



Coping Strategies Against Information Disorder

Sprievodca

Usmernenia pre pracovníkov s mládežou



**Spolufinancovaný
Európskou úniou**

Financované Európskou úniou. Vyjadrené názory a stanoviská sú však len názorami autora (autorov) a nemusia nevyhnutne odrážať názory a stanoviská Európskej únie alebo Európskej výkonnej agentúry pre vzdelávanie a kultúru (EACEA). Európska únia ani EACEA za ne nemôžu niesť zodpovednosť.

Autori:

Eliane Smits van Waesberghe & Tim Paulusse - Verwey-Jonker Instituut (hlavní editori)

Leen D'Haenens & Joyce Vissenberg - KU Leuven

Tzvetalina Genova - Medzinárodný inštitút manažmentu

Wolfgang Eisenreich - Wissenschaftsinitiative Niederösterreich

Sonja Bercko Eisenreich - Integra Institute

Alenka Valjašková - QUALED

Pantelis Balaouras - Connexions

Vyhlasenie o autorských právach:



Toto dielo je licencované licenciou Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Môžete:

- zdieľať - kopírovať a šíriť materiál na akomkoľvek médiu alebo v akomkoľvek formáte
- adaptovať - remixovať, transformovať a rozširovať materiál

za týchto podmienok:

- Uvádzanie autora - musíte uviesť príslušné údaje, poskytnúť odkaz na licenciu a uviesť, či boli vykonané zmeny. Môžete tak urobiť akýmkoľvek primeraným spôsobom, ale nie spôsobom, ktorý by naznačoval, že poskytovateľ licencie schvaľuje vás alebo vaše použitie.
- Nekomerčné - materiál nesmiete používať na komerčné účely.
- ShareAlike - Ak materiál remixujete, transformujete alebo ho rozširujete, musíte svoj príspevok šíriť pod rovnakou licenciou ako originál.

Kapitola 3

Technológie a nástroje

Cieľová skupina

Tieto usmernenia sú zamerané na tzv. "pracovníkov prvého kontaktu". Ide o zastrešujúci pojem pre všetky osoby, ktoré sú v priamom kontakte s osobami ohrozenými informačnými hrozbami, so zameraním na skupiny v odbornom vzdelávaní a tréningu. Príklady ľudí, ktorí spadajú pod tento zastrešujúci pojem, sú: vychovávatelia, učitelia, školitelia, poradcovia a konzultanti pre mládež, sociálni pracovníci a pracovníci s mládežou. Tento zoznam však nie je úplný. Do rozsahu tohto projektu patria aj ďalšie osoby pracujúce v oblasti vzdelávania, kultúry, sociálnej alebo zdravotnej starostlivosti.

3.1 Úvod.....	1
3.2 Vyhľadávače a algoritmy	3
3.3 Online stratégie dezinformačných a extrémistických organizácií	4
Najbežnejšie metódy	4
Trolling a doxxing	5
Začlenenie dezinformácií do hlavného prúdu.....	6
Vplyvné osoby.....	6
Irónia, satira a mémy	6
3.4 Manipulovaný obsah.....	8
Porozumenie deepfakes: syntetická mediálna manipulácia.....	9
Odhaľovanie a zmiernovanie deepfake videí: technologické prístupy ..	10
3.5 Referencie.....	11

3.1 Úvod

Koncept dezinformácií (vymyslených správ) bol diskutovaný v predchádzajúcich kapitolách.

Správy zvyčajne šíria profesionálni poskytovatelia spravodajstva vrátane verejnoprávnych médií, komerčných spravodajských médií, nezávislej profesionálnej žurnalistiky alebo amatérski používatelia (v prípade platforiem sociálnych médií). Správy sú prístupné a dostupné v rôznych formátoch:

- **Digitálne správy:** Digitálne správy: správy šírené prostredníctvom internetových kanálov vo formáte digitálnych médií (text, obrázky, zvuk, video).
- **Tlačové médiá:** Noviny a časopisy obsahujúce textový a obrazový obsah.
- **Vysielanie:** Televízia a rozhlas s video a audio obsahom.

