Caiet CONSTITUȚIONAL II

Coordonatori
Robert-Marius Cazanciuc
Cristian Pîrvulescu

MONITORUL OFICIAL
Editură și Tipografie
BUCUREȘTI, 2023

Descrierea CIP a Bibliotecii Naţionale a României
Caiet constituţional / coord.: Robert-Marius Cazanciuc, Cristian Pîrvulescu. - Bucureşti : Monitorul Oficial R.A., 2021-
6 vol.
ISBN 978-606-035-114-6
Vol. 2. - 2023. - ISBN 978-606-035-112-2

I. Cazanciuc, Robert-Marius (coord.)
II. Pîrvulescu, Cristian (coord.)

342.4(498)

cuprins

Inteligența artificială și drepturile omului

Robert-Marius Cazanciuc

Îmi face o deosebită plăcere, doamnelor și domnilor profesori, să vă spun bună dimineața, aici, la Muzeul Unirii din Iași, care ne găzduiește cu amabilitate, pentru a participa la Conferința „Inteligența Artificială și Drepturile Omului", o temă nu numai de reflecție filosofică sau juridică, ci și una de largă dezbatere asupra avantajelor și, deopotrivă, a riscurilor pentru om, pentru societate și chiar pentru stat, a utilizării inteligenței artificiale în forme și modalități care aduc atingere valorilor ocrotite de stat.

Salut și pe cei care, din motive obiective, nu au putut să vină astăzi cu noi aici, dar sunt prezenți, grație tot inteligenței artificiale. Mă bucur să văd în sală un număr foarte mare de studenți interesați de problematica inteligenței artificiale și a drepturilor omului și sper din toată inima să nu vă pară rău la finalul acestei întâlniri.

În urmă cu un an de zile, la 1 iulie 2021, am organizat la Palatul Patriarhiei, într-un cadru la fel de frumos ca și aici, o dezbatere referitoare la Constituția României, pentru că era un moment aniversar, în acel an se împlineau trei decenii de la adoptarea Constituției. Am invitat la acea dezbatere câteva personalități ale lumii academice și universitare, alături de doi membri marcanți ai Adunării Constituante din perioada 1990 – 1991, pentru a ne oferi analize și perspective asupra evoluției societății românești în cadrul instituțional și politic deschis de această Constituție. Au fost invitați să participe atunci sociologi,

politologi, istorici, juriști, profesori de economie, care au abordat evoluția țării românești după '89 în multiple domenii de activitate.

Ne-am întrebat la vremea aceea și ulterior, vorbind despre Constituție, dacă ar fi nevoie, la 30 de ani de la adoptarea ei, să fie revizuită sau o altă Constituantă să elaboreze o nouă Lege fundamentală. Și, ca să îl citez pe dl profesor Valeriu Stoica, „Da, să se schimbe, primesc, dar să rămână așa cum e." Cu alte cuvinte, este foarte bună Constituția României și dacă de-a lungul vremii au fost lucruri care au fost considerate derapaje sau de neînțeles, este pentru că noi, cei care am citit Constituția, am forțat uneori interpretarea ei și n-am vrut să înțelegem spiritul unor prevederi ale acesteia.

Am spus la un moment dat că această Constituție a venit poate mult prea devreme pentru societatea românească, a fost foarte bine scrisă de oameni cu multă înțelepciune, cu multă experiență. Din păcate, după aceea, în viața de zi cu zi, în Parlament, în Guvern, chiar și în sistemul judiciar, mai ales în domeniul cercetării penale, nu am avut aceiași parteneri care să citească și să înțeleagă Constituția cu bună-credință, așa cum a fost scrisă, sau, mai degrabă, în spiritul ei. Una dintre provocările pe care le-am discutat atunci și ulterior a fost următoarea: dacă am accepta ideea de revizuire constituțională, care ar fi natura acestei revizuiri, ar fi nevoie de o revizuire mai amplă a Constituției sau doar a anumitor prevederi ale acesteia? Ar trebui să ne gândim la forma de organizare și funcționare a instituțiilor politice prevăzute de Constituție sau numai la relațiile dintre aceste instituții? Ce ar trebui, eventual, să revizuim, dacă s-ar pune problema, la un moment dat, să existe o voință politică de revizuire a Constituției?

Au fost două încercări nefinalizate de revizuire a Constituției, în 2011 și 2014, și una reușită în 2003. Practic, Constituția României, care a fost adoptată în '91, a fost modificată în 2003 pentru a putea permite României să acceadă în Uniunea Europeană și în NATO. Cele două încercări de revizuire inițiate în 2011 și 2014 nu au fost duse până la

capăt, pentru că au fost mai degrabă folosite ca instrumente politice în bătăliile interne.

Revizuirea din 2003 a venit pe o nevoie reală a României de aderare la Uniunea Europeană și la NATO. Întrebarea este: în prezent, suntem într-un moment în care există o nevoie reală de modificare a Constituției României, din perspectiva experienței dobândite ca țară membră a acestor organizații? Mă gândesc că un text constituțional care să reglementeze raporturile României ca stat membru cu drepturi depline cu Uniunea Europeană și cu NATO, care să reflecte și realitățile noi, de natură economică, social-culturală și politică din cadrul acestor organisme, ar fi binevenit.

Am împărțit pe trei categorii posibile motivele de revizuire: unele, să le spunem, minore, de redactare, și aici îl avem pe fostul ministru al comunicațiilor, dl Petrescu, căruia îi mulțumesc pentru prezență, și o să îi dau un exemplu pe care îl cunoaște foarte bine. I-am spus la un moment dat că noi avem un text în Constituție, privind secretul corespondenței, care folosește termenul de telegrame, utilizat ca formă de comunicare dintre cetățeni. L-am întrebat dacă se mai folosește cumva acest instrument de comunicare. Am întrebat Poșta Română anul trecut, într-o scrisoare oficială, câte telegrame s-au trimis în România în ultimul timp. Poșta Română mi-a răspuns că ultimul poștalion cu telegrame a plecat în urmă cu 10 ani. Deci de 10 ani nu s-a expediat nicio telegramă în România. Are rost să păstrăm în Constituție un cuvânt care nu se mai folosește în mod curent? Este o întrebare la care trebuie să răspundem.

Vorbind despre integrarea României în UE și în NATO, Constituția folosește termenul de preaderare. România nu mai este în faza de preaderare, am dobândit deja o experiență ca cetățeni europeni care ne deschide drumul spre oricare țară a Uniunii Europene.

Avem apoi o altă categorie de prevederi, să le spunem inovative, care ar putea fi introduse în Legea fundamentală, și cred că toată lumea este de acord. De pildă, unele constituții conțin texte care reglementează

protecția apei, ca resursă fundamentală a vieții. Ar fi util ca în Constituția României să existe o dispoziție referitoare la protecția apei? Ar trebui să avem în Legea fundamentală un text care să stabilească protecția drepturilor fundamentale față de riscurile utilizării inteligenței artificiale? Iată o întrebare la care nu vom avea un răspuns astăzi, nici nu ne-am propus asta, ci ne-am propus să facem o dezbatere cu specialiști pe fiecare domeniu în parte, pentru ca în momentul în care vom spune că trebuie să revizuim Constituția, să fim pregătiți pe fiecare abordare în parte, cu elemente științifice care să ajute decidentul politic să ducă revizuirea acolo unde este nevoie pentru următorii 20-30 de ani, pentru că nu se modifică Constituția în fiecare zi, chiar dacă la un moment dat am avea o clasă politică cu foarte multă efervescență.

Aceasta este dorința noastră, de a pune pe masă un suport științific de dezbateri, de abordări, care să ajute decidentul politic.

Să spunem că vom avea până atunci modificări tehnice, de natură redacțională, minore ca importanță, unde toată lumea este de acord, și unele inovative, unde nu cred că ar avea cineva ceva împotrivă, cum ar fi, de pildă, reglementarea unor aspecte ale inteligenței artificiale în Constituție.

Avem apoi o serie de propuneri de modificare a Legii fundamentale, de certă natură politică, cum ar fi, de exemplu, problema desemnării premierului după alegeri. Cine ar trebui să desemneze premierul? Partidul care se află pe locul întâi sau partidul care poate să formeze o coaliție? Aceasta este o dezbatere politică foarte interesantă. Nu despre acest lucru vrem însă să vorbim acum.

Dorim să discutăm despre chestiunile inovative pornind de la această dezbatere – inteligența artificială este una dintre ele –, dar și despre multe alte idei, pe care sperăm să le aducem în atenția unui grup de specialiști, care să-și propună, pur și simplu, să ofere informație de calitate, să provoace un interes în societate despre un subiect care cu siguranță va veni mai devreme sau mai târziu peste noi și atunci trebuie să ne găsească pregătiți.

Cei care scriu de foarte mulți ani despre Constituție, cum este dl profesor Cristian Ionescu, aici de față, vorbesc despre un moment constituțional. Suntem sau nu într-un moment constituțional de revizuire a Constituției? Aceasta este o întrebare la care încercăm să găsim un răspuns.

„Vorba lungă, sărăcia omului". Mă opresc aici și îl rog să ia cuvântul și pe dl. profesor Cristian Pîrvulescu, care a acceptat această provocare încă de anul trecut, de a participa la aceste întâlniri-dezbateri, din bucuria de a fi parte la acest tip de demers academic, dar sunt convins că și din dorința de a nu lăsa lucruri interesante pentru cetățeni fără a fi analizate pe îndelete, deoarece se știe foarte bine că într-o conjunctură politică, într-un fel sau altul, poți să treci, să adopți anumite lucruri doar pentru că ai la un moment dat o anumită majoritate. Noi nu vrem să se întâmple acest lucru, ca, într-un context politic dat, să recurgi la o revizuire constituțională fără să ai o bază de documentare și informare solidă. Sunt foarte multe lucruri care s-au decis ori s-au modificat fără o analiză profundă, serioasă și care produc consecințe, din păcate negative, după foarte mulți ani.

Dezbaterile de anul trecut au fost ca și aici, într-un format academic, care să stimuleze exprimarea liberă a opiniilor, tocmai ca oamenii să poată vorbi deschis, dar am decis de comun acord la vremea respectivă să publicăm expunerile prezentate. Avem aici lucrarea *Caiet Constituțional. 1991 – 2021*, publicată la Editura Monitorului Oficial, în care sunt consemnate intervențiile de la vremea respectivă. Sper, dacă sunteți toți de acord, ca și intervențiile și discuțiile ce vor avea loc în cadrul acestei conferințe să fie publicate sub același titlu – *Caiet Constituțional.*

Ce rol și funcție mai pot avea drepturile fundamentale în era digitală?

Cristian Pîrvulescu

Drepturile fundamentale în fața revoluției digitale

În acest început de mileniu, România și Europa se află în toiul celei mai importante transformări de la revoluția industrială: intrarea în era digitală. Soluțiile digitale pot deschide noi perspective în educație și cercetare, în tehnologie, precum și în economie în ansamblu. În plus, ele pot avea un impact pozitiv asupra tranziției ecologice și pot contribui la dezvoltarea unei societăți deschise și democratice. Așa s-a întâmplat și cu precedenta revoluție industrială care a însoțit primul val al democratizării din secolul al XIX-lea. Și, ca de fiecare dată într-o astfel de situație, transformarea digitală oferă oportunități, dar presupune și provocări sau riscuri. Prima revoluție industrială, cea a mașinismului și a cărbunelui, a dus la o dezvoltare economică fără precedent, dar și la pauperizarea celor care se ocupau cu agricultura și la o poluare care a generat criza ecologică de astăzi. Iar a doua revoluție industrială, cea a petrolului, a pus democrația liberală și parlamentarismul clasic în dificultate, creând condiții sociale și economice pentru dezvoltarea mișcărilor totalitare de toate tendințele. Astăzi, pe măsură ce Europa își consolidează suveranitatea digitală, Uniunea Europeană trebuie să se asigure că drepturile fundamentale ale cetățenilor săi sunt protejate. Or din această perspectivă există o situație periculoasă,

deoarece mai multe studii arată că pandemia de COVID-19 a accelerat transformarea digitală, testând în același timp protecția și garanțiile libertăților și drepturilor fundamentale.

După 2021, în conformitate cu strategia de consolidare a aplicării *Cartei drepturilor fundamentale a Uniunii Europene*, Comisia Europeană s-a orientat, în cadrul raportului anual pe care-l realizează pentru a monitoriza aplicarea Cartei, spre o analiză tematică. În cea mai recentă ediție a acestui raport, Comisia a făcut un bilanț al progreselor digitale și s-a concentrat asupra drepturilor fundamentale din epoca digitală. În acest sens, organizații ale societății civile și organisme naționale independente pentru apărarea drepturilor omului sunt parteneri esențiali ai instituțiilor europene și ai statelor membre atunci când vine vorba de promovarea și protejarea drepturilor fundamentale, a democrației și a statului de drept.

În perspectiva unui viitor care asigură protejarea reală a drepturilor omului și libertăților cetățenești, este necesară nu doar crearea unui cadru european eficace pentru protecția drepturilor omului în era digitală, ci și realizarea unui schimb de bune practici care pot sprijini efortul de a crea un sector digital reglementat pentru a asigura conformitatea cu drepturi fundamentale, de exemplu, reglementarea și protejarea vieții private și a secretului comunicațiilor electronice. În această direcție, dezvoltările juridico-constituționale, dar și cele cibernetice trebuie centrate pe om și pe asigurarea demnității sale printr-o reglementare clară și puternică a identității digitale și prin dezvoltarea unei inteligențe artificiale fiabile și democratice.

Inteligența artificială și dreptul la nediscriminare

Tehnologiile privind inteligența artificială (IA) oferă multe beneficii, dar prezintă totodată și riscuri de discriminare fără precedent. Acestea pot să fie, de exemplu, rezultatul unei atitudini părtinitoare a sistemelor ce folosesc inteligența artificială, care poate rezulta din intervenția prejudecăților și a stereotipurilor care se strecoară – voluntar sau

involuntar – fie în datele utilizate pentru a genera un sistem bazat pe inteligența artificială, fie în cadrul acestuia. Inteligența artificială este din ce în ce mai des folosită pentru a lua decizii care uneori au repercusiuni semnificative asupra vieții cetățenilor, de exemplu atunci când este utilizată în procedurile de recrutare a personalului, în educație, în publicitatea politică direcționată (cum ne-a arătat-o scandalul Cambridge Analytica), în furnizarea de informații personalizate utilizatorilor serviciilor publice sau în stabilirea unor profiluri de risc în domeniul asigurărilor. Iar o inteligență artificială care este părtinitoare riscă să conducă spre soluții discriminatorii dacă va reflecta sau perpetua prejudecățile societale preexistente referitoare la rasă, sex, sex biologic, vârstă și cultură. Tocmai de aceea *Agenda digitală a UE* a făcut o prioritate din dezvoltarea unei inteligențe artificiale care să respecte drepturile omului și libertățile cetățenești. În același timp, societatea civilă a propus o inițiativă care să permită dezvoltarea unor forme de IA mult mai incluzive[1]. În lumina acestor inițiative și pornind de la implementarea Regulamentului general privind protecția datelor (GDPR – *General Data Protection Regulation*), dezvoltarea și utilizarea IA este și ar putea fi și mai bine reglementată în vederea garantării dreptului la nediscriminare, de exemplu în utilizarea sistemelor IA pe piața muncii și a sistemelor de recunoaștere facială.

Protecția drepturilor fundamentale și moderarea conținutului platformelor online

Pe măsură ce cetățenii UE interacționează din ce în ce mai mult online, beneficiile serviciilor oferite de platformele digitale devin atât de importante încât nu mai pot fi evitate. În ultimele două decenii, platformele online au devenit principalele mass-media, asigurând schimbul de informații, inclusiv de știri. În același timp, pericolele pe care le reprezintă informațiile distribuite pe aceste platforme devin, proporțional cu numărul lor, substanțiale.

[1] A se vedea „Artificial Intelligence and Inclusion Compendium of Promising Initiatives", publicat de UNESCO cu ocazia *Mobile Learning Week 2020*.

În contextul ideologic actual dominat de iliberalism, noile media – așa cum s-a întâmplat și în alte momente istorice cheie când au fost introduse noi instrumente de transmitere a informațiilor – au un impact important asupra calității și imparțialității difuzării informațiilor. Într-un articol din 2015 Robert Epstein și Ronald Robertson[2] constatau, pe baza unui experiment, existența unui așa-numit *efect de manipulare al motorului de căutare* (SEME – *Search Engine Manipulation Effect*). În urma cercetării coordonate de cei doi, s-a putut confirma ipoteza conform căreia clasamentele făcute de motoarele de căutare sunt părtinitoare și pot schimba preferințele de vot ale alegătorilor indecișI (cercetarea indică un număr semnificativ de aproximativ 20% dintre indecișI care și-au schimbat preferința de vot ca urmare a informațiilor „servite" de motoarele de căutare). Pe de altă parte, Epstein și Robertson constatau că această proporție poate fi mult mai mare în cazul anumitor grupuri demografice izolate, dar și că efectul acestor clasificări nu este evident, astfel încât cei afectați de acest tratament al informațiilor nu realizează manipularea.

Motoarele de căutare pe internet, de tipul Google sau Facebook, pot avea un impact semnificativ asupra alegerilor consumatorilor[3]. Cetățenii, reduși la rolul de simpli consumatori, au impresia că primesc informații neutre atunci când preiau automat rezultatele căutărilor ierarhizate prin intermediul unor algoritmi sofisticați, dar nu este tocmai așa. Ierarhiile rezultate în urma căutărilor sunt adaptate profilului fiecărui individ. Epstein și Robertson nu cred că acestea ar fi neutre, ba constată că sunt adaptate preferințelor fiecărui consumator, iar în cazul alegerilor, elector. Și, dacă experimentul realizat pe trei grupe de subiecți arăta

[2] „The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections", publicat pe site-ul American Institute for Behavioral Research and Technology.

[3] Pe 9 octombrie 2017 *The New York Times* prelua o știre Reuters care făcea publice rezultatele cercetărilor Google privind posibila implicare a unor grupuri de influență rusești în campania electorală din 2016 din SUA: „Google has discovered Russian operatives spent tens of thousands of dollars on ads on its YouTube, Gmail and Google Search products in an effort to meddle in the 2016 U.S. presidential election, a person briefed on the company's probe told Reuters on Monday".

că algoritmii de căutare pot schimba cu ușurință preferințele de vot ale alegătorilor indecișI cu de la 20% – în cazul grupurilor deschise – până la 80% – în cazul unor grupuri demografice izolate –, impactul politic al motoarelor de căutare devine esențial.

