

Definiții ale termenilor cheie din prezentarea prof. dr. Marten Risius

Definiții bazate pe informații de la Google AI și ChatGPT

Camere de ecou

Spații digitale (de ex., grupuri pe rețele sociale) în care oamenii văd aproape exclusiv informații care le confirmă propriile convingeri; astfel se amplifică dezinformarea. Camerele de ecou îi reunesc pe cei cu opinii similare și limitează schimbul cu „exteriorul”.

Doxing (doxare)

Publicarea pe internet a datelor personale ale unei persoane (de ex., adresă, număr de telefon), adesea ca mijloc de intimidare sau pedepsire.

Vigilantism digital (justiție pe cont propriu)

„Justiție” administrată de utilizatori pe internet — când aceștia îi pedepsesc singuri pe alții (de ex., prin atacuri în masă, shaming sau ură), în loc să apeleze la instituții legitime.

Atacuri organizate cu ransomware

Atacuri ale unor grupuri de hackeri care criptează date și cer răscumpărare — folosite și pentru sabotarea mass-media și a canalelor informaționale.

Restricționare din umbră (shadow banning)

Limitarea invizibilă a conturilor sau postărilor fără ca utilizatorii afectați să observe (de ex., reducerea razei de acoperire), pentru a frâna dezinformarea.

Chatboți cu contra-narative

Chatboți bazați pe inteligență artificială care răspund informațiilor false cu contraargumente sau verificări ale faptelor.

Detoxificare bazată pe LLM

Utilizarea modelelor lingvistice mari (LLM) pentru a detecta și reduce conținutul toxic sau manipulator ori pentru a-l rescrie într-un ton mai neutru.

Prebunking (demontare preventivă)

Educație preventivă care informează dinainte despre strategiile tipice ale știrilor false, astfel încât oamenii să recunoască mai ușor manipularea.

Agenți IA fiducieri

Sisteme de inteligență artificială concepute să acționeze în interesul utilizatorului (de ex., filtre împotriva dezinformării), asemenea unui „fiduciar digital”.

Red herring (pistă falsă)

O abatere în argumentare — se introduce un subiect irelevant pentru a devia atenția de la problema reală sau de la știrea falsă.

Cherry-picking (selectare părtinitoare)

Alegerea intenționată doar a elementelor care convin cel mai mult, ignorând restul.

Dihotomii false

Erori logice în care o situație este prezentată greșit ca o alegere doar între două opțiuni extreme și exclusiviste, deși există mai multe posibilități și nuanțe.

Ad hominem

Tactică/eroare retorică în care este atacată persoana, nu ideea, pentru a slăbi argumentele adversarului și a influența publicul.

Polarizare

În context politic — fie o diferențiere socială care duce la conflict, fie accentuarea divergențelor de opinie. Adesea cele două se suprapun.

Imitație

Imitarea sau copierea comportamentului, stilului ori conținutului altora.

Argumente „pantă alunecoasă”

Argumente retorice care susțin că o acțiune inițială aparent inofensivă (A) duce inevitabil la un lanț de consecințe negative, culminând cu un rezultat moralmente inacceptabil (B).

Decontextualizare

Înlăturarea deliberată a informațiilor, textelor, imaginilor sau videoclipurilor din contextul lor inițial pentru a le da un sens nou, adesea înșelător.

Troll

O persoană — sau uneori un program (bot) — care publică intenționat conținut provocator, ofensator sau înșelător pentru a perturba discuțiile și a provoca sau hărțui alți utilizatori.

Boți

Programe automate care execută în mod autonom sarcini repetitive mai rapid decât un om.

Clonare

Crearea de copii identice.

Domenii

Un domeniu este un nume unic, ușor de citit de către oameni, care funcționează ca adresă pentru localizarea unui site, în locul unui șir numeric greu de reținut (adresa IP).

Gatekeeper (filtru de acces)

O persoană care controlează accesul la ceva (de ex., paznicul de la intrarea într-un club) sau, mai abstract, controlează cine obține acces la o categorie ori un statut.