Alin Popescu

AI Act
Viitorul deciziilor noastre

Reguli explicate, studii de caz şi exemple practice, pe înţelesul tuturor

Universul Juridic
Bucureşti
-2024-

Cuprins

Introducere

Inteligența artificială va schimba multe lucruri, în bine, în societatea și ecosistemul planetei noastre. Pentru prima dată în istorie, am putea intra într-o eră de aur, în care resursele vor fi distribuite mai eficient, bolile vor fi detectate și tratate mai rapid, iar cunoașterea va fi accesibilă tuturor.

Nu e nicio îndoială că AI va juca un rol crucial în combaterea schimbărilor climatice. S-ar putea să fie singurul instrument pe care îl vom putea folosi, cu succes, în tratarea unei probleme fundamentale pentru planeta noastră. Orașele inteligente vor utiliza AI pentru a gestiona mai bine resursele, de la distribuția energiei până la gestionarea deșeurilor.

În cercetare și inovare, AI va accelera descoperirile științifice. De la explorarea spațiului cosmic până la descoperirea de noi materiale, inteligența artificială va deschide orizonturi noi, rezolvând probleme complexe într-un ritm fără precedent.

Cred cu tărie toate aceste lucruri. Cred că AI va fi singurul prieten pe care unii oameni l-ar putea avea, în momente dificile, mâna de ajutor pe care un bătrân o așteaptă la greu, umărul pe care mulți își vor plânge necazurile, majordomul perfect pentru cei ocupați și intermediarul care ne va ajuta, pentru prima dată în istoria noastră, să vorbim cu alte viețuitoare de pe Pământ.

Însă, pe măsură ce pășim în această eră de aur, trebuie să ne păstrăm ochii larg deschiși. Beneficiile inteligenței artificiale vin la pachet cu responsabilități juridice, etice și sociale. Această carte spune povestea luptei permanente dintre dorințele și coșmarurile noastre. Dintre lumea imaginată de oameni și lumea în care oamenii sunt simpli participanți.

Realitatea pe care o știm noi, astăzi, va fi diferită în doi, trei, cinci ani de acum înainte. Poate nu ne va fi evident acest fapt, dar el se va întâmpla.

Cu cât vom accepta realitatea mai repede, cu atât ne vom putea adapta mai uşor la viitor.

În vremurile care vin, ar fi bine să ne amintim un lucru simplu: mai de folos ne sunt câteva cunoştinţe temeinice decât o mulţime de informaţii superficiale. De-a lungul timpului, am făcut astfel de alegeri de mai multe ori. Ne-am transformat puterile şi voinţa, pe măsură ce tehnologia ne-a influenţat modul de viaţă.

Când a apărut limbajul vorbit, am învăţat acel „cod" straniu pe care îl presupune deprinderea vorbirii. A fost, dacă vreţi, una dintre primele noastre aventuri în lumea imaginară. La fel ca în cazul unui mare model de limbaj, a cărui poveste o vom spune în primele capitole ale cărţii, inventarea vorbirii ne-a obligat să creăm, pe lângă cuvinte, şi contextul general necesar pentru înţelegerea lor. Cuvântul „foc" a devenit o idee abstractă. Vorbeam despre ceva ce nu ne putea arde, încălzi ori pregăti mâncarea.

Când s-a inventat scrierea, am renunţat treptat la tradiţia veche a povestirilor „vorbite", transmise din generaţie în generaţie, din bătrân în bătrân, ca un semnal de alarmă ori amintire a unor vremuri legendare. Odată cu scrierea, miturile s-au transformat în stană de piatră, la fel ca eroii ghinionişti din vechime, şi au fost închise între paginile unor cărţi care nu le mai permit să trăiască, să evolueze şi să devină mai puternice. Am câştigat ceva extraordinar, dar am şi pierdut ceva preţios.

Când a apărut tiparul, când cartea s-a răspândit ca focul, a murit ceva din partea supranaturală a vieţii noastre. Ne-am despărţit de unii zei, de unele frici şi de multe năluci. În schimb, am ridicat bârfa şi vorbitul pe la spate la rang de artă. Am eliberat puterea şi am democratizat-o mai mult decât oricând până atunci. Ne-am reconstruit viaţa în jurul unor poveşti pe care cărţile le-au purtat în toate colţurile lumii. Am făcut toate aceste lucruri şi, ca în alte momente ale trecutului nostru, am primit ceva şi am dat ceva la schimb.

