Securitatea cognitivă. Analiza operațiunilor automatizate de influență bazate pe reclamele deepfake

- REZUMATUL TEZEI DE DOCTORAT -

Conducător științific: Prof. univ. dr. Alina Bârgăoanu

Doctorand: Mihaela Pană

Această teză este fundamentată pe evoluția legăturii interdependente dintre comunicare și tehnologie, din care derivă și problema de cercetare referitoare la transformarea platformelor de publicitate online în vectori de atac, în cadrul operațiunilor automatizate de influență bazate pe reclamele deepfake. Motivația cercetării urmărește să ofere perspective de cunoaștere și soluții de contracarare a reclamelor deepfake, o formă de publicitate online transformată în fenomen perturbator în contextul unui climat informațional predominat de crize suprapuse. Folosind instrumente algoritmice de identificare a cazurilor și de analiză, lucrarea explorează tendințele și interdependențele operaționale dintre atacurile informaționale, regăsite sub forma publicității false (*fake advertising*) prin reclamele deepfake, și cele cibernetice, prezentate sub forma utilizării malițioase a tehnologiei, având ca obiectiv identificarea și caracterizarea amprentei digitale asociate campaniilor de influență cu daune la nivel cognitiv.

IMPACTUL ȘI CONTRIBUȚIILE ORIGINALE

Cadrul teoretic argumentează perspectiva tehnologică a comunicării și conturează dimensiunea cibernetică a influenței, introduce în discuție problematica reclamelor deepfake din perspectiva publicității false ca amenințare hibridă de securitate și aduce în atenție nevoia dezvoltării conceptului de securitate cognitivă în sfera politicilor publice.

Originalitatea cercetării empirice constă în abordările inovative ale celor două perspective definitorii problemei de cercetare. Din perspectiva comunicării, analiza calitativ-cantitativă evidențiază exploatarea platformelor de publicitate și corelează conținutul promovat de reclamele deepfake cu teorii din domeniul psihologiei precum declanșatoarele cognitive reflectate în factorii psihologici ai ingineriei sociale (Gragg, 2003), dar și tipurile de frici din ierarhia lui Karl Albrecht (2007).

Din perspectiva tehnologică, inovația cercetării este conferită de strategia de identificare a reclamelor înșelătoare prin conceptul de „honeypot algoritmic", descris prin accesarea sistematică a reclamelor înșelătoare într-un comportament disimulat de interes pentru acest tip de conținut pentru a influența efectul algoritmic al camerei de ecou în scopul afișării reclamelor deepfake. De asemenea, studiul introduce modelul PART pentru interpretarea datelor rezultate în urma verificărilor OSINT ce stau la baza determinării tiparelor din modul de operare al operațiunilor de influență derulate prin reclamele deepfake și modelul de analiză statistică algoritmică bazat pe verificarea încrucișată a rezultatelor oferite de platformele cu servicii publice bazate pe inteligență artificială, capabile să genereze și să asiste în realizarea de analize statistice.

În plus, corpusul de analiză a reprezentat o bază de date relevantă cu 400 de cazuri înregistrate pe parcursul anilor 2023-2025, necesară specialiștilor din domeniul securității cibernetice pentru determinarea infrastructurii digitale a operațiunii Doppelgänger în România și în Europa. Raportarea imediată către Google, Meta și Directorul Național de Securitate Cibernetică a permis cunoașterea timpurie a amenințării pentru identificarea soluțiilor de contracarare (ex. mecanismul guvernamental #NoFake), iar rezultatele analizei de conținut au conturat parametri amenințării cognitive generată de reclamele deepfake distribuite în spațiul informațional românesc.

CONTEXTUL CERCETĂRII

Alegerea problemei de cercetare a fost determinată de următoarele patru aspecte:

- apariția unor noi concepte asociate cu domeniul securității informaționale în anul 2022, precum FIMI pentru a defini interferențele străine în mediul informațional și cadrul DISARM conceput pentru a le identifica și contracara;
- intensificarea reclamelor deepfake pe parcursul anului 2023 în social media, cu țintă predilectă pe audiența din România;
- intrarea în vigoare a Regulamentului privind serviciile digitale (DSA) în februarie 2024, care prevede inclusiv măsuri de transparență și obligă companiile de tehnologie să facă o prioritate din securitatea informațională;
- anticiparea unor posibile acțiuni ostile în perspectiva celor două procese electorale majore din România, alegerile europarlamentare din iunie 2024 și alegerile prezidențiale din noiembrie 2024.

Evoluția tehnologică în era digitală a transformat comunicarea pe toate palierele, iar acțiunile malițioase de influență au devenit tot mai sofisticate prin combinarea diverselor

tipuri de amenințări care exploatează vulnerabilitățile din lanțul utilizării tehnologiei, de la erorile tehnice din sistemele informatice până la slăbiciunile oamenilor determinate de emoții și prejudecăți.

În formele eficiente de comunicare actuale, conținutul multimedia urmează modelul strategiilor de marketing digital pentru a livra mesajul cu precizie, în cel mai scurt timp, adecvat categoriilor de utilizatori țintă. Conceptualizarea fake news și a componentelor dezordinii informaționale au fost un prim pas în cercetarea alterării mediului informațional. Evoluția tehnologică a Inteligenței Artificiale (AI) permite generarea unei noi forme de conținut fals de tip deepfake, care poate fi creat foarte rapid, cu costuri reduse și mult mai dificil de identificat. În plus, noua abordare strategică a războiului informațional implică și capacitatea de utilizare a instrumentelor și mecanismelor de marketing și publicitate online, evoluând spre dimensiunea operațiunilor automatizate de influență.