V tejto časti sa zameriavame predovšetkým na prvý formát, digitálne správy. Predložené informácie sa však môžu vzťahovať aj na iné formáty, ktoré využívajú video, zvuk a obrázky.

Digitálne spravodajstvo je prístupné prevažne prostredníctvom spravodajských platforiem prevádzkovaných profesionálnymi poskytovateľmi spravodajstva. Používatelia môžu k týmto spravodajským platformám pristupovať buď bezplatne, alebo prostredníctvom predplatených služieb. Tieto platformy ponúkajú webové kanály v rámci populárnych prehliadačov a spravodajské kanály na platformách sociálnych médií, čo používateľom umožňuje dostávať personalizované aktualizácie správ.

Spravodajský kanál je webová stránka alebo obrazovka, ktorá sa často aktualizuje a zobrazuje najnovšie správy alebo informácie. Personalizované informačné kanály sú služby integrované do webových prehliadačov (webové kanály) a platforiem sociálnych médií, ktoré používateľom poskytujú správy na základe ich osobných preferencií. Tieto preferencie sa určujú na základe rôznych faktorov, ako je prihlásenie sa na odber webových kanálov, návšteva konkrétnych webových stránok a iné.

Na platformách sociálnych médií používatelia dostávajú aj zdieľané médiá od iných používateľov, ktoré môžu obsahovať prípady dezinformácií.

Personalizované správy sa zobrazujú v prehliadači používateľa alebo na platforme sociálnych médií prostredníctvom služieb a algoritmov sociálnych médií. Používatelia si nemusia byť plne vedomí toho, ako tieto algoritmy fungujú, pretože zostáva nejasné, či je výber správ založený výlučne na preferenciách používateľa alebo na iných kritériách. Napríklad platformy sociálnych médií prezentujú správy na základe poskytovateľov správ, ktorých používateľ sleduje, čitateľských návykov jeho priateľov alebo nedávno kliknutých článkov, čo môže tvoriť určitý druh preferencií. Následne sa zoznam správ filtruje, čo znamená, že sa nezobrazujú všetky správy, ale skôr tie, ktoré algoritmus považuje za najzaujímavejšie pre používateľa. Mnohí tvrdia, že sa tak vytvára informačná bublina, v ktorej sú správy vyberané algoritmom a nemusia obsahovať správy, ktoré používateľa zaujímajú, ale algoritmus ich nezradí. Preto je dôležité nespoliehať sa len na osobné informačné kanály, ale radšej navštevovať profesionálne platformy poskytovateľov správ,

ktoré sú dôveryhodné. Vždy je dôležité kriticky hodnotiť dôveryhodnosť a presnosť správ bez ohľadu na formát alebo spôsob ich poskytovania.

3.2 Vyhľadávače a algoritmy

Mnohé webové stránky na internete sa snažia udržať svojich používateľov v kontakte a maximalizovať ich čas strávený na platforme. Platí to najmä pre platformy sociálnych médií, ktoré využívajú rôzne stratégie na zlepšenie používateľského zážitku. Jednou z takýchto stratégií je zobrazovať používateľom obsah, ktorý zodpovedá ich záujmom. Webové stránky však nemôžu čítať myšlienky, aby poznali preferencie používateľov. Na riešenie tohto problému sa používajú algoritmy na analýzu údajov používateľov a poskytovanie personalizovaného obsahu.

Algoritmy sú zložité vzorce, ktoré sledujú, merajú a vypočítavajú preferencie obsahu jednotlivca. Môžu zahŕňať faktory, ako je čas sledovania konkrétnych typov videí používateľom, čas strávený na konkrétnom príspevku alebo akcie zapojenia, ako je zanechanie lajkov alebo komentárov. Analýzou týchto údajov algoritmy určujú typ obsahu, ktorý používateľov udrží v pozornosti dlhší čas. Tieto informácie sa potom používajú na spravovanie a odporúčanie podobného obsahu používateľovi.

Hoci sa tento prístup môže zdať logický a neškodný, používanie algoritmov má aj svoje potenciálne nevýhody. Uprednostňovanie obsahu, ktorý používateľ považuje za zaujímavý, môže vytvoriť filtrované informačné bubliny, v ktorých sú používatelia vystavení len určitým názorom. Filtrované bubliny obmedzujú rozmanitosť obsahu, čo môže viesť k vzniku alebo k posilneniu existujúcich bublín.