Acest dans al progresului şi al sacrificiului s-a repetat de nenumărate ori în istoria noastră: când am inventat electricitatea, telegraful, telefonul, radioul ori televiziunea, când am inventat internetul ori când ne-am dus în spaţiu. Dacă te gândeşti la toţi aceşti paşi şi la toate cele care s-au întâmplat după fiecare dintre ei, vei vedea că roata vieţii toarce la fel, împletind într-o ţesătură complexă lucruri pe care le primim şi lucruri de care ne despărţim.

Da, inteligenţa artificială e un pas nou pe o scară ce urcă fără oprire. De fapt, probabil c-ar fi corect să spun că e un pas în evoluţia planetei noastre, ca ecosistem, pentru că implicaţiile ei se vor simţi dincolo de graniţele noastre, ca specie. Ce va trebui să decidem în anii următori va fi dacă aceasta este o evoluţie a tehnologiei noastre sau o evoluţie a noastră, ca specie umană. Distincţia dintre cele două idei e mai importantă decât pare la prima vedere.

De multă vreme am încetat să cred că inteligenţa artificială e doar o tehnologie. A început aşa, dar s-a transformat în altceva. Pe măsură ce trece timpul, ea devine reflexia unor nevoi şi dorinţe intime pe care noi, ca oameni, nu am putut niciodată să le exprimăm atât de clar şi limpede. E partenerul de viaţă pierdut demult, e fiica nenăscută ori prietenul din copilărie cu care nu am vorbit cât ne-am fi dorit. E versiunea noastră mai puternică, e sfătuitorul nostru, e consultantul ori profesorul de care am avut mereu nevoie. Cu abilitatea extraordinară de a mima inteligenţa, empatia ori cunoaşterea, AI-ul nu mai e, în mintea unui om obişnuit, doar o tehnologie. E speranţă. Şi, oricât de puternic ai fi, atunci când vrei să reglementezi speranţa, oamenii se vor împotrivi.

Cu această idee în minte, înţelegerea AI Act devine mai mult decât o simplă analiză a unei reglementări. Devine o explorare a modului în care societatea noastră alege să navigheze în aceste ape noi şi potenţial tulburi.

AI Act nu este doar un set de reguli tehnice. Este o oglindă a anxietăţilor, speranţelor şi valorilor noastre colective în faţa unei tehnologii care promite să redefinească graniţele dintre om şi maşină. Este o încercare de a trasa linii în nisip, ştiind foarte bine că marea le va şterge şi rescrie constant.

În paginile care urmează, vom descoperi împreună nu doar prevederile acestei legi, ci şi motivaţiile din spatele lor. Vom încerca să înţelegem de ce anumite practici sunt interzise, în timp ce altele sunt încurajate. Vom explora tensiunea constantă dintre inovaţie şi reglementare, dintre beneficiile promise de AI şi riscurile pe care le aduce cu sine.

Nu vom intra în detalii tehnice obositoare sau în jargon juridic complicat. În schimb, vom povesti. Vom folosi exemple din viaţa de zi cu zi, analogii familiare şi scenarii ipotetice pentru a aduce la viaţă conceptele abstracte ale AI Act.

Pe parcursul acestei călătorii, vom descoperi că AI Act nu are toate răspunsurile. Şi asta e în regulă. Scopul său nu este să ofere soluţii definitive, ci să creeze un cadru flexibil care să evolueze odată cu tehnologia.

Aşadar, dragă cititorule, pregăteşte-te pentru o călătorie fascinantă în lumea reglementării inteligenţei artificiale. Nu va fi mereu uşor, dar promit că va fi mereu interesant. Şi cine ştie? Poate că, la finalul acestei cărţi, vei înţelege nu doar mai multe despre AI Act, ci şi despre tine însuţi şi despre societatea în care trăim. Pentru că, în esenţă, AI Act nu este doar despre inteligenţa artificială. Este despre noi, oamenii, şi despre viitorul pe care alegem să ni-l construim împreună.

AI sau n-AI probleme?

