FUNKY CITIZENS

Kit de verificare

Partea V

AI și Dezinformarea

Inteligența artificială are potențialul de a fi cel mai bun prieten al dezinformării, însă, paradoxal, poate fi și cel mai bun prieten al oamenilor care se străduiesc să combată dezinformarea. În lumea digitală în care trăim, instrumentele de inteligență artificială pot fi considerate o sabie cu două tăișuri.

Pe de o parte, acestea pot **amplifica dezinformarea**, ducând la răspândirea ei mai rapidă și la o scală mai mare. Pe de altă parte, pot fi folosite pentru **a detecta și a combate dezinformarea**, ajutându-ne să ne protejăm împotriva acestei amenințări digitale.

Așadar, depinde de oameni cum aleg să folosească aceste instrumente și care este scopul lor final. Trăim într-un moment în care tehnologia în sine nu este nici bună, nici rea, ci este doar un instrument care poate fi folosit în ambele moduri.

Ce este deepfake?

Pentru a explica mai simplu, gândiți-vă la deepfake ca la o modalitate de a crea videoclipuri sau înregistrări audio care par să prezinte oameni vorbind sau acționând în moduri care nu au avut loc niciodată în realitate. Tehnologia este atât de avansată încât poate face ca aceste conținuturi să arate foarte autentice, folosind fețe și voci care pot părea familiare în mai puțin de 5 minute.

Trebuie să fie o preocupare semnificativă deoarece pot manipula conținut vizual și audio pentru a crea aparențe realiste, dar false.

În contextul alegerilor, acestea pot avea consecințe masive asupra proceselor democratice și pot destabiliza încrederea oamenilor în democrație.

Deepfake în România

În România, fenomenul deepfake-urilor este în creștere. Chiar dacă nu au fost folosite într-un mod abuziv pentru a influența alegerile sau pentru a influența societatea, acestea au fost folosite pentru **fraude financiare**. Atât timp cât deepfake-urile deja există, nu trebuie să nu ne pregătim pentru cele mai negre scenarii și prin care să influențeze comportamentele electorale ale oamenilor.

În România, **site-uri de știri sunt copiate și falsificate** pentru a promova conținut sub egida acestor branduri, ceea ce subminează încrederea publicului în sursele legitime de informație. Aceste metode sofisticate de manipulare a informației reprezintă un pericol semnificativ pentru integritatea informațiilor publice și necesită o vigilență sporită din partea consumatorilor de media.

La ce ne putem aștepta și care sunt mizele?

Manipularea discursurilor politice: Ne putem aștepta la fabricarea de videoclipuri false cu politicieni care fac declarații sau susțin poziții pe care aceștia nu le-au avut niciodată. Deepfake-urile pot fi folosite pentru a schimba discursul public și pentru a distrage atenția de la probleme reale prin crearea de controverse false sau scandaluri. **Miza**: afectează opiniile și deciziile alegătorilor cu privire la un subiect sau o persoană și poate schimba agenda politică.

Crearea de dezinformare despre candidați: Se pot fabrica evenimente sau situații compromițătoare cu anumiți candidați. **Miza:** decredibilizează candidații în ochii alegătorilor.

Imitarea surselor de încredere: Liderii de opinie sau știrile pot fi supuse la deepfake. **Miza**: oferă credibilitate acestui tip de manipulare și influențează alegătorii.

Slovacia: Cu 2 zile înaintea votului pentru alegerile prezidențiale a apărut un deepfake despre candidatul pro-Uniunea Europeană, acuzându-l de fraudă electorală.

SUA: Un telefon primit de la "Joe Biden" îndemna alegătorii să nu voteze la primarele democrate și să-și păstreze votul pentru alegerile generale din Noiembrie 2024.

Reversul deepfake-urilor

Politicienii și nu numai pot profita de această ocazie pentru a nega orice au făcut sau nu au făcut, caracterizându-l drept deepfake. Astfel, populația nu mai știe în ce ar trebui să aibă încredere și în ce nu

Situații de deepfake

Deepfake-urile audio: Pot apărea chiar cu o zi înainte de alegerile prezidențiale din turul II. Consecințe: Timpul de reacție este foarte limitat, iar conținutul fals se propagă rapid pe social-media, scăzând încrederea cetățenilor în procesul de vot și afectând decizia acestora cu privire la un candidat.

Deepfake-urile în discursurile oficiale: Un discurs fals atribuit unui lider politic poate fi distribuit online, în care acesta pare să facă promisiuni nerealiste sau să adopte poziții extremiste. Consecințe: Populația poate deveni confuză și dezamăgită, iar încrederea în autorități poate scădea dramatic.

Deepfake-urile în publicitatea electorală: Un spot publicitar fals cu un candidat care promite lucruri imposibile poate fi distribuit masiv în zilele de dinaintea alegerilor. Consecințe: Alegătorii pot fi induși în eroare și își pot schimba votul pe baza informațiilor false.