Mai mult, într-un interviu acordat Politico în august 2015, același Epstein observa deja că Donald Trump se afla pe o poziție foarte bună în ierarhia motoarelor de căutare. Or aceste ierarhii nu erau generate în niciun caz de preferințe economice, cum sunt tentați să creadă mulți analiști ai fenomenului politic, ci de predispozițiile politice ale consumatorilor de informații de pe internet.

Apare evident că utilizarea platformelor amplifică problemele societale, cum ar fi polarizarea sau diseminarea de conținut ilicit, care de foarte multe ori încalcă grav drepturile fundamentale. În plus, criza COVID-19 a scos la iveală diseminarea de conținut online care nu este ilegal în sine, cum ar fi dezinformarea și teoriile conspirației, dar care poate submina discursul public democratic, încrederea în instituții, care poate afecta grav sănătatea publică și siguranța cetățenilor. În aceste condiții populismul, prin polarizarea crescândă a dezbaterii politice și prin erodarea încrederii publicului în procesele democratice, folosește dezinformarea ca o armă politică redutabilă.

Dar dacă democrația se confruntă cu multe astfel de provocări, inițiativele menționate în „Compendium of Promising Initiatives: Mobile Learning Week 2020" al UNESCO evidențiază potențialul pozitiv al IA în educație, în special în privința accelerării progresului cunoașterii și a reducerii decalajelor în ceea ce privește accesul digital, genul, capacitatea în vederea creării unor societăți ale cunoașterii incluzive.

În aceste condiții, în fața creșterii exponențiale a conținutului și a schimburilor online, problema moderării platformelor devine din ce în ce mai acută și, la rândul său, ridică multe întrebări cu privire la responsabilitățile asociate cu apărarea anumitor drepturi și valori fundamentale în mediul online. Unele platforme se bazează pe soluții automate pentru a înlocui moderarea conținutului uman, de exemplu

folosind sisteme care utilizează IA pentru a identifica și elimina conținutul discriminatoriu sau instigator la ură. Cu toate acestea, automatizarea moderării online poate prezenta multe riscuri pentru drepturile fundamentale, inclusiv în privința libertății de exprimare, a dreptului de a obține informații, precum și a dreptului la viață privată și nediscriminare. Această automatizare ar putea da naștere la practici de cenzură pe scară largă și solicită remedii eficiente în cazul ștergerii abuzive a conținutului de la utilizatori, care nu de puține ori nu pot contesta decizia care îi afectează.

Una dintre dezvoltările legislative recente în privința conținuturilor comunicării online din UE este *Digital Services Act* care-și propune să reglementeze modul în care companiile de tehnologie își moderează platformele. În acest sens este necesară construirea unui spațiu de dialog între reprezentanții instituțiilor europene și ai statelor membre cu actorii societății civile și alte părți interesate, unde se vor putea discuta modalitățile de moderare a conținutului pe platformele online și reglementările în acest domeniu. De aceea, unii au sugerat o „abordare holistică, non-tehnologică"[4], pentru a asigura condițiile ca reglementările să nu restrângă controlul democratic și să întărească poziția dominantă pe piață a marilor companii de tehnologie. Organizațiile societății civile au subliniat că, în loc să se concentreze doar pe înăsprirea regulilor privind moderarea conținutului, reglementările ar trebui să abordeze cauza principală a proliferării conținutului ilegal sau dăunător online, să cunoască modelul de afaceri al acestor companii, bazat pe date și publicitate.

Toate aceste avantaje și dezavantaje arată că, departe de a-și fi epuizat rolul, drepturile fundamentale în era digitală sunt mai importante ca oricând în trecut. Iar o revizuire constituțională, chiar dacă va exista un cadru de reglementări europene bine structurat și protectiv, nu va putea face abstracție de dezvoltările tehnologice și implicațiile lor societale.

[4] Cum este cazul raportului EDRi (European Digital Rights – o rețea de ONG-uri, experți, avocați și cadre universitare care lucrează pentru apărarea și promovarea drepturilor digitale în Europa) „A losing game: moderating online content fuels Big Tech power"), publicat în 2021.

Inteligența artificială - o nouă frontieră geopolitică

Dragoș Tudorache

Inteligența artificială (IA) schimbă lumea. Trebuie să înțelegem această schimbare, să învățăm să o gestionăm în interesul nostru, să nu ne fie teamă de efectele ei. Deși avem deja în jurul nostru multe exemple pozitive ale utilizării IA în fiecare domeniu al activității umane, există în continuare un nivel însemnat de rezistență în rândul publicului larg. De cealaltă parte, au existat situații în care aplicații ale noilor tehnologii de IA au exacerbat prejudecăți preexistente în societate și au dus la discriminare, ceea ce a alimentat și mai mult rezistența la schimbare.

Uniunea Europeană a înțeles mizele majore ale impactului IA asupra societăților noastre, asupra valorilor care stau la baza democrației. De aceea, demersul european de reglementare a IA și, în general, strategia digitală a UE pentru viitoarea decadă trec dincolo de valențele strict tehnice. E important să înțelegem că tehnologia IA este, înainte de toate, neutră, ceea ce înseamnă ca ea nu poate fi bună sau rea, ci felul în care este utilizată produce efecte bune sau rele. Același algoritm poate fi folosit pentru a diagnostica mai repede o tumoare sau pentru a identifica și accelera distribuirea de conținut toxic online.

Sunt cel puțin trei perspective majore în abordarea europeană privind reglementarea IA.

În primul rând, este vorba de o perspectivă societală. IA și procesele de digitalizare schimbă tot în societate. De la modurile în care interacționăm între noi, în care gestionăm relațiile cu administrația publică sau în care

statul interacționează cu cetățenii săi, până la modurile în care serviciile sunt furnizate sau în care se iau deciziile care privesc drepturile noastre; de la modurile în care funcționează democrația, de la alegeri, până la felul în care ne informăm. Toate aceste aspecte sunt influențate atât în bine, cât și în rău de această tehnologie și atunci nu ne poate fi indiferent cum este utilizată. În al doilea rând este vorba despre un motiv economic. IA schimbă fundamental lanțurile valorice din zona economică, din moment ce o mare parte din valoarea produselor și serviciilor viitorului nu va mai fi dată de ceea ce vedem fizic, ci de tehnologia care stă la baza produselor și serviciilor pe care le vom accesa. În al treilea rând, IA a devenit un element geopolitic, tocmai ca urmare a impactului său profund din economie și societate. Toate puterile lumii au un interes direct de a obține un element de superioritate în modul în care dezvoltă și controlează aceste tehnologii, dar și în modul în care vor reuși să impună standarde globale bazate pe valorile lor. În mijlocul acestei curse între două modele de evoluție a tehnologiei ne aflăm acum, în plan geopolitic. Primul model are la bază valorile democratice, precum modelul european care este promovat împreună cu partenerii noștri care înțeleg democrația și valorile în același fel, țări ca SUA, Canada, Japonia, Coreea de Sud sau Australia. Celălalt model este promovat de China, care investește masiv și folosește IA nu în interesul individului sau pentru promovarea drepturilor sale, ci pentru a asigura un control cât mai strict al societății.

Aceste trei motive sunt la baza deciziei Parlamentului European de a-și asuma, încă din primul an de mandat, înființarea unei Comisii speciale pentru inteligența artificială în era digitală (AIDA) pe care am avut plăcerea să o conduc și care s-a uitat dincolo de orizont, nu doar unde ne-am dori să ne aflăm după următorii cinci-zece ani. Activitatea Comisiei a culminat cu un raport pe care l-am adoptat în această primăvară și care consider că elaborează bine jaloanele viitorului în ceea ce privește impactul inteligenței artificiale, astfel încât să-i simțim efectele benefice și să-i controlăm potențialul negativ.

Legea privind inteligența artificială (*AI Act*) se află într-o fază avansată în acest moment, după primirea a peste 3.300 de amendamente din

partea colegilor, un record pentru acest mandat. Legea este prima, la nivel global, care folosește o abordare atât de complexă și detaliată cu privire la această tehnologie. Nu trebuie să avem aroganța de a crede că doar noi, europenii, știm mai bine. De aceea, este important să lucrăm în strânsă colaborare prin procese bilaterale cu partenerii noștri care au aceleași valori. Chiar dacă avem tradiții normative și legislative diferite, important este să ajungem la același rezultat, chiar și prin procese diferite.

Legea privind IA prevede un sistem de guvernanță centralizat, la nivel european, pentru a nu risca să repetăm greșelile trecutului, precum în cazul GDPR, când am avut practici nealiniate între Statele Membre. Pentru a evita fragmentarea pieței unice și a asigura nivelul de expertiză necesar pentru o bună guvernanță, acest model centralizat va răspunde cel mai bine nevoilor și va include, totodată, un nivel de servicii descentralizate care permit interacțiuni granulare la nivelul Statelor Membre.

Subiectele principale din lege care vor fi dezbătute intens în următoarele luni sunt: definiția, domeniile de aplicare și aplicațiile cu un risc ridicat.

În acest moment, nu există o definiție asumată și acceptată unanim cu privire la ce înseamnă, de fapt, IA. Definiția poate fi una largă, care cuprinde cât mai multe elemente și care ar rămâne valabilă și peste zeci de ani, sau una strictă, clară, care stabilește din start modalitățile și procesele tehnologice reglementate de lege. Opțiunea mea este pentru o definiție largă și permisivă. Consider că este nevoie de reglementări precise atunci când elaborezi ceea ce este interzis sau ceea ce are un risc ridicat – acolo este nevoie de precizie în definiție, nu în definiția largă. Legea este gândită pentru a reglementa scopul și felul în care este utilizată tehnologia – nu tehnologia în sine, ci efectele sale și modul în care este utilizată în domenii practice și clar definite.

Un alt subiect conex care va atrage dezbateri ample și care este susținut de un lobby intens din partea industriei digitale este cel legat de includerea IA cu un scop general în legislație, mai exact acei algoritmi care nu au încă un scop dedicat și care sunt mașinării puternice, capabile să proceseze și să ducă la rezultate în funcție de

datele cu care este hrănit algoritmul, dar care nu sunt îndreptate către un anumit scop sau altul.

Vor exista dezbateri intense și legate de modul în care vor fi reglementate aplicațiile care folosesc IA și care aduc atingere drepturilor noastre fundamentale. Articolul 5 din lege prevede domeniile interzise de aplicare a tehnologiei IA, acelea pe care nu le dorim în societate. Comisia Europeană a propus patru astfel de domenii de aplicare, la care am adăugat și eu unul, cel de *predictive policing,* care se referă la instrumentele pe care organele de cercetare penală le folosesc pentru a anticipa criminalitatea pe bază de caracteristici sociale, personale, biometrice sau pentru a anticipa recurența în comportamente criminale. Au existat astfel de exemple atât în UE, cât și în SUA, care au dovedit un potențial negativ major, exacerbând prejudecăți deja existente în societate. Zona de utilizare a biometriei, a recunoașterii biometrice în spații publice și a supravegherii folosind biometria în spațiile publice beneficiază de o majoritate largă în Parlamentul European pentru a fi încadrată la interdicții.

Un alt subiect deschis dezbaterilor este lista de aplicații care utilizează IA și au un risc ridicat în ceea ce privește drepturile și libertățile individului. Acestea sunt prevăzute într-o anexă, pentru că o anexă este mai flexibilă și permite adaptarea listei mai ușor, din punct de vedere procedural. Rămânând în zona de protejare a drepturilor cetățenilor, am introdus în lege dreptul individual de a sesiza un organ administrativ sau instanță atunci când un consumator simte că este afectat sau că drepturile sale au fost afectate de orice tip de produs sau serviciu care are în spate IA.

Impactul pe care inteligența artificială îl va avea, în continuare, asupra societății va crește și mai mult în următorii ani. Reglementarea tehnologiilor în era digitală necesită o optică pe termen lung și, din acest motiv, Legea privind inteligența artificială este construită astfel încât să fie adaptată viitorului, să facă față obstacolelor încă necunoscute. Astfel, ne va permite să beneficiem în continuare de efectele pozitive ale tehnologiei fără să ne compromitem drepturile fundamentale sau valorile.

Robustețe, siguranță și transparență pentru inteligența artificială de încredere

Prof. dr. ing. Adina Magda Florea,
Universitatea POLITEHNICA din București
adina.florea@upb.ro

Inteligența artificială (IA) are ca scop construirea de sisteme care manifestă un comportament inteligent, analizează mediul și iau decizii cu un anumit grad de autonomie, pentru realizarea unor sarcini specifice. Descoperirile asociate primei revoluții industriale au permis înlocuirea muncii fizice umane sau animale cu mașini. În epoca actuală, inteligența artificială înlocuiește, din ce în ce mai mult, activități umane care presupun efectuarea de raționamente, recunoaștere de imagini, dialog în limbaj natural, planificare automată, învățare automată. Aplicațiile bazate pe IA sunt astăzi omniprezente, de la telefoane inteligente la roboți, mașini autonome și asistenți inteligenți, la traducere automată, sinteza opiniilor din volume imense de texte și postări din rețelele sociale, pentru a da numai câteva exemple. Putem astfel spune deja că suntem în epoca inteligenței artificiale și aceasta este doar la început.

Despre inteligența artificială

Începutul domeniului inteligenței artificiale este considerat anul 1956, odată cu organizarea conferinței de la Dartmouth College, în Hanover, New Hampshire, SUA, unde au participat cercetători de renume în știința calculatoarelor care au introdus pentru prima dată termenul de inteligență artificială și au prezentat primul program de IA numit *The Logic Theorist*, capabil să demonstreze teoreme în calculul propozițional. Printre momentele importante care au marcat dezvoltarea domeniului se pot menționa: victoria programului de șah Deep Blue asupra lui Gary Kasparov (campion mondial) în 1997, programul Watson care în 2011 a concurat cu campioni mondiali în cadrul unui joc de cultură generală, Jeopardy, bazat pe întrebări în limbaj natural din diferite domenii și a obținut locul I, victoria programului AlphaGo împotriva unuia dintre cei mai buni jucători de Go, Lee Seedol, în patru din cinci meciuri, și tot în 2011 prima rețea neurală profundă pentru recunoașterea obiectelor din imagini care a deschis drumul dezvoltărilor explozive în domeniul învățării profunde (*deep learning)*.

Sistemele de IA pot fi grupate în două categorii principale care se referă la capacitatea de raționament a sistemului și, respectiv, la capacitatea de a învăța. Capacitatea de raționament a sistemului este realizată prin modele simbolice care includ reprezentarea cunoștințelor și utilizarea acestora în luarea deciziilor, planificare, căutare și optimizare. Capacitatea de învățare a sistemului poate fi realizată prin modele cum ar fi rețele neurale cu învățare profundă, sisteme cu suport vectorial, algoritmi genetici și altele. Aceste tehnici permit unui sistem de IA să învețe cum să rezolve probleme care nu pot fi specificate cu exactitate sau pentru care metoda de rezolvare nu poate fi descrisă prin reguli de raționament simbolic.

De-a lungul timpului au fost propuse numeroase definiții ale inteligenței artificiale, fără ca o definiție sau alta să fie unanim acceptată. În 2019,

Grupul de experți la nivel înalt pe probleme de IA al Comisiei Europene[1] a formulat următoarea definiție: „Sistemele de inteligență artificială sunt sisteme software (și, eventual, hardware) proiectate de oameni, care, dacă li se dă un obiectiv complex, acționează în dimensiunea fizică sau digitală, percepând mediul prin intermediul preluării datelor, prin interpretarea datelor structurate sau nestructurate colectate, prin raționament cu privire la cunoștințe sau prin prelucrarea informațiilor obținute din aceste date și prin decizia asupra celei/celor mai bune acțiuni care trebuie întreprinse pentru a realiza obiectivul dat. Sistemele de IA pot să utilizeze reguli simbolice sau să învețe un model numeric și, de asemenea, își pot adapta comportamentul analizând modul în care mediul este afectat de acțiunile lor anterioare."

Recent, propunerea *Artificial Intelligence Act*[2], care are ca scop stabilirea unui regulament privind norme armonizate ale utilizării inteligenței artificiale la nivelul Uniunii Europene, a formulat următoarea definiție a IA: „Un sistem de inteligență artificială reprezintă un software care poate, pentru un anumit set de obiective definite de om, să genereze rezultate precum conținut, predicții, recomandări sau decizii care influențează mediile cu care interacționează și care este dezvoltat cu una sau mai multe dintre următoarele tehnici și abordări: (1) învățare automată, inclusiv învățarea supervizată, nesupervizată și prin recompensă, folosind o mare varietate de metode, inclusiv învățarea profundă; (2) abordări logice și bazate pe cunoștințe, inclusiv reprezentarea cunoștințelor, programarea inductivă (logică), bazele de cunoștințe, motoarele de inferență și deductive, raționamentul (simbolic) și sistemele expert; (3) abordări statistice, estimare bayesiană, metode de căutare și optimizare."

Indiferent de o definiție sau alta, tehnologiile inteligenței artificiale evoluează rapid, aduc și vor aduce din ce în ce mai mult beneficii de o importanță majoră, atât la nivel economic, cât și social, având

[1] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

[2] https://artificialintelligenceact.eu/

capacitatea de a oferi avantaje competitive importante companiilor și economiei. Cu toate acestea, integrarea sistemelor de decizie bazate pe IA în viața curentă, oricât de rațional ar fi comportamentul acestora, și utilizarea pe scară largă a tehnologiilor IA creează noi riscuri sau pot avea consecințe negative pentru indivizi sau societate.

Evitarea riscurilor și a impactului negativ al inteligenței artificiale

La nivelul Uniunii Europene dar și la nivel mondial au existat, în ultimii ani, o preocupare constantă și multe inițiative de analiză și stabilire a unor acțiuni sau cadre de reglementare pentru limitarea potențialelor riscuri în utilizarea IA, cât și pentru asigurarea faptului că inteligența artificială are un impact pozitiv asupra societății, fără a exacerba sau a duce la noi amenințări asupra libertăților și drepturilor fundamentale ale oamenilor.

