

CRJM – Centrul de Resurse Juridice din Moldova

Dezinformare și deepfakes – posibile implicații pentru democrație

Jomir Dumitru · Tuesday, May 4th, 2021

Trăim în epoca post-adevărului (post-truth), un climat cultural și politic în care există tot mai puțin respect pentru adevăr. Acest concept presupune că oamenii – și mai ales politicienilor – le pasă mai puțin de ceea ce spun și dacă ceea ce spun este adevărat.[1]

În spatele unei asemenea atitudini formate în timp, nu stă neapărat ignoranța, ci eforturile concertate prin care spațiul public, mai cu seamă cel online, este invadat constant de informații false, pătinoare sau neverificate, a.k.a. Fake news!. Pe lângă metodele deja devenite „tradiționale” precum paginile web de știri false sau trolii, o nouă tehnologie care amenință să consolideze și să întrească era post-adevărului sunt tehnologiile *deepfakes*, numite convențional și *synthetic media*.

Ce sunt deepfakes și cum sunt create?

Deepfakes sunt videoclipuri false create prin intermediul aplicațiilor digitale și sistemelor automate. Imaginile video reale sunt combinate pentru a crea conținut nou, cu declarații sau acțiuni care nu s-au întâmplat în realitate.

Exemplu de DeepFake: Discurs al fostului președinte american, Barack Obama

Conceptul de bază din spatele tehnologiei este recunoașterea facială (vezi articolul integral aici). Videoclipurile false pot fi create folosind o tehnică de învățare automată și inteligență artificială numită „rețea contradictorie generativă” sau acronimul din engleză GANs. În cuvinte simple, modul în care acționează această tehnologie este următorul: Un GANs scanează fotografii sau video-uri cu o anumită persoană. Tehnologia stabilește puncte de reper pentru a mapa fața persoanei în funcție de anumite caracteristici specifice precum colțurile ochilor și gurii, nările și conturul liniei maxilarului, printre altele, după care poate crea conținut video nou cu o combinație de sunete sau mișcări inexistente.[2] Pe lângă imaginile video, tehnologia GANs poate fi utilizată pentru a genera sunete noi din cele existente, sau un text nou dintr-un text existent.

De ce trebuie să fim precauți în legătură cu Deepfakes?

Tehnologia **deepfake** nu mai este limitată doar în cercurile academice ci disponibilă pe internet.

Aplicații software precum FakeApp sau FaceSwap, Snapchat sau Facebook Messenger, utilizează această tehnologie în care fața reală a persoanei este suprapusă cu fața unei alte persoane sau animații.[3] Chiar dacă în prezent, majoritatea video-urilor false sunt de o calitate care ușor poate fi depistată ca fiind falsă, potrivit experților, această tehnologie avansează zilnic.[4] Astfel foarte curând ne va fi complicat să putem discerne dacă un anumit video este sau nu în realitate fals. Rezultatele unui **deepfake** vor fi destul de convingătoare, aceste video-uri fiind foarte dificil de identificat ca fiind false.

Riscul principal de utilizare al tehnologiei deepfake este că acesta poate permite la un moment dat oricărui utilizator să videoclipuri despre oricine, făcând și spunând orice.[5] Implicațiile unei asemenea tehnologii pentru dezinformare sunt majore, mai ales dacă nu doar imaginea video poate fi replicată dar și vocea persoanei, până la exactitate. Iată câteva exemple de utilizare a tehnologiei deepfake:

- Panică publică indusă prin urgențe „false” precum calamități naturale sau stare de asediu și războaie;
- Discursuri politice manipulatorii sau false;
 - Atacuri asupra unor persoane publice, apăratori ai drepturilor omului sau jurnaliști.

Ce putem face în legătură cu Deepfakes?

