Revista Universul Juridic ● nr. 10, octombrie 2023, pp. 69-77

INTELIGENȚA ARTIFICIALĂ – CE PROBLEME ATRAGE LIPSA CADRULUI LEGISLATIV

Av. Raluca ANDERCO

Abstract

The age of generative AI is here: only six months after OpenAI's ChatGPT burst onto the scene, as many as half the employees of some leading global companies are already using this type of technology in their workflows, and many other companies are rushing to offer new products with generative AI built in.

But, as those following the burgeoning industry and its underlying research know, the data used to train the large language models (LLMs) and other transformer models underpinning products such as ChatGPT, Stable Diffusion and Midjourney comes initially from human sources — books, articles, photographs and so on — that were created without the help of artificial intelligence. ace with Generative AI.

Now, as more people use AI to produce and publish content, an obvious question arises: What happens as AI-generated content proliferates around the internet, and AI models begin to train on it, instead of on primarily human-generated content?

Keywords: *AI, deepfake, artificial intelligence, intellectual property rights, European regulations*

Epoca Inteligenței Artificiale generativă este aici: la doar şase luni după ce ChatGPT a apărut pe scenă, până la jumătate din angajații unor companii de top la nivel mondial folosesc deja acest tip de tehnologie în fluxurile lor de lucru şi multe alte companii se grăbesc să ofere produse noi cu Inteligența Artificială generativă încorporată, sau să integreze AI în produse (a se vedea situația ADOBE, precum şi recenta campanie COCA COLA).

Folosită corect, AI poate aduce reale beneficii, însă este important de subliniat câteva aspecte care pot ridica reale probleme.

Ştirile recente din SUA arată că vocile cântăreților celebri sunt reproduse din ce în ce mai uşor şi mai des, cu ajutorul AI (e.g. Kanye West, Drake), situația care ridică următoarea întrebare:

Cui se vor plăti drepturile de autor pentru o melodie ce reproduce vocea unui artist celebru? Artistului sau celor care reproduc vocea cu AI?

O altă consecință a AI este fenomenul deepfake, prin care i se încalcă victimei drepturile fundamentale ale persoanei fizice: dreptul la viață privată, dreptul la demnitate, dreptul la propria imagine, dreptul la confidențialitate.

Pe măsură ce tot mai mulți oameni folosesc inteligența artificială pentru a produce şi a publica, apare încă o întrebare evidentă:

Ce se întâmplă pe măsură ce conținutul generat de inteligența artificială proliferează pe internet?

* * *

Epoca Inteligenței Artificiale (denumită AI, în continuare) este o realitate, iată că, la doar câteva luni după ce ChatGPT a apărut, până la jumătate din angajații unor companii de top la nivel mondial folosesc deja acest tip de tehnologie în fluxurile lor de lucru şi multe alte companii se grăbesc să ofere produse noi cu Inteligența Artificială, sau să integreze AI în produse (spre exemplu această situație se regăseşte la compania ADOBE, precum şi în recenta campanie COCA COLA).

Folosită corect, AI poate aduce reale beneficii, însă este important de subliniat câteva aspecte care pot ridica reale probleme.

Ştirile recente din SUA arată că vocile cântăreților celebri sunt reproduse din ce în ce mai uşor şi mai des, cu ajutorul AI (e.g. Kanye West, Drake), situația care ridică următoarea întrebare:

Cui se vor plăti drepturile de autor pentru o melodie ce reproduce vocea unui artist celebru, artistului sau celor care reproduc vocea cu AI? Cine reglementează şi stabileşte întinderea drepturilor de proprietate intelectuală?

Prin conținutul său, AI poate aduce atingere inclusiv drepturilor fundamentale ale persoanei, prin reproducerea unui conținut discriminatoriu sau a înfățişării sale fizice, a vocii sale sau utilizarea unei asemenea reproduceri,

O altă consecință a AI este fenomenul deepfake[1], prin care i se încalcă victimei drepturile fundamentale ale persoanei fizice: dreptul la viață privată, dreptul la demnitate, dreptul la propria imagine, dreptul la confidențialitate.