Cartea albă privind inteligența artificială - O abordare europeană axată pe excelență și încredere[3], elaborată de Comisia Europeană în 2020, afirmă că este necesară o abordare europeană comună a domeniului pentru a ajunge la o scară suficientă și a evita fragmentarea pieței unice. În plus, se subliniază impactul major al IA asupra întregii societăți și necesitatea de a construi la nivelul Uniunii Europene ***inteligența artificială de încredere***, ancorată în valori și drepturi fundamentale cum ar fi demnitatea umană, protecția vieții private, nediscriminare. Impactul sistemelor IA trebuie considerat nu numai dintr-o perspectivă individuală, dar și din perspectiva societății ca un întreg. Grupul de experți la nivel înalt pe probleme de IA al Comisiei Europene a publicat în 2019 *Ghidul de Etică pentru IA de încredere*[4] și, în 2020, *Lista de evaluare pentru IA de încredere (ALTAI)*[5] care conține un set de

[3] https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[4] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[5] https://altai.insight-centre.org/

întrebări legate de etică și de principii pe care organizațiile trebuie să le considere în dezvoltarea sistemelor de IA. Ideile și cerințele legate de inteligența artificială de încredere au fost mai departe elaborate în *Planul coordonat revizuit al dezvoltării IA la nivel european,* publicat în 2021. În luna aprilie a aceluiași an s-a publicat *Proposal of the European Parliament and of the Council of a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*, urmând ca în 2023 să se facă pașii legislativi corespunzători la nivelul Parlamentului European și al Consiliului Uniunii Europene pentru aprobarea acestui cadru legal flexibil și robust de reglementare a utilizării tehnologiilor IA în economie și societate.

În noiembrie 2021, cele 193 de state membre ale UNESCO au adoptat *Recomandările privind etica inteligenței artificiale*[6], primul instrument la nivel mondial de stabilire a standardelor privind etica inteligenței artificiale care să protejeze, dar și să promoveze drepturile omului și demnitatea umană, și care să acționeze ca un ghid etic normativ pentru respectarea statului de drept în lumea digitală.

Printre alte inițiative la nivel mondial privind aspecte etice ale inteligenței artificiale și limitarea riscurilor aduse de această nouă tehnologie se pot menționa: *AI Policy Observatory*[7] al OECD (Organisation for Economic Co-operation and Development), *Global Initiative on Ethics of Autonomous and Intelligent Systems*[8] elaborată de IEEE Standards Association, IEEE fiind cea mai mare asociație profesională de inginerie electronică, electrică și calculatoare din lume, sau *Guidance for Regulation of Artificial Intelligence Applications*[9] (OMB AI Memorandum) publicat de Oficiul de Management și Buget al Statelor Unite în anul 2020. La nivelul Pactului Nord-Atlantic, în *NATO Artificial Intelligence*

[6] https://unesdoc.unesco.org/ark:/48223/pf0000381137

[7] https://oecd.ai/en/

[8] https://standards.ieee.org/industry-connections/ec/autonomous-systems/

[9] https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

Strategy[10] sunt incluse principiile NATO pentru utilizarea responsabilă a IA în apărare, principii comune care vor conduce la dezvoltarea și implementarea aplicațiilor de IA cu impact în apărare, în toate țările aliate. În mediul privat, giganți tehnologici precum IBM, Microsoft, Oracle, HP și alții au statuat de asemenea cadre de reglementare ale utilizării responsabile și etice ale IA.

Inteligența artificială poate avea un impact major și în educație, unde promovarea utilizării sale echitabile, incluzive și transparente este o necesitate. Rolul IA în educație poate fi văzut atât din punctul de vedere al utilizării tehnicilor IA pentru dezvoltarea unui învățământ personalizat, centrat pe student, pentru atingerea unor rezultate mai bune ale învățării și a unei evaluări mai eficiente a procesului de învățare, cât și din punctul de vedere al necesității dezvoltării abilităților necesare în contextul IA la nivelul curriculumului și al creării oportunităților de învățare pe tot parcursul vieții pentru forța de muncă existentă. În acest sens, în 2019 a fost elaborat și adoptat de către UNESCO documentul-cadru *Consensul de la Beijing referitor la inteligența artificială și educație*[11] cu ocazia Conferinței Internaționale *Planning Education in the AI Era: Lead the Leap*.

La nivelul Uniunii Europene, în 2021, s-au enunțat principii etice de utilizare a IA în educație în cadrul documentului *Artificial intelligence in education, culture and the audiovisual sector*[12].

Inteligența artificială de încredere

Pentru ca inteligența artificială să devină o tehnologie pusă în slujba oamenilor și a binelui pentru societate, este nevoie să avem încredere în această nouă tehnologie disruptivă. Strategia privind inteligența artificială a Comisiei Europene vizează „dezvoltarea și utilizarea IA etice

[10] https://www.nato.int/cps/en/natohq/official_texts_187617.htm

[11] https://unesdoc.unesco.org/ark:/48223/pf0000368303

[12] https://www.europarl.europa.eu/doceo/document/A-9-2021-0127_EN.html

și sigure, promovând o abordare centrată pe om în context global". Grupul de experți la nivel înalt pentru inteligența artificială (AI HLEG)[13] care sprijină implementarea strategiei europene privind inteligența artificială a identificat ***4 principii etice*** care reprezintă fundamentul inteligenței artificiale de încredere: *Respectul pentru autonomia umană* – respectul pentru libertatea și autonomia oamenilor și implicarea acestora în deciziile luate de sistemele de IA; *Prevenirea vătămării* – sistemele de IA nu trebuie să producă vătămarea oamenilor, nu trebuie utilizate în mod malițios și trebuie să fie sigure; *Corectitudinea* – beneficii și costuri egale pentru toți, cât și lipsa discriminării sau a părtinirii; și *Explicabilitatea* – capacitatea sistemelor de IA de a explica deciziile luate și de a le comunica celor afectați de aceste decizii, precum și **7 cerințe-cheie** pentru realizarea IA de încredere: *Implicarea umană și supraveghere*; *Robustețe tehnică și siguranță*; *Confidențialitatea și guvernarea datelor*; *Transparență*; *Diversitate, nediscriminare și corectitudine*; *Bunăstarea mediului și a societății* și *Responsabilitate*.

Vom discuta în continuare două dintre aceste cerințe-cheie ale inteligenței artificiale de încredere, care ridică provocări comunităților științifice, tehnice și de business implicate în dezvoltarea cercetărilor în inteligență artificială și a produselor și serviciilor bazate pe IA, respectiv robustețe tehnică și siguranță, și transparență.

Robustețea tehnică și siguranța sistemelor de inteligență artificială

Robustețea tehnică este strâns legată de conceptul de inteligență artificială de încredere. O serie de provocări stau în fața sistemelor de IA din acest punct de vedere: evitarea acțiunilor defectuoase, a deciziilor riscante luate de algoritmii IA care afectează oameni, a accidentelor în funcționarea produselor bazate pe IA dezvoltate (de exemplu, roboți, mașini autonome etc.) sau a carențelor în proiectare care pot cauza rău oamenilor, reproductibilitatea rezultatelor, utilizarea unor

[13] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

seturi de date de calitate, documentarea adecvată a produselor sau serviciilor pentru asigurarea trasabilității, cât și măsuri care să permită supravegherea și analiza umană a deciziilor.

Pentru evitarea accidentelor sau a deciziilor incorecte, deoarece IA este bazată pe algoritmi și este știut faptul, demonstrabil matematic, că nu se poate face o verificare totală a lor, în special în cazul algoritmilor de învățare profundă, este necesară o testare extinsă, pe diferite seturi de date și luând în calcul diferite schimbări ale mediului de operare, precum și prezența altor agenți (umani sau artificiali). Proiectarea sistemelor de IA trebuie făcută în așa fel încât să considere, pe cât posibil, riscurile care ar putea exista în funcționarea acestora și să ofere utilizatorilor siguranță în mediile în care aceștia interacționează cu un sistem de IA. De exemplu, în viitorul nu foarte îndepărtat vehiculele autonome vor fi o prezență constantă pe șosele pentru transportul de persoane sau de mărfuri și operarea lor trebuie să fie sigură din punct de vedere tehnic. Pe lângă aceste daune neintenționate, IA poate fi utilizată în scopuri malițioase sau pentru a provoca rău. În această categorie se înscriu atacurile de securitate cibernetică, atât atacuri asupra programelor bazate pe IA, cât și dezvoltarea unor virusuri informatice cu tehnici de IA. Pentru a combate aceste tendințe, comunitatea de inteligență artificială trebuie să coopereze cu cea de securitate cibernetică pentru a dezvolta soluții de protecție în aceste situații. Să ne imaginăm, de exemplu, un program bazat pe IA care decide autonom dacă să acorde asistență medicală unei persoane; dacă acest program este corupt decizia poate fi total incorectă sau chiar fatală pentru persoana în cauză. Sau situația în care un program de IA operează autonom conducerea unei infrastructuri critice, funcționarea incorectă în acest caz putând afecta viața a zeci sau sute de oameni.

Un loc important în lista daunelor intenționate pe care persoane malițioase le pot realiza cu IA pot ocupa și acțiunile prin care se influențează opiniile sau chiar deciziile persoanelor, de exemplu: manipularea opiniilor pe rețelele sociale prin distribuirea de știri false cu aparență de autenticitate, manipularea convingerilor și exacerbarea

concepțiilor greșite si a prejudecăților, realizarea de personaje false (*deep fake*) prin care persoane publice foarte cunoscute sunt imitate la nivel de imagine și/sau voce pentru a influența opiniile sau deciziile.

Reproductibilitatea rezultatelor, prin aceasta înțelegând capacitatea de a produce aceleași rezultate în aceleași condiții de funcționare, este de asemenea importantă pentru robustețea tehnică a sistemelor de IA, permițând o descriere exactă a funcționalității, o analiză corectă a deciziilor și rezultate echitabile pentru orice instanță de utilizare. Este adevărat că acest obiectiv este extrem de dificil de realizat în cazul sistemelor cu învățare automată sau profundă. De aceea, este necesară, încă din faza de proiectare, considerarea tuturor scenariilor cu pericole posibile și introducerea unor mecanisme de evitare a lor. În *Cartea albă privind inteligența artificială* este clar menționat și faptul că sistemele de IA trebuie să fie dezvoltate luând în considerare „cerințe care să asigure reziliența sistemelor de IA atât împotriva atacurilor deschise, cât și a încercărilor mai subtile de manipulare a datelor sau chiar a algoritmilor, precum și măsurile de atenuare necesare în astfel de cazuri".

Un aspect care influențează fundamental robustețea tehnică a sistemelor bazate pe învățare automată, în particular pe învățare profundă, este calitatea datelor utilizate în antrenarea acestor sisteme. Multe dintre aceste date sunt produse de oameni, de exemplu postări pe rețelele sociale, imagini și videoclipuri, înregistrări ale unor activități ale persoanelor cum ar fi împrumuturi, date și informații personale etc. În anumite cazuri oamenii pot fi părtinitori sau afectați de prejudecăți și dacă se introduc date preferențiale într-un algoritm de învățare automată, acesta va produce decizii părtinitoare (*biased*). Acest aspect este evident legat și de cerința-cheie a nediscriminării și corectitudinii pentru inteligența artificială de încredere De exemplu, s-a constatat că mulți algoritmi de identificare facială recunosc în mod incorect fețe afro-americane și asiatice cu de 10 până la 30 de ori mai des decât în cazul recunoașterii incorecte a fețelor caucaziene. Alte exemple

de decizii părtinitoare pe baza unor date neadecvate sunt: predicție incorectă a unui nivel de criminalitate mai mare în zone urbane mai sărace, părtinire în selecția automată a candidaților pentru angajarea pe un anumit post influențată de date anterioare în care preferința a fost dată candidaților de sex masculin, traduceri automate de texte dintr-o limbă în alta care neintenționat iau în calcul existența unor prejudecăți sociale (*o bir asci* care înseamnă „sunt bucătar" în limba turcă, unde substantivele sunt neutre ca gen, este tradus în alte limbi în care substantivele au gen „ea este o bucătăreasă"). Utilizarea unor metode corecte de colectare a datelor, a tehnicilor de *debiasing* (eliminarea părtinirii), augmentarea sintetică a datelor, precum și găsirea unor metrici pentru a cuantifica părtinirea sunt subiecte de cercetare care preocupă la ora actuală comunitatea științifică.

Ținând cont de toate aceste aspecte care pot influența robustețea tehnică și siguranța sistemelor bazate pe inteligență artificială, se impune cu necesitate o evaluare corectă a riscurilor potențiale asociate cu utilizarea acestor sisteme pentru diferite domenii de aplicabilitate. Nivelul măsurilor de siguranță necesare depinde, evident, de nivelul de risc, care, la rândul lui, depinde de capabilitățile sistemului. De exemplu, dacă deciziile sistemului pot afecta viața umană, măsurile de siguranță trebuie să fie maxime.

Cadrul de reglementare al utilizării IA la nivelul Uniunii Europene *Artificial Intelligence Act* (menționat anterior) urmează o abordare bazată pe risc și diferențiază utilizările IA care creează (i) un risc inacceptabil, (ii) un risc ridicat și (iii) un risc scăzut sau minim. Ca exemple de aplicații cu risc inacceptabil putem menționa: manipularea persoanelor prin tehnici subliminale dincolo de conștiința lor, exploatarea vulnerabilităților unor grupuri vulnerabile specifice, cum ar fi copiii sau persoanele cu dizabilități, în scopul de a distorsiona comportamentul acestora într-o manieră care este susceptibilă de a le provoca acestora sau altor persoane vătămare fizică sau morală, evaluarea socială bazată pe inteligență artificială în scopuri generale,

realizată de autoritățile publice, practici de manipulare sau exploatare care afectează adulții și care ar putea fi facilitate de sistemele de inteligență artificială, acestea fiind acoperite de legislația existentă privind protecția datelor, protecția consumatorilor și serviciile digitale care garantează că persoanele fizice sunt informate corespunzător și au libertatea de a nu fi supuse profilării sau altor practici care le-ar putea afecta comportamentul.

În cadrul utilizării inteligenței artificiale în aplicații cu risc ridicat se menționează: identificarea biometrică și clasificarea automată a persoanelor, operarea infrastructurilor critice, sistemele de recrutare și selecție pentru diferite joburi, accesul la servicii publice esențiale, sisteme destinate a fi utilizate de autoritățile de aplicare a legii pentru a face evaluări individuale în scopul evaluării riscului unei persoane fizice de săvârșire sau recidivă sau a riscului pentru potențialele victime ale infracțiunilor penale și altele.

Transparența sistemelor de inteligență artificială

Transparența sistemelor de IA implică mai multe aspecte care trebuie considerate: explicarea proceselor de decizie și a mecanismelor prin care sistemul a ajuns la o anumită decizie, transparență asupra datelor utilizate în procesele de învățare, inclusiv modul în care datele au fost colectate și etichetate (proces aflat în strânsă corelare cu necesitatea eliminării părtinirii datelor, descrisă anterior) cât și, pentru multe cazuri, transparența modelului de afaceri care utilizează un anumit produs sau serviciu bazat pe IA.

Modelele de învățare profundă au reprezentat, fără îndoială, unul dintre cei mai puternici factori de progres ai utilizării pe scară largă a inteligenței artificiale. Acestea sunt însă modele de tip *black box* în care procesul de învățare (în particular actualizarea ponderilor rețelei neurale și utilizarea acestora în decizia rețelei) sunt departe de a fi transparente pentru proiectant sau utilizator. Există pericolul de a lua și utiliza decizii care nu sunt justificabile, legitime sau care nu permit

explicarea modului prin care s-a ajuns la aceste decizii. Explicațiile care justifică ieșirea sistemului de IA sunt cruciale, de exemplu, în medicină, transport, securitate, finanțe. Ori de câte ori un sistem de IA are un impact semnificativ asupra vieții oamenilor, ar trebui să fie posibilă oferirea unei explicații adecvate a procesului decizional al sistemului. O astfel de explicație trebuie să fie adaptată la nivelul de cunoștințe al părții interesate, de exemplu utilizator obișnuit, autoritate de reglementare sau cercetător/proiectant al sistemului.

Capacitatea de explicare a modelului poate fi identificată ca fiind diferită de interpretabilitatea modelului. Interpretabilitatea reprezintă descrierea modelului unui sistem de IA în termeni accesibili oamenilor, pe când explicabilitatea se referă la capacitatea de explicare a funcționării modelelor pentru a înțelege modul de raționament și de luare a deciziilor în scopul obținerii încrederii utilizatorilor în sistem sau pentru a permite justificarea deciziilor, dar și pentru a oferi cercetătorilor o bază de corectare și/sau îmbunătățire a sistemului. Paradigmele care stau la baza explicabilității se încadrează în domeniul numit *eXplainable AI* (XAI), această abordare fiind larg recunoscută ca o caracteristică crucială pentru implementarea practică a modelelor de inteligență artificială. Capacitatea de explicare a funcționării unui model opac, de tipul celor utilizate în învățarea profundă, reprezintă un element-cheie în realizarea inteligenței artificiale de încredere și pentru implementarea corectă și responsabilă a tehnologiilor de IA pe scară largă în organizații. Dezvoltările spectaculoase ale modelelor de învățare profundă din punctul de vedere al acurateței și performanței de rezolvare a problemelor, existente în prezent, au condus la sisteme cu o complexitate din ce în ce mai mare și care sunt din ce în ce mai opace. La ora actuală, se impune cu necesitate extinderea acestor sisteme cu capacitatea de explicare și interpretare a modelelor utilizate.