Democrațiile moderne se confruntă astfel cu o avalanșă de amenințări care afectează implicit și domeniul cognitiv, prin combinarea atacurilor informaționale cu operațiunile cibernetice în care sunt exploatate obiceiurile de consum a tehnologiei cu scopul de a modela percepțiile și de a eroda coeziunea societală care stă la baza securității naționale. Adversarii urmăresc prin aceste tehnici hibride să influențeze modul în care oamenii gândesc și acționează, iar modul de operare este conceptualizat prin noțiunea de *război cognitiv*, în care mintea umană a devenit câmpul de luptă. Apărarea în fața acestor amenințări complexe, identificate atât în cazul interferențelor electorale, cât și în campaniile cu informații false legate de pandemii medicale și conflicte armate, necesită o abordare interdisciplinară pe măsura gradului de sofisticare a pericolelor, care să combine expertize din domeniul comunicării, al psihologiei și sociologiei, dar și din domeniul securității cibernetice. Această abordare se regăsește în conceptul emergent de „*securitate cognitivă*”.

Exploatarea platformelor de publicitate în scopuri ostile este pusă în legătură directă cu acțiunile de criminalitate cibernetică, fie că este vorba despre infractori motivați financiar care recurg la escrocherii pentru a înșela utilizatorii mai puțin conștienți despre existența amenințărilor cibernetice sau actori statali care sponsorizează acest tip de operațiuni de influență urmărind obținerea avantajelor strategice prin interferența în politica națiunilor adversare. Drept urmare, analiza are în vedere natura polimorfă a motivației actorilor implicați, care poate oscila între obiective infracționale specifice criminalității cibernetice și agende strategice, caracteristice acțiunilor sponsorizate de state, acestea din urmă fiind, în anumite cazuri, deliberat camuflate sub aparența unor atacuri autonome de natură infracțională.

Concluziile cercetării contribuie la clarificarea dilemei privind clasificarea publicității false ca fraudă cibernetică sau operațiune de influență cu daune la nivel cognitiv. În plus, dezvăluirea noilor tehnici utilizate în hackingul cognitiv ajută la înțelegerea eficacității rapoartelor publicitare în contextul Regulamentului privind serviciile digitale (DSA), iar datele obținute în urma monitorizării servesc ca bază pentru cercetări aprofundate pentru contracararea operațiunilor bazate pe manipularea informațiilor și ingerințele străine (*Foreign Information Manipulation and Interference* – FIMI) într-o abordare care implică expertiză în comunicare și securitate cibernetică.

ÎNTREBĂRILE ȘI OBIECTIVELE DE CERCETARE

Principalele întrebări de cercetare se referă la *modul în care platformele de publicitate sunt transformate în vectori de atac în cadrul operațiunilor automatizate de influență* și urmărește *determinarea parametrilor prin care reclamele deepfake pot fi identificate și contracarate*. Pentru atingerea acestor obiective, cercetarea empirică urmărește o serie de obiective specifice în analiza amenințărilor hibride care combină elemente precum fake news, deepfake și malvertising în conținutul distribuit sub formă de reclame prin intermediul platformelor de publicitate online.

ÎNTREBARE	OBIECTIV
Cum operează campaniile de influență bazate pe reclamele online cu conținut deepfake?	Identificarea tacticilor și tehnicilor folosite în campaniile publicitare cu reclame deepfake;
Cum pot fi identificate operațiunile cibernetice de influență cu instrumente OSINT și cadre de analiză FIMI?	Crearea unui model de analiză care să valorifice observațiile extrase cu instrumente OSINT și cadraje de analiză FIMI;
Care sunt tiparele și tendințele actuale referitoare la utilizarea platformelor de publicitate online în cadrul operațiunile automatizate de influență?	Determinarea amenințărilor hibride care afectează spațiul informațional, cibernetic și cognitiv prin intermediul platformelor de publicitate online;
Cum funcționează mecanismele psihologice bazate pe factori declanșatori ai fricilor și emoțiilor prin reclamele deepfake?	Determinarea parametrilor atacurilor cognitive din reclamele deepfake.

STRUCTURA TEZEI

Pentru a răspunde întrebărilor de cercetare și pentru a îndeplini obiectivele acestei teze, am structurat lucrarea în două părți esențiale: încadrări conceptuale și cercetare empirică. În capitolul dedicat conceptelor teoretice sunt tratate perspectivele tehnologice ale comunicării, teoriile falsului în mediul online, dar și modul în care dezinformarea online se poziționează în prezent între fake news și deepfake. De asemenea, în acest capitol am argumentat și modul în care studiile de actualitate abordează publicitatea online ca instrument de promovare și vector de atac cognitiv în dimensiunea cibernetică a influenței. Un alt subcapitol tratează legătura amenințărilor cibernetice cu înșelătoriile în online, dar și automatizarea operațiunilor cibernetice de influență și conflictul high-tech ca parte din evoluția războiului hibrid. O altă reprezentare teoretică se referă la încadrarea problemei de cercetare în domeniul războiului cognitiv, cu accept pe rolul publicității false în tacticile de atac. De asemenea, există și o abordare defensivă în legătură cu aceste fenomene ostile cuprinsă în sinteza modelelor, cadrajelor și instrumentelor utilizate în analiza operațiunilor de influență, dar și în prefigurarea conceptului de securitate cognitivă, ca abordare necesară în politicile de securitate.