[1] Deepfake este o formă a Inteligenței Artificiale (AI), ai cărei algoritmi[1] sunt folosiți pentru a schimba fețele oamenilor din conținuturile digitale şi a crea, astfel, un fals cu aspect realist. Dacă o persoană vrea să pună anumite cuvinte în gura unui personaj, acest lucru devine posibil prin crearea unui deepfake.

Pe măsură ce tot mai mulți oameni folosesc inteligența artificială pentru a produce şi a publica, apare încă o întrebare evidentă:

Ce se întâmplă pe măsură ce conținutul generat de inteligența artificială proliferează pe internet?

Un grup de cercetători din Marea Britanie şi Canada au analizat chiar această problemă şi au publicat recent o lucrare despre munca lor în jurnalul cu acces deschis. Ceea ce au descoperit este îngrijorător pentru tehnologia AI actuală şi viitorul acesteia, astfel că se consideră că utilizarea conținutului generat de modele în cursuri de instruire provoacă pagube ireversibile.

Apare aşa zisul fenomen de „Umplere a internetului cu bla bla", adică aşa cum am împrăştiat oceanele cu gunoi de plastic şi am umplut atmosfera cu dioxid de carbon, aşa suntem pe cale să umplem internetul cu balast.

Această „poluare" cu date generate de inteligență artificială are ca rezultat că modelele să obțină o percepție distorsionată a realității. Chiar şi atunci când cercetătorii au instruit modelele să nu producă prea multe răspunsuri repetate, ei au descoperit că încă s-a produs colapsul modelului, deoarece modelele ar începe să inventeze răspunsuri eronate pentru a evita repetarea datelor prea des.

Privind în mod specific distribuțiile de probabilitate pentru modelele AI, text-la-text şi imagine-la-imagine, cercetătorii au concluzionat că „învățarea din datele produse de alte modele provoacă colapsul modelului - un proces degenerativ prin care, în timp, modelele uită adevăratul subiect".

„De-a lungul timpului, greşelile în datele generate se compun şi, în cele din urmă, forțează modelele care învață din datele generate să perceapă greşit realitatea şi mai departe", a scris unul dintre autorii principali ai lucrării, Ilia Shumailov, într-un e-mail către VentureBeat: „Am fost surprinşi să observăm cât de repede se produce colapsul modelului: modelele pot uita rapid majoritatea datelor originale din care au învățat inițial".

Cu alte cuvinte, pe măsură ce un model de antrenament al AI este expus la mai multe date generate de AI, are performanțe mai slabe în timp, producând mai multe erori în răspunsurile şi conținutul pe care îl generează şi producând mult mai puțină varietate neeronată a răspunsurilor sale.

Ted Chiang, aclamat autor SF - „Povestea vieții tale", roman care a inspirat filmul Arrival şi Scriitor la Microsoft, a publicat recent un articol în The New Yorker postulând că, toate copiile AI ale copiilor ar avea ca rezultat o calitate degradată, asemănând problema cu artefactele crescute vizibile atunci când copiați o imagine JPEG în mod repetat.

Un alt mod de a gândi problema este ca în filmul de comedie SF Multiplicity din 1996, cu Michael Keaton, în care un om umil se clonează pe sine şi apoi clonează clonele, fiecare dintre acestea având ca rezultat scăderea exponențială a nivelurilor de inteligență şi creşterea prostiei.

Există multe alte aspecte care vor duce la implicații mai grave, cum ar fi **discriminarea bazată pe gen, etnie sau alte atribute sensibile**, a spus Shumailov, mai ales dacă AI învață în timp să producă, să zicem, o rasă în răspunsurile sale, în timp ce „uitarea" altora există.

Din fericire, există modalități de a evita prăbuşirea modelului, cercetătorii evidențiază două moduri specifice.

Primul este prin păstrarea unei copii de prestigiu a setului de date original produs exclusiv sau nominal de oameni şi evitarea contaminării cu date generate de AI. Apoi, modelul ar putea fi reantrenat periodic pe aceste date, sau reîmprospătat în întregime cu acestea, începând de la zero.

A doua modalitate de a evita degradarea calității răspunsului şi de a reduce erorile sau repetările nedorite de la modelele AI este introducerea de seturi de date noi, curate, generate de oameni, înapoi în formarea lor.