Okrem filtrovaných bublín môžu algoritmy zosilňovať extrémnosť obsahu odporúčaním čoraz viac špecializovaných, okrajových a extrémnych príspevkov. Cieľom je udržať používateľov v kontakte, ale to môže viesť k tomu, že používatelia budú presmerovaní do online priestoru bez rôznorodých názorov, čo vyústí do tzv. králičej nory.

Filtrované bubliny a králičie nory vystavujú používateľov radikálnemu obsahu a komunitám, ktoré sú s ním spojené. Tieto online priestory poskytujú živnú pôdu pre rozvoj, rast, skresľovanie a šírenie misinformácií a dezinformácií.

Ako používatelia postupujú králičou dierou, extrémne výroky a nepravdivé informácie sa normalizujú. Táto normalizácia ďalej uľahčuje šírenie a prijímanie misinformácií a dezinformácií z radikálnych zdrojov.

Pochopením vplyvu algoritmov na skúsenosti používateľov môžeme lepšie pochopiť riziká spojené s bublinami, komorami ozvien a králičími dierami. Tieto poznatky sú nevyhnutné na orientáciu v online prostredí a na riešenie výziev, ktoré prinášajú informačné hrozby.

3.3 Online stratégie dezinformačných a extrémistických organizácií

V predchádzajúcich častiach sme skúmali, ako sa informačný neporiadok vytvára a udržiava v rámci komôr ozvien a filtrovaných názorových bublín. Ako sa však nepravdivé informácie dostávajú k ľuďom mimo týchto priestorov?

Najbežnejšie metódy

Misinformácie a dezinformácie sa môžu šíriť prostredníctvom nespočetných foriem komunikácie. Najčastejšie sa však šíria rôznymi kanálmi, ako sú platformy sociálnych médií, webové stránky, e-mail a ústne podanie. Najčastejšie metódy, ktorými sa šíria informačné hrozby, úzko súvisia so siedmimi kategóriami problematického obsahu, ktoré sú uvedené v *Kapitole 1: Pochopenie "falošných správ"*:

- **Satira alebo paródia:** Niektoré misinformácie a dezinformácie sú vytvorené na zábavné účely alebo ako satira, ale môžu byť nesprávne interpretované ako skutočné správy. Satirické webové stránky alebo účty na sociálnych sieťach môžu uverejňovať humorné alebo prehnané príbehy, ale čitatelia, ktorí si neuvedomujú ich satirický charakter, ich môžu mylne považovať za faktické informácie.
- **Clickbait:** články, ktoré obsahujú dezinformácie, často používajú senzačné alebo zavádzajúce titulky, aby upútali pozornosť a získali viac kliknutí alebo zobrazení. Ich cieľom je využiť zvedavosť alebo emócie ľudí na zvýšenie návštevnosti webových stránok a generovanie príjmov prostredníctvom reklamy.
- **Skreslené informácie:** Ide o skresľovanie alebo dezinterpretovanie skutočných správ selektívnym uvádzaním faktov, zavádzaním alebo vynechávaním kľúčových informácií. Môže ísť o vytrhávanie výrokov z kontextu, úpravu obrázkov alebo videí alebo prekrúcanie významu príbehu tak, aby zodpovedal konkrétnemu príbehu.
- **Vydávanie sa za niekoho:** dezinformácie môžu zahŕňať aj vydávanie sa za renomované spravodajské zdroje alebo verejné osobnosti, aby sa nepravdivým informáciám dodala dôveryhodnosť. To sa môže uskutočniť vytvorením falošných webových stránok alebo účtov na sociálnych médiách, ktoré napodobňujú legitímne zdroje, čím čitateľov oklamú a presvedčia ich, že informácie sú dôveryhodné.
- **Politická manipulácia:** Politická manipulácia: dezinformácia sa niekedy vytvára alebo šíri s cieľom ovplyvniť verejnú mienku alebo voľby. Môže ísť o šírenie nepravdivých informácií o politických kandidátoch, manipuláciu nálad verejnosti alebo využívanie existujúcich predsudkov a rozdelenia spoločnosti.