Capitolul de cercetare empirică conține cele două etape: studiul de caz referitor la reclamele deepfake observate pe platforma YouTube pentru exemplificarea modului de exploatare a publicității programatice și analiza cantitativ-calitativă a reclamelor deepfake distribuite prin platformele de publicitate online Google și Meta. Profilul comportamental rezultat din corelarea tacticilor observate prin mai multe forme de analiză cu cadrul DISARM în prima parte a cercetării a determinat grila aplicată în analiza de conținut din cea de-a doua parte a cercetării. Astfel, din analiza datelor înregistrate, a rezultat tiparul de acțiune în exploatarea platformelor de publicitate și tendințele referitoare la conținutul înșelător promovat și la tacticile și tehnicile din modul de operare. Pe de altă parte, analiza legăturilor dintre conținutul înșelător și declanșatoarele cognitive a evidențiat mecanismele psihologice instrumentate în reclamele deepfake, precum exploatarea fricii și a triggerelor cognitive.

CONCEPTELE ESENȚIALE

Încadrarea conceptuală pornește cu perspectiva tehnologică a comunicării și propune o versiune de cartografiere a teoriilor de la determinismul tehnologic la operațiunile de influență, făcându-se astfel legătura cu studiile privind dimensiunea cibernetică a influenței și formele de publicitate bazate pe noile tehnologii în care se încadrează activitatea de cercetare științifică. De asemenea, tipologiile amenințărilor hibride și perspectivele legăturii dintre

atacurile cibernetice și atacurile informaționale susțin cadrul teoretic, iar cercetarea empirică este ancorată în capitolele dedicate războiului cognitiv și instrumentelor de analiză ale operațiunilor de influență care țintesc domeniul cognitiv.

Conținutul deepfake este definit ca **dezinformare audiovizuală** (Farid et al., 2019) sau **media sintetică** (Monteith et al., 2024), fiind un **produs al inteligenței artificiale generative** (McKinsey & Co, 2023), utilizat sub diverse versiuni ale amenințărilor emergente din operațiunilor automate de influență (Goldstein et al., 2023).

Fake advertising este conceptul care definește publicitatea falsă, asociată **conținutului publicitar intenționat fals, înșelător sau neautentic, difuzat cu scopul de a înșela sau manipula publicul** (Dan et al., 2022). Utilizarea conținutului deepfake generat cu AI în domeniul publicității este cunoscută sub denumirea de **deepfake advertising** (Agarwal & Nath, 2023; McCreary, 2024).

Operațiuni cibernetice de influență definesc operațiunile care afectează stratul logic al spațiului cibernetic cu intenția de a influența atitudinile, comportamentele sau deciziile publicului țintă (Brangetto & Veenendaal, 2016; Libick, 2007). Relaționarea sistemelor algoritmice și a mecanismelor de marketing cu **manipularea algoritmică** (Hacker, 2023) și cu **operațiunile automatizate de influență** (Goldstein et al., 2023).

Războiul cognitiv este definit ca „**înarmarea opiniei publice** de către o entitate externă în scopul influențării politicilor publice și guvernamentale și destabilizării instituțiilor publice” (Bernal et al., 2020).

Securitatea cognitivă este un concept nou aflat în curs de cercetare pentru o definire comprehensivă. Una dintre abordările definitorii surprinde situația în care tehnologia scapă de sub controlul uman și devine o armă folosită cu scopuri ostile, într-un duel al algoritmilor care necesită intervenție pentru a proteja utilizatorii și pentru a asigura funcționarea tehnologiei cu riscuri limitate (Andrade & Yoo, 2019a; Michael Krigsman, 2019; Terp & Lesser, 2022).

METODOLOGIA

Metodologia de cercetare cuprinde multiple forme de analiză de conținut și studii de caz, distribuite în etape succesive care răspund unor întrebări de cercetare specifice.

Prima etapă de cercetare include studiul de caz referitor la un advertiser din rețeaua Google Ads care promovează reclame deepfake cu același mesaj pe YouTube, identificate în luna mai 2023. Această primă analiză urmărește să răspundă la întrebarea de cercetare: *Cum poate fi determinat modul de operare în operațiunile de influență bazate pe reclame*

deepfake? Pentru a consolida practica anticipării și răspunsului integrat în cazul amenințărilor cognitive generate de publicitatea falsă în mediul online, analiza multiplă a studiului de caz include analiza mesajului și a identității digitale, analiza facilitată de instrumente din sfera Open-Source Intelligence (OSINT) pentru colectarea dovezilor tehnice, interpretată prin propunerea modelului conceput pe *Purposes, Actions, Results, Technique* (PART) și pe aplicarea cadrului inovator de analiză strategică *Disinformation Analysis and Risk Management* (DISARM), folosit pentru anticiparea operațiunilor de manipulare a informațiilor și ingerințele străine (*Foreign Information Manipulation and Interference* – FIMI), chiar și atunci când aceste operațiuni nu par să aibă un scop specific, identificabil imediat.

Observațiile rezultate în cadrul acestei analize aprofundate conduc la necesitatea monitorizării sistematice a formelor de publicitate online în cea de-a doua etapă de cercetare, pentru a identifica și alte cazuri de *hacking cognitiv* derulate prin reclame înșelătoare. Astfel, concluziile primei etape de cercetare sunt valorificate în conceperea grilei de analiză aplicată în cea de-a doua etapă de cercetare, al cărei scop este *determinarea parametrilor prin care reclamele deepfake pot fi identificate și contracarate.* Categoriile cu variabile fac referire la conținutul deepfake asociat cu site-uri fabricate pentru a redirecționa utilizatorii în timpul campaniei active spre alte fluxuri de știri false (*fake news*), sub care se ascund diverse programe malware, amenințare cibernetică cunoscută sub denumirea de *malvertising*. Analiza de conținut este fragmentată în două direcții de colectare a datelor – aspecte privind identificarea, cu elementele specifice platformei și agenților de publicitate, și aspectele privind indexarea tacticilor și tehnicilor folosite, din care fac parte analiza de conținut și evaluarea impactului cognitiv pe care aceste reclame îl au în mediul online.