Deşi puternice, aceste tehnici de antrenament pot duce la o varietate de probleme şi efecte negative dacă nu sunt aplicate cu atenţie.

Generarea de stereotipuri (bias) şi discriminare

Ambele sunt probleme critice în dezvoltarea şi aplicarea sistemelor de inteligenţă artificială. Acestea apar atunci când un model AI învaţă, reflectă şi chiar amplifică prejudecăţile şi stereotipurile prezente în datele folosite pentru a-l antrena.

Să spunem că o bancă foloseşte un sistem AI pentru a decide cine ar trebui să primească un împrumut. Modelul a fost antrenat pe date istorice despre împrumuturi anterioare. Cu toate acestea, dacă, în trecut, banca a avut tendinţa de a favoriza angajaţii din anumite industrii în deciziile de acordare a împrumuturilor (poate din cauza prejudecăţilor umane inconştiente sau a practicilor istorice de discriminare), atunci aceste prejudecăţi vor fi prezente în datele de antrenament.

Ca urmare, modelul AI poate învăţa să asocieze a fi lucrător în industria X cu o probabilitate mai mare de a primi un împrumut. Când este aplicat pentru a lua decizii despre noii solicitanţi de împrumut, poate favoriza, în mod sistematic, acel tip de angajaţi, chiar dacă alţi solicitanţi sunt la fel de calificaţi.

Acesta este un exemplu de stereotip algoritmic – când prejudecăţile din datele istorice duc la decizii discriminatorii ce sunt luate de sistemele AI. În timp, se poate crea un cerc vicios, în care prejudecăţile trecutului sunt perpetuate şi chiar amplificate. Prejudecăţile pot apărea pe baza oricărei caracteristici, de la vârstă şi statut socio-economic până la locaţie geografică şi mult mai mult. Şi pot afecta orice domeniu în care sunt

folosite sisteme AI pentru a lua decizii **importante**, fie că este vorba de recrutare, justiţie penală, asistenţă **medicală sau** altele.

Dacă un sistem AI folosit pentru a prezice riscul de recidivă în justiţia penală este antrenat pe date în care minorităţile au fost în mod istoric supuse unei supravegheri poliţieneşti excesive şi unor rate mai mari de arestare (din cauza prejudecăţilor sistemice), atunci ar putea asocia, în mod incorect, apartenenţa la o minoritate cu un risc mai mare de criminalitate.

Aceste probleme sunt deosebit de îngrijorătoare deoarece sistemele AI sunt, adesea, percepute ca *„obiective"* şi *„imparţiale"*. În realitate, ele sunt doar la fel de bune ca datele pe care sunt antrenate şi pot codifica prejudecăţile umane într-un mod care este departe de a fi clar sau transparent.

Invadarea vieţii private

Poate pentru că, în trecut, am fost unul dintre oamenii care au lucrat constant cu legislaţia privind protecţia datelor din Europa, acest domeniu mi-e mai apropiat decât multe altele. Motiv pentru care m-am gândit mult la implicaţiile AI-ului şi modurile în care lucrurile pot merge prost, odată ce un AI foloseşte şi combină datele personale ale cuiva în tot felul de rezultate creative.

Pe măsură ce AI devine mai avansat şi omniprezent, capacitatea sa de a colecta, analiza şi utiliza date personale a crescut exponenţial. Deşi acest lucru poate aduce beneficii (de exemplu, servicii personalizate şi mult mai eficiente), el ridică, de asemenea, întrebări profunde cu privire la confidenţialitatea datelor şi la potenţialele abuzuri ce ar putea apărea de aici.

Unii dintre cei mai vizibili utilizatori ai AI sunt platformele de social media şi motoarele de căutare. Aceste sisteme se bazează pe colectarea datelor

despre comportamentul nostru online (la ce postări dăm like, pe ce linkuri dăm click, ce căutări facem etc.). Sistemele lor folosesc, apoi, tehnici sofisticate de AI, cum ar fi învăţarea automată şi prelucrarea limbajului natural, pentru a analiza aceste date şi a construi profiluri detaliate ale intereselor, preferinţelor şi chiar trăsăturilor noastre de personalitate.