Există diferite strategii pentru ca un sistem de IA să devină explicabil, de exemplu extragerea cunoștințelor din grafuri de cunoștințe sau din modele ale limbajului natural, analiza influenței caracteristicilor

datelor de intrare și identificarea celor mai importante caracteristici care au influențat decizia sau rezultatul, vizualizarea grafică intuitivă și analiza datelor corelate cu predicția realizată, utilizarea exemplelor adversariale și investigarea modului în care acestea afectează deciziile sistemului (acest aspect fiind legat și de robustețea tehnică descrisă anterior). Dacă modelele care utilizează reprezentări simbolice ale cunoștințelor specifice domeniului (reguli de producție, grafuri de cunoștințe sau ontologii) sunt ușor interpretabile și explicabile, nu același lucru se poate spune despre modelele bazate pe rețele neurale profunde. Metodele XAI pentru aceste modele pot fi dependente de model, fiind specifice unei anumite arhitecturi neurale (de exemplu *Layer-wise Relevance Propagation, Gradient-weighted Class Activation Mapping*), sau independente de model, putând fi aplicate mai multor arhitecturi de rețele profunde (de exemplu *Shapley Values*, metodă provenită din teoria matematică a jocurilor). Astfel de abordări, deși aflate la început de drum, pot fi de un real folos cercetătorilor și proiectanților de sisteme cu învățare profundă, punând în evidență caracteristici esențiale ale funcționării sistemului. Cu toate acestea, ele sunt mai puțin utile managerilor sau utilizatorilor aplicațiilor bazate pe IA care nu dețin cunoștințe tehnice specifice. Astfel, comunitatea de cercetători în inteligență artificială este pusă în fața unor provocări importante pentru a găsi atât modalități de inspectare și justificare a funcționării sistemelor la nivel tehnic, cât și abordări care să permită explicarea acestora unui public țintă fără cunoștințe tehnice în domeniul inteligenței artificiale.

Cercetători din cadrul Centrului Internațional de Excelență în Inteligență Artificială al Universității POLITEHNICA din București au început, de curând, să dezvolte tehnici de explicabilitate atât pentru recunoașterea obiectelor și a persoanelor din imagini prin vedere computerizată, cât și pentru explicarea modului în care o rețea neurală recunoaște activități umane din secvențe video, și au efectuat studii de utilizator cu diferite categorii de persoane pentru a evalua satisfacția acestora

referitoare la diverse metode de explicare a deciziilor unui sistem de învățare profundă.

O altă problemă importantă legată de transparența sistemelor de inteligență artificială este aceea de a informa utilizatorii dacă interacționează cu o persoană sau cu un software de IA. Oamenii au dreptul de a fi informați asupra faptului că interacționează cu un sistem de IA. Aceasta înseamnă că sistemele bazate pe inteligență artificială trebuie să fie identificabile ca atare. De exemplu, utilizatorii ar trebui să fie informați dacă interacționează cu o persoană sau cu un agent conversațional cu IA atât pe rețele de socializare, cât și în multiple aplicații de suport al clienților. În plus, utilizatorilor trebuie să li se furnizeze „informații clare în ceea ce privește capacitățile și limitările sistemelor de IA, în special scopul în care sunt concepute sistemele, condițiile în care se preconizează că acestea pot funcționa conform scopului prevăzut și nivelul de precizie preconizat în ceea ce privește atingerea scopului specificat". Transparența este legată și de un alt aspect important, și anume responsabilitatea, adică posibilitatea de a răspunde pentru acțiunile sau deciziile luate și de a identifica entitățile responsabile. Bineînțeles că lipsa de transparență și imposibilitatea explicării deciziilor luate nu permit respectarea cerinței de responsabilitate.

Concluzii

Tehnologiile inteligenței artificiale au și vor avea în viitor, din ce în ce mai mult, un rol fundamental benefic pentru oameni, economie și societate. Inteligența artificială este o tehnologie transformatoare și perturbatoare, cu o evoluție facilitată de disponibilitatea unor cantități enorme de date digitale, progrese tehnologice majore în puterea de calcul și capacitatea de stocare, precum și de inovații științifice și inginerești semnificative în metode și instrumente de IA. Sistemele de IA vor continua să aibă un impact asupra societății și cetățenilor în moduri pe care încă nu ni le putem imagina.

În acest context, nu este suficient doar să dezvoltăm sisteme avansate bazate pe IA, ci trebuie ca aceste sisteme să fie de încredere: cei care creează sistemele trebuie să fie siguri că sistemul se comportă conform specificațiilor, cei care utilizează sistemul pentru a lua decizii trebuie să aibă încredere în rezultatele și expertiza sistemului, iar cei care sunt afectați de deciziile unui sistem de IA trebuie să aibă încrederea că sistemul este sigur și corect. Pentru ca inteligența artificială să devină de încredere, trebuie ca cetățenii să se simtă în siguranță, să fie convinși că beneficiază de siguranță în utilizarea sistemelor de IA și că se bucură în continuare de toate drepturile și libertățile lor fundamentale. În plus, lipsa încrederii în IA poate fi un factor limitativ în adoptarea pe scară largă a inteligenței artificiale. Specialiștii în IA și factorii de decizie trebuie să fie parte integrantă a buclei de decizie a sistemelor, să poată prelua controlul în situații critice limită – dorim sisteme autonome, dar decizia de a ceda controlul în contexte specifice rămâne aceea a oamenilor, iar aceste sisteme trebuie să devină parteneri de încredere ai oamenilor.

Trebuie să ne asigurăm că riscurile și alte efecte adverse potențiale asociate tehnologiilor inteligenței artificiale sunt gestionate în mod corespunzător, astfel încât oamenii să profite cu încredere și pe deplin de beneficiile sistemelor bazate pe aceste tehnologii. Este de netăgăduit necesitatea unui cadru legislativ corespunzător de reglementare a utilizării sistemelor de IA, care să protejeze drepturile fundamentale ale oamenilor, inclusiv protecția datelor cu caracter personal și asigurarea nediscriminării. Un astfel de cadru de reglementare trebuie să fie flexibil, adaptabil la evoluții ulterioare ale tehnologiei și, în același timp, să nu îngrădească dezvoltarea pe mai departe a inteligenței artificiale, cu potențialul ei imens de a schimba în bine viețile oamenilor, și utilizarea acesteia pe scară largă în industrie și societate spre beneficiul tuturor.

Inteligența artificială vs drepturile omului. Riscuri vs oportunități

Conf. univ. dr. Laura Maria Stănilă,
Facultatea de Drept, Universitatea de Vest din Timișoara
Director, Centrul de Cercetări în Științe Penale
Avocat, Baroul Timiș

I. Aspecte introductive cu privire la inteligența artificială (IA) – factor-cheie al sistemului de justiție penală

IA este un nou actor în zona relațiilor sociale, întrucât pare a fi un panaceu universal. IA simplifică procesul decizional pentru oameni, datorită algoritmilor săi matematici perfecți, obiectivi, are ca scop declarat să ne facă viața mai ușoară și mai plăcută, să ne degreveze de sarcinile împovărătoare și să ne permită să dedicăm mai mult timp sarcinilor și lucrurilor care „contează"[1]. Prin „inteligență" înțelegem „capacitatea de a face predicții despre viitor și de a rezolva sarcini complexe", această abilitate fiind demonstrată de algoritmii pe baza cărora funcționează telefoanele moderne (*smartphones*), tabletele, laptopurile, dronele, vehiculele autonome sau roboții. Toate aceste *lucruri* ar putea prelua sarcini, „variind de la întreținerea gospodăriei, asigurarea asistenței și companiei de orice fel pentru oameni (inclusiv

[1] M. Risse, *Human Rights and Artificial Intelligence. An Urgently Needed Agenda*, Carr Center for Human Rights Policy, USA 2018, p. 2, https://carrcenter.hks.harvard.edu/files/cchr/files/human-rightsai_designed.pdf (accesat la 22.09.2022).

companie sexuală), până la realizarea activităților de poliție sau de război"[2].

Sistemele IA urmăresc să îmbunătățească modalitățile de acțiune în activitatea companiilor și autorităților publice din întreaga lume, determinând schimbări constante ale acestora, conducând inevitabil la interferențe cu problematica drepturilor omului. Doctrina a arătat că „reglementările privind protecția datelor cu caracter personal adoptate până acum. precum și constituirea de garanții pentru responsabilitate și transparență ar putea fi capabile să atenueze unele dintre cele mai nocive utilizări ale AI cunoscute astăzi, dar este nevoie de mai multă muncă pentru a proteja drepturile omului, pe măsură ce tehnologia AI devine mai sofisticată și se extinde în alte domenii"[3].

Dar sub denumirea de „inteligență artificială" (IA), există o sumă de sisteme, dispozitive, tehnologii și algoritmi. De fapt, conceptul de „inteligență artificială" a fost calificat ca fiind un concept „umbrelă" sub care sistemele A pot fi clasificate în două categorii:

a) sisteme bazate pe cunoștințe și

b) sisteme având capacitatea de a îmbunătăți continuu performanța decizională.

a) Sistemele bazate pe cunoștințe includ „sisteme expert" care utilizează logica formală și reguli codificate pentru a se angaja în raționament. De exemplu, un software comercial de calculare a impozitelor, algoritmi de asistență pentru luarea deciziilor de diagnostic al sănătății. Aceste tipuri de sisteme IA iau decizii optime pe baza unor reguli definite într-un anumit domeniu și, de obicei, nu pot învăța sau nu pot folosi automat informațiile pe care le-au acumulat de-a lungul timpului pentru a îmbunătăți calitatea procesului decizional.

[2] *Ibidem*.

[3] L. Andersen, 2018, *Human Rights in the Age of Artificial Intelligence*, Access Now, p. 37, https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf (accesat la 22.09.2022).

b) Sistemele având capacitatea de a îmbunătăți continuu performanța decizională utilizează învățarea statistică pentru a realiza „învățarea automată" (*machine learning*) și „învățarea profundă" (*deep learning*). Ar intra în această categorie, spre exemplu, vehiculele autonome, sistemele de recunoaștere facială (*facial recognition systems*) utilizate de forțele de poliție, sistemele de traducere automată (*automated translation systems – language systems*), algoritmii care ne spun ce să urmărim în continuare pe serviciile de streaming video etc. Această categorie de sisteme IA a fost criticată pentru faptul că nu este fiabilă la nivel individual: „De exemplu, sistemele *deep learning computer vision* pot clasifica o imagine aproape la fel de precis ca un om; cu toate acestea, ocazional vor face greșeli pe care niciun om nu le-ar face - cum ar fi confundarea unei fotografii a unei broaște țestoase cu o armă".[4]

IA sub formă de „rețele neuronale" (*neuronal networks*) este din ce în ce mai utilizată în tehnologii precum autovehicule cu conducere autonomă pentru a putea identifica și recunoaște obiectele și obstacolele de pe carosabil. Astfel de sisteme ar putea fi folosite și pentru identificarea explozivilor în liniile de securitate a spațiilor publice, dar, după cum au arătat studiile, acestea mai fac și greșeli. O echipă de cercetare a arătat că, nu numai că a fost capabilă să păcălească o rețea neuronală să creadă că un pistol este un alt obiect, dar a putut de fapt determina sistemul IA să clasifice un anumit obiect fizic ca orice alt obiect. Echipa de cercetare a schimbat ușor textura obiectului, astfel încât o bombă a fost clasificată drept legumă și a putut chiar să facă un obiect complet invizibil pentru IA. Consecințele utilizării unui astfel de sistem IA ar fi dezastruoase dacă s-ar realiza implementarea sa la scară largă[5].

[4] A. Conner-Simons, „Fooling neural networks w/3D-printed objects", MIT Computer Science, https://www.csail.mit.edu/news/fooling-neural-networks-w3d-printed-objects (accesat la 22.09.2022).

[5] *Ibidem*.

O altă clasificare doctrinară[6] vine în sprijinul publicului mai puțin cunoscător în acest domeniu extrem de tehnic și împarte sistemele IA în patru categorii mai cuprinzătoare:

1) *sisteme care gândesc ca oamenii* (de exemplu, arhitecturi cognitive și rețele neuronale);
2) *sisteme care acționează ca oamenii* (de exemplu, cele care trec testul Turing, reprezentarea cunoștințelor, raționamentul automat și învățarea automată);
3) *sisteme care gândesc rațional* (de exemplu, algoritmi de soluționare logică, inferențe, cei care realizează optimizarea);
4) *sisteme care acționează rațional* (de exemplu, agenți software inteligenți și roboți încorporați care ating obiectivele prin percepție, planificare, raționament, învățare, comunicare, luarea deciziilor și acționare).

Ca orice nouă tehnologie care nu a trecut încă testul timpului, IA prezintă atât oportunități, cât și riscuri privind utilizarea sa generală sau specifică. Confruntați cu asimetriile la nivel de informații ale procesului decizional algoritmic, cetățenii se tem că nu vor avea nicio putere să își apere drepturile și siguranța, iar entitățile colective sunt preocupate de problema insecurității juridice. Deși, pe de o parte, IA poate contribui la protejarea securității cetățenilor și îi poate ajuta să se bucure de drepturile lor fundamentale, pe de altă parte, poate avea efecte nedorite sau chiar poate fi utilizată în scopuri malefice. În viziunea Comisiei Europene, „este necesar ca aceste preocupări să fie luate în considerare. În plus, pe lângă lipsa investițiilor și a competențelor, lipsa de încredere este un factor principal care împiedică adoptarea pe scară mai largă a AI"[7].

[6] S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall Series, în *Artificial Intelligence*, Englewood Cliffs, N.J. Prentice Hall, 1995, p. 31-53. https://www.cin.ufpe.br/~tfl2/artificial-intelligence-modern-approach.9780131038059.25368.pdf (accesat la 22.09.2022).

[7] Comisia Europeană, *Cartea albă privind inteligența artificială - O abordare europeană axată pe excelență și încredere*, Bruxelles, 19.02.2020, COM(2020) 65 final, www.ec.europa.eu.

Raportat la domeniul specific al justiției penale, IA pare să aibă o *triplă valență*, unele aspecte având caracter actual, altele fiind potențiale, dar extrem de realizabile în viitorul nu atât de îndepărtat. Astfel, am arătat cu altă ocazie[8] că IA poate deveni, la modul cel mai plauzibil, subiect de drept în general și subiect de drept penal în special; de asemenea, IA reprezintă, la ora actuală, un mijloc eficient de comitere a unei tipologii de infracțiuni cu privire la care există deja codificări și eforturi concertate de combatere[9]; nu în ultimul rând, IA poate deveni un instrument extrem de util și chiar un mijloc de realizare a justiției penale, fiind deja implementat în anumite faze ale procesului penal din unele state, și de asemenea, un instrument de management eficient al actului de justiție prin aplicații specifice ale digitalizării în procesul penal (ex. dosarul electronic, semnătura electronică, arhivarea cu ajutorul cloudului și al algoritmilor de clasificare etc.).

II. Domenii de utilizare IA în sistemul de justiție penală

Cu referire la cea de-a treia valență a IA - instrument util și mijloc de realizare a justiției penale – la ora actuală au fost identificate cinci domenii specifice: utilizarea IA în faza de investigație penală, utilizarea IA pentru evaluarea riscului penal al infractorilor, poliția predictivă (*predictive policing*), utilizarea IA pentru managementul dosarului și utilizarea IA pentru previzionarea deciziei judiciare, ca instrument la dispoziția

[8] L.M. Stănilă, *Inteligența artificială, dreptul penal și sistemul de justiție penală. Amintiri despre viitor,* Universul Juridic, București, 2020.

[9] Este vorba despre categoria infracțiunilor informatice (*cybercrime*) a căror combatere reprezintă obiectul Convenției de la Budapesta din 2001 privind criminalitatea informatică (ratificată de România prin Legea nr. 64/2004 și publicată în Monitorul Oficial al României, Partea I, nr. 343 din 20 aprilie 2004), dar și al unor directive europene recente: Directiva 2013/222/JAI privind atacurile împotriva sistemelor informatice, https://eur-lex.europa.eu/legal-content/RO/TXT/PDF/?uri=CELEX:32013L0040&from=en (accesat la 22.09.2022); Directiva (UE) 2019/713 privind combaterea fraudelor și a contrafacerii în legătură cu mijloacele de plată fără numerar, https://eur-lex.europa.eu/legal-content/RO/TXT/?uri=CELEX:32019L0713 (accesat la 22.09.2022).; Directiva 2011/93/UE a Parlamentului European și a Consiliului din 13 decembrie 2011 privind combaterea abuzului sexual asupra copiilor, a exploatării sexuale a copiilor și a pornografiei infantile și de înlocuire a Deciziei-cadru 2004/68/JAI a Consiliului, https://eur-lex.europa.eu/legal-content/RO/TXT/?uri=celex%3A32011L0093 (accesat la 22.09.2022).

practicienilor, în special a avocaților inculpatului și a persoanelor vătămate ori responsabile civilmente care stau în procesul penal.

A) Utilizarea IA în faza de investigație penală

Având în vedere că fenomenul infracțional a dobândit noi trăsături și forme, conturându-se un fenomen expansiv al criminalității informatice, este evident că investigarea infracțiunilor necesită o updatare a metodelor, tehnicilor și instrumentelor pe care organele judiciare să le folosească cu succes în scopul identificării și tragerii la răspundere penală a infractorilor, necesită o specializare a lucrătorilor din sistem în domeniul IT și, de asemenea, necesită o evaluare a impactului acestor metode, instrumente și tehnici noi raportate la standardul unui proces echitabil, astfel cum este acesta configurat de art. 6 din CEDO.

Creșterea utilizării IA în sistemul de justiție penală și deschiderea organelor judiciare către astfel de instrumente se datorează, în primul rând, unei schimbări de paradigmă, accentul politicii penale actuale fiind pus pe acțiuni preventive referitoare la criminalitatea de orice fel. În altă ordine de idei, faza de investigație a procesului penal necesită multă muncă, prelucrarea unei cantități mari de date, un demers dificil de procesare a acestora, elemente care generează adeseori erori umane. Dovezile sunt colectate de către organele judiciare pentru a susține acuzațiile și a demonstra conduita infracțională și vinovăția infractorului, însă noile tipuri specifice de infracționalitate (ex. infracționalitate cibernetică, infracționalitate transnațională, crima organizată) fac eforturile organelor judiciare extrem de dificile. Agenții de aplicare a legii sunt nevoiți să găsească noi modalități, instrumente și mijloace care să le ușureze munca pentru a obține rezultatele scontate în lupta împotriva fenomenului criminalității. Din această perspectivă, în mod evident, IA este un instrument foarte util, care ajută actorii judiciari să gestioneze toate dificultățile.