Specialiștii recomandă câteva pași care merită urmați atunci când credem că avem de-a face cu un video care portretizează o declarație neobișnuită sau informație posibil falsă. Ca și în cazul informațiilor scrise, pretins false, trebuie să urmăm câteva pași de control[6]:

1. **Să analizăm cu atenție de fiecare dată sursa** – specialiștii recomandă de fiecare dată să atragem atenție originii conținutului video. Câteva întrebări care neapărat trebuie adresate sunt: A fost încreștat de către un utilizator cunoscut sau într-un cont aleatoriu pe o rețea socială? Este sau nu conținut sponsorizat? Cine pretinde că este proprietarul acesteia?. A fost publicat de către un portal de știri cu reputație bună sau un site/portal necunoscut?
2. **Să verificăm dacă video-ul mai poate fi găsit online și în altă parte:** Verificăm unde mai poate fi găsit videoclipul, dacă mai este (sau nu) de găsit online, pentru a vedea dacă sunt mai multe știri cu acest video din partea unor publicații cu reputație bună sau persoane credibile (jurnaliști, formatori de opinie);
3. **Să ne abținem de la concluzii** până nu primim confirmări suplimentare. La această etapă este foarte important să nu distribuim mai departe video-ul. Acest lucru este valabil mai ales în situațiile de ultimă oră, în care informațiile se mișcă rapid și sunt adesea greșite sau interpretate greșit în primele ore de la difuzare, până ce sunt verificate.

Suplimentar pașilor descriși, putem apela la anumite programe sau organizații media care verifică în regim online dacă conținutul video are semne a fi fals, sau dacă este acesta este original.

De exemplu, programul **Reality Defender**[7] scanează și verifică dacă conținutul este fals, manipulatorii sau original. Un raport asupra fiecărui videoclip trimis rezumă concluziile făcute a specula intenția manipulatorilor.

Un alt mecanism, **Platforma Stop-fals**[8], o campanie împotriva informațiilor false și tendențioase, desfășurat de Asociația Presei Independente (API) din Republica Moldova stochează o bază de date și cu ajutorul specialiștilor verifică autenticitatea conținutului video. Oricine poate sesiza

platforma stop-fals dac? întâlne?te con?inut posibil fals, inclusiv video.

R?mâne?i al?turi de CRJM pentru a afla mai multe detalii despre noile tehnologii ?i drepturile omului în era digital?. Ne pute?i urm?ri pe pagina noastr? web CRJM.org ?i re?elele de socializare Facebook, OK.ru, Twitter, LinkedIn.

Acest articol face parte dintr-o serie de publica?ii non-academice realizate de Centrul de Resurse Juridice din Moldova (CRJM) în cadrul proiectului „Program de capacitate în drepturi digitale” sus?inut de Centrul Interna?ional pentru Drept non-profit (ICNL). Opiniile exprimate apar?in CRJM ?i nu reflect? în mod necesar pozi?ia ICNL.

[1] <https://dictionary.cambridge.org/dictionary/english/post-truth>.

[2] How to Spot a Deepfake Like the Barack Obama?Jordan Peele Video”, Craig Silverman,

BuzzFeed News: <https://www.buzzfeed.com/craigsilverman/obama-jordan-peelee-deepfake-video-debunk-buzzfeed>; Mal-uses of AI-generated Synthetic Media and Deepfakes: Pragmatic Solutions Discovery Convening (2018).

[3] Deepfakes: It’s Not What it Looks Like”, Sam Gregory, Atlantic Council:

https://www.youtube.com/watch?v=Qh_6cHw50I0&feature=youtu.be; “Mal-uses of AI-generated Synthetic Media and Deepfakes: Pragmatic Solutions Discovery Convening” (2018).

[4] Fake Videos of Real People and How to Spot Them”, Supasorn Suwajanakorn, TED:

<https://www.youtube.com/watch?v=o2DDU4g0PRo>.

[5] <https://www.reuters.com/article/us-chertoffdonahoe-hacking-commentary-idUSKCN1NH1W2>

[6] Commentary: For Election Hackers a New and More Dangerous Tool”, Michael Chertoff and

Eileen

Donahoe,

Reuters:

<https://www.reuters.com/article/us%20chertoffdonahoe%20hackingcommentary/commentary%20for%20election%20hackers%20a%20new%20and%20more%20dangerous%20tool-idUSKCN1NH1W2>

[7] Reality Defender: <https://rd2020.org/index.html>.

[8] Platforma Stop-Fals (Asocia?ia Presei Independente): <https://semnale.stopfals.md/ro/about>.

This entry was posted on Tuesday, May 4th, 2021 at 4:28 pm and is filed under [Blog](#). You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. Both comments and pings are currently closed.