Din eșantionul de analiză fac parte 400 de cazuri, care reprezintă conturi de advertiseri/ promoteri care au administrat reclame prin platformele Google Ads și Meta Ads, identificate în urma unei monitorizări sistematice derulată în perioada noiembrie 2023 – ianuarie 2025 și grupate în clustere tematice în funcție de subiectul promovat. Numitorul comun a tuturor cazurilor analizate determină întreaga arhitectură a reclamelor deepfake și tiparul acestei amenințări persistente derulată prin conturi false sau pagini compromise care combină mesaje structurate pe mecanisme psihologice, cu conținut deepfake și domenii fake ce redirecționează spre site-uri de *phishing*, cu conținut de tip fake news, care impersonează branduri media și celebrități, asociate și cu potențial conținut malware.

INSTRUMENTELE DE ANALIZĂ

Pe parcursul cercetării au fost folosite o serie de instrumente de analiză OSINT, platforma Nvivo a fost folosită pentru indexarea și organizarea datelor colectate pe baza grilei de analiză, OpenAI ChatGPT Deep Research a fost folosit pentru procesarea automatizată a analizei statistice, iar Google AI Studio a fost util în procesul de verificare a raționamentului OpenAI ChatGPT Deep Research pe baza căruia au fost obținute rezultatele statistice.

VALIDAREA TEHNICĂ

Validarea tehnică a cercetării a inclus verificări în procesul de identificare a cazurilor și verificarea rezultatelor analizei algoritmice.

Procesul de identificare a cazurilor a fost validat tehnic prin verificarea încrucișată a unui rezultat prin metode alternative – rezultatul observației directe sau al raportării crowd-sourcing a fost verificat prin interogarea bibliotecii cu reclame a platformei de publicitate în care a fost identificat și validarea prin raportare oficială - informațiile indexate au fost raportate atât platformei unde au fost detectate reclamele pentru a fi blocate, cât și Directorului Național pentru Securitate Cibernetică din România pentru evaluarea amenințării și eliminarea pericolelor cibernetice din spațiul informațional.

Validarea algoritmică încrucișată a constat în verificarea rezultatelor obținute cu ajutorul OpenAI ChatGPT Deep Research printr-un sistem similar de analiză, precum Google AI Studio.

CRITERIILE DE SELECȚIE ȘI VARIABILE ANALIZATE

Identificarea profilului comportamental și a modului de operare din prima etapă de cercetare a determinat condițiile de selecție a cazurilor - conturile agenților de publicitate incluse în corpus: (1) identitate falsă: conturile folosesc nume false, nume generice sau pseudonime; (2) tehnici de redirecționare a link-urilor promovate către site-uri fake compromise / deturnate; (3) elemente de comunicare simbolică privind identitatea națională / autoritatea statului; (4) conținut fals de tip fake news - text fabricat sau copiat din surse originale; (5) conținut deepfake - foto, audio sau video; (6) avertizare privind amenințări cibernetice: phishing și/sau malware; (7) tehnici de ascundere și/sau de ștergere a amprentei digitale.

Structura grilei de analiză vizează pe de-o parte 14 aspecte specifice cazurilor în etapa de identificare - variabile cu caracteristici specifice conturilor agenților pe platformele de publicitate și pe de altă parte elementele descriptive ale conținutului promovat urmărind **cine**

promovează? în legătură cu originea conturilor care promovează conținutul, **ce promovează?** în legătură cu detaliile privind conținutul promovat și **cum promovează?** în legătură cu aspecte privind modul de operare și cu asocieri privind declanșatoarele cognitive și exploatarea fricilor.

INTERPRETAREA REZULTATELOR

Analiza cazurilor identificate a evidențiat corelații și tipare comportamentale importante în cadrul campaniilor malițioase cu reclame deepfake. Studiul acestor reclame, care integrează atât datele despre contul agentului de publicitate, detaliile legate de momentul identificării, cât și analiza de conținut, relevă faptul că acestea diferă ca mărime, durată și limbaj, însă toate urmăresc în esență același scop: manipularea opiniei publice în favoarea unor agende ascunse. Evidențierea legăturilor dintre caracteristicile reclamelor insidioase poate contribui la dezvoltarea unui cadru proactiv de combatere a acestor amenințări hibride la adresa societății informaționale.

Rezultatele analizei arată că variabilele esențiale, precum durata campaniei, volumul de reclame, audiența/impactul sau limbile comunicate, se află într-o interdependență cu conținutul promovat, structurat în marea majoritate a cazurilor pe combinația mecanismelor psihologice. Toate aceste aspecte reflectă strategia modului de operare și pot determina parametri unui sistem de alertare timpurie în cazurile operațiunilor cibernetice de influență în care sunt exploatate platformele de publicitate.

Majoritatea covârșitoare a campaniilor publicitare identificate au fost persistente în timp, semn că administratorii acestor reclame dispun de o infrastructură care le permite prezența continuă pentru a-și consolida influența. Aceste campanii de lungă durată, mai ales cele cu număr mare de reclame și multilingve, sunt cel mai probabil operațiuni orchestrate la nivel înalt, care pot fi asociate cu grupări de criminalitate cibernetică asociate cu actori statali. În schimb, campaniile de scurtă durată sugerează fie experimente pentru a testa audiența, fie reacții rapide la evenimente pentru a marca agenda informațională cu scopul de a surprinde răspunsul publicului. Pe baza dovezilor înscrise în tiparul de operare, campaniile derulate de conturile agenților de publicitate lansate recent, cu număr redus de reclame, pot constitui potențiale semnale timpurii ale unor noi narațiuni care urmează să fie propagate masiv.