Cu toate acestea, aşa cum subliniază cercetătorii, acest lucru ar necesita un fel de mecanism de etichetare în masă sau efort din partea producătorilor de conținut sau a companiilor de inteligență artificială pentru a face diferența între conținutul generat de inteligență artificială şi conținutul generat de oameni.

În prezent, nu există un astfel de efort de încredere sau pe scară largă online.

„Pentru a opri prăbuşirea modelului, trebuie să ne asigurăm că grupurile minoritare din datele originale sunt reprezentate corect în seturile de date ulterioare", a spus Shumailov pentru VentureBeat, continuând:

„În practică, este complet non-trivial. Datele trebuie să fie copiate cu atenție şi să acopere toate cazurile de colț posibile. În evaluarea performanței modelelor, utilizați datele pe care se aşteaptă să lucreze modelul, chiar şi cele mai improbabile cazuri de date. Rețineți că acest lucru nu înseamnă că datele improbabile ar trebui să fie supraeşantionate, ci mai degrabă că ar trebui reprezentate în mod corespunzător. Pe măsură ce progresul vă determină să vă reeducați modelele, asigurați-vă că includeți date vechi şi noi. Acest lucru va creşte costul instruirii, dar vă va ajuta să contracarați colapsul modelului, cel puțin într-o anumită măsură."

Ce pot face industria AI şi utilizatorii în acest sens?

În timp ce toate aceste ştiri sunt îngrijorătoare pentru tehnologia actuală de inteligență artificială şi pentru companiile care încearcă să monetizeze cu ea, în special pe termen mediu şi lung, există o linie fină pentru creatorii de conținut uman: cercetătorii concluzionează că într-un viitor plin de gen, Instrumentele AI şi conținutul lor, conținutul creat de oameni vor fi şi mai valoroase decât este în prezent – chiar dacă doar ca sursă de date de antrenament impecabile pentru AI.

Aceste constatări au implicații semnificative pentru domeniul inteligenței artificiale, subliniind necesitatea unor metodologii îmbunătățite pentru a menține integritatea modelelor generative în timp.

La toate acestea se adaugă lipsa cadrului legislativ!

Un prim pas a fost făcut, şi iată că Parlamentul şi Consiliul Uniunii Europene au propus un Regulament privind inteligența artificială („Regulamentul" sau „AI Act").

În contextul progreselor recente, scopul Regulamentului este reglementarea şi implementarea unor mecanisme de supraveghere şi control a sistemelor de inteligență artificială („AI").

Regulamentul este structurat în funcție de următoarele principii care stau la baza redactării lui:

a) Riscurile sistemelor AI: Regulamentul clasifică sistemele de AI în funcție de nivelul de risc asociat utilizării lor ca: (i) inacceptabil, (ii) risc mare, (iii) risc mediu şi (iv) risc mic. O astfel de abordare permite stabilirea unor cerințe şi obligații diferite pentru diferite tipuri de sisteme de AI, în funcție de potențialul lor de a aduce atingere drepturilor fundamentale şi siguranței cetățenilor.

b) Responsabilitatea furnizorilor de sisteme de AI: Regulamentul stabileşte obligații şi responsabilități clare pentru furnizorii de sisteme de AI, în scopul de a asigura conformitatea cu principiile etice şi normele legale în materie de AI. Acest lucru include, de exemplu, obligația de a realiza o evaluare a conformității, a furniza documentație adecvată şi a păstra fişierele de jurnalizare.

c) Protecția drepturilor fundamentale şi a intereselor publice: Unul dintre obiectivele principale ale AI Act este protejarea drepturilor fundamentale ale cetățenilor şi promovarea intereselor publice, precum protecția datelor personale, nediscriminarea şi transparența. Prin stabilirea unor cerințe stricte pentru sistemele de AI cu risc mare şi interzicerea celor cu risc inacceptabil, regulamentul vizează prevenirea abuzurilor şi asigurarea că sistemele de AI respectă valorile şi principiile democratice ale UE.

d) Cooperarea între statele membre: AI Act promovează cooperarea între statele membre ale UE în domeniul reglementării AI şi încurajează schimbul de bune practici şi informații. Acesta prevede, de asemenea, înființarea unor autorități naționale de supraveghere şi crearea unui Comitet European pentru Inteligență Artificială, pentru a asigura aplicarea coerentă a regulamentului în întreaga UE.