Resurse

- **AI detection tool for audio deepfakes**
- **AI Speech Classifier**
- **AI or not**
- **GPTZero**
- **Hive AI Detector**
- **Deepfake-O-Meter**
- **Vera AI**
- **Hive Moderation**
- **Deepfake Total**
- **Deepware**
- **Loccus.ai**
- **TrueMedia.org**

Tips and tricks:

1. **Sincronizarea buzelor:** Asigurați-vă că mișcarea buzelor este în sincronizare perfectă cu sunetul. Uneori, buzele nu se potrivesc exact cu ceea ce se spune.
2. **Clipitul ochilor:** Verificați dacă persoana din video clipește normal. Deepfake-urile adesea au clipiri nefirești sau foarte rare.
3. **Anomalii vizuale:** Căutați margini pixelate, mișcări nefirești sau schimbări ciudate în iluminare.
4. **Calitatea sunetului:** Ascultați cu atenție vocea. Dacă sună robotică sau distorsionată, este posibil să fie un deepfake.
5. **Expresii faciale și mișcări:** Observați expresiile faciale și mișcările capului. Dacă par rigide sau nefirești, pot indica utilizarea tehnologiei deepfake.
6. **Context și sursă:** Verificați contextul și sursa videoclipului. Deepfake-urile sunt adesea partajate fără informații clare despre proveniența lor.
7. **Detalii incoerente:** Căutați detalii incoerente sau nepotrivite, cum ar fi iluminarea diferită pe față sau fundal, schimbări în culoarea pielii sau elemente de fundal care se mișcă neobișnuit.
8. **Absența emoțiilor:** Evaluați dacă persoana din video exprimă emoții naturale. Deepfake-urile pot avea dificultăți în a reda subtilitățile emoționale.
9. **Utilizarea excesivă a tehnologiei de editare:** Fiți atenți la efecte de editare excesive sau la suprapuneri de text și grafică care pot încerca să distragă de la detalii anormale ale video-ului.

Termeni specifici AI&Deepfake (1)

Prejudecată algoritmică	Atunci când algoritmii produc rezultate care sunt prejudiciate în mod sistematic, din cauza unor ipoteze greșite în procesul de învățare automată.
Astroturfing	O falsă mișcare de tip grassroot, adesea orchestrată de interese politice, corporative sau alte interese speciale.
Bias	Erori sistematice care ar putea afecta validitatea informațiilor.
Chatbot	Un program informatic conceput pentru a simula conversația cu utilizatorii, adesea utilizat în campaniile de dezinformare.
Spionaj cibernetic	Utilizarea rețelelor informatice pentru a obține acces ilicit la informații confidențiale.
Deep learning	Un subset al inteligenței artificiale care permite mașinilor să își îmbunătățească performanțele, pe baza rezultatelor anterioare.
Deepfake	Media sintetică în care asemănarea unei persoane este înlocuită cu cea a altcuiva.
Criptare	Metoda prin care informațiile sunt convertite într-un cod secret pentru a preveni accesul neautorizat.
Tehnologia „Face-Swap”	Tehnologie care permite înlocuirea fețelor în imagini video sau imagini.
Generative Adversarial Network (GAN)	Algoritmi AI utilizați pentru a genera imagini realiste, videoclipuri și alt conținut.
Inteligență artificială generativă	Un tip de inteligență artificială care se concentrează pe generarea de date noi, mai degrabă decât pe simpla analiză și clasificare a datelor existente

Termeni specifici AI&Deepfake (2)

Modele mari de limbaj	Numite și Large Language Models, reprezintă un tip de model de inteligență artificială conceput pentru a înțelege și a genera texte asemănătoare celor umane pe baza unor cantități mari de date.
Modele de învățare automată	Numite și Machine Learning Models, algoritmii permit calculatoarelor să îndeplinească sarcini fără a fi programate în mod explicit.
Metadate	Date care furnizează informații despre alte date, cum ar fi creatorul, data și locația unei părți de conținut.
Microtargetare	Utilizarea analizei datelor pentru a identifica interesele unor persoane sau ale unor grupuri foarte mici de persoane cu interese similare.
Prelucrarea limbajului natural (NLP)	Algoritmi care înțeleg limbajul uman.
Phishing	Încercări frauduloase de a obține informații sensibile, adesea prin e-mailuri înșelătoare.
Pseudonimitate	Starea de identitate mascată, în care o persoană se poate angaja în activități online fără a-și dezvălui identitatea reală, dar poate fi totuși responsabilă pentru acțiunile sale.
Media sintetică	Termen generic pentru a descrie materiale video, imagini, text sau voce care au fost generate integral sau parțial cu ajutorul algoritmilor inteligenței artificiale.
Rețea privată virtuală (VPN)	O rețea care permite utilizatorilor să trimită și să primească date prin rețele publice sau partajate ca și cum dispozitivele lor informatice ar fi conectate direct la o rețea privată.

Sursă: Syntehic Media Exposed: A Comprehensive Guide to AI Disinformation Detection

Ai ajuns la final.
Spor la verificat articole!