známa osoba propaguje produkt na neznámom reklamnom účte,
- odporúčanie nie je na jej oficiálnom profile,
- reklama používa silný tlak, zľavu alebo sľub rýchleho výsledku,
- hlas je podobný, ale intonácia pôsobí umelo,
- pohyb úst nesesí presne so zvukom,
- tvár má inú ostrosť alebo kvalitu než zvyšok videa.

Odpoved'

Ide o deepfake endorsement scam. Najsilnejším znakom nie je iba technická chyba vo videu, ale kombinácia neovereného zdroja, komerčného tlaku, chýbajúceho potvrdenia celebrity a podozrivého produktu.

Incident 3: Klonovaný hlas a podvod s urgentnou žiadosťou o peniaze

Odkazy

<https://www.newindianexpress.com/nation/2026/Jan/09/madhya-pradeshs-first-ai-voice-cloning-fraud-reported-in-indore-play-school-head-loses-entire-savings>

<https://the420.in/ai-voice-cloning-fraud-madhya-pradesh-indore-play-school-owner-duped/>

<https://obnews.co/Flow/News/id/13406657.html>

Popis

Majiteľka materskej školy v Indore v Indii mala prísť o 97 500 rupií po tom, čo podvodník údajne použil AI klonovanie hlasu jej príbuzného. Volajúci tvrdil, že niekto potrebuje urgentnú operáciu srdca, a obeť presvedčil na prevod peňazí cez QR kódy.

Otázky

- Aké emócie útočník využíva?
- Prečo známy hlas už nemusí byť dôkaz identity?
- Ako by ste overili podobný telefonát?

Príznaky

- urgentná zdravotná alebo finančná kríza,
- tlak na okamžitý prevod,
- platba cez QR kód alebo neznámy účet,
- volajúci nechce, aby ste zložili a zavolali späť,
- hlas znie povedome, ale používa nezvyčajné frázy,
- volajúci sa vyhýba osobným otázkam.

Odpoveď

Samotný hlas už nestačí ako dôkaz identity. Správny postup je ukončiť hovor, zavolať späť na známe číslo a overiť situáciu cez ďalšiu dôveryhodnú osobu.

Incident 4: Falošné AI videá údajnej lúpeže v Louvre

Odkazy

<https://www.scmp.com/news/world/article/3333036/ai-generated-scenes-frances-2025-louvre-heist-circulate-hong-kong>

Popis

Po lúpeži šperkov v Louvre sa online šírili AI-generované videá, ktoré údajne zobrazovali samotnú lúpež. Podľa AIID videá obsahovali morphing artefakty, miznúce objekty, čiastočné vodoznaky Sora a scény, ktoré nesedeli so skutočnou Apollo Gallery.

Otázky

- Čo by ste porovnali so skutočnými fotografiami miesta?
- Je dôležitejší vizuálny artefakt alebo nesúlad s realitou?
- Ako overiť video z krízovej udalosti?

Príznaky

- objekty sa počas videa menia alebo miznú,
- postavy a predmety sa deformujú,
- obraz má „morphing“ efekt,
- objaví sa čiastočný vodoznak generatívneho nástroja,
- architektúra nesedí so skutočným miestom,
- video pochádza zo sociálnej siete bez jasného pôvodného zdroja.

Odpoveď

Ide o AI-generované falošné videá. Overenie treba robiť porovnaním so skutočnými fotografiami miesta, oficiálnymi správami a dôveryhodnými médiami.

Incident 5: AI-upravená fotografia policajného dôkazu

Odkazy

<https://www.boston.com/news/local-news/2025/07/02/maine-police-department-apologizes-ai-altered-evidence-photo/>

Popis

Policajné oddelenie vo Westbrook v štáte Maine zverejnilo fotografiu dôkazov z drogovej razie, ktorá bola neúmyselne zmenená AI editačným nástrojom. Zmenili sa detaily balenia a vznikli falošné vizuálne prvky. Polícia neskôr použitie AI priznala a ospravedlnila sa.