Limba română domină ca mijloc de comunicare în reclamele deepfake, ceea ce indică o focalizare puternică pe influențarea audienței din România. Totodată, prezența semnificativă a limbii ruse confirmă existența unei influențe externe (posibil rusești) active, în concordanță cu contextul geopolitic în care România și țările vecine au fost ținte ale

propagandei rusești. Acest fapt este susținut și de campaniile multilingve (română-rusă-ucraineană-engleză) direcționate pe audiența segmentată pe mai multe țări europene, ceea ce semnalează eforturi concertate transnațional, ele reprezentând cazurile cele mai complexe, dar totodată fiind și cele mai vizate de contramăsuri, așa cum sugerează statusul reclamelor la finalul perioadei de monitorizare.

Corelațiile cantitative determină parametri care pot contribui la mecanismele de alertă asupra acestei amenințări. Volumul mare de reclame duce la audiență mare, duratele mai lungi ale reclamelor active permit acumularea de audiență, iar multilingvismul extinde aria de acțiune. Importanța acestor corelații rezidă în faptul că ele oferă indicii despre eficacitatea tacticilor: dacă un actor dorește să atingă audiență foarte mare, datele arată că ar trebui să investească în multe reclame, fie să le mențină active cât mai mult timp (sau să combine cele două variante pentru a atrage cât mai multe accesări pe site-ul promovat prin intermediul reclamelor). Astfel, relațiile obținute din analiza corelațiilor pot constitui euristici de detectare timpurie. De exemplu, un nou cont de promotor/advertiser care în primele săptămâni difuzează deja zeci de reclame în paralel, în mai multe limbi, poate fi un indiciu de comportament neautentic coordonat ce merită investigat imediat, înainte ca mesajul promovat să atingă audiențe de milioane de utilizatori. De asemenea, aceste indicii ale comportamentului pe rețeaua de publicitate se pot asocia cu tacticile de influență recurente care includ conținutul specific și țintirea audienței căreia îi este destinat: imitarea media locale autentice la grupuri țintă segmentate la nivel de localitate/județ, propagandă regională integrată cu conținut distribuit prin segmentarea audienței la nivel de țară, exploatarea momentelor-cheie din agenda media și diversificarea canalelor pentru o arie mai mare de acoperire. Toate aceste tactici au fost documentate și în rapoarte internaționale asupra operațiunilor de influență; de exemplu, Facebook a observat tendința de *"platform diversification"*, *"retail IO"* (campanii țintite, mai mici), precum și uzul unor firme intermediare pentru a rula astfel de campanii – fenomene pe care datele noastre le reflectă în contextul României.

Corelațiile dintre categoria, durata, volumul, audiența și limba reclamelor înșelătoare evidențiază două categorii predominante de operațiuni de influență active în spațiul online românesc: campanii locale susținute, ce par a proveni din interior, și campanii transfrontaliere de mare amploare, asociate probabil influenței străine având în vedere discrepanța dintre originea conturilor de promotori și limba comunicată în reclamele pe care le sponsorizează. Ambele categorii folosesc reclame plătite pentru a manipula percepții, însă se diferențiază prin volum, limbaj și public țintă. Identificarea tiparelor recurente (spre exemplu: persistent

vs. spontan, multilingv vs. monolingv) poate sprijini eforturile viitoare de detecție automată a operațiunilor de influență, deoarece aceste caracteristici compun amprenta digitală a comportamentului neautentic. În același timp, din perspectiva politicilor publice și de securitate, rezultatele sugerează necesitatea unei vigilențe sporite atât față de narațiunile importate din exterior, cât și față de campaniile indigene de dezinformare, mai ales atunci când se desfășoară în limba și în contextul local și pot trece mai ușor drept legitime.

Datele rezultate confirmă existența unor corelații semnificative și a unor tendințe definitorii în modul de operare al campaniilor de influență prin reclame înșelătoare. Folosirea imaginii trucate a unor personalități publice este una dintre cele mai puternice unelte de manipulare din aceste reclame, fie că este vorba despre politicieni de prim rang, vedete media și experți respectați, aceștia fac parte din categoriile de oameni care pot influența puternic opinia publică. Această constatare accentuează gravitatea fenomenului: nu numai brandurile instituționale sunt deturnate, ci și identitatea persoanelor reale – un abuz care poate fi considerat defăimător și care poate avea implicații juridice din perspectiva identității folosite fără consimțământ.

Analiza de față a evidențiat un ecosistem complex de reclame deepfake, care îmbină în mod strategic elemente de dezinformare, inginerie socială și tehnici cibernetice avansate. Din perspectiva cantitativă, am observat că fenomenul implică un număr mare de entități false (zeci de țări de origine, sute de identități false, numeroase teme abordate). Dominanța unor surse externe (Ucraina, țări din Asia) și omniprezența identităților false arată caracterul coordonat internațional al operațiunilor – sugerând existența unor rețele bine puse la punct. Din punct de vedere calitativ, reclamele deepfake se remarcă prin versatilitate și adaptabilitate: abordează subiecte de actualitate (financiare, sociale, medicale și politice), folosesc imaginea celor mai de încredere figuri publice și branduri reprezentative pentru audiența vizată și apasă exact "butonul" emoțional potrivit pentru publicul țintă (frica de sărăcie/boală, lăcomia de câștig, patriotismul, viciile, etc.). Ele combină în mod ingenios tactici multiple – de la *fake news* media și *deepfake* la *phishing* tehnic – creând un atac multi-vector asupra utilizatorului, iar această redundanță a tacticilor face reclamele extrem de eficiente în tentativa de a păcăli publicul țintă.