- **Výmysly:** misinformácie a dezinformácie môžu byť úplne vymyslené a nemajú žiaden základ v realite. Zahŕňa vytváranie falošných príbehov, citátov alebo udalostí s cieľom zavádzať čitateľov alebo divákov.
- **Zosilnenie prostredníctvom sociálnych médií:** Platformy sociálnych médií zohrávajú významnú úlohu pri šírení dezinformácií. Falošné príbehy sa môžu rýchlo stať virálnymi, pretože používatelia ich zdieľajú a opätovne uverejňujú, často bez overenia presnosti informácií. Algoritmy používané platformami sociálnych médií môžu tiež prispieť k posilneniu tým, že podporujú obsah založený skôr na angažovanosti než na presnosti.

Existujú aj ďalšie bežné metódy šírenia dezinformácií, o ktorých by sa malo hovoriť podrobnejšie: trolling, doxxing a mainstreaming. To je uvedené v nasledujúcich dvoch podkapitolách.

Trolling a doxxing

Jednou z často používaných stratégií je trolling. Trolling je definovaný ako zámerné používanie nezdvorilosti, agresivity, klamstva a manipulácie v online komunikácii s cieľom vyvolať konflikt alebo pobavenie. Trollovia podnecujú online konflikty klamaním, manipuláciou alebo agresivitou. Vykoľajújú konverzácie pre vlastné pobavenie, čím sa v podstate dopúšťajú digitálneho šikanovania.

V malom meradle sa trolling môže javiť ako relatívne neškodný, pretože je len nepríjemný. Keď však organizované skupiny trollov majú spoločný konkrétny cieľ, táto nepríjemnosť sa môže rýchlo zmeniť na dezinformačnú epidémiu. Príkladom je využívanie trollov v sociálnych médiách Ruskom ako "zbrane". Rusko použilo rozsiahlu sieť trollov na globálne šírenie dezinformácií vo viacerých jazykoch s cieľom kontrolovať online diskurz okolo Ruska. Títo trollovia nielenže šíрили nepravdivé informácie, ale zamierovali sa aj na používateľov sociálnych médií s príspevkami, ktoré sa odchyľovali od naratívu, ktorý mali propagovať. V dôsledku toho sa mnohí používatelia sociálnych médií zdržali diskusií o Rusku, čím účinne umožnili trollom kontrolovať naratív pomocou svojich misinformácií.

Doxxing, ďalšia forma internetového šikanovania, zahŕňa zverejňovanie osobných údajov alebo identity osôb online bez ich súhlasu. Hoci táto taktika nezahŕňa konkrétne šírenie dezinformácií, je to ďalšia stratégia, ktorú internetoví trollovia používajú na kontrolu rozprávania o konkrétnej téme, podobne ako trolling. Doxxing sa môže používať na zastrašovanie používateľov sociálnych médií, čím sa potláča ich ochota zverejňovať obsah, ktorý je v rozpore s preferovaným naratívom trola.

Pochopenie vplyvu trollingu a doxxingu je kľúčové pre rozpoznanie rôznych taktík používaných na manipuláciu a kontrolu online naratívov. Tieto stratégie nielenže prispievajú k šíreniu nepravdivých informácií, ale predstavujú aj výzvu pri podpore otvoreného a informovaného digitálneho prostredia.

Začlenenie dezinformácií do hlavného prúdu

Dôležitou stratégiou šírenia dezinformácií a extrémistického obsahu je ich normalizácia alebo "mainstreaming". V tomto procese zohráva kľúčovú úlohu ich zverejňovanie. Vystavenie misinformáciám a dezinformáciám môže viesť k pretrvávajúcim mylným predstavám ľudí o konkrétnych témach, čo normalizuje falošné predstavy v ich myšliach. K tomuto vystaveniu môže dôjsť v rôznych formách.