Otázky

- Prečo je AI úprava dôkazovej fotografie problém aj vtedy, keď nie je úmyselná?
- Ktoré detaily by ste na dôkazovej fotografii kontrolovali?
- Ako by mala organizácia označovať AI-upravený obsah?

Príznaky

- nezmyselné texty alebo logá na obaloch,
- neprirodzené hrany predmetov,
- opakujúce sa alebo príliš vyhladené detaily,
- predmety vyzerajú ako dodatočne doplnené,
- chýba originál fotografie,
- chýba transparentný postup úpravy.

Odpoveď

Ide o AI-upravený obsah, nie klasický deepfake. Dôležité je ukázať, že AI môže narušiť dôveryhodnosť aj neúmyselne.

Incident 6: AI-asistovaný phishing cez Google kontext

Odkazy

https://www.theregister.com/2025/01/27/google_confirms_action_taken_to/

<https://www.technadu.com/google-tightens-security-after-sophisticated-phishing-scam-targets-workspace-users/572395/>

<https://cybernews.com/security/hackers-leverage-google-phone-number-subdomains/>

<https://nypost.com/2025/01/30/tech/gmail-warns-users-to-secure-accounts-after-malicious-ai-hack-confirmed/>

<https://neuron.expert/news/gmail-security-warning-for-25-billion-usersai-hack-confirmed/10847/de/>

Popis

Zakladateľ Hack Clubu Zach Latta bol cieľom sofistikovaného phishingu. Útočníci mali spoofovať číslo Google Assistant, použiť hlasové prvky na vydávanie sa za podporu Google a posilniť dôveryhodnosť e-mailom z Google Workspace domény. Útok bol zastavený, keď si obeť všimla nezrovnalosti.

Otázky

- Prečo je tento útok presvedčivejší než obyčajný phishing?
- Ktoré prvky vytvárajú dôveru?
- Ako overiť telefonát od podpory veľkej služby?

Príznaky

- legitímne vyzerajúce prvky sú použité v neštandardnom procese,
- telefonát vytvára tlak na okamžitú reakciu,
- volajúci žiada citlivý úkon mimo bežného administrátorského rozhrania,
- hlas pôsobí profesionálne, ale odpovede sú šablónové,
- drobné nezrovnalosti v odkaze, čase alebo identite volajúceho.

Odpoved'

Ide o AI-asistovaný sociálny inžiniering. Overenie musí prebehnúť cez oficiálne rozhranie služby, nie cez kontakt alebo odkaz dodaný volajúcim.

Incident 7: Politické deepfaky v indických voľbách

Odkazy

<https://www.wired.com/story/indian-elections-ai-deepfakes/>

Popis

Počas indických volieb v roku 2024 boli používané AI deepfaky na personalizované oslovovanie voličov. Voliči nemuseli vedieť, že komunikujú so syntetickou verziou politika a nie s reálnym človekom.

Otázky

- Je politický deepfake prijateľný, ak s ním kandidát súhlasí?
- Ako má byť označený syntetický volebný obsah?
- Ako overiť, či politické video pochádza z oficiálnej kampane?

Príznaky

- video je personalizované pre konkrétnu skupinu alebo jazyk,
- chýba jasné označenie, že ide o syntetický obsah,
- rovnaké gesto alebo mimika sa opakuje vo viacerých jazykových verziách,
- video nie je na oficiálnom kanáli kampane,
- neexistuje záznam z reálneho vystúpenia.

Odpoveď

Aj deepfake vytvorený so súhlasom môže byť manipulatívny, ak nie je jasne označený. Pri politickom obsahu je zásadná transparentnosť.

Incident 8: Deepfake investičný pump-and-dump podvod

Odkazy

<https://www.ynetnews.com/business/article/hkzaj8kr11>

Popis

Koordinovaná podvodná kampaň v Izraeli mala používať deepfake napodobnenia verejných osobností a celebrit, napríklad politikov, spevákov alebo technologických lídrov, na propagáciu akcie OST. Podvod sa šíril cez platené reklamy, WhatsApp skupiny a falošné investičné platformy.