Un alt aspect important în determinarea tiparului înșelător este sincronizarea categoriilor studiate în grila de analiză: pretinsa origine a agenților de publicitate, conținutul promovat și tacticile prin care mesajul este construit să activeze audiența. Maniera sofisticată în care aceste elemente se susțin reciproc este semnalul unei operațiuni orchestrate printr-o infrastructură complexă. Originea ascunsă permite tactici agresive fără repercusiuni,

conținutul mincinos se bazează pe informații false, iar tacticile permit acoperirea originii și a informațiilor false. Împreună, aceste elemente conturează o adevărată campanie de dezinformare menită să susțină o fraudă integrată, care nu poate fi confundată cu acțiunile disparate de fraudă informatică.

Una dintre caracteristicile de bază ale acestor reclame înșelătoare face referire la capacitatea de manipulare a psihologiei umană prin declanșarea de temeri, dorințe și părtiniri cognitive pentru a influența comportamentul utilizatorilor. Codarea calitativă evidențiază mai mulți factori declanșatori cognitivi și apeluri ale fricii utilizate sistematic în campanii. Pârghiile psihologice dominante au fost autoritatea, reciprocitatea, frica și construirea încrederii. Acești factori au fost adesea identificați într-o abordare complexă care crește semnificativ puterea de persuasiune prin implicarea mai multor declanșatoare simultan. De exemplu, o înșelătorie bazată pe ideea de investiții folosește autoritatea (garanția de stat pentru a câștiga încredere și abuzul de putere pentru a stimula acțiunea), induce frica (de sărăcie sau îngrijorările legate de inflație, șomaj și criza financiară) și susține dorința de câștig rapid și fără efort prin recompensă reciprocă (un bonus sau limitarea timpului de acces).

Cazurile agenților de publicitate ale căror reclamele descriu guvernul ca o forță amenințătoare sugerează o preocupare pentru înțelegerea sensibilităților audienței: unele segmente de public se tem de stat mai mult decât au încredere în el, motiv pentru care au fost țintite cu mesaje care promovat sentimentul de revoltă față de abuzul autorităților. De asemenea, cazurile care exploatează frica de război sau de ocupație teritorială și-au construit comunicarea pe extinderea conflictului dincolo de granițele Ucrainei pentru a speria oamenii să-și protejeze bunurile sau să urmeze recomandările din mesajele naționaliste.

Astfel, corelațiile între frici și tactici au evidențiat tiparele care leagă mecanismele psihologice de mesajele false comunicate. De exemplu, reclamele care exploatează teama de sărăcie au folosit declanșatori de autoritate și au uzurpat identitatea demnitarilor și a unor instituții financiare. De asemenea, a fost folosită și retorica naționalistă, sugerând că investiția nu este doar o acțiune pentru binele personal, ci și una patriotică. În cazul fricii de boală, corelația a fost cu uzurparea identității autorității medicale și declanșatoare emoționale puternice pentru a determina utilizatorii să cumpere tratamentele promovate. În cele mai multe cazuri, declanșatoarele cognitive erau incluse într-o serie de avertismente înfricoșătoare declanșator de frică, urmate de speranța unei recuperări miraculoase (declanșator de exaltare).

De asemenea, se remarcă o dualitate psihologică în care declanșatorii de autoritate sunt folosiți fie să amplifice frica (când figura de autoritate avertizează asupra unui pericol),

fie pentru a atenua frica (când figura de autoritate exprimă încredere și oferă siguranță). Coeficientul ridicat de autoritate (321 de reclame) și frica (în special frica de sărăcie în 289 de reclame) sugerează că înșelăciunile financiare au fost orchestrate pe combinația frică-autoritate: publicul a fost mai întâi intimidat, apoi salvat prin oportunitatea susținută de autoritate.

O altă corelație este între utilizarea declanșatoarelor relaționale și temerile specifice, observate cu precădere în reclamele din domeniul medical. Declanșatoarele „relații înșelătoare" (116 reclame) au apărut adesea în clusterele pe teme de sănătate și de recuperare a banilor pierduți în înșelătoriile cibernetice, bazate pe scenarii de construire a încrederii. Spre exemplu, reclamele care promit recuperarea fraudei trebuie să convingă victima înșelătoriei să aibă încredere și comunică acest lucru pe un ton empatic („îți înțelegem durerea, suntem aici pentru a te ajuta"), exploatând consecințele emoționale (frica de a nu recupera banii, speranța că există o salvare).

În concluzie, profilul psihologic al acestor reclame deepfake este unul care provoacă teamă și apoi oferă o (falsă) soluție salvatoare, combinând apelul fricii (în special temerile economice și de sănătate) cu prejudecățile cognitive (încrederea în autoritate, reciprocitate, sentimentul de urgență) pentru a reduce scepticismul și gândirea critică. Această sinergie de frică și autoritate este un semn distinctiv al operațiunilor de influență eficiente și este evidentă în setul de date.

MODELE OPERAȚIONALE ALE ATACURILOR COGNITIVE BAZATE DE PUBLICITATEA DEEPFAKE

Corelând constatările analitice din etapa analizei de conținut cu datele operaționale din etapa de identificare, se disting mai multe modele operaționale ale campaniilor publicitare care au evoluat pe parcursul monitorizării.

Exploatarea platformelor de publicitate: majoritatea reclamelor deepfake studiate (89% dintre cazuri) au fost difuzate prin intermediul rețelei Meta (Facebook/Instagram), iar acest fapt evidențiază că publicitatea distribuită pe rețelele sociale a fost principalul vector pentru operațiunea de influențare. Diversitatea multimedia și țintirea unor categorii demografice specifice sunt printre cauzele pentru care rețeaua Meta este favorită în campaniile ostile. De asemenea, prezența pe mai multe platforme (Meta și Google) indică faptul că unele dintre cele mai prolifice campanii s-au desfășurat pe ambele platforme. Această abordare multiplatformă îngreunează detectarea și reprezintă dovada unei operațiuni coordonate.