Astfel, între aplicațiile IA utilizate deja în procesul penal în această fază de investigare amintesc doar câteva:

a.1. *Chatbots*

Chatboturile sunt aplicații software utilizate pentru a desfășura o conversație online (text sau vorbire) în loc să ofere contact direct cu un agent uman. Este de notorietate cazul controversatului chatbot *Sweetie* care s-a dovedit a fi un instrument extrem de eficient în identificarea prădătorilor sexuali din mediul online și în deconspirarea rețelelor și comunităților de pedofili care acționează în mediul online, dar care nu satisface nici standardul dreptului la un proces echitabil, nici construcția tradițională a infracțiunii și angajării răspunderii penale pentru comiterea unei infracțiuni care se fundamentează pe conceptul de victimă umană. Pe scurt, acest chatbot, creat de o companie olandeză privată, a fost calificat drept cel mai bun instrument[10] în investigarea infracțiunilor de pornografie infantilă, deoarece nu ar exista, în cazul utilizării sale, victime umane, iar faptele penale nu sunt comise efectiv[11]. Crearea și utilizarea acestor instrumente inovatoare este condiționată de acceptarea factorilor de decizie și de cooperarea dintre aceștia și sectorul privat, care, deși este principalul actor în domeniul inteligenței artificiale, este axat exclusiv pe profit. „Atâta timp cât opiniile învechite ale factorilor de decizie privind prevenirea criminalității și abordarea centrată pe profit a sectorului privat prevalează față de metodele neconvenționale de combatere a infracționalității, aceste tipuri de soluții teoretice sunt condamnate să rămână doar pe hârtie".[12] *Sweetie* nu mai poate fi însă folosit de către autoritățile din Olanda deoarece, ca urmare a unui studiu[13] desfășurat de către profesorii de la Universitatea din Leiden în anul 2016, s-a constatat că nu este compatibil cu exigențele sistemului juridic olandez. Însă acest raport a determinat adoptarea unei noi

[10] Terre des Hommes, „First conviction for child abuse in Belgium thanks to Sweetie", 9.04.2015, https://www.terredeshommes.nl/en/news/first-conviction-child-abuse-belgium-thanks-sweetie (accesat la 12.09.2022).

[11] A se vedea videoul cu privire la *Sweetie*: https://youtu.be/aGmKmVvCzkw?t=10 (accesat la 13.09.2022).

[12] K.V. Açar, „Webcam Child Prostitution: An Exploration of Current and Futuristic Methods of Detection", *International Journal of Cyber Criminology*, ianuarie – iunie 2017, vol. 11(1): 98–109, DOI: 10.5281/zenodo.495775 (accesat la 12.09.2022).

[13] B. W. Schermer, I. Georgieva, S. Van der Hof, B-J. Koops, *Legal Aspects of Sweetie 2.0*, Leiden/Tilburg: TILT, 2016, p. 10.

legislații în Olanda, privitoare la criminalitatea informatică – *Legea privind criminalitatea informatică III (Wet Computercriminaliteit III)* – adoptată la 21 septembrie 2018 și care a intrat în vigoare la 1 martie 2019. Acest act normativ a reușit să îmbunătățească legislația procedurală și substanțială penală olandeză prin modificarea Codului penal olandez (DCC) și a Codului de procedură penală olandez (DCCP), permițând instanțelor și poliției să acceseze calculatoarele în mod disimulat și la distanță, atât computere personale, cât și telefoane mobile sau servere, pentru a investiga infracțiuni grave, cum ar fi pornografia infantilă, traficul de droguri și utilizarea premeditată a armelor de foc, și oferind agenților de investigații puterea de a aplica diverse tactici de investigare, cum ar fi să facă anumite date inaccesibile, să copieze fișiere și să acceseze diferite canale de comunicare. Acest lucru va face mai dificilă utilizarea internetului de către infractori cu scopul de a evita descoperirea faptelor de către autorități.

a.2. Analiza Big Data de către companii VoIP[14]

Companiile VoIP pot realiza două tipuri de analiză Big Data: analiza metadatelor și analiza conținutului datelor. *Analiza metadatelor* presupune colectarea și analiza unor atribute ale comunicațiilor, cum ar fi data, creatorul și adresele IP, fără a compromite în mod sever confidențialitatea comunicațiilor, dar contribuie la obținerea de indicii sau chiar probe cu privire la comiterea unei infracțiuni (ex. cazul în care o companie VoIP citește „semnalele" sau indicatorii de infracționalitate - un locuitor al unui oraș foarte sărac discută cu mai mulți străini din țări relativ mai bogate – apoi dezvăluie adresele IP și alte informații utile precum adrese de e-mail autorității de aplicare a legii pentru investigații suplimentare.[15] *Analiza conținutului datelor* se bazează pe

[14] *Voice-over-IP (VoIP)* este o metodă eficientă de comunicare. VoIP implică trimiterea de transmisii vocale sub formă de pachete de date utilizând Protocolul Internet (IP), prin care vocea utilizatorului este transformată într-un semnal digital, comprimată și defalcată într-o serie de astfel de pachete. Pachetele sunt apoi transportate prin rețele IP private sau publice și reasamblate și decodate de entitatea care le primește. A se vedea U. Varshney, A. Snow, M. McGivern, C. Howard, „Voice over IP", Communications of the ACM, 45(1): 89-96, 2002, p. 89, https://dl.acm.org/doi/10.1145/502269.502271 (accesat la 12.09.2022).

[15] K.V. Açar, *art. cit.*

o analiză de conținut a datelor concrete și expune informații specifice despre conversațiile VoIP realizate între părți: texte, fișiere audio și video, punând astfel probleme legate de confidențialitate, protecția datelor personale și prezumția de nevinovăție. Există mecanisme și garanții legale care să permită o astfel de analiză și utilizarea sa în procesul penal, cum ar fi, spre exemplu, limitarea duratei, autorizarea prealabilă din partea judecătorului și existența unei căi de atac.

a.3. Software specializat: face recognition (ex. FACEFIRST), language recognition (ex. BRAINS)

Există multe exemple de software concepute în mod expres pentru a fi utilizate în faza de investigare a procesului penal: *Face First*[16], *Analist Notebook* by IBM[17], *HOLMES 2 (Home Office Large Major Enquiry System)*[18], *BRAINS, FLINTS (Forensic-Led Intelligence System)*[19]. FLINTS a fost folosit pentru prima dată de poliția britanică în 1999 pentru gestionarea probelor medico-legale. A fost concepută și o aplicație Google care blochează termenii de căutare ce sunt asociați pornografiei

[16] De exemplu, *FaceFirst* oferă supraveghere, control acces, recunoaștere facială, date biometrice etc. Software-ul este capabil de performanțe impresionante: poate identifica o persoană făcând până la 75 de milioane de comparații în 0,1 secunde; primește alerte pozitive în timp real; notifică orice număr de sisteme și dispozitive mobile, astfel încât să poată fi utilizate cu succes în prevenirea criminalității. Poate detecta figura unei persoane în înregistrări video, poate detecta trăsături faciale, poate produce clasificări (în funcție de sex, vârstă, etnie), poate identifica mai multe figuri simultan, poate semnala pericolul pe care îl prezintă o persoană etc. A se vedea https://www.facefirst.com.

[17] https://www.ibm.com/security/intelligence-analysis/i2/law-enforcement (accesat la 22.09.2022).

[18] HOLMES 2 (*Home Office Large Major Enquiry System*) este un sistem de tehnologia informației utilizat de poliția din Marea Britanie pentru a facilita investigațiile în descoperirea omorurilor și a fraudei. A fost dezvoltat de Unisys, fiind susținut de Inițiativa Financiară Privată. Denumirea aleasă face referire la personajul scriitorului Arthur Conan Doyle, Sherlock Holmes. Mai multe informații pot fi găsite la adresa http://www.holmes2.com/holmes2/index.php (accesat la 22.09.2022).

[19] E. Nissan, *Computer Applications for Handling Legal Evidence, Police Investigation and Case Argumentation*, vol. 1, Springer, 2012, p.767-836. În ceea ce privește FLINTS și beneficiile sale, a se vedea, de asemenea, A.R.W. Jackson, J.M. Jackson, *Forensic Science,* ediția a doua, Pearson, 2008, p. 7.

infantile sau crearea *unor* baze de date pentru urmărirea imaginilor cunoscute cu conținut sexual și extragerea lor offline[20].

a.4. Sisteme expert – profiling (HOLMES)

Acestea sunt programe de calculator care utilizează tehnologii IA pentru a simula judecata și comportamentul unui om sau al unei organizații care au acumulat cunoștințe și expertiză într-un anumit domeniu. Aceste sisteme încorporează o bază de cunoștințe care conține experiență acumulată și un motor de inferență care reprezintă un set de reguli pentru aplicarea bazei de cunoștințe fiecărei situații particulare care este evaluată cu ajutorul programului. Utilizarea lor uzuală este pentru derularea de activități de profilare a actelor infracționale, a comportamentului și modului de raționament infracțional, a comportamentului și modului de raționament al investigatorului și pentru a realiza profiluri specifice ale infractorilor – *profiling*.

a.5. Analiza ADN prin intermediul softurilor specializate

IA joacă un rol important în analiza ADN, datorită capacității sale de a accelera semnificativ compararea secvențelor genetice, ADN-ul colectat trebuind să fie comparat cu ADN-ul conținut într-o bază de date extrem de mare, pe fondul tendinței crescânde a sistemului judiciar de a utiliza baze de date ADN ale site-urilor comerciale care colectează probe pentru a identifica genealogia unei persoane în funcție de cost și la cerere. De exemplu, poliția din California a decis să utilizeze astfel de date în 2018 pentru a-l captura pe Joseph James DeAngelo, în vârstă de 72 de ani, un criminal în serie poreclit „Golden State Killer", care a comis mai multe crime și violuri între 1974 și 1986. Baza de date utilizată pentru identificarea suspectului a fost preluată de pe GEDMatch, un site folosit de amatori și profesioniști pentru a crea arbori genealogici. Însă, raportat la contextul legal și al drepturilor și libertăților cetățenilor, analiza ADN poate genera date care pot

[20] Thorn, amintit de K. Schweizer, „Avatar Sweetie exposes sex predators", The Age, 26 aprilie 2014, https://www.theage.com.au/world/avatar-sweetie-exposes-sex-predators-20140425-379kf.html (accesat la 22.09.2022).

contribui la încălcarea unor drepturi și libertăți și pot perpetua modele discriminatorii[21] (ex. din secvențele ADN se poate deduce rasa infractorului).

B) Evaluarea riscului penal

Evaluarea riscului penal a constituit un punct focal al justiției penale și obiect al preocupării actorilor implicați pentru a asigura obiectivitatea deciziei judiciare și acuratețea sancțiunii penale aplicate la cazul concret. Conceptul de justiție individualizată este dominant în dreptul penal modern, iar evaluarea riscului personal (periculozitatea infractorului) s-a realizat tradițional de către organele judiciare, în baza experienței profesionale și a principiilor și regulilor procedurii judiciare. Fără a echivala cu o înlocuire a evaluării umane, evaluarea statistică, obiectivă, bazată pe recunoașterea unor tipare comportamentale, tinde să câștige teren. Instrumentele actuariale, algoritmice, de învățare automată bazate pe IA promit să ofere capacități predictive precise și evaluări obiective, consecvente ale riscurilor penale. Această tendință de politică penală judiciară s-a realizat pe fondul dezvoltării doctrinare a doctrinei actuariale, care are la bază modelul mecanic de evaluare a riscurilor în care infractorii sunt analizați în funcție de o serie de itemi care au fost cel mai puternic asociați cu recidiva în stadiile de probă. Apoi, scorurile totale sunt încrucișate cu tabelele de risc actuarial.

Instrumentele predictive bazate pe IA sunt programe IA care, în baza inputului (date cu care este hrănit sistemul și prin intermediul cărora acesta învață să recunoască tipare) realizează o evaluare a riscului de reiterare a comportamentului infracțional (output) bazat pe media statistică. În variantele cele mai sofisticate un asemenea sistem poate prezice recidiva specializată (ex. pentru infracțiuni sexuale sau care implică violența). Outputul servește ca fundament pentru decizia judiciară în diferite faze ale procesului penal, atât în faza predecizională (oportunitatea arestării preventive ca urmare a evaluării riscului lăsării

[21] M. K. Cho, P. Sankar, „Forensic genetics and ethical, legal and social implications beyond the clinic", Nature Genetics 2004; vol. 36 (11 Suppl): S8–12, doi: 10.1038/ng1434.

în libertate), cât și în cea postdecizională (pentru alegerea modalității de executare a pedepsei cu suspendare sau în detenție, acordarea liberării condiționate, alegerea unor programe de reintegrare socială pe parcursul executării pedepsei etc.). Aceste instrumente sunt deja utilizate atât de judecători în procesul decizional în diverse state ale lumii, cât și de organele administrative de la locul de deținere pentru a asigura un management eficient al riscului în spațiul de deținere. *COMPAS - Correctional Offender Management Profiling for Alternative Sanctions, IORNS - Inventory of Offender Risk, Needs, and Strengths, LSI-R (Level of Service Inventory-Revised), LS/CMI (Level of Service/ Case Management Inventory)* și *LS/RNR (Level of Service/Risk, Need, Responsivity)* reprezintă doar câteva exemple implementate în SUA care au deja o tradiție de utilizare în sistemul american de justiție penală[22]. Problema cu utilizarea acestor instrumente a fost că au generat rezultate care s-au dovedit a fi părtinitoare, prin perpetuarea unor modele discriminatorii care se strecuraseră în datele de input, fapt care a dus chiar la contestarea utilizării acestora în justiție. Este de notorietate cazul *Loomis*[23] din Wisconsin în care judecătorul de primă instanță a ordonat realizarea unei evaluări a riscului prin utilizarea softului COMPAS, pe baza căreia l-a condamnat pe inculpatul Eric Loomis la închisoare. Deși în apel acesta a invocat caracterul discriminatoriu și excesiv de general al rezultatului, la care a adăugat necesitatea individualizării pedepsei raportat la persoana infractorului, iar nu la date statistice, atât instanța de apel, cât și Curtea Supremă din Wisconsin mai târziu au respins toate argumentele inculpatului, arătând că raportul COMPAS nu ar putea fi niciodată singura probă pe care să se fundamenteze decizia de condamnare, instanțele judecătorești putând acționa discreționar și obține toate informațiile necesare pentru a respinge evaluarea unui soft de acest gen, în fiecare caz în parte.

[22] A se vedea L.M. Stănilă, *op. cit.*, p. 157 și urm.

[23] Curtea Supremă din Wisconsin, *State of Wisconsin v. Eric L. Loomis*, N3. 2015AP157–CR, Decizia din 13 iulie 2016 (881 N.W.2d 749 Wis., 2016), https://caselaw.findlaw.com/wi-supreme-court/1742124.html (accesat la 22.09.2022).

C) Poliție predictivă

Acest domeniu presupune utilizarea de predicții statistice cu privire la potențialitatea producerii unor activități infracționale, prin utilizarea unor programe informatice. Poliția predictivă cunoaște două modalități, *predicția locației* și *predicția persoanei.*

c.1. Predicția locației presupune utilizarea softului specializat pentru interpretarea de date retroactive de infracționalitate pentru a face previziuni referitoare la momentul și locația unde este posibil să se comită ulterior infracțiuni (ex. magazine de băuturi alcoolice, baruri, parcuri unde s-au comis anumite infracțiuni în trecut). Plecând de la aceste predicții, infracționalitatea ar putea fi prevenită prin implementarea unor măsuri suplimentare: patrulare în zonă, creșterea efectivelor polițienești, supraveghere permanentă etc.

c.2. Predicția persoanei presupune utilizarea softului specializat pentru a facilita identificarea potențialilor agresori sau victime cu ajutorul datelor statistice. Spre exemplu, *PredPol*[24] este un algoritm dezvoltat de Universitatea din California, Los Angeles (UCLA) la cererea Departamentului de Poliție din Los Angeles în 2012 care utilizează seturi de date colectate pe o perioadă cuprinsă între 2 și 5 ani care nu cuprind date cu caracter personal, demografice, etnice sau socioeconomice, eliminând astfel riscul producerii unei predicții părtinitoare. Cu toate acestea, studiile arată că, în ciuda precauțiilor celor care îl utilizează și îl hrănesc cu date noi, opiniile părtinitoare sunt totuși prezente și perpetuate în predicții.[25]

D) Managementul dosarului

Așa cum am punctat deja, IA se dovedește a fi un instrument foarte bun pentru managementul dosarelor, atât în cadrul instanțelor, cât

[24]A se vedea www.predpol.com.

[25] A. Sankin, D. Mehrotra, S. Mattu, D. Cameron, A. Gilbertson, D. Lempres, J. Lash, „Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them", Gizmodo, 2 decembrie 2021, https://gizmodo.com/crime-prediction-software-promised-to-be-free-of-biases-1848138977 (accesat la 22.09.2022).

și în ceea ce privește activitatea avocaților, notarilor, practicienilor în insolvență etc. Dosarul electronic[26] este o aplicație care permite părților vizionarea în timp real pe o platformă online special creată a tuturor actelor și lucrărilor efectuate în cadrul unei cauze, părțile putând vizualiza și descărca documentele încărcate în baza de date a instantei, create în cadrul acesteia sau primite la dosarul unei cauze, în măsura în care acestea din urmă au putut fi digitalizate, în limita resurselor avute la dispozitie. Documentele electronice se semnează cu semnătura electronică[27], iar cererile în justiție pot fi realizate și depuse online, scutind părțile de efortul fizic specific deplasării la sediul organului judiciar și reducând foarte mult timpul necesar efectuării unor acte în instanță. Există și alte aplicații *blockchain* de tipul *smart contracts* – contracte inteligente, dar ele nu fac obiectul prezentului studiu. Ca să concluzionez, IA este extrem de utilă întrucât face viața juriștilor mult mai ușoară și îi ajută să își eficientizeze activitatea.

E) Previzionarea deciziei judiciare

Algoritmii pot fi utilizați, cu serioase rezerve legate de etica profesională, dar cu multiple avantaje din punctul de vedere al anticipării soluției litigiului și al eficientizării activității cu clienții, în ceea ce privește predicțiile deciziilor judiciare. Astfel au fost creați algoritmi care prezic rezultatul cauzelor judecate la Curtea Supremă a SUA cu o precizie de 70%[28] și, respectiv, în cazul altui model, de 75%. Pentru comparație, experții umani au înregistrat o precizie de 59%.[29]

[26] https://portal.just.ro/300/SiteAssets/SitePages/acasa_default/Instrucțiuni%20de%20utilizare%20a%20aplicației%20Dosar%20Electronic%20JS2.pdf (accesat la 22.09.2022).