Otázky

- Prečo je deepfake známej osoby účinný pri investičnom podvode?
- Ako overíte investičné odporúčanie celebrity?
- Ktoré znaky ukazujú, že nejde iba o video, ale o širší podvodný systém?

Príznaky

- známa osoba odporúča konkrétnu akciu alebo kryptomenu,
- odporúčanie nie je potvrdené na jej oficiálnom kanáli,
- reklama sľubuje rýchly alebo garantovaný zisk,
- používateľ je presmerovaný do WhatsApp/Telegram skupiny,
- existuje falošná investičná platforma,
- rovnaký scenár sa objavuje s rôznymi celebritami.

Odpoveď

Ide o AI-posilnený finančný podvod. Pri investičnom obsahu je rozhodujúce overenie mimo videa: regulátor, oficiálny profil osoby, dôveryhodné finančné médiá a odmietnutie akéhokoľvek garantovaného zisku.

Incident 9: Deepfake investičné reklamy vo Švédsku

Odkazy

<https://www.tv4.se/artikel/6YQRkByHXwnifKU4fHbRN1/finansprofiler-utnyttjas-a-v-bedragare-5000-svenskar-har-blivit-lurade>

<https://www.sverigesradio.se/artikel/smasparare-lurade-pa-halv-miljard-bedragare-ai-fejkar-finansprofiler>

<https://omni.se/profilen-rasar-mot-meta-jag-kanner-stor-frustration/a/yELAAJ>

<https://www.sverigesradio.se/artikel/thousands-of-small-savers-in-sweden-connected-out-of-millions-in-ai-powered-deepfake-investment-scams>

Popis

Vo Švédsku malo byť najmenej 5 000 investorov podvedených cez AI-generované deepfake investičné reklamy na sociálnych sieťach. Podvodníci napodobňovali známe finančné osobnosti, napríklad Gabriela Mellqvista a Güntera Mårdera, a propagovali falošné akciové tipy v pump-and-dump schéme. Straty sa uvádzajú približne 500 miliónov SEK.

Otázky

- Prečo sú známe finančné osobnosti vhodným cieľom pre deepfake podvody?
- Ako by ste overili, či investičné odporúčanie pochádza od skutočnej osoby?
- Ktoré znaky ukazujú, že nejde iba o reklamu, ale o koordinovaný podvod?

Príznaky

- známa osoba odporúča konkrétnu akciu alebo investíciu cez platenú reklamu,
- odporúčanie nie je potvrdené na oficiálnom profile danej osoby,
- reklama sľubuje rýchly zisk alebo „overený tip“,
- používateľ je presmerovaný do súkromnej skupiny alebo na falošnú investičnú platformu,
- rovnaký scenár sa opakuje s viacerými známymi osobnosťami,
- video pôsobí presvedčivo, ale hlas, mimika alebo pohyb pier nie sú úplne prirodzené.

Odpoveď

Ide o AI-posilnený investičný podvod. Pri finančných odporúčaníach treba overovať informáciu mimo videa: oficiálny profil osoby, regulátor, dôveryhodné finančné médiá a reálnu licenciu platformy.

Incident 10: Zneužitie tvárí a hlasov influenceriek vo falošných reklamách

Odkazy

<https://www.washingtonpost.com/technology/2024/03/28/ai-women-clone-ads/>

<https://www.nytimes.com/2024/05/20/world/asia/china-russia-deepfake.html>

Popis

Podvodníci mali použiť nástroje HeyGen a ElevenLabs na vytvorenie deepfake videí influenceriek Michel Janse, Olgy Loiek, Shadé Zahrai a Carrie Williams. Videá propagovali falošné alebo urážlivé produkty a správy. V niektorých prípadoch bola zneužitá aj cudzia hlasová stopa.