Persistența campaniilor publicitare: operațiunile cu reclame deepfake tind să fie de lungă durată, aproximativ 77% din campaniile publicitare au fost active timp de peste 30 de zile. Longevitatea activității publicitare malițioase sugerează că aceste conturi au reușit să evite detectarea pentru a rămâne cât mai mult timp în atenția audienței. Tacticile de ocolire a detecției au inclus modificări frecvente ale contului pentru a evita închiderea sau ascunderea reclamelor deepfake în versiuni multiple ale unei reclamele aparent autentice. Evoluția permanentă a metodelor de ascundere a conținutului malițios este motivul pentru care o parte semnificativă (peste 60 %) din conturile identificate nu au fost eliminate definitiv de pe platforme, în pofida raportării, până la finalul perioadei de monitorizare. În cele mai multe cazuri, reclamele deepfake au fost eliminate sau dezactivate, dar conturile au persistat. Doar 36% dintre conturi au fost "șterse" până în 2025, fie prin aplicarea măsurilor de contracarare de către platformă, fie prin eliminarea de către operatorii ostili ca măsură de precauție. Concluzia operațională sugerează că aplicare măsurilor de către platforma de publicitate a fost inconsecventă sau întârziată: multe conturi malițioase au supraviețuit mult după raportare. Această longevitate le-a permis totuși actorilor ostili să difuzeze volume mari de reclame și să ajungă la audiențe mari (unele depășind un milion de impresii) înainte de orice intervenție.

Amploarea campaniilor: distribuția numărului de reclame a variat de la un cont la altul, existând diferențe semnificative între conturile cu număr mare de reclame și cele cu un număr mic. Cu toate acestea, există câteva observații în privința modului de operare. Conturile cu reclame puține sunt de regulă cele care difuzează același conținut pentru scurtă perioadă de timp. Conturile cu un volum de peste 1000 de reclame fiecare sunt, în esență, operațiuni de influență la scară industrială și corespund adesea entităților străine care desfășoară campanii multilingve țintind o audiență extinsă la nivel regional sau global din RO, RU, UA și UE. Aceste campanii mari au acoperit probabil internetul cu variante ale anunțurilor lor deepfake, încercând mesaje diferite și optimizând atingerea. Acest tipar sugerează comportamentul neautentic coordonat în care câteva grupuri cu resurse suficiente pot produce conținut la scară largă.

Segmentarea audienței: multe dintre conturile studiate dețineau la momentul identificării campanii publicitare cu o audiență de peste zeci de milioane de utilizatori. În 83 de cazuri, reclamele distribuite aveau un impact estimat în rețeaua de publicitate la peste un milion de persoane, iar în alte 95 de cazuri între 500.000 și un milion. Doar un sfert din campanii au rămas sub 50.000 de spectatori. Campaniile multilingve, în special, au vizat utilizatori din mai multe țări sau comunități etnice. De exemplu, unele anunțuri au fost

difuzate simultan în română și rusă, vizând atât cetățenii români, cât și populația moldoveană sau ucraineană de limbă română sau minoritatea rusofonă. Au existat chiar și cazuri cu conținut în română, rusă, ucraineană și engleză la un loc (19 campanii au avut această combinație), indicând o strategie de maximizare a razei de acțiune din regiune. Această abordare multilingvă sugerează că actorii acționau în cunoștință de cauză la nivel global, reambalând aceeași înșelătorie pentru diferite localități. De asemenea, complică atribuirea - o campanie multilingvă ar putea fi condusă de o echipă internațională sau de un singur actor cu competențe lingvistice diverse.

O altă observație se referă la segmentarea în funcție de gen în legătură cu tematica promovată. Astfel, reclamele referitoare la investiții financiare aveau ca public țintă bărbații cu vârsta între 35și 64 de ani, pe când reclamele la tratamentele miraculoase vizau femeile între 35 și 64 de ani.

Tactici de legitimitate: conform datelor de identificare, aproape toate conturile agenților de publicitate erau identități "neverificate" în cadrul rețelelor de publicitate, adesea fiind conturi sau pagini de Facebook nou create. Cu toate acestea, au existat și 14 cazuri cu identități "verificate" - ceea ce înseamnă că administratorii acestor conturi fie au compromis pagini/conturi Facebook legitime, fie au manipulat procesul de verificare a identității de către platforma de publicitate. Un alt model este acela că mai multe conturi false s-au dat drept companii sau ONG-uri sau au trecut de testul timpului prin maturizarea conturilor de tip sock-puppet (este modul prin care operatorii pot lăsa conturile inactive pentru a evita suspiciunea asupra conturile nou create în cadrul unei campanii programate la un interval mai mare de timp). Această răbdare operațională este revelatoare pentru eforturile de influență coordonate.

Detectarea și eliminarea: modul în care au fost descoperite conturile false care răspândeau reclamele deepfake indică o eficiență a observării directe de către un operator uman, în detrimentul detecției automate a platformei și de către organismele de supraveghere externe sau de comunitate. Instrumentele de transparență a platformelor impuse de reglementările europene au ajutat la aplicarea grilei de analiză, dar în mod clar nu sunt suficiente pentru a preveni difuzarea reclamelor înșelătoare. Modelul operațional în acest caz tinde mai degrabă spre eliminarea reactivă decât spre prevenirea proactivă.