[27] Introdusă în România prin Legea nr. 455/2001 privind semnătura electronică, republicată în Monitorul Oficial al României, Partea I, nr. 316 din 30 aprilie 2014.

[28] K. Ward, D. M. Katz, „Using data to predict Supreme Court decisions", Michigan State University, 4 noiembrie 2014, https://msutoday.msu.edu/news/2014/using-data-to-predict-supreme-courts-decisions/ (accesat la 22.09.2022).

[29] T. W. Ruger, P. T. Kim, A. D. Martin, K. M. Quinn, „The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking", în *Columbia Law Review*, vol. 104, nr. 4 (mai 2004), p. 1150-1210, https://www.jstor.org/stable/4099370 (accesat la 22.09.2022).

III. IA și drepturile omului

Contextul discuției privind IA și drepturile omului în sistemul de justiție penală constituie, de departe, cea mai „la modă" dezbatere doctrinară la ora actuală. Aceasta deoarece, parafrazând o idee care circulă în mediul online, „băieții răi s-au digitalizat, băieții buni încă nu, deoarece au o serie de limitări impuse de principiile Statului de drept". Firește că IA aduce multiple beneficii în sistemul de justiție penală și firește că reprezintă un instrument extrem de util și eficient în lupta pentru combaterea fenomenului infracțional, însă întreaga construcție a sistemului de justiție penală este centrată pe om și pe respectarea drepturilor și libertăților acestuia, fiind rezultatul a secole de evoluție socială și juridică.

Pe de altă parte, chestiunea drepturilor omului în relație cu utilizarea IA trebuie analizată în două direcții: pe de o parte se impune a fi evaluat impactul utilizării IA raportat la exercitarea drepturilor și libertăților în lumea (realitatea) fizică, iar pe de altă parte se impune o analiză distinctă a modului în care se exercită drepturile omului în mediul online.

A) Impactul utilizării IA în sistemul de justiție penală asupra drepturilor și libertăților omului care se exercită în realitatea fizică

În mod evident, raportat la multiplele utilizări ale IA în sistemul de justiție penală pe care le-am prezentat anterior, există riscuri însemnate pentru drepturile fundamentale ale omului, inclusiv în ceea ce privește protecția datelor cu caracter personal, a confidențialității acestora și nediscriminarea. Fiind introduși într-un sistem, utilizarea algoritmilor poate genera probleme de securitate cibernetică, probleme asociate cu aplicațiile IA în infrastructurile critice sau legate de utilizarea malițioasă a IA (spre exemplu, atacuri cibernetice asupra unor platforme și sisteme informatice utilizate în sistemul de justiție penală). Utilizarea IA poate afecta, de asemenea, dreptul la libertatea de exprimare, libertatea de întrunire, demnitatea umană, nediscriminarea bazată pe sex, origine rasială sau etnică, religie sau credință, dizabilitate, vârstă

sau orientare sexuală, după caz, în anumite domenii, protecția datelor cu caracter personal și a vieții private sau chiar dreptul la o cale de atac judiciară efectivă și un proces echitabil. Aceste riscuri pot avea cauze multiple: pot fi generate de proiectarea deficientă a sistemelor IA, de o utilizare nediligentă a acestora, fără a corecta eventualele opinii părtinitoare care se strecoară în datele de input (de exemplu, sistemul este instruit folosindu-se numai sau în principal date referitoare la persoane de sex masculin, ceea ce duce la rezultate părtinitoare în ceea ce privește persoanele de sex feminin). Prin faptul că analizează cantități mari de date și realizează conexiuni între acestea, IA poate fi utilizată în vederea retragerii și dezanonimizării datelor cu privire la anumite persoane, creând noi riscuri de protecție a datelor cu caracter personal chiar și în ceea ce privește seturile de date care în sine nu includ date cu caracter personal (metadate). IA este folosită și de intermediarii online pentru a prioritiza informațiile pentru utilizatorii lor și pentru a efectua moderarea conținutului. Datele prelucrate, modul în care sunt concepute aplicațiile și domeniul de intervenție umană pot afecta drepturile la libera exprimare, protecția datelor cu caracter personal, confidențialitatea și libertățile politice.

Anumiți algoritmi IA, cei care se utilizează în vederea evaluării riscului penal și în realizarea de predicții în ceea ce privește reiterarea comportamentului infracțional (riscul de recidivă) pot genera opinii părtinitoare discriminatorii, fapt demonstrat științific prin numeroasele studii desfășurate în ultimii 20 de ani.

În cazurile în care opinia părtinitoare nu ar fi putut fi prevenită sau anticipată în faza de proiectare a sistemului IA, riscurile nu vor decurge dintr-un defect în proiectarea originală a sistemului, ci, mai degrabă, din impactul practic al corelațiilor sau modelelor pe care sistemul le identifică într-un set mare de date. „Caracteristicile specifice ale multor tehnologii IA, inclusiv opacitatea („efectul de *black box*"), complexitatea, imprevizibilitatea și comportamentul parțial autonom pot face dificilă verificarea respectării și pot împiedica aplicarea efectivă a normelor legislației UE existente pentru a proteja drepturile fundamentale. Autoritățile

de aplicare a legii și persoanele afectate ar putea să nu aibă mijloacele necesare pentru a verifica modul în care a fost luată o anumită decizie cu implicarea IA și, prin urmare, dacă au fost respectate regulile relevante. Persoanele fizice și juridice se pot confrunta cu dificultăți de acces efectiv la justiție în situațiile în care astfel de decizii le pot afecta negativ"[30].

De departe, cele mai dezbătute riscuri pentru drepturile omului au fost cele identificate în utilizarea instrumentelor IA de evaluare a riscului penal, prezentate anterior.

Punctual, aceste sisteme generează probleme legate de:

a) Precizia rezultatelor

Studiile au relevat doar un anumit grad de precizie, fiind multiple dezbateri doctrinare cu privire la acest aspect în ultimii 20 de ani, mai mult de 50% dintre persoanele evaluate ca prezentând risc penal ridicat fiind incorect clasificate cu astfel de algoritmi[31]. Într-un alt studiu[32] s-a arătat că numai 20% dintre persoanele evaluate ca prezentând risc ridicat de recidivă violentă au reiterat efectiv comportamentul criminal. De asemenea, algoritmul a făcut greșeli raportat la inculpații afro-americani și albi în moduri diferite. Inculpații afro-americani au

[30] Comisia Europeană, *Cartea albă privind inteligența artificială - O abordare europeană axată pe excelență și încredere*, Bruxelles, 19.02.2020, COM(2020) 65 final, p. 13, www.ec.europa.eu.

[31] A se vedea D. Kehl, P. Guo, S. Kessler, „Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessment in Sentencing", Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2017, http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041 (accesat la 22.09.2022); B. Büchel, „Artificial intelligence could reinforce society's gender equality problems", 1 martie 2018, The Conversation UK, http://theconversation.com/artificial-intelligence-could-reinforce-societys-gender-equality-problems-92631 (accesat la 22.09.2022); V. Southerland, „With AI and Criminal Justice, the Devil is in the Data", ACLU, 9 aprilie 2018, https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data (accesat la 22.09.2022); T. Douglas, J. Pugh, I. Singh, J. Savulescu, S. Fazel, „Risk assessment tools in criminal justice and forensic psychiatry: The need for better data", European Psychiatry, 2017, nr. 42: 134-137. DOI: 10.1016/j.eurpsy.2016.12.009 (accesat la 22.09.2022).

[32] J. Angwin, J. Larson, S. Mattu, L. Kirchner, „Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks", ProPublica, 23 mai 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accesat la 22.09.2022).

fost evaluați greșit ca infractori viitori, într-o proporție de aproape două ori mai mare decât inculpații albi. 23,5% dintre inculpații albi au fost etichetați cu risc mare de recidivă, dar nu au recidivat, față de 44,9% afro-americani care nu au reiterat comportamentul infracțional. În același timp, 47,7% dintre inculpații albi au fost desemnați eronat ca prezentând risc scăzut, dar au recidivat, în timp ce 28,0% dintre afro-americanii evaluați cu risc scăzut au reiterat comportamentul criminal.

b) Legalitatea utilizării IA în procesul penal – dreptul la un proces echitabil – nediscriminarea

Toate sistemele penale actuale se fundamentează pe principiul aplicării sancțiunii penale condiționat de comiterea unei infracțiuni, pe principiul individualizării răspunderii penale raportat la împrejurările concrete ale comiterii faptei și caracteristicile persoanei infractorului și pe stabilirea și aplicarea sancțiunii penale doar de către judecător, după parcurgerea unui proces penal fundamentat pe noțiunea de echitabilitate și justețe. Condamnarea unei persoane și aplicarea sancțiunii penale reprezintă o sarcină importantă pentru judecători, care trebuie îndeplinită cu obiectivitate, răspundere și observarea regulilor și principiilor de drept penal. Însă, în contextul schimbării de paradigmă și poziționării riscului în centrul politicilor penale legislative și judiciare a statelor, utilizarea unor metode, mijloace și instrumente (printre care și cele bazate pe IA) care promit să identifice riscurile referitoare la infractor și infracțiune, sistemul judiciar tinde să se deschidă tot mai mult spre utilizarea acestora. În sistemul judiciar al multor state, așa-numitul „risc de recidivă" a devenit un punct focal atât în operațiunea de condamnare realizată de către instanțe, cât și în evaluările și anticipările reformatorilor[33]. Această preocupare obstinată

[33] Conference of Chief Justices, Conference of State Court Administrators, *Resolution 12: In Support of Sentencing Practices that Promote Public Safety and Reduce Recidivism*, National Center for State Courts, adoptată la 1 august 2007, http://www.ncsc.org/sitecore/content/microsites/csi/home/~/media/microsites/files/csi/education/handout%2031%20ccj%20resolution%2012.ashx (accesat la 22.09.2022).

pentru anticiparea riscului criminogen a fost denumită în literatura de specialitate străină „anxietate judiciară"[34].

Însă procedura penală stabilește expres și limitativ atât mijloacele și metodele care pot fi utilizate în procesul penal pentru dovedirea vinovăției persoanei despre care se pretinde a fi comis o infracțiune, cât și condițiile utilizării acestora, ca parte a conceptului de proces echitabil și a garanțiilor necesare pentru respectarea drepturilor și libertăților fundamentale ale persoanelor părți în proces. Ideea de bază este una simplă: ori de câte ori se pune problema utilizării unei asemenea metode/mijloc/instrument care ar putea periclita drepturile fundamentale, este necesară autorizarea unui magistrat imparțial și trebuie prevăzută o cale de atac împotriva deciziei acestuia, ca garanție specifică statului de drept. Doar că pătrunderea acestor instrumente în sistemul de justiție penală s-a realizat la fel de accelerat și de precipitat, pe cât s-a realizat și dezvoltarea IA în ultimele trei decenii. Cu alte cuvinte, în scopul asigurării unei lupte eficiente împotriva fenomenului infracțional, autoritățile judiciare au utilizat prima dată astfel de mijloace IA orbite de rezultatele excelente obținute, și abia apoi le-au analizat compatibilitatea cu sistemul de drept penal și procesual penal.[35] Tactică pe care, personal, o consider nesănătoasă și imperativ corectibilă.

Decizia în Cazul *Loomis* (2016) amintit mai sus, în ciuda faptului că nu validează vreo nelegalitate a utilizării IA în procesul judiciar, stabilește totuși că instanțele trebuie să procedeze cu prudență atunci când folosesc astfel de rapoarte IA de evaluare a riscului infracțional. Curtea din Wisconsin a explicat că scorurile de risc generate de algoritm nu pot fi folosite „pentru a determina dacă un infractor va fi încarcerat" sau „pentru a determina severitatea sentinței". Prin urmare, judecătorii care

[34] A. M. Barry-Jester, B. Casselman, D. Goldstein, „The New Science of Sentencing", Marshall Project, 4 august 2015, https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing (accesat la 22.09.2022); în același sens a se vedea M. Levin, „The Conservative Case for Pretrial Justice Reform and Seven Solutions for Right-Sizing the Solution", Right on Crime, 29 iunie 2015, http://rightoncrime.com/2015/06/the-conservative-case-for-pretrial-justice-reform-and-seven-solutions-for-right-sizing-the-solution (accesat la 22.09.2022).

[35] A se vedea *Sweetie, COMPAS* etc.

utilizează astfel de rapoarte trebuie să explice modul în care au stabilit și individualizat pedeapsa, cuantumul acesteia, precum și modul de executare. Însă, prin Decizia *Loomis* s-a mai stabilit și faptul că rapoartele COMPAS ar trebui să includă cinci avertismente pentru a fi avute în vedere de către judecători: natura privată a COMPAS împiedică divulgarea modului în care se calculează scorurile de risc; scorurile COMPAS nu sunt în măsură să identifice persoanele cu risc sporit, deoarece aceste scoruri se bazează pe date generale relevante pentru un anumit grup; deși COMPAS se bazează pe un eșantion național de date, nu a existat un studiu de validare încrucișată cu populația din statul Wisconsin; studiile utilizate pentru a valida COMPAS de obicei evaluează precizia modelului predictiv și, de obicei, implică testarea modelului pe un set de date care nu este utilizat în estimarea inițială, cum ar fi o populație locală. Astfel un studiu de validare din California a constatat că, în evaluările efectuate de COMPAS pentru acest stat, s-au ridicat întrebări dacă scorurile COMPAS clasifică în mod disproporționat infractorii aparținând minorităților ca prezentând un risc mai mare de recidivă; COMPAS a fost conceput și dezvoltat special pentru a facilita activitatea Departamentului de corecții în efectuarea evaluărilor după condamnare.

Aceste avertismente au subliniat clar poziția justiției de a insufla atât un scepticism general cu privire la precizia algoritmilor informatici (chiar dacă analizează cazul concret al COMPAS), cât și un scepticism mai bine orientat în ceea ce privește evaluarea de către astfel de algoritmi a riscurilor prezentate de infractorii aparținând minorităților, fiind, de fapt, un demers de temperare a entuziasmului actual pentru utilizarea instrumentelor IA de evaluare a riscului penal în vederea condamnării unei persoane și de încurajare a scepticismului judiciar în privința importanței rapoartelor IA în operațiunea de condamnare.

În fine, Consiliul Europei[36] arată că tehnicile de prelucrare automată a datelor (în special algoritmi) pot afecta următoarele drepturi fundamentale:

- dreptul la un proces echitabil
- dreptul la viață privată și protecția datelor cu caracter personal
- libertatea de exprimare
- dreptul la un remediu efectiv
- dreptul la nondiscriminare
- dreptul la servicii sociale și publice
- dreptul la alegeri libere
- dreptul la viață în contextul armelor inteligente și al dronelor operate algoritmic
- dreptul la viață în contextul sănătății și al cercetării conexe
- dreptul la opinie și libertatea de gândire, conștiință și religie în contextul sistematizării opiniilor prin algoritmi.

Deși statele sunt îndemnate în general la prudență atunci când vine vorba de acceptarea utilizării IA în diferite domenii ale vieții sociale, cu toate acestea Consiliul Europei sugerează o deschidere a intervenției judiciare în cazul combaterii unor forme de infracționalitate specifică, cum ar fi combaterea abuzului sexual asupra copiilor, și arată că recomandă legiuitorilor naționali să gândească *outside the box*. S-a subliniat necesitatea de a se dispune de instrumente adecvate și specifice pentru a lupta împotriva abuzului online asupra copiilor, inclusiv posibilitatea autorităților competente de a exploata datele colectate în timpul investigațiilor. Statele membre au fost încurajate să dezvolte și să aplice metode de investigație inovatoare, precum și să

[36] Consiliul Europei, *Algoritmii și drepturile omului. Studiu privind dimensiunile drepturilor omului în ceea ce privește tehnicile de prelucrare automată a datelor (în special algoritmi) și posibilele implicații de reglementare*, 2018, DGI(2017)12, pregătit de Comitetul de experți privind intermediarii de internet. Council of Europe, [*Algorithms and Human Rights. Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications*, 2018, DGI(2017)12, prepared by the Commitee of Experts on Internet Intermediaries], https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5, (accesat la 22.09.2022).

ia în considerare alocarea unor resurse specializate de aplicare a legii pentru combaterea abuzului asupra copiilor și a exploatării lor sexuale, iar întreprinzătorii privați (inclusiv furnizorii de servicii online) au fost îndemnați să asigure accesul legal la dovezi digitale pentru forțele de ordine și alte autorități competente. În plus, furnizorii de servicii online au fost invitați să elimine sau să dezactiveze accesul la conținuturi identificate ca materiale de abuz sexual asupra copiilor online cât mai curând posibil, după ce au luat cunoștință de acest conținut.[37]

Mesajul pare ușor contradictoriu: să fim prudenți în utilizarea IA, dar nu avem ce face, avem nevoie de aceasta. Sau, daca îmi este permis, „scopul scuză mijloacele, atunci când este nobil"[38].

B) Drepturile omului în realitatea virtuală (online)

Așa cum s-a exprimat Curtea Europeană a Drepturilor Omului, „Internetul a devenit acum unul dintre principalele mijloace prin care indivizii își exercită dreptul la libertatea de a primi și de a împărtăși informații și idei, oferind astfel instrumente esențiale pentru participarea la activități și discuții referitoare la probleme politice și de interes general. (...) Mai mult, în ceea ce privește importanța site-urilor de internet în exercitarea libertății de exprimare, în lumina accesibilității acestora și a capacității sale de a stoca și comunica cantități mari de informații, Internetul joacă un rol important în îmbunătățirea accesului publicului la știri și în facilitarea difuzării informațiilor în general"[39]. La utilizarea Internetului ca rețea în toate domeniile vieții sociale se adaugă desfășurarea relațiilor umane în lumile virtuale create cu ajutorul IA.

Întrucât o mare parte a existenței noastre se realizează în mediul virtual, este important să înțelegem dacă spațiul cibernetic asigură drepturi și libertăți asociate identității noastre digitale, chestiune care este

[37] Consiliul Europei, *Concluziile Consiliului privind combaterea abuzurilor sexuale asupra copiilor*, 12862/19, 8 octombrie 2019, https://data.consilium.europa.eu/doc/document/ST-12862-2019-INIT/en/pdf, (accesat la 22.09.2022).

[38] Dicton aparținând lui Machiavelli, în lucrarea *Il Principe*. A se vedea N. Machiavelli, *Principele*, Humanitas, București, 2019.