Otázky

- Ako by influencer zistil, že jeho tvár alebo hlas boli zneužitý?
- Prečo je kombinácia tváre a hlasu presvedčivejšia než samotný obrázok?
- Ako môže takýto deepfake poškodiť reputáciu obete?

Príznaky

- osoba hovorí veci, ktoré nezodpovedajú jej bežnému štýlu alebo hodnotám,
- video sa objaví na cudzom účte alebo v reklame bez kontextu,
- hlas znie podobne, ale intonácia je plochá alebo neprirodzená,
- tvár pôsobí vložené do záberu,
- pohyb pier nesedí so zvukom,
- chýba potvrdenie na oficiálnom profile osoby.

Odpoved'

Ide o deepfake zneužitie identity. Odhalenie je často založené na kontexte: kde bolo video zverejnené, či ho osoba potvrdila a či obsah zodpovedá jej bežnej komunikácii.

Incident 11: Deepfake realitného makléra v romantickom podvode

Odkazy

<https://www.local10.com/news/local/2025/04/22/deepfake-videos-of-miami-beach-man-used-to-catfish-woman-in-uk/>

<https://www.local10.com/espanol/2025/04/22/videos-deepfake-de-un-hombre-de-miami-beach-fueron-usados-para-estafar-sentimentalmente-a-una-mujer-en-reino-unido/>

Popis

Podvodník mal použiť podobu realitného makléra z Miami, Andresa Asiona, na romantický podvod voči žene v Spojenom kráľovstve. Využitie mali byť falošné videá s jeho tvárou a klonovaným hlasom. Obeť komunikovala s podvodníkom približne rok a až po ceste do Miami sa ukázalo, že reálny Asion s tým nemal nič spoločné.

Otázky

- Prečo je romantický podvod s videom presvedčivejší než podvod cez text?
- Aké overenie by malo prebehnúť pred tým, než obeť pošle peniaze?
- Prečo môže byť verejne známa osoba alebo podnikateľ vhodnou identitou pre podvodníka?

Príznaky

- vzťah sa rozvíja online, ale osobné stretnutie sa stále odkladá,
- osoba žiada peniaze, darčkové karty alebo investície,
- videá sú krátke, jednostranné alebo nereagujú prirodzene na otázky,
- hlas a pohyb pier nemusia byť úplne synchronizované,
- komunikácia je emocionálne intenzívna a tlačí na dôveru,
- identita je založená na verejne dostupných fotografiách a videách.

Odpoveď

Ide o deepfake romance scam. Pri online vzťahoch je dôležité overiť osobu cez živý videohovor, nepredvídateľné otázky a nezávislý kontakt. Žiadosť o peniaze je zásadný varovný signál.

Incident 12: Deepfake Jasona Momou v romantickom podvode

Odkazy

<https://www.wisbechstandard.co.uk/news/25522148.cambridgeshire-widow-homeless-scammers-took-500-000/>

<https://www.thesun.co.uk/news/37478449/widow-jason-momoa-ai-videos-scam/>

<https://www.cryptopolitan.com/british-widow-loses-600k-to-ai-scam/>

<https://cryptorank.io/news/feed/b9f0d-british-widow-loses-600k-to-ai-scam>

<https://www.ibtimes.co.uk/jason-momoa-cons-widow-out-500k-british-granny-loses-everything-ai-deepfake-aquaman-1759047>

Popis

Britská vdova mala prísť o viac než 600 000 USD v romantickom podvode, kde podvodníci údajne použili AI-generované videá herca Jasona Momou. Obeť uverila, že komunikuje so známym hercom, predala svoj dom a posielala peniaze.

Otázky

- Prečo je predstava vzťahu so známou osobnosťou rizikový scenár?
- Aké správanie by malo okamžite vyvolať podozrenie?
- Ako overiť, či známa osoba skutočne komunikuje so súkromnou osobou?