Aceste profiluri pot fi folosite pentru a ne livra publicitate *extrem* de personalizată (poate că acest cuvânt nu permite un termen de comparaţie în limba română, dar voiam să punctez faptul că personalizarea e ceva ce nici măcar nu ne putem imagina). Dacă un sistem AI ştie, de exemplu, că eşti interesat de alergare şi eşti în grupa de vârstă 25-34 de ani, ar putea să-ţi arate reclame pentru pantofi de alergare destinaţi tinerilor profesionişti. Deşi acest lucru poate părea inofensiv sau chiar convenabil, acţiunea ridică îngrijorări cu privire la cât de mult din viaţa noastră personală este monitorizată şi exploatată în scopuri comerciale, adesea fără consimţământul sau înţelegerea noastră deplină.

Problema merge dincolo de publicitatea personalizată. Datele colectate de sistemele AI pot fi folosite pentru a lua decizii cu impact major asupra vieţii, cum ar fi aprobarea unui împrumut, o cerere de asigurare sau o oportunitate de angajare. Dacă aceste date sunt inexacte, incomplete sau folosite în moduri discriminatorii, ar putea duce la prejudicii semnificative pentru indivizi.

O îngrijorare înrudită este utilizarea AI pentru supraveghere în masă. Multe guverne şi organisme de aplicare a legii folosesc acum sisteme AI pentru a monitoriza şi analiza cantităţi vaste de date, de la filmări de supraveghere şi recunoaştere facială până la monitorizarea reţelelor sociale şi a metadatelor telefonice. Deşi aceste sisteme sunt, adesea, folosite pentru prevenirea infracţiunilor sau contracararea terorismului, ele ridică, de asemenea, întrebări profunde despre dreptul la viaţă privată şi potenţialul de abuz şi supraveghere excesivă.

Dezinformarea şi manipularea

Am întâlnit mulţi oameni care nu cred că pot fi manipulaţi vreodată. Sunt convinşi că sistemul lor de valori şi credinţe e atât de solid încât ideea asta cu dezinformarea şi manipularea se aplică doar altora. Ori de câte ori sunt în faţa unor astfel de oameni, continui să îi întreb câteva lucruri simple. Câte dintre informaţiile pe care le citesc sau ascultă au fost verificate cu ochii şi urechile lor? În câte situaţii se bazează pe afirmaţiile altor oameni atunci când *„cred"* că ştiu un lucru?

Noi, oamenii moderni, am ajuns să nu mai experimentăm aproape nimic singuri. Mai toată viaţa noastră e trăită privind prin fereastra unor tehnologii care ne intermediază cuvintele, sursele de informare sau relaţiile cu alţi oameni. Chiar şi legea, acel set de reguli precum cel despre care vorbim şi în această carte, e cunoscută de un număr infim de oameni. Majoritatea aud lucruri în diverse publicaţii ori reţele sociale.

Legislaţia a devenit, astăzi, un set de reguli auzite şi interpretate din ce credem că au spus cei din jurul nostru. Pentru că nimeni nu mai citeşte, în realitate, ce scrie în textul unei legi. Adevărul e că ar fi greu de citit, din moment ce, odată cu trecerea timpului, legile au devenit mai complicate, mai lungi şi aproape imposibil de înţeles. De ce le-ar citi cineva? Din păcate, cu toate că nu le citim, o regulă importantă spune că nu ne putem folosi de necunoaşterea legii pentru a fugi de răspundere. Cu alte cuvinte, ni se aplică fie că le înţelegem de la sursă, fie că le aflăm de pe net, de la influenceri şi oameni care-şi dau, şi ei, cu părerea, printre alte subiecte mondene.

În tot contextul ăsta, am auzit, în jurul meu, mituri interesante. Cineva mi-a spus că nu te poţi recăsători de mai mult de cinci ori, că ar fi ilegal. Nu e, dar ar fi interesant de aflat de unde a pornit mitul. Şi astfel de idei circulă, din ce în ce mai multe, în jurul nostru. Am ajuns să ne certăm unii cu alţii pe lucruri pe care nu le putem înţelege, cu adevărat, cu mintea unui nespecialist. Cum ar fi scandalul în jurul vaccinurilor ARN mesager

sau cel în jurul încălzirii climatice pe care, încă, unii n-o văd decât ca pe o păcăleală (am întâlnit oameni care dădeau gerul năpraznic drept contra-exemplu al unei încălziri climatice).