În esență, AI Act urmăreşte să creeze un cadru legal armonizat pentru sistemele de AI utilizată pe teritoriile statelor membre, care să protejeze cetățenii şi consumatorii, fără a afecta inovația şi dezvoltarea acestui domeniu, care pot fi împiedicate de o reglementare excesivă.

Propunerea de Regulament se va aplica următoarelor categorii de destinatari:

a) furnizorilor care introduc pe piață sau pun în funcțiune sisteme de AI în Uniune, indiferent dacă aceştia furnizori se află fizic sau sunt stabiliți în cadrul Uniunii sau într-o țară terță;

b) utilizatorilor de sisteme AI care se află fizic sau sunt stabiliți în cadrul Uniunii;

c) furnizorilor şi utilizatorilor de sisteme AI care se află fizic sau sunt stabiliți într-o țară terță, în cazul în care rezultatele produse de sistem sunt utilizate în Uniune;

d) importatorilor şi distribuitorilor de sisteme AI;

e) producătorilor de produse care introduc pe piață sau pun în funcțiune un sistem de AI împreună cu produsul lor şi sub propriul nume sau marcă;

f) reprezentanților autorizați ai furnizorilor, care sunt stabiliți în Uniune.

Nu este surprinzător faptul că legiuitorul european a ales să se concentreze pe creatorii, dar şi comercianții acestor sisteme, care oferă acces publicului la acestea prin intermediul **API**-urilor („**Application Programming Interface**").

API-urile reprezintă un set de reguli prin care publicul poate interacționa cu sistemele de inteligență artificială. În acest mod, de exemplu, Chat GPT poate accesa internetul. Însă în lipsa acestor interfețe, ChatGPT nu are acces direct la internet atunci când este folosit de cetățeni. Alte sisteme de inteligență artificială aflate în competiție cu ChatGPT, precum sistemul Bard creat de Google, au acces direct la internet atunci când interacționează cu consumatorii.

Clasificarea riscului sistemelor de AI în AI Act are implicații semnificative pentru dezvoltatorii de software, dar şi pentru utilizatorii de sisteme de AI. Este important de menționat că, în ceea ce priveşte sistemele de AI cu risc inacceptabil, Regulamentul impune o interdicție privind dezvoltarea sau utilizarea lor de către consumatori.

Documentul atenționează dezvoltatorii că ar trebui să fie conştienți de implicațiile etice şi legale ale proiectelor lor şi să ia măsuri pentru a preveni crearea unor sisteme care ar putea fi considerate cu risc inacceptabil.

Dezvoltatorii de software care lucrează la sistemele de AI cu risc mediu sau mic nu sunt supuşi unor cerințe specifice, dar trebuie să respecte în continuare legislația UE şi națională existentă. Dezvoltatorii ar trebui să fie conştienți de impactul potențial al sistemelor lor de AI asupra siguranței şi drepturilor fundamentale şi să ia în considerare aceste aspecte în procesul de dezvoltare.

Propunerea de regulament are în vedere sisteme de inteligență artificială folosite în domenii deja supuse reglementări, precum aviația, automobilele, bărcile, ascensoarele, echipamentele medicale sau utilaje industriale. În aceste situații, cerințele pentru AI cu risc ridicat sunt integrate în procesul actual de evaluare a conformității, realizat de autoritățile de reglementare sectoriale, precum Agenția

Europeană pentru Siguranța Aviației pentru avioane, sau printr-o combinație de organizații terțe aprobate şi o entitate centrală a UE, aşa cum se întâmplă în cazul dispozitivelor medicale. Astfel, companiile din afara UE care comercializează produse reglementate în spațiul european se supun deja procedurii de evaluare a conformității. AIA modifică doar detaliile acestui control, dar nu şi domeniul sau procesul în sine.

Principala critică are în vedere impactul pe care acest Regulament îl poate avea asupra sistemelor open-source. Un sistem open-source este program sau o tehnologie a cărei sursă de cod este deschisă publicului şi poate fi utilizată, modificată sau distribuită de oricine, fără restricții. Criticii AI Act susțin că există un risc ca cerințele impuse asupra dezvoltatorilor acestor sisteme ar putea crea o povară administrativă şi financiară pentru dezvoltatorii open-source, care de multe ori lucrează în mod voluntar şi nu dispun de resursele necesare pentru a se conforma acestor standarde. Dacă o companie ar implementa un sistem de IA open-source care ar duce la un rezultat dezastruos există un risc semnificativ ca aceasta să încerce să îşi decline responsabilitatea dând în judecată dezvoltatorii open-source care au construit produsul.