[39] *Cengiz ș.a. c. Turcia*, hotărârea din 1 decembrie 2015, www.echr.coe.int.

de interes în ceea ce privește realizarea justiției penale și cu ajutorul mediului online. Am arătat deja că dosarul electronic, platformele digitale, algoritmii de clasificare, arhivele electronice, toate reprezintă elemente ale unui sistem de justiție penală actual, aflat, și din punctul de vedere al infrastructurii, în dezvoltare accelerată.

În timp ce internetul și digitalizarea sunt prezente în toate domeniile activităților sociale, urmărind să le îmbunătățească, să crească viteza de acțiune și de rezolvare a sarcinilor și să asigure obiectivitatea reacțiilor, este evident că drepturile și libertățile exercitate în procesul de socializare în mediul virtual sunt în dezbatere acerbă. De fapt, Curtea Europeană a Drepturilor Omului (CtEDO) se confruntă cu un număr tot mai mare de cauze care se referă la încălcarea prevederilor Convenției Europene a Drepturilor Omului și a Libertăților Fundamentale (CEDO) în mediul online și în legătură cu internetul.

O serie de documente europene aduc în discuție problematica drepturilor digitale ca sistem de drepturi paralel cu acela al drepturilor „clasice, tradiționale", insistând pe nevoia unei codificări distincte a acestora, după modelul CEDO.

Spre exemplu, în *Recomandarea CM/Rec(2014)6 și Memorandumul explicativ al Comitetului de Miniștri către statele membre cu privire la Ghidul drepturilor omului pentru utilizatorii de Internet* se arată că „nimeni nu ar trebui să fie supus unei interferențe ilegale, inutile sau disproporționate în exercitarea drepturilor omului și a libertăților fundamentale atunci când folosește Internetul"[40]. Conform paragrafului 14 din *Comentariile la Recomandarea CM/Rec(2014)6,* accesul persoanelor și al comunităților utilizatorilor la Internet și utilizarea optimă a acestuia necesită eforturi pentru a le informa și împuternici să își exercite drepturile și libertățile în mediile online. În 2011 Comitetul de Miniștri a subliniat, prin *Declarația privind principiile guvernanței*

[40] Consiliul Europei, *Ghidul drepturilor omului pentru utilizatorii de Internet*, *Recomandarea CM/Rec(2014)6 a Comitetului de Miniștri către statele membre cu privire la Ghidul drepturilor omului pentru utilizatorii de Internet și expunerea de motive*, adoptată de Comitetul de Miniștri la 16 aprilie 2014 la cea de-a 1197-a reuniune a adjuncților miniștrilor, punctul 3.

Internetului, necesitatea abordării internetului centrată pe oameni și bazată pe drepturile omului, conform căreia utilizatorii pot să își exercite drepturile și libertățile pe Internet[41]. De asemenea, *Strategia de guvernare a Internetului 2016-2019 a Consiliului Europei* oferă noi termeni și definiții printre care și *Magna Carta Digitală* (*Digital Magna Carta*) - o declarație a drepturilor pentru internet, al cărei scop este, printre altele, dezvoltarea unor legi pozitive care să protejeze și să extindă drepturile utilizatorilor la un *web* deschis, gratuit și universal[42].

Nu în ultimul rând, *Ghidul drepturilor omului pentru utilizatorii de Internet*[43], adoptat la 16 aprilie 2014 de Comitetul de Miniștri al Consiliului Europei, arată că „(...) drepturile omului, care sunt universale și indivizibile, precum și standardele conexe, prevalează asupra termenilor și condițiilor generale impuse utilizatorilor de Internet de către orice actor din sectorul privat. "

În concluzie, sistemul de justiție penală în era tehnologiei presupune coexistența a două categorii de drepturi: drepturile specifice privind derularea procesului penal și care sunt cele cuprinse de art. 6 din CEDO, drepturi conexe acestora (ex. protecția datelor cu caracter personal) și drepturi specifice digitalizării care reies din necesitatea utilizării unei infrastructuri digitalizate care asigură realizarea justiției penale. Recunoașterea acestor drepturi, concilierea lor, precum și garantarea acestora constituie deziderate ale statului de drept,

[41] *Comentarii la Recomandarea CM/Rec(2014)6 a Comitetului de Miniștri către statele membre cu privire la Ghidul drepturilor omului pentru utilizatorii de Internet*, https://rm.coe.int/16804d5b31 (accesat la 22.09.2022).

[42] Consiliul Europei, *Guvernarea Internetului – Strategia Consiliului Europei 2016-2019 - Democrație, drepturile omului și statul de drept în lumea digitală (Internet Governance – Council of Europe Strategy 2016-2019 - Democracy, human rights and the rule of law in the digital world)*, adoptată la cea de-a 1252-a reuniune a Comitetului de Miniștri din 30 martie 2016; „Strategia de guvernare a internetului 2016-2019/Glosar de termeni" („Internet Governance Strategy 2016-2019/Glossary of Terms"), p. 17, 21, https://rm.coe.int/internet-governance-strategy-2016-2019-updated-version-06-mar-2018/1680790ebe (accesat la 22.09.2022).

[43] Consiliul Europei, *Ghidul drepturilor omului pentru utilizatorii de Internet*, *Recomandarea CM/*Rec(2014)6 *a Comitetului de Miniștri către statele membre cu privire la Ghidul drepturilor omului pentru utilizatorii de Internet și expunerea de motive*, adoptată de Comitetul de Miniștri la 16 aprilie 2014 la cea de-a 1197-a ședință.

dovadă a civilizației și nivelului de evoluție al unei societăți, care se fundamentează pe valori sănătoase și ancorate în realitate.

IV. Concluzii

Parafrazând expresia celebră a unui jurnalist american – „software is eating the world"[44] – oare putem afirma la ora actuală că IA devorează sistemul de justiție penală? Pătrunderea sa, deloc insidioasă, în diferitele sectoare ale justiției penale pare că îndreaptă lucrurile în această direcție. Dar o asemenea interpretare nu ne caracterizează, deschiderea și fascinația personală pentru tehnologie și beneficiile acesteia fiind de notorietate.[45] Nu putem vorbi de „devorare", dar putem invita la moderație, responsabilitate și transparență. Nu trebuie să interzicem IA în sistemul de justiție penală, dar putem să gândim și să creăm mecanisme eficiente care să elimine opiniile părtinitoare și să maximizeze beneficiile utilizării algoritmilor în acest domeniu. Obișnuiesc să spun studenților: „nu mă tem de IA, ci mă tem de oamenii care o creează și o folosesc". Factorul uman este cheia acestei dileme sociale și juridice. Oamenii sunt cei care creează premisele încălcării drepturilor omului în utilizarea IA în sistemul de justiție penală și tot ei sunt cei care ar trebui să răspundă.

Urmărirea penală se confruntă cu vremuri noi care aduc o mulțime de provocări, cum ar fi profesionalizarea infractorilor, proliferarea fenomenului infracțional adunând noi caracteristici precum tehnologia și componenta transnațională sau mediul de ofertă pe internet. Aceste provocări obligă anchetatorii și alte organe judiciare să se adapteze și să fie creative în lupta lor împotriva criminalității. Inteligența artificială și toate aplicațiile ei ar putea fi cheia pentru aplicarea cu succes a legii penale, dar utilizarea unor astfel de instrumente noi ar putea

[44] „Software-ul devorează lumea". A se vedea M. Andreessen, „Why Software Is Eating The World", *Wall Street Journal*, 20 august 2011, https://www.wsj.com/articles/SB10001424053111903480904576512250915629460, (accesat la 22.09.2022).

[45] A se vedea L.M. Stănilă, *op.cit.* Monografia *Inteligența artificială, dreptul penal și sistemul de justiție penală. Amintiri despre viitor* reprezintă rezultatul a trei ani de cercetare juridică, pe parcursul cărora am publicat pe această temă peste 10 articole în reviste de specialitate române și străine.

contesta cadrul legislativ al statelor. În ciuda deschiderii și inventivității designerilor și programatorilor IA, a bunelor lor intenții și a obiectivelor nobile, instrumentele de investigație futuriste nu pot fi folosite în faza de investigare în absența unui cadru legal foarte solid. Utilizarea unor astfel de instrumente poate pune în pericol drepturile omului și standardele procesului echitabil, așa că necesită precauție extremă. Stă în puterea noastră să creăm algoritmi performanți și utili care să nu genereze opinii părtinitoare și stă în puterea noastră să îi utilizăm cu bună-credință și în realizarea scopurilor sociale și juridice. Stă în puterea noastră să creăm un cadru legal complex și comprehensiv care să stabilească cazurile și condițiile utilizării IA în sistemul de justiție penală și garanțiile necesare în ceea ce privește respectarea drepturilor și principiilor statului de drept.

Ne plasăm în favoarea abordărilor inovatoare ale fenomenului criminalității, militām pentru o justiție penală modernă și eficientă, dar cu respectarea principiilor legale și a drepturilor fundamentale ale omului. Viitorul ne va arăta dacă recursul la instrumentele IA a fost o alegere bună... doar că viitorul depinde de noi, iar nu de IA!

Principii și valori etice în inteligența artificială (IA)

Prof. univ. dr. Călin Enăchescu,
Universitatea de Medicină, Farmacie, Științe și Tehnologie „George Emil Palade" din Târgu Mureș

I. Introducere

Inteligența artificială (IA) este una dintre problemele centrale ale erei tehnologiilor convergente, cu implicații profunde pentru umanitate, cultură, societate și mediu[1].

Acesta este motivul pentru care cercetarea, dezvoltarea și implementarea sistemelor IA trebuie să fie însoțite de o reflecție etică. Sistemele IA nu sunt neutre, ci părtinitoare, din cauza datelor de antrenare și a parametrilor care influențează procesul de antrenament[2]. De asemenea, deciziile sistemelor IA, în special cele bazate pe învățarea automată, nu pot fi complet controlate. Mai mult, IA este o tehnologie distribuită, care necesită o abordare multidisciplinară, multiculturală, deschizând întrebări despre ce tip de viitor ne dorim pentru umanitate. Această reflecție trebuie să abordeze principalele provocări în dezvoltarea și utilizarea tehnologiilor IA legate de părtinirile încorporate în date și algoritmi, protecția vieții private a persoanelor și datele personale, lipsa diversității, problemele distribuției echitabile a beneficiilor și riscurilor,

[1] European Commission High-Level Expert Group on AI (2019-06-26). „Policy and investment recommendations for trustworthy Artificial Intelligence". Shaping Europe's digital future – Comisia Europeană. Arhivată pe baza originalului la 26,02,2020. Recuperată în 2019, 16,03.2020.

[2] https://www.weforum.org/agenda/2016/10/top-10-ethical– issues-in-artificial-intelligence/.

responsabilitate, impactul asupra ocupării forței de muncă și viitorul muncii, drepturile omului[3].

Posibilitatea de a crea mașini inteligente ridică o serie de întrebări și de probleme etice, care se referă atât la asigurarea faptului că astfel de mașini nu dăunează oamenilor și altor ființe, cât și la statutul moral al mașinilor în sine[4].

Dezvoltarea sistemelor care înglobează IA generează întrebări legale și etice importante ale căror răspunsuri afectează producătorii și consumatorii de tehnologie IA. Aceste întrebări au implicații legate de informatică, drept, politici publice, etica profesională și filozofie și vor necesita expertiza specialiștilor din aceste domenii[5].

Algoritmii de IA permit computerului să analizeze date pentru a detecta tipare și a dobândi cunoștințe sau abilități fără a fi necesară o programare specifică[6].

Mecanica necunoscută a modului exact în care funcționează algoritmii și incapacitatea noastră de a prezice impactul macrosocial al acestora sunt provocări etice importante[7], iar absența reglementărilor este o preocupare etică în continuă creștere la nivel mondial.

II. Elemente generale de inteligență artificială

Scopul inteligenței artificiale (IA) este de a dezvolta algoritmi sau metode pentru sistemele de calcul, care să simuleze modul de gândire uman[8]. Evident, această afirmație nu trebuie luată ca o definiție rigu-

[3] Rodney Brooks „The Seven Deadly Sins of Predicting the Future of AI", Rodney Brooks: Robots, AI, and Other Stuff, 2017.

[4] https://www.ibm.com/cloud/learn/ai-ethics

[5] High-Level Expert Group on Artificial Intelligence AI HLE: „Ethics Guidelines for Trustworthy AI", Comisia Europeană, 2019.

[6] C. Enăchescu, *Calculul neuronal*, Editura Casa Cărții de Știință, Cluj-Napoca, 300 pagini, ISBN 978-973-133-460-8, 2009.

[7] Jacob Metcalf, Emily F. Keller, and Danah Boyd, „Perspectives on Big Data, Ethics, and Society", 23 mai 2016, Council for Big Data, Ethics, and Society, 2016.

[8] J. Hertz, A. Krogh, R.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley Publishing Co., 1992.

roasă a conceptului de IA. Ceea ce trebuie totuși remarcat în această afirmație este utilizarea termenului de „gândire" și nu de „inteligență", cu scopul de lărgi câmpul aplicațiilor care pot fi considerate ca aparținând IA, ca de exemplu percepția, prelucrări de limbaj, robotică etc.[9]

Un sistem IA trebuie să fie capabil să efectueze trei lucruri principale[10]:
a) memorare de cunoștințe;
b) aplicarea cunoștințelor dobândite (memorate) pentru a rezolva probleme;
c) dobândirea de noi cunoștințe prin experiență.

De asemenea, un sistem IA este constituit din trei componente:
a) reprezentare;
b) învățare;
c) gândire.

Să analizăm fiecare dintre componentele unui astfel de sistem IA[11]:

1. Reprezentare Una dintre cele mai distincte trăsături ale unui sistem IA este posibilitatea de a utiliza un *limbaj* constituit din *simboluri*, cu ajutorul cărora se pot construi structuri pentru a reprezenta două elemente:

- cunoștințe generale despre o problemă de rezolvat.
- cunoștințe specifice despre soluția problemei de rezolvat.

Simbolurile trebuie descrise, de obicei, în termeni cât mai familiari, pentru a face reprezentarea simbolică a unui sistem IA cât mai ușor de înțeles de către un subiect uman. De aceea, claritatea simbolisticii folosite de sistemele IA le face atât de utile în cadrul procesului de comunicație om – mașină.

În terminologia IA, prin termenul de „cunoștințe" înțelegem, de fapt, o altă formă de exprimare pentru noțiunea de dată. Dar dacă privim din punctul de vedere al unei *reprezentări declarative*, cunoștințele reprezintă o mulțime statică de fapte, reunită cu o mulțime de proceduri generale

[9] C., Enăchescu, *Elemente de inteligență artificială. Calculul neuronal*, Editura Universitatea Tehnică Cluj, 1997.

[10] C. Enăchescu, *Bazele teoretice ale rețelelor neuronale*, Editura Casa Cărții de Știință, Cluj-Napoca, 1998.

[11] *Ibidem*.

de prelucrare și manipulare a faptelor. De fapt, trăsătura caracteristică a reprezentărilor declarative o constituie faptul că aceste reprezentări conțin un înțeles intrinsec prin prisma unui utilizator uman, independent de utilizarea lor în cadrul unui sistem IA. Într-o *reprezentare procedurală,* cunoștințele sunt incluse într-un cod executabil care acționează de fapt în afara înțelesului acestor cunoștințe[12].

Ambele tipuri de cunoștințe, declarative și procedurale, sunt necesare pentru a putea rezolva majoritatea problemelor.

2. Gândire În cea mai simplă definiție, putem spune despre gândire că reprezintă abilitatea de a rezolva probleme. Dar pentru ca un sistem să poată fi calificat ca un sistem dotat cu gândire, acesta trebuie să satisfacă anumite condiții[13]:

- sistemul trebuie să fie capabil să exprime și să rezolve o gamă largă de probleme și de tipuri de probleme;
- sistemul trebuie să fie capabil să extragă din informațiile memorate *informații explicite* și *informații implicite*;
- sistemul trebuie să posede un mecanism de control care să determine, atunci când o soluție a fost obținută, care operație să fie aplicată unei probleme particulare sau când trebuie oprită orice activitate relativ la problema de rezolvat.

Rezolvarea problemelor poate fi privită, de fapt, ca o problemă de *căutare* (*searching*). Metoda clasică de a aborda o problemă de căutare este de a folosi *reguli, date* și *control*.[14] Regulile acționează asupra datelor, iar controlul acționează asupra regulilor. În practică, de cele mai multe ori însă cunoștințele disponibile sunt limitate (de exemplu, în diagnosticul medical), putând fi incomplete sau inexacte. În astfel de situații, se folosesc *proceduri de gândire probabilistice*, permițând astfel sistemelor IA să ia în considerare și nedeterminarea.

[12] C. Enăchescu, *Calculul neuronal*, ed. cit.

[13] C. Enăchescu, *Bazele teoretice ale rețelelor neuronale*, ed. cit.

[14] J. Hertz, A. Krogh, R.G. Palmer, *op. cit*.

3. Învățare Procesul de învățare poate fi reprezentat grafic prin intermediul schemei din Figura 1. După cum se vede din schemă, *mediul înconjurător* furnizează anumite informații *elementului de învățare*, care la rândul său utilizează această informație pentru a îmbogăți și îmbunătăți conținutul unei *baze de cunoștințe*, iar în final *elementul de procesare* utilizează baza de cunoștințe pentru a efectua sarcina dorită[15].

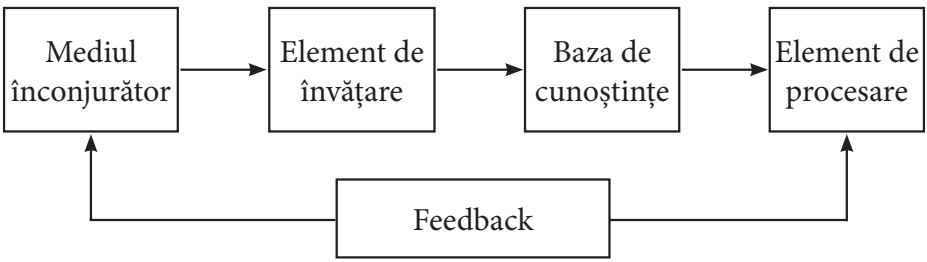

Fig.1. Schema pentru reprezentarea procesului de învățare[16]

Informația furnizată de mediul înconjurător sistemului de învățare (mașinii) este, în general, imperfectă, elementul de învățare neștiind să umple golurile lăsate de informațiile lipsă sau să ignore elementele neesențiale. De aceea, mașina lucrează mai mult pe baza deducțiilor, ajustându-și permanent comportamentul pe baza feedbackului obținut de la elementul de procesare.