Príznaky

- celebrita kontaktuje bežného človeka cez súkromný účet,
- vzťah sa rýchlo stáva emocionálne intenzívnym,
- osoba žiada peniaze alebo finančnú pomoc,
- videá sú krátke, bez prirodzenej interakcie,
- profil nie je oficiálne overený,
- výhovorky bránia reálnemu stretnutiu alebo otvorenému živému hovoru.

Odpoveď

Ide o deepfake romantický a finančný podvod. Reálna celebrita nebude žiadať súkromnú osobu o peniaze cez neoficiálny účet. Overenie musí prebehnúť cez oficiálne kanály.

Incident 13: Deepfake Elton John a „push button“ online zárobok

Odkazy

<https://www.news5cleveland.com/money/consumer/dont-waste-your-money/deepfake-elton-john-video-scams-northeast-ohio-man-out-of-20-000>

Popis

Muž zo severovýchodného Ohia bol údajne nalákaný deepfake videom napodobňujúcim Eltona Johna do podvodu s jednoduchým online zárobkom. Podvod sľuboval príjem z internetového obchodu. Obeť bola následne vedená k otvoreniu účtov a kreditných kariet, pričom vznikli poplatky približne 20 000 USD.

Otázky

- Prečo známa osobnosť zvyšuje dôveryhodnosť podvodu?
- Aké sľuby sú typické pre podvody s rýchlym zárobkom?
- Čo by ste overili pred otvorením účtov alebo kariet?

Príznaky

- známa osoba propaguje jednoduchý spôsob zárobku,
- reklama používa frázy typu „stačí kliknúť“ alebo „bez skúseností“,
- používateľ má otvoriť účty, karty alebo poskytnúť údaje,
- video je len vstupná brána do komunikácie s „poradcami“,
- chýba overiteľná firma a jasné obchodné podmienky,
- propagácia nie je na oficiálnom profile celebrity.

Odpoveď

Ide o deepfake scam kombinovaný so sociálnym inžinierstvom. Kľúčovým znakom je nereálny sľub jednoduchého zisku a následná požiadavka na finančné kroky.

Incident 14: Deepfake reklamy s Tomom Hanksom a Gayle King

Odkazy

<https://www.nytimes.com/2023/10/02/technology/tom-hanks-ai-dental-video.html>

<https://www.usatoday.com/story/entertainment/celebrities/2023/10/02/tom-hanks-ai-advertisement-without-permission/71030195007/>

<https://www.engadget.com/tom-hanks-calls-out-dental-ad-for-using-ai-likeness-of-him-161548459.html>

<https://www.dailymail.co.uk/news/article-12585775/Gayle-King-AI-deepfake-weight-loss.html>

<https://www.theguardian.com/film/2023/oct/02/tom-hanks-dental-ad-ai-version-fake>

Popis

Reklamy mali použiť AI verzie známych osobností, napríklad Toma Hanksa a Gayle King, bez ich súhlasu. Tom Hanks upozornil, že reklama na dentálny plán s jeho podobou je AI fake. Gayle King upozornila na zmanipulované video propagujúce službu na chudnutie.

Otázky

- Prečo sú zdravotné a dentálne služby častým cieľom falošných reklám?
- Ako zistíte, či celebrita naozaj podporuje produkt?
- Čo je silnejší signál: technická chyba vo videu alebo chýbajúce potvrdenie osoby?

Príznaky

- reklama používa známu tvár, ale nie je na oficiálnom kanáli osoby,
- produkt je zdravotný, finančný alebo „zázračný“,
- video je krátke a orientované na okamžité kliknutie,
- hlas alebo mimika pôsobia neprirodzene,
- chýbajú dôveryhodné informácie o poskytovateľovi,
- známa osoba neskôr upozorní, že reklama je falošná.

Odpoveď

Ide o deepfake reklamy bez súhlasu osoby. Pri zdravotných službách je nutné overiť poskytovateľa, licenciu, oficiálnu stránku a vyjadrenie celebrity.