Pentru că este direct vizat de acest Regulament, după ce i-am furnizat definiția IA cu risc ridicat, l-am întrebat pe Chat GPT dacă consideră că se încadrează în această categorie. Iată răspunsul: „*Conform AI Act, sistemele de IA cu risc inacceptabil sunt cele care pot provoca daune care nu pot fi compensate prin beneficiile obținute şi care pot afecta grav siguranța sau drepturile fundamentale ale persoanelor. Exemplele furnizate de AI Act includ sistemele de IA care permit «practicarea manipulării comportamentale» sau sistemele care permit «evaluarea socială» într-un context în care evaluarea nu este permisă în conformitate cu dreptul Uniunii sau al dreptului intern.*

Chat GPT este un model de limbaj avansat dezvoltat de OpenAI, care poate genera texte coerente şi relevante în funcție de instrucțiunile primite. Deşi poate fi utilizat într-o varietate de aplicații, inclusiv asistență virtuală, redactare şi generare de conținut, nu este conceput în mod specific pentru a manipula comportamentul oamenilor sau pentru a evalua indivizii într-un context social".

AI Act prevede şi o serie de sancțiuni semnificative pentru încălcarea obligațiilor legale impuse destinatarilor. Violarea unei practici interzise ar putea atrage amenzi de până la 40 de milioane de euro sau 7% din venitul global anual al unei companii, în funcție de care este mai mare, în creştere de la 30 de milioane de euro sau 6% din venitul anual global. Astfel de sancțiuni depăşesc semnificativ gama de amenzi existente în domeniul GDPR, care ajung până la 4% din venitul global al unei companii. Sancțiunile pentru furnizorii de modele de bază care încalcă Legea AI ar putea ajunge până la 10 milioane de euro sau 2% din venitul anual, în funcție de care este mai mare.

În plus, față de sancțiunile financiare, pot exista şi alte tipuri de sancțiuni. De exemplu, companiile care încalcă AI Act ar putea fi interzise de la aplicarea ulterioară a inteligenței artificiale în anumite contexte, ar putea fi supuse unor revizuiri

mai frecvente şi mai riguroase din partea autorităților, sau ar putea fi forțate să elimine sau să modifice aspecte specifice ale sistemelor lor AI.

În cazul în care o încălcare a Legii AI are ca rezultat daune materiale sau fizice, companiile pot fi, de asemenea, obligate să acorde despăgubiri părților afectate. În cazul în care încălcarea a fost intenționată sau a rezultat din neglijență gravă, aceasta ar putea conduce şi la sancțiuni penale, în funcție de legislația specifică a statului membru.

Acest set de sancțiuni reflectă seriozitatea cu care atât Comisia Europeană, cât şi Parlamentul European percep potențialele pericole pe care inteligența artificială le poate aduce societății umane. Aceste măsuri reprezintă un pas semnificativ în asigurarea faptului că tehnologia emergentă nu numai că avansează, dar o face într-un mod care păstrează interesele şi drepturile fundamentale ale oamenilor în centrul dezvoltării şi aplicării sale.

AI Act reprezintă un pas important în reglementarea sistemelor de IA în UE, care încearcă, cel puțin declarativ, să găsească un echilibru între inovație şi protecția drepturilor fundamentale. Clasificările de risc şi consecințele practice ale acestora pentru dezvoltatori şi utilizatori au scopul de a crea unui mediu de încredere şi transparență în utilizarea sistemelor de IA, promovând responsabilitatea şi conformitatea cu principiile etice şi normele legale. Rămâne de văzut dacă aceste mecanisme legislative de control vor fi suficient de puternice să reducă sau chiar să stopeze potențialele efecte negative ale acestor sisteme. Prin introducerea unui cadru legal pentru clasificarea sistemelor de IA în funcție de nivelul de risc, AI Act încurajează dezvoltatorii şi utilizatorii să adopte o abordare responsabilă şi transparentă în procesul de creare, implementare şi utilizare a tehnologiilor bazate pe IA.