În concluzie, corelarea analizei de conținut cu datele operaționale conturează imaginea unei operațiuni de influență bine organizate și persistente, în cadrul căreia s-a distribuit conținut sofisticat de tip deepfake pe platforme relevante, țintind audiențe largi de public. Operatorii au acționat pe scară largă, adesea la nivel internațional, și și-au adaptat

strategiile (în ceea ce privește limbajul, tematica și tehnica) la diferite contexte. Utilizarea deepfake-urilor și a persuasiunii psihologice a făcut parte dintr-un cadru sofisticat de tactici, tehnici și proceduri: de la inginerie socială (impostură, autoritate) la exploatări tehnice (phishing, domenii false) și la securitate operațională (acoperirea urmelor). Acest lucru demonstrează modul în care anunțurile înșelătoare de tip deepfake combină elemente ale campaniilor de dezinformare și ale criminalității informatice, estompând granița dintre dezinformare și fraudă.

Constatările acestei cercetări subliniază necesitatea unor contramăsuri la acel nivel de sofisticare, într-o manieră care să combine aplicarea politicilor platformelor, educarea utilizatorilor cu privire la amenințările informaționale de tip deepfake și cooperarea internațională pentru depistarea actorilor transfrontalieri implicați.

În esență, reclamele deepfake studiate reprezintă un exemplu elocvent de amenințare emergentă la adresa spațiului informațional, cu daune financiare și cognitive la nivelul societății. Prin structura lor, combină ce e mai nociv din două lumi: *fake news* și *scam-urile online*, potențate de instrumente digitale moderne de propagare. Tendința observată este de creștere a sofisticării, iar datoria securității informaționale este să anticipeze pericolul pe care aceste reclame îl reprezintă, mai ales atunci când tehnologia îngreunează modul în care comunicările legitime pot fi deosebite de cele false (măsură ce tehnologia *deepfake* avansează și devine accesibilă în masă). În acest context, cunoașterea tacticilor evidențiate în această analiză este esențială pentru a putea dezvolta contra-măsuri eficiente și pentru a educa publicul să nu pice în capcanele cibernetice.

Rezultatele acestei cercetări dovedesc că reclamele deepfake nu sunt simple “reclame mincinoase”, ci fac parte dintr-un mix toxic de dezinformare, fraudă cibernetică și abuz de imagine, care necesită răspuns pe măsură. Prin înțelegerea *operatorilor* care produc conținutul promovat și *modul în care acesta ajunge în atenția persoanelor susceptibile să îl acceseze*, măsurile de combatere pot fi mai eficiente. Această analiză a scos la lumină tiparele de acțiune pentru ca aceste cunoștințe să fie folosite proactiv în demersurile de protejare a spațiului informațional.

VIITOAREA AGENDĂ DE CERCETARE

Cercetarea de față poate constitui o bază pentru dezvoltarea unor **sisteme de avertizare timpurie** pornind de la parametri operaționali rezultați în urma corelațiilor dintre variabilele studiate. **Baza de date** obținută poate contribui la extinderea analizei despre infrastructura partajată (care include ID-uri de pe rețelele sociale, IP-uri și nume de domeniu

promovate) prin crearea unui flux al amenințării în platforma de cyber threat intelligence **Open CTI** (contribuție la integrarea datelor în investigațiile comunității de securitate cibernetică). De asemenea, concluziile pot ajuta la **conceperea unei politici de conștientizare publică a amenințării generată de publicitatea falsă**, care să vină în sprijinul entităților responsabile de creșterea nivelului de cultură de securitate. Clasificarea amenințării generată de reclamele deepfake ca instrument al războiului cognitiv atrage după sine nevoia de popularizare a **conceptului de securitate cognitivă** la nivelul Sistemului Național de Apărare prin implementarea unui program cu măsuri de protecție internă.

RECOMANDĂRI PENTRU COMUNICAREA PUBLICĂ

Concluziile acestei cercetări sunt relevante în mod special pentru specialiștii din domeniul comunicării publice, oferind argumente privind necesitatea unei abordări integrate a elementelor de securitate cibernetică în dezvoltarea culturii media și a educației digitale. Un răspuns adecvat în fața amenințărilor cognitive bazate pe reclamele false se poate construi pe o perspectivă tridimensională: umană, tehnică și instituțională, regăsită în conceptul de securitate cognitivă.

Din perspectiva umană, poate fi dezvoltată dimensiunea legislativă și de reglementare și cea de comunicare publică. Ajustarea legislației ar putea presupune **introducerea unor norme clare privind răspunderea platformelor** pentru difuzarea de reclame deepfake, **etichetarea obligatorie** a conținutului sintetic și m**ecanisme de raportare rapidă** a reclamelor suspecte de către utilizatori, cu timpi de reacție stabiliți prin lege.

Optimizarea comunicării publice ar implica derularea unor campanii de conștientizare a pericolului generat de reclamele deepfake, pentru a putea fi mai ușor de recunoscut de către utilizatori, și încurajarea dezmințirii publice cu reacții prompte din partea brandurilor și persoanelor publice a căror imagine este folosită în mod abuziv.

Cooperarea tehnică pe tema combaterii publicității deepfake ar presupune **standardizarea internațională** pentru detectarea automată a conținutului deepfake (similar cu standardele pentru spam sau malware din domeniul securității cibernetice), **baze de date comune** între state și companii tech pentru semnalarea conturilor și a domeniilor false și **parteneriate public-privat** pentru dezvoltarea unor instrumente open-source de identificare a deepfake-urilor și a rețelelor de conturi false existente pe platformele digitale. Fiind un fenomen infracțional fără granițe, colaborarea internațională în combaterea fraudelor cibernetice ar trebui să-și concentreze atenția și pe indicatorii oferiți de publicitatea online pentru a stopa activitatea rețelelor umane din spatele celor virtuale.