Vplyvné osoby

Jednou z bežných foriem vystavenia je šírenie informácií medzi jednotlivcami, keď sa jednotlivci delia o informácie s ostatnými. K tomu môže dôjsť prostredníctvom osobných interakcií alebo vo väčšom rozsahu cez vplyvné osoby na sociálnych sieťach. Influenceri, ktorí majú značný dosah v rôznych skupinách, môžu nevedomky alebo zámerne zdieľať nepravdivé informácie, čo má vplyv na veľký počet jednotlivcov. Takéto rozsiahle pôsobenie vedie k normalizácii dezinformácií medzi rôznorodým publikom.

Irónia, satira a mémy

Extrémistickí jednotlivci a organizácie často používajú humor, satiru a iróniu na šírenie svojich myšlienok.

Satira môže byť mocným nástrojom na spochybnenie útlakom operujúcich ideológií, zmenu naratívov alebo normalizáciu špecifických názorov v rámci hlavného prúdu. V oblasti misinformácií sa satira využíva v rôznom spektre. Parodické webové stránky ako The Onion alebo De Speld uverejňujú nefaktický obsah na humorné účely bez zámeru klamať verejnosť. Niektorí jednotlivci a skupiny však využívajú satiru a iróniu so zlým úmyslom zdiskreditovať mainstreamovú žurnalistiku, vedu alebo propagovať extrémistické myšlienky a dezinformácie. Využívaním satiry a humoru sa takýto obsah stáva prístupnejším a prijateľnejším v politickom diskurze, čím sa vystavuje širšiemu publiku.

Extrémistický obsah často oslovuje mladých ľudí ako forma zábavy alebo hľadania senzácií. Mladí ľudia, ktorých poháňa hľadanie zmyslu, majú tendenciu vyhľadávať intenzívne a nové zážitky, čím sú náchyľnejší na extrémistické myšlienky a s nimi spojené dezinformácie.

Mémy, ktoré sú široko zdieľaným humorným kultúrnym obsahom, slúžia ako ďalší spôsob šírenia extrémistických ideológií. Mémy (memečká) majú rôzne formáty vrátane obrázkov, videí, zvukových klipov, emotikonov a symbolov. Hoci mémy same osebe nie sú vo svojej podstate škodlivé, extrémisti ich využívajú na normalizáciu svojich myšlienok. Hravá povaha mémov umožňuje extrémistom zamaskovať, vyvrátiť alebo poprieť škodlivosť ich posolstiev. Takýto "ostrý" alebo provokatívny obsah sa stáva prijateľnejším a v prípade konfrontácie s obvineniami zo sexizmu, rasizmu alebo xenofóbie ho tvorcovia môžu ľahko odmietnuť ako "len vtip". Toto stieranie hraníc medzi hravou zlomyseľnosťou a problematickým obsahom vytvára nejednoznačnosť, vďaka ktorej je náročné rozoznať nevinné vtipy od extrémistických

posolstiev. Internetovú kreslenú postavičku Pepe the Frog, ktorá bola pôvodne vytvorená ako neškodný vtip, si privlastnili online beloší rasistmi. To spôsobilo zmätok medzi používateľmi internetu, pretože extrémistické iterácie tohto meme sa ocitli v zmesi s neškodnými. K normalizácii extrémistického obsahu dochádza, keď je týmto posolstvom vystavených viac jednotlivcov, čím sa stierajú hranice medzi tým, čo je prijateľné a čo nie.

3.4 Manipulovaný obsah

Z technického hľadiska sú všetky informácie alebo "správy" v médiách kombináciou textu, obrazu, zvuku a videa. Problém však spočíva v určení, či sú informácie autentické alebo nie. Je potrebné poznamenať, že dezinformácie môžu využívať pravé obrázky, ale manipulovať s príbehom a skresľovať skutočné fakty.