Incident 15: Deepfake politické reklamy propagujúce falošné vládne príspevky

Odkazy

<https://www.nytimes.com/2025/10/01/technology/facebook-political-ads-deepfakes-scams.html>

<https://www.techtransparencyproject.org/articles/meta-awash-in-deepfake-scams-ads>

<https://www.japantimes.co.jp/news/2025/10/02/world/politics/us-deepfake-political-ads-meta/>

Popis

Vyšetovanie Tech Transparency Project identifikovalo deepfake reklamy na Facebooku, ktoré napodobňovali Donalda Trumpa, Elona Muska, Alexandriu Ocasio-Cortez, Elizabeth Warren, Bernieho Sandersa a Karoline Leavitt. Reklamy propagovali falošné vládne príspevky alebo „rebates“ vo výške 5 000 USD.

Otázky

- Prečo je spojenie politika a štátneho príspevku presvedčivé?
- Ako overiť, či štátna dávka alebo príspevok naozaj existuje?
- Prečo sú seniori a zraniteľné skupiny častým cieľom takýchto reklám?

Príznaky

- politik alebo verejná osoba oznamuje finančný benefit cez reklamu,
- reklama smeruje na neoficiálnu stránku,
- používateľ má rýchlo vyplniť osobné údaje,
- text sľubuje konkrétnu sumu bez jasných podmienok,
- chýba stránka oficiálnej vládnej inštitúcie,
- rovnaký podvod používa rôzne politické osobnosti.

Odpoveď

Ide o deepfake reklamy kombinované s podvodom na osobné údaje. Reálne štátne príspevky sa overujú cez oficiálne domény a úrady, nie cez reklamu na sociálnej sieti.

Incident 16: Deepfake Elon Musk a investičný podvod v Kanade

Odkazy

<https://www.ctvnews.ca/ottawa/article/all-our-dreams-are-gone-ottawa-couple-scammed-out-of-177k/>

https://youtu.be/V-Mk79wSIJA?si=o5TVXjAi0TS0DW_v

Popis

Dôchodcovský pár z Ottawy prišiel o 177 023 CAD po tom, čo ich nalákalo údajné deepfake video Elona Muska propagujúce falošnú investičnú platformu. Podvod pokračoval telefonátmi, vzdialeným prístupom do počítača a kryptomenovými prevodmi.

Otázky

- Prečo je meno Elona Muska často používané v investičných podvodoch?
- Ktorý moment by mal byť jasným varovaním?
- Prečo je vzdialený prístup do počítača extrémne rizikový?

Príznaky

- známa technologická osobnosť propaguje investíciu,
- platforma sľubuje vysoký alebo garantovaný výnos,
- používateľ má zadať kontaktné údaje a čakať na telefonát,
- „poradca“ žiada vzdialený prístup do počítača,
- peniaze sa posielajú cez kryptomeny,
- výber peňazí je komplikovaný alebo nemožný.

Odpoveď

Ide o deepfake investičný podvod. Najsilnejšie varovné znaky sú garantovaný výnos, kryptomenové prevody a vzdialený prístup do zariadenia.

Incident 17: Pokus o impersonáciu CEO spoločnosti WPP

Odkazy

<https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam>

<https://www.ft.com/content/308c42af-2bf8-47e4-a360-517d5391b0b0>

<https://www.designrush.com/news/wpp-ceo-addresses-deepfake-scam-attempt>

<https://nypost.com/2024/05/10/business/ceo-of-wpp-falls-victim-to-deepfake-scam/>

<https://www.telegraph.co.uk/business/2024/05/10/wpp-boss-mark-read-impersonated-ai-in-elaborate-deepfake/>

Popis

Podvodníci sa pokúsili napodobniť CEO spoločnosti WPP Marka Readu. Použili falošný WhatsApp účet s jeho fotografiou, AI klonovaný hlas a verejné zábery z YouTube na zorganizovanie Microsoft Teams stretnutia. Cieľom bolo získať peniaze a osobné údaje. Útok bol neúspešný.