Evident, stofa din care e creată o societate puternică poartă în ea mugurii unei dezbateri permanente. Dialogul, dezbaterea, toate creează progres. Doar că, în zilele noastre, am sentimentul că nu e dialog şi nici dezbatere ce ni se întâmplă. Dimpotrivă, pe baza efectului „rabbit hole" de care am vorbit mai devreme, pe baza acelei bule în care ne afundăm, cu toţii, pe măsură ce trece timpul, nimeni nu mai foloseşte dialogul pentru scopurile sale clasice.

Adam Grant spunea[14] că a observat mai multe tipare interesante la oamenii din jurul său. Modul preot, adică individul care încearcă să-i convertească pe ceilalţi la părerea sa. Modul procuror, cel în care îi acuzi pe ceilalţi, care sunt de altă părere decât a ta. Şi, în final, modul politician, care e activat atunci când îi asculţi doar pe cei care sunt de acord cu tine. Sunt curios dacă îţi sună cunoscut vreunul dintre ele.

În tot acest context, adăugăm, precum benzina pe foc, un alt actor care n-are nicio treabă cu dezbaterea în sine, ci cu mecanismele de inflamare a opiniei publice. Pe măsură ce modelele de limbaj precum GPT, Gemini, LlaMa ori Claude devin din ce în ce mai sofisticate, la fel devine şi capacitatea lor de a genera texte convingătoare, dar potenţial înşelătoare. Aceste texte pot fi folosite pentru a răspândi ştiri false, pentru a influenţa opiniile politice sau pentru a promova anumite produse sau ideologii, totul sub aparenţa de a fi generate de oameni.

Să presupunem că un actor politic rău intenţionat doreşte să influenţeze alegerile parlamentare dintr-o ţară. Ar putea folosi un model GPT pentru a genera sute sau mii de postări pe reţelele sociale, comentarii pe bloguri

[14] *Think Again: The Power of Knowing What You Don't Know*, Ed. Viking, 2 februarie 2021, https://www.amazon.com/Think-Again-Power-Knowing-What/dp/1984878107.

şi chiar articole care par a fi scrise de alegători reali. Aceste texte ar putea răspândi informaţii false despre un candidat, ar putea exagera beneficiile politicilor celuilalt candidat sau ar putea, pur şi simplu, să inunde spaţiul online cu *„zgomot"* menit să genereze confuzie şi dezbinare.

Forţa acestui tip de dezinformare este că este scalabilă şi greu de detectat. În loc să angajeze o armată de „troli" pentru a scrie manual postări, un singur actor ar putea folosi AI pentru a genera un volum masiv de conţinut, personalizat pentru diferite platforme şi audienţe. Şi, deoarece textul este generat de un model sofisticat, ar putea fi foarte greu de distins de cel scris de un om real.

O problemă similară este utilizarea AI pentru a crea recenzii false ale produselor sau pentru a promova scheme de tip pump-and-dump în finanţe. Imaginaţi-vă mii de recenzii convingătoare, dar complet false, apărând peste noapte pentru un produs nou lansat. Sau gândiţi-vă la o campanie coordonată AI care promovează o criptomonedă obscură, doar pentru a-i face preţul să crească înainte ca iniţiatorii să-şi vândă participaţiile şi să lase alţi investitori să se descurce cu ce-o urma.

Chiar şi atunci când intenţia nu este explicit răuvoitoare, AI ar putea fi folosită pentru a amplifica involuntar dezinformarea. Un site de ştiri ar putea folosi AI pentru a genera automat articole despre subiecte de actualitate, dar dacă modelul a fost antrenat pe date părtinitoare sau inexacte, ar putea perpetua acele inexactităţi la scară.

Aceste scenarii reprezintă o ameninţare nu doar pentru indivizi, ci şi pentru însuşi discursul public şi procesul democratic. Dacă nu putem avea încredere că lucrurile pe care le citim online sunt autentice şi corecte, devine incredibil de dificil să avem dezbateri semnificative şi să luăm decizii informate. Riscăm să intrăm într-o lume în care *„adevărul"* este, pur şi simplu, povestea pe care modelul AI cel mai sofisticat o poate genera.