V minulosti sa všeobecne chápalo, že text môže napísať ktokoľvek, zatiaľ čo obrázky, zvuky a videá sa považovali za viac či menej autentické a na ich úpravu boli potrebné profesionálne zručnosti. S nedávnym technologickým pokrokom však aj obrázky, audio a video môžu podliehať manipulácii. Môžu to dosiahnuť profesionáli alebo aplikácie využívajúce systémy umelej inteligencie, ako je napríklad technológia deepfake. V dôsledku toho je potrebné rozlišovať, či bol zvuk alebo video skutočne zachytené mikrofónom alebo videokamerou, alebo či je výsledkom odbornej úpravy alebo systémov umelej inteligencie (generatívna umelá inteligencia a syntetické médiá: hlasové klony, **deepfake** videá). Okrem toho by malo byť technicky možné identifikovať pôvodný zdroj, výrobcu alebo vydavateľa obrazového, zvukového alebo video zdroja. Dôvodom je, že zdroje sa môžu zdieľať, kopírovať alebo opätovne šíriť mnohokrát v rámci celosvetového webu a sociálnych médií. Preto môže byť pre bežných používateľov náročné identifikovať pôvodný zdroj a výrobcu, aj keď majú podozrenie, že správa môže byť dezinformáciou.

Umožniť používateľom rozlišovať medzi skutočnými a vymyslenými správami si vyžaduje niekoľko krokov. Viac informácií o tom nájdete v *kapitole 2: Činnosti a zručnosti*. Odporúča sa prečítať si túto kapitolu, aby ste sa dozvedeli zložité detaily, ktoré stoja za rozpoznávaním falošných informácií. Tu však uvádzame krátke, veľmi zjednodušené zhrnutie:

- **Krok 1: Zvýšenie informovanosti používateľov:** Používatelia si musia byť vedomí, že správy môžu byť vymyslené. Uskutočňovanie osvetových aktivít je kľúčové na informovanie používateľov o tom, čo predstavujú vymyslené správy a ako sa môžu chrániť pred ich dôsledkami.
- **Krok 2: Overenie spoľahlivosti vydavateľa:** Zvýšené povedomie o vymyslených správach núti používateľov spochybňovať spoľahlivosť zdrojov a vydavateľov správ. Je nevyhnutné zvážiť formát média, či už ide o televízny kanál, časopis, noviny (online alebo tlačené) alebo platformu sociálnych médií. Mediálne kanály, ktoré umožňujú jednoduché zdieľanie alebo ďalšie šírenie správ, môžu byť menej spoľahlivé. Naopak, mediálne kanály, ktoré uľahčujú identifikáciu a overovanie zdrojov, bývajú spoľahlivejšie.

Pokiaľ ide o správy šírené cez internet, poskytovatelia služieb, ako sú informačné kanály a siete sociálnych médií, by mali využívať nové technológie na overovanie spoľahlivosti zdrojov a sledovanie pôvodného vydavateľa a zdroja. Technológia blockchain je jednou z takýchto technológií, ktorá môže tieto snahy uľahčiť.

Dodržiavaním týchto krokov a využívaním technológií môžu používatelia získať možnosť orientovať sa v digitálnom prostredí, rozlišovať skutočné správy od dezinformácií a prijímať informované rozhodnutia o informáciách, s ktorými sa stretávajú.

Porozumenie deepfakes: syntetická mediálna manipulácia

Podľa definície v slovníku Cambridge Dictionary ide o " *videozáznamy alebo zvukové nahrávky, ktoré nahrádzajú tvár alebo hlas niekoho iného tak, aby sa zdali byť skutočné* ".

V článku "Deepfake explained" z roku 2020 autorka Meredith Somers uvádza, že "(*deepfake*) sa vzťahuje na špecifický druh syntetických médií, kde je osoba na obrázku alebo videu nahradená podobizňou inej osoby". Ďalej sa vysvetľuje, že "termín 'deepfake' po prvýkrát vymyslel koncom roka 2017 používateľ Redditu s rovnakým menom. Tento používateľ vytvoril na internetovej stránke na zhromažďovanie správ a informácií priestor, kde zdieľal pornografické videá, ktoré využívali technológiu výmeny tvárí s otvoreným zdrojovým kódom".