Otázky

- Prečo je kombinácia WhatsApp, Teams a známej fotografie presvedčivá?
- Čo by mal zamestnanec overiť pred vykonaním pokynu od vedenia?
- Aké interné pravidlo by útok zastavilo?

Príznaky

- vysokopostavený manažér komunikuje z nového alebo neštandardného účtu,
- požiadavka sa týka peňazí, novej firmy alebo citlivých údajov,
- komunikácia sa presúva medzi viacerými kanálmi,
- hlas znie podobne, ale interakcia je strojená,
- tlak na dôvernosť alebo rýchlosť,
- chýba štandardné interné schválenie.

Odpoveď

Ide o neúspešný AI-asistovaný firemný podvod. Obrana: spätné overenie cez známy kontakt, dvojité schvaľovanie a pravidlo, že fotka a hlas v online hovore nie sú dôkaz identity.

Incident 18: Deepfake Taylor Swift a falošný Le Creuset giveaway

Odkazy

<https://malwaretips.com/blogs/ree-drummond-le-creuset-giveaway/>

<https://bartlesvilleradio.com/pages/news/405352023/ree-drummond-warns-of-scam>

<https://www.newson6.com/story/658e070673910c065af946ff/pioneer-woman-ree-drummond-warns-of-artificial-intelligence-scam->

<https://malwaretips.com/blogs/le-creuset-cookware-giveaway-scam/>

<https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html>

Popis

Podvodníci používali deepfake videá Taylor Swift a ďalších celebrit na propagáciu falošných darčkových akcií na riad Le Creuset. Reklamy tvrdili, že používatelia môžu získať drahý riad zadarmo alebo za malý poplatok za dopravu. Obete mohli byť následne prihlásené do drahých predplatných služieb.

Otázky

- Prečo je „darček zadarmo“ účinný podvodný prvok?
- Ako overíte, či značka skutočne organizuje giveaway?
- Čo je podozrivé na platbe za „iba dopravu“?

Príznaky

- známa osoba rozdáva drahý produkt cez reklamu,
- akcia nie je potvrdená značkou Le Creuset,
- používateľ má vyplniť dotazník alebo zaplatiť malý poplatok,
- reklama vytvára tlak „zásoby sa mieniajú“,
- hlas celebrity je syntetický alebo neprirodzene plynulý,
- stránka má skryté predplatné alebo nejasné obchodné podmienky.

Odpoveď

Ide o deepfake giveaway scam. Lacný poplatok za dopravu je často vstup do drahšieho podvodu. Overenie musí prebehnúť na oficiálnom profile značky alebo celebrity.

Incident 19: Falošný hlas Joea Bidena v robocall kampani

Odkazy

<https://www.theguardian.com/us-news/2024/jan/22/biden-fake-robocalls-new-hampshire>

<https://www.nbcnews.com/politics/2024-election/fake-joe-biden-robocall-tells-new-hampshire-democrats-not-vote-tuesday-rcna134984>

<https://www.reuters.com/world/us/fake-biden-robocall-not-expected-affect-nh-primary-official-2024-01-23/>

<https://cyberscoop.com/biden-new-hampshire-robo-call-deepfake/>

<https://cyberscoop.com/lingo-life-biden-ai-robocall/>

Popis

Pred primárkami v New Hampshire v roku 2024 boli voličom rozosielané robocally s AI-generovaným hlasom prezidenta Joea Bidena. Hlas vyzýval demokratických voličov, aby nehlasovali v primárkach. Vyšetrovanie neskôr spojilo incident s politickým konzultantom a tvorcom AI hlasu.

Otázky

- Prečo je hlas verejnej osoby v telefonáte silný nástroj manipulácie?
- Ako by mal volič overiť politický telefonát?
- Prečo je tento incident nebezpečný pre voľby?