O analiză recentă a Stanford University, referitoare la perioada 2016–2022, releva că s-a început cu o lege cu rezonanță în materie şi s-a ajuns deja la 37. Prezența AI în procedurile reglementate din 81 țări a crescut aproape de şapte ori. În capul listei se află Spania cu 273 de mențiuni de gen, urmată de Canada cu 211, Marea Britanie cu 140 şi SUA cu 138. Şi toate acestea în condițiile în care se aşteaptă cadrul juridic pertinent al UE considerat cel mai complet şi adecvat domeniului. Aşadar, a început drumul spre o reglementare adecvată a aplicațiilor AI şi toată lumea priveşte spre marile puteri economice în căutarea şi urmarea unei orientări generale, care să evite discrepanțele şi neconcordanțele reglementare care ar afecta grav comerțul internațional prin distorsionarea competitivității între țări şi marile blocuri economice.

În data de 16.05.2023, Sam Altman, PDG al OpenAI, creatorul robotului conversațional ChatGPT, a fost audiat în congresul S.U.A., de membrii comisiei pentru probleme juridice a Senatului, într-un exercițiu mai degrabă politic şi mediatic destinat pentru uzul opiniei publice asupra inteligenței artificiale şi a posibilelor sale implicații politico-juridice.

Unele din problemele abordate în cadrul audierii sunt în dezbatere şi în cadrul Uniunii Europene, ca de pildă: în caz de daună, care e responsabilitatea fabricanților de softuri precum OpenAI? Aceasta trebuie să fie limitată, precum cea a rețelelor sociale care adăpostesc conținuturi, sau orice cetățean trebuie să poată urmări un editor de AI.

În timp ce Parlamentul European examinează Artificial Intelligence Act, titularii de drepturi de proprietate intelectuală îşi manifestă din ce în ce mai mult îngrijorările în privința bulversărilor pe care le-ar putea angaja AI privind profesiile de creatori. Potrivit Comisiei Europene, acestea nu ar avea temei, întrucât chiar dacă a crescut „complexitatea şi importanța interacțiunii între inteligența artificială şi dreptul de autor", ea a ajuns totuşi la concluzia că dreptul actual asigură un echilibru necesar. S-a apreciat în acest sens că excepțiile prevăzute de directiva (UE) 2019/790 privind dreptul de autor şi drepturile conexe în piața unică digitală vizând căutarea de texte şi de date „sunt pertinente în contextul IA" şi asigură „un echilibru între două elemente: a proteja titularii drepturilor, în special artiştii, şi a facilita explorarea de texte şi de date, în special de către dezvoltatorii de IA".

În România, lipsa acută a unui cadrul legislativ care să reglementeze AI, are repercusiuni în materia răspunderii civile, în materia drepturilor de proprietate intelectuală şi a gravelor încălcări ale drepturilor fundamentale nepatrimoniale.

Într-un astfel de mediu, este dificil de anticipat întinderea prejudiciilor create de folosirea fără limite şi reglementări a AI.

Ne întoarcem astfel la normele generale de drept, depăşite în mod clar de realitatea AI. Ce e de făcut?

Se pune întrebarea dacă în fața diversității de atitudini şi strategii de abordare a acțiunii pertinente din partea diferitelor state, nu s-ar impune totuşi ideea unui cadru unificat al reglementării AI, indiferent de dilemele şi impreviziunile pe care le implică. În aşteptarea marii revoluții juridice pe care o presupune cea ştiințifico-existențială la începuturile căreia asistăm deja, se poate recurge la o abordare treptată care să presupună mai întâi actualizarea reglementărilor existente, urmată de noi dezvoltări normative bazate inclusiv şi mai ales pe colaborarea interstatală, aşa încât obiectivele să fie concertate pentru ca în acest mod să se atingă rezultatele eficiente şi relevante. Şi în cele din urmă să se ajungă la adoptarea unor norme şi reglementări comune, nediferențiate, ale inteligenței (umane şi artificiale), una sigură supusă unei guvernanțe responsabile şi eficiente. Prin urmare, un cod juridic (universal) al inteligenței terestre, unul deja al prezentului şi mai ales al viitorului!