Deepfake našli uplatnenie v rôznych odvetviach a používajú sa na rôzne účely. Medzi niektoré významné príklady patria:

- **Vydieranie:** deepfake podvrhy možno použiť na vytvorenie falošného usvedčujúceho materiálu, čo môže viesť k vydieraniu. Okrem toho, keďže je čoraz ťažšie rozlíšiť deepfakes od pravého obsahu, obeť skutočného vydierania môžu tvrdiť, že dôkazy sú falošné, čo im zabezpečí hodnoverné popretie.
- **Pornografia:** deepfake pornografia získala na internete značný význam. V správe holandského startupu Deeptrace, ktorý sa zaoberá kybernetickou bezpečnosťou, sa odhaduje, že približne 96 % všetkých online deepfakes bolo pornografických.
- **Politika:** Deepfake sa využívajú na skresľovanie známych politikov vo videách, šírenie dezinformácií a manipuláciu s verejným mienením. Medzi príklady patria deepfake videá, v ktorých vystupujú Barack Obama, Donald Trump, Volodymyr Zelenskyj a Vladimir Putin.
- **Herectvo/filmy:** Špekuluje sa o využití deepfake na vytvorenie digitálnych hercov v budúcich filmoch. Hoci digitálne vytvorenie alebo pozmenenie ľudí sa už vo filmoch objavilo, deepfake by mohli prispieť k novým pokrokom v tejto oblasti.
- **Sociálne médiá:** Používatelia využívajú deepfake na rôznych platformách sociálnych médií. Jednotlivci nahrádzajú tváre v populárnych filmových alebo seriálových scénach svojimi vlastnými, čím vytvárajú personalizované videá. Platformy ako Facebook prijali opatrenia na odhalenie a označenie deepfake ako falošných, čím sa znížila ich priorita v kanáloch používateľov.

Odhaľovanie a zmierňovanie deepfake videí: technologické prístupy

Výskumníci aktívne skúmajú metódy na odhaľovanie a riešenie problému falošného zvuku a videa. Využívajú sa rôzne prístupy:

- **Algoritmická detekcia:** Jeden z prístupov zahŕňa vývoj algoritmov, ktoré dokážu identifikovať manipulovaný obsah. Tieto algoritmy analyzujú rôzne vizuálne a zvukové podnety s cieľom odhaliť nezrovnalosti alebo anomálie, ktoré naznačujú prítomnosť deepfake. Využitím techník strojového učenia a umelej inteligencie môžu tieto algoritmy časom zlepšiť svoje detekčné schopnosti.
- **Technológia blockchain:** Ďalšia technika navrhuje využitie technológie blockchain na overenie zdroja médií. Blockchain je "digitálna účtovná kniha", ktorá zaznamenáva transakcie v sieti počítačov bezpečným, transparentným a voči manipulácii odolným spôsobom. Využíva decentralizáciu a kryptografiu na zabezpečenie dôveryhodnosti bez potreby centrálného orgánu. V tomto scenári by videá pred zobrazením na platformách sociálnych médií museli prejsť overením prostredníctvom blockchainovej "účtovnej knihy". Zabezpečením toho, aby sa schvaľovali len videá z dôveryhodných zdrojov, by sa mohlo obmedziť šírenie potenciálne škodlivých deepfake médií.
- **Digitálne podpisy:** Niektorí navrhujú digitálne podpisovanie všetkých videí a snímok zachytených kamerami vrátane kamier smartfónov ako prostriedok boja proti falošným snímkam. Išlo by o priradenie jedinečných digitálnych podpisov ku každému médiu, čo by umožnilo vystopovať každú fotografiu alebo video až k pôvodnému vlastníkovi. Hoci tento prístup môže pomôcť pri sledovaní pôvodu obsahu, existujú obavy týkajúce sa možného zneužitia, ako napríklad zameranie sa na disidentov alebo porušenie súkromia.