Príznaky

- automatický telefonát vyzýva ľudí, aby nehlasovali,
- správa vyvoláva zmätok alebo mení volebný postup,
- chýba potvrdenie z oficiálnej kampane alebo úradu,
- hlas znie známo, ale obsah je nečakaný,
- caller ID môže byť spoofované,
- správa sa šíri krátko pred voľbami.

Odpoveď

Ide o AI-generovaný hlas použitý na volebnú manipuláciu. Pri voľbách treba overovať informácie cez oficiálnu volebnú komisiu, štátne úrady a dôveryhodné médiá.

Incident 20: Falošné LinkedIn profily s GAN fotografiami

Odkazy

<https://www.npr.org/2022/03/27/1088140809/fake-linkedin-profiles>

<https://www.siliconrepublic.com/machines/research-1000-computer-generated-images-linkedin-deepfake>

<https://gizmodo.com/move-over-global-disinformation-campaigns-deepfakes-ha-1848716481>

<https://www.thequint.com/tech-and-auto/tech-news/deepfakes-invade-linkedin-delhi-firm-offers-ready-to-use-ai-made-profiles>

Popis

Výskumníci zo Stanfordu upozornili LinkedIn na viac než tisíc neautentických profilov, ktoré používali fotografie pravdepodobne generované pomocou GAN. Mnohé profily boli odstránené pre porušovanie pravidiel proti falošným profilom a nepravdivým informáciám.

Otázky

- Prečo sú falošné LinkedIn profily užitočné pre útočníkov?
- Ako rozpoznať AI-generovanú profilovú fotografiu?
- Aké riziko predstavuje prijatie kontaktu od falošného profilu?

Príznaky

- tvár je dokonale vycentrovaná a oči sú v podobnej polohe ako pri GAN obrázkoch,
- pozadie je neurčité alebo rozmazané,
- náušnice, okuliare, zuby alebo vlasy sú asymetrické,
- profil má málo reálnych aktivít,
- pracovná história je všeobecná alebo nekonzistentná,
- profil rýchlo smeruje ku komerčnej ponuke, odkazu alebo získavaniu informácií.

Odpoveď

Ide o AI-generované alebo AI-asistované falošné identity. Pri profesijných sieťach treba overovať históriu profilu, spoločné kontakty, firemnú e-mailovú adresu a konzistentnosť pracovnej stopy.

Záver

Workshop účastníkom poskytol komplexný pohľad na to, ako môže byť umelá inteligencia zneužitá na rôzne formy kybernetických útokov. Diskutované boli štyri hlavné oblasti:

- Adversarial AI/ML, ktorá sa snaží cielene manipulovať výstupy AI systémov,
- Temná AI, ktorá umožňuje automatizované vyhľadávanie zraniteľností a škodlivé rozhodovanie,
- Deepfake, ktorá využíva AI na tvorbu falošných, avšak veľmi realistických multimediálnych výstupov,
- AI-generovaný sociálny inžiniering, ktorý napodobňuje ľudskú komunikáciu s cieľom vylákať citlivé informácie.

Účastníci si vyskúšali aj praktické techniky na identifikáciu falošných videí a spoznali možnosti, ako rozpoznať manipulatívny obsah generovaný AI.

Záverom workshopu je uvedomenie si, že AI nie je len nástrojom pre pokrok, ale aj potenciálnym rizikom, ak sa dostane do nesprávnych rúk. Prevencia, vzdelávanie a rozvoj kritického myslenia sú kľúčom k obrane pred týmito hrozbami.

Účastníkom sa odporúča naďalej sledovať vývoj v oblasti umelej inteligencie, špeciálne v oblasti bezpečnostných opatrení, techník odhaľovania deepfake obsahu a nástrojov na detekciu podvodných komunikácií. Priebežné vzdelávanie a tréning sú najlepšou obranou proti čoraz sofistikovanejším formám AI-útokov.

Tento dokument bol vypracovaný pre projekt Nitrianske kompetenčné centrum pre kybernetickú bezpečnosť (17R05-04-V01-00003)