3.5 Referencie

- Aro, J. (2016). The Cyberspace War: Propaganda and Trolling as Warfare Tools. *European View*, 15(1), 121–132. <https://doi.org/10.1007/s12290-016-0395-5>
- Cambridge English Dictionary: Meanings & Definitions*. (2023). <https://dictionary.cambridge.org/dictionary/english>
- Daniels, J. (2018). The Algorithmic Rise of the “Alt-Right.” *Contexts*, 17(1), 60–65. <https://doi.org/10.1177/1536504218766547>
- Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association*, 43(2), 97–116. <https://doi.org/10.1080/23808985.2019.1602782>
- Greene. (2019). “Deplorable” Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies. *Studies in American Humor*, 5(1), 31–69. <https://doi.org/10.5325/studamerhumor.5.1.0031>
- Hardaker, C. (2013). “Uh. . . not to be nitpicky,,,,,but. . .the past tense of drag is dragged, not drug.” *Journal of Language Aggression and Conflict*, 1(1), 58–86. <https://doi.org/10.1075/jlac.1.1.04har>
- IED. (2022, August 23). *How Do Social Media Algorithms Work*. Institute of Entrepreneurship Development. <https://ied.eu/blog/technology-blog/how-do-social-media-algorithms-work/>
- Johnson, D., & Johnson, A. (2023, June 15). What are deepfakes? How fake AI-powered audio and video warps our perception of reality. *Business Insider*. <https://www.businessinsider.com/guides/tech/what-is-deepfake?international=true&r=US&IR=T>
- Levy, G., & Razin, R. (2019). Echo Chambers and Their Effects on Economic and Political Outcomes. *Annual Review of Economics*, 11, 303–328. <https://doi.org/10.1146/annurev-economics-080218-030343>
- Lewis, B., & Marwick, A. E. (2017). Media Manipulation and Disinformation Online. *Data & Society Research Institute*. <https://www.posiel.com/wp-content/uploads/2016/08/Media-Manipulation-and-Disinformation-Online-1.pdf>

- McNealy, J. (2015). Readers react negatively to disclosure of poster's identity. *Newspaper Research Journal*, 38(3).
<https://doi.org/10.1177/0739532917722977>
- Munn, L. (2019). Alt-right pipeline: Individual journeys to extremism online. *First Monday*. <https://doi.org/10.5210/fm.v24i6.10108>
- Sample, I. (2020, January 13). What are deepfakes – and how can you spot them? *The Guardian*. <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>
- Schumpe, B. M., Bélanger, J. J., Moyano, M., & Nisa, C. F. (2020). The role of sensation seeking in political violence: An extension of the Significance Quest Theory. *Journal of Personality and Social Psychology*, 118(4), 743–761.
<https://doi.org/10.1037/pspp0000223>
- Seth, S. (2023, September 11). The World's Top 10 News Media Companies. *Investopedia*. <https://www.investopedia.com/stock-analysis/021815/worlds-top-ten-news-companies-nws-gci-trco-nyt.aspx>
- Somers, M. (2020, July 21). Deepfakes, explained. *MIT Sloan*.
<https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining “Fake news”: A Typology of Scholarly Definitions. *Digital Journalism*, 6(2), 137–153.
<https://doi.org/10.1080/21670811.2017.1360143>
- Van Puffelen, M. (2021). Rechtsextremisme: Geweld met een rechtsextremistisch motie. In *DSP-groep*. DSP-groep. <https://www.dsp-groep.nl/wp-content/uploads/18MP-Rechtsextremisme-DSP-2021.pdf>
- Van Wonderen, R. (2023). *Rechts-extremistische Radicalisering op Sociale Media Platformen*. Verwey-Jonker Instituut.
- Van Wonderen, R. (2023). *Richtlijn / onderbouwing Radicalisering*. Verwey-Jonker Instituut.
- Van Wonderen, R. & Peeters, M. (2021). *Werken aan weerbaarheid tegen desinformatie en eenzijdige meningsvorming. Evaluatie lesprogramma Under Pressure*. Utrecht: Verwey-Jonker Instituut. https://www.verwey-jonker.nl/wp-content/uploads/2022/04/120550_Werken-aan-weerbaarheid-tegen-desinformatie-eenzijdige-meningsvorming.pdf.

Wasike, B. (2022). When the influencer says jump! How influencer signaling affects engagement with COVID-19 misinformation. *Social Science & Medicine*, 315, 115497. <https://doi.org/10.1016/j.socscimed.2022.115497>

Wolfowicz, M., Weisburd, D., & Hasisi, B. (2021). Examining the interactive effects of the filter bubble and the echo chamber on radicalization. *Journal of Experimental Criminology*, 19(1), 119–141. <https://doi.org/10.1007/s11292-021-09471-0>