Importanța bazelor de cunoștințe, precum și dificultățile unui proces de învățare au condus la dezvoltarea mai multor metode pentru a augmenta bazele de cunoștințe. Mai concret, dacă există experți într-un anumit domeniu, este mult mai ușor să beneficiem de experiența lor într-o formă compilată, decât să duplicăm această experiență. Aceasta este, de fapt, ideea ce se află la baza *sistemelor expert*.

Până în acest moment am prezentat doar elemente constitutive ale unor mașini IA simbolice. Ceea ce ne va interesa în continuare este cum să comparăm aceste sisteme expert cu rețelele neuronale ca

[15] C. Enăchescu, *Calculul neuronal*, ed. cit

[16] *Ibidem*.

niște modele cognitive? Pentru a răspunde la această întrebare ne vom folosi de modelul descris de Călin Enăchescu (*Bazele teoretice ale rețelelor neuronale*, Casa Cărții de Știință, Cluj-Napoca, 1998), care presupune trei niveluri:

1. Nivelul explicativ. În cadrul IA clasic, efortul principal este concentrat pe construcția *reprezentări simbolice*. De obicei, aceste reprezentări sunt discrete și arbitrare, de exemplu proprietăți abstracte, în locul unor imagini analogice. Din punctul de vedere al procesului cognitiv, este evident că nu ne putem pune problema unei reprezentări mentale, modelarea procesului cognitiv făcându-se pe baza unei procesări secvențiale reprezentărilor simbolice[17].

În cadrul calculului neuronal, procesele cognitive sunt total diferite de cele din IA clasic. Scopul calculului neuronal este de a construi modele *paralele de procesare distribuită* (PDP – *Parallel Distributed Processing*). Aceste modele PDP presupun că procesarea informației se face prin interacțiunea unui număr mare de neuroni, fiecare neuron trimițând semnale excitatorii sau inhibitorii către alți neuroni ai rețelei neuronale de care aparțin[18]. Mai mult chiar, rețelele neuronale pun un mare preț pe explicarea neuro-biologică a fenomenelor cognitive[19].

2. Metoda de procesare. În IA clasic, modul de procesare este secvențial, ca la calculatoarele clasice von Neumann. Chiar dacă nu există o ordine predeterminată, operațiile trebuie efectuate în maniera pas-cu-pas. Acest mod de procesare secvențial are ca sursă de inspirație natura secvențială a limbajului natural, trebuind să observăm că IA clasic s-a născut la puțin timp după mașina von Neumann[20].

Pe de altă parte, *procesarea paralelă* reprezintă una dintre trăsăturile definitorii ale rețelelor neuronale. Paralelismul este esențial pentru modul de procesare a informațiilor de către o rețea neuronală, reprezentând sursa principală a flexibilității lor. Paralelismul poate fi masiv

[17] C. Enăchescu, *Elemente de Inteligență Artificială. Calculul neuronal*, ed. cit.

[18] C. Enăchescu, *Calculul neuronal*, ed. cit.

[19] C. Enăchescu, *Bazele teoretice ale rețelelor neuronale*, ed. cit.

[20] C. Enăchescu, *Elemente de Inteligență Artificială. Calculul neuronal*, ed. cit.

în cadrul rețelelor neuronale (sute de mii de neuroni), ceea ce le conferă acestora o remarcabilă *robustețe*. Procesul de calcul fiind distribuit relativ la un număr mare de neuroni, deviația calculelor generate de un număr mic de neuroni nu îl afectează. Date de intrare zgomotoase, deteriorate sau incomplete pot fi folosite totuși de rețeaua neuronală. O rețea neuronală parțial deteriorată poate funcționa satisfăcător, învățarea unei rețele neuronale netrebuind să fie perfectă; performanțele rețelei neuronale se degradează continuu și nu abrupt. Astfel, sistemele PDP aproximează flexibilitatea unui sistem continuu, în contrast evident cu rigiditatea sistemelor IA tradiționale bazate pe simbolica discretă[21].

O altă trăsătură demnă de menționat a paralelismului o reprezintă faptul că cunoștințele nu sunt reprezentate prin expresii declarative, ci prin structura și nivelul de activare al rețelei neuronale.

Calculul secvențial reprezintă trăsătura fundamentală a IA clasic, în timp ce calculul paralel caracterizează rețelele neuronale.

3. Structura de reprezentare. După cum am văzut, IA clasic are la bază reprezentarea simbolică care posedă o *structură cvasilingvistică*. Ca și expresiile limbajului natural, expresiile din IA clasic sunt în general complexe, fiind construite din simboluri simple într-o manieră sistematică. Cu ajutorul unei mulțimi limitate de simboluri, noi expresii pline de conținut pot fi construite pe baza analogiei dintre structurile semantice și sintactice[22].

Natura și structura reprezentării reprezintă o problemă crucială a rețelelor neuronale artificiale. Într-o rețea neuronală reprezentarea este distribuită. Totuși, trebuie să subliniem faptul că cele mai multe rețele neuronale propuse drept candidați pentru reprezentarea structurală distribuită au mai degrabă un caracter ad-hoc; ele rezolvă problema relativ la o clasă particulară într-o manieră ce nu permite o extindere simplă.

În concluzie, putem defini IA clasic ca fiind manipularea formală a unui limbaj algoritmic și reprezentarea datelor după modelul *top-down*.

[21] C. Enăchescu, *Bazele teoretice ale rețelelor neuronale*, ed. cit.

[22] C. Enăchescu, *Calculul neuronal*, ed. cit.

Pe de altă parte, putem defini rețelele neuronale ca fiind procesoare simple distribuite ce posedă o capacitate naturală de a învăța, modul lor de operare fiind mai degrabă de tip *bottom-up*. De aceea, pentru a implementa aplicații cu caracter cognitiv, cea mai bună soluție ar fi construirea unui puternic model structural conexionist, care să îmbine puterea ambelor direcții din IA: IA clasic și rețelele neuronale. Astfel am fi în stare să combinăm trăsăturile pozitive ale rețelelor neuronale – adaptivitate, robustețe, uniformitate – cu cele ale IA simbolic – reprezentare, inferență și universalitate. Un astfel de hibrid poate fi eficient dacă reușim să stabilim cu precizie domeniile de eficiență a celor două constituente[23]:

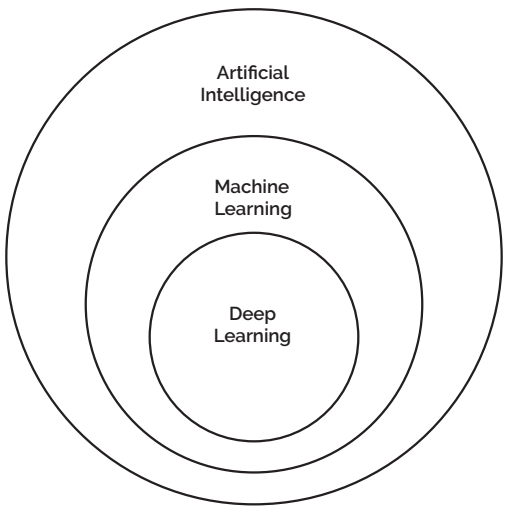

Fig. 2.: Structura IA[24]

În Figura 2 se poate vedea o reprezentare a structurii unui sistem IA. Cea mai importantă componentă este învățarea automată (*machine learning*)[25] – ramură a IA care se ocupă cu proiectarea și dezvoltarea de

[23] S. Haykin, *Neural Networks. A Comprehensive Foundation,* IEEE Press, Macmillan, 1994.

[24] *Ibidem*.

[25] Michael Bernico, *Deep Learning Quick Reference: Useful hacks for training and optimizing deep neural networks with TensorFlow and Keras*, Packt Publishing, Limited, ProQuest Ebook Central, 2018.

algoritmi și metode care permit sistemelor informatice să învețe. Cele mai importante tipuri de învățare automată sunt descrise în M. McCord Nelson, W.T. Illingworth, *A Practical Guide to Neural Nets*, Addison-Wesley, Redwood, CA, 1991; Giuseppe Bonaccorso. *Mastering Machine Learning Algorithms: Expert techniques to implement popular machine learning algorithms and fine-tune your models,* Packt Publishing, Limited, ProQuest Ebook, 2018:

- învățare supervizată *(supervised learning);*
- învățare nesupervizată *(unsupervised learning);*
- învățare prin întărire *(reinforcement learning).*

III. Cele mai presante 7 probleme etice ale inteligenței artificiale[26]

1. Șomaj și inechități sociale[27]

- O preocupare importantă a oamenilor relativ la IA este posibila pierdere a locurilor de muncă. *Oare ar trebui să ne străduim să dezvoltăm și să integrăm pe deplin IA în societate dacă IA ar avea drept consecință pierderea locurilor de muncă și, probabil, a mijloacelor de trai?*
- Potrivit noului raport al McKinsey Global Institute, până în anul 2030, aproximativ 800 de milioane de oameni își vor pierde locurile de muncă din cauza roboților IA. Argumente în favoarea IA:
  - Crearea de noi locuri de muncă mai specializate care să profite de capacitatea umană unică ce implică funcții cognitive superioare, analiză și sinteză.
  - *o* IA poate crea mai multe locuri de muncă pentru specialiștii care vor fi însărcinați cu crearea și mentenanța acestor roboți.
  - Problema legată de pierderea locurilor de muncă generează inechități sociale. Companiile plătesc salarii, impozite, alte cheltuieli pentru forța de muncă umană.
  - Ce se întâmplă dacă introducem IA în fluxul economic? Roboții nu sunt plătiți și nu plătesc taxe, determinând menținerea unor costuri de producție reduse. Acest lucru ar încuraja

[26] https://kambria.io/blog/the-7-most-pressing-ethical-issues-in-artificial-intelligence/

[27] *Ibidem.*

managementul să obțină profituri mai mari generate de forța de muncă IA, ceea ce ar duce la o și mai mare inechitate socială.

2. Sistemele IA nu sunt perfecte![28]

- IA nu este imună la greșeli, iar învățarea automată necesită timp și putere de calcul pentru a deveni utilă. Dacă procesul de învățare este corect, dacă datele folosite în procesul de învățare sunt corecte, atunci IA poate funcționa eficient.
- Cu toate acestea, dacă utilizăm date corupte sau apar erori în procesul de învățare, IA poate fi dăunătoare.
- IA poate face greșeli! Dar un sistem IA face greșeli mai mari sau mai puține decât oamenii? Câte vieți au fost curmate de deciziile greșite luate de oameni? Este mai bine sau mai rău atunci când IA face o greșeală similară cu cea a unui operator uman?

3. Ar trebui să li se permită sistemelor IA să ucidă?

- Sistemele IA nu sunt create pentru a face ceea ce vrem noi să facă – fac ceea ce învață să facă.
- Dronele există de peste un deceniu și sunt larg utilizate în război (vezi Ucraina). Aceste aeronave pilotate de la distanță pot trage rachete, deși etica ar cere ca oamenii să ia deciziile finale de atac.
- Este mai bine să folosești IA pentru a ucide decât să pui oamenii să facă acest lucru? Ce-ar fi dacă am folosi roboții doar pentru descurajare, mai degrabă decât pentru violență reală?

4. Sistemele IA pot fi corupte![29]

- Dacă acceptăm ideea că mașinile inteligente bazate pe IA pot face greșeli, atunci există posibilitatea de corupere prin crearea unor consecințe periculoase datorate urmăririi unor obiective aparent inofensive.
- În prezent, experții spun că tehnologia IA actuală nu este încă capabilă să realizeze această posibilitate extrem de

[28] *Ibidem.*

[29] *Ibidem.*

periculoasă a conștiinței de sine; cu toate acestea, viitoarele supercomputere IA ar putea.

- Un posibil scenariu de mare actualitate, legat de pandemie, este cazul în care un sistem IA este utilizat să studieze structura genetică a unui virus pentru a crea un vaccin de neutralizare. Un sistem IA corupt ar putea decide o soluție în care virusul să devină o armă, în loc să producă un vaccin benefic.
- Este ca și cum am deschide o cutie a Pandorei modernă! Din nou intervine etica, determinând abordarea unor preocupări legitime de prevenire a unui astfel de scenariu.

5. Momentul singularității!

- Va fi posibilă o evoluție a sistemelor IA care să ducă la o superioritate față de ființele umane? Ce se întâmplă dacă devin mai inteligente decât oamenii și apoi încearcă să ne controleze? Vor deveni oamenii inutili în raport cu sistemele IA?
- Momentul în care dezvoltarea tehnologică a sistemelor IA depășește inteligența umană este denumit „singularitate tehnologică"[30].
- Mulți cercetători (Hawkins, Musk) cred că acest lucru va reprezenta sfârșitul erei umane și că ea ar putea avea loc încă din anul 2030, luând în considerare ritmul de creștere a inovației tehnologice. În acest scenariu, momentul singularității ar putea duce la extincția umană – motiv pentru care creșterea performanțelor IA este înfricoșătoare pentru mulți oameni[31].

6. Cum ar trebui să tratăm IA?[32]

- Ar trebui să li se acorde roboților drepturi similare cu ale omului sau cetățenie?

[30] Luke Muehlhauser și Louie Helm. „Intelligence Explosion and Machine Ethics", în *Singularity Hypotheses: A Scientific and Philosophical Assessment,* editori Amnon Eden, Johnny Søraker, James H. Moor și Eric Steinhart. Berlin: Springer, 2012, arhivat la 7.05.2015 în arhiva Wayback Machine.

[31] Elisabeth Costa și David Halpern, „The behavioural science of online harm and manipulation, and what to do about it: An exploratory paper to spark ideas and debate", The Behavioural Insights Team Report, 1-82, 2019.

[32] „The 7 Most Pressing Ethical Issues in Artificial Intelligence" – https://kambria.io/blog/the-7-most-pressing-ethical-issues-in-artificial-intelligence/

- Dacă dezvoltăm roboți capabili să fie „conștienți", le dă acest lucru drepturi similare cu ale oamenilor sau animalelor?[33]
- Dacă roboții primesc drepturi, atunci cum le clasificăm statutul social?

7. Sisteme IA părtinitoare[34]

- Sistemele IA au început să fie utilizate în recunoașterea facială și vocală, având un impact direct asupra oamenilor. Acestea sunt vulnerabile la prejudecățile și erorile introduse de creatorii lor umani, iar datele utilizate pentru a instrui aceste sisteme IA pot conține prejudecăți.
- Sistemele IA nu au o busolă morală și nici un set de principii așa cum au oamenii.
- Cu toate acestea, chiar și compasul nostru moral și principiile noastre nu aduc întotdeauna beneficii umanității în ansamblu; deci cum ne asigurăm că sistemele IA nu au aceleași defecte ca și creatorii lor?[35]
- Dacă sistemele IA dezvoltă o anumită prejudecată față de rasă, sex, religie sau etnie, atunci vina va fi în mare parte datorată modului în care au fost instruite.

IV. UNESCO – Recomandări pentru etica IA. Mesaje-cheie[36]

Respectarea deplină a dreptului internațional, în special a drepturilor omului. Drepturile omului *(human rights)* sunt concepte clar definite în dreptul internațional. O abordare bazată pe drepturile omului se aplică pe tot parcursul ciclului de viață al sistemelor IA pentru a se asigura că inegalitățile existente nu sunt aprofundate, iar drepturile persoanelor și comunităților nu sunt afectate în mod negativ[37].

[33] J.R. Anderson, *The Architecture of Cognition*, Harvard University Press, 1983.

[34] „The 7 Most Pressing Ethical Issues in Artificial Intelligence" – https://kambria.io/blog/the-7-most-pressing-ethical-issues-in-artificial-intelligence/

[35] Tristan Harris, „How Technology Is Hijacking Your Mind – from a Magician and Google Design Ethicist", Thrive Global, 2016.

[36] https://unesdoc.unesco.org/ark:/48223/pf0000381137

[37] John R. Searle, „Consciousness in Artificial Intelligence", Google's Singularity Network, Talks at Google (YouTube video), 2015.

Aceasta este o condiție *sine qua non* pentru obiectivul comun, realizarea Agendei 2030 UNESCO. Prin urmare, trimiterile existente la drepturile omului trebuie menținute, iar recomandarea actuală ar putea fi chiar consolidată (de exemplu, protecția datelor, viața privată, libertatea de exprimare).

UNESCO sprijină cu fermitate o abordare pe deplin incluzivă a mai multor părți interesate, inclusiv guverne, instituții academice, companii și cetățeni.

Egalitatea de gen[38]. Recomandarea are în vedere inegalitățile în materie de putere și propune soluții adecvate cu privire la modul de remediere a acestora. Pentru un impact complet sunt necesare:

- integrarea în toate domeniile politice, educație și cercetare;
- abordarea diferențelor de gen în ceea ce privește dezvoltarea competențelor digitale și accesul la educația formală;
- abordarea egalității din perspectiva mai multor surse de discriminare care se pot intersecta și care pot consolida exclusivitatea.

Comunicare și informare (CI)[39]. Având în vedere impactul major al noilor tehnologii, inclusiv asupra comunicării și informațiilor, și pentru a se ține cont de recomandare în toate domeniile de mandat ale UNESCO, CI are un spațiu politic separat.

Stabilirea unui echilibru între riscuri și beneficii[40]*.* IA este un set de tehnologii cu impact puternic asupra vieții umane și, pentru a fi benefic, are nevoie de o abordare bazată pe riscuri. În cazul în care riscurile legate de utilizarea IA sunt mai mari, răspunsul în materie de reglementare ar trebui să fie mai puternic.

- În prezent, există tehnologii IA care prezintă riscuri pentru drepturile omului și etica centrată pe om, care trebuie examinate cu atenție înainte de utilizare.
- În același timp, trebuie subliniat faptul că IA aduce multe beneficii omenirii în facilitarea protecției mediului, producerea de

[38] https://unesdoc.unesco.org/ark:/48223/pf0000381137

[39] *Ibidem*.

[40] *Ibidem*.

alimente, optimizarea logisticii etc. Aceste beneficii ar trebui consolidate prin inovare.

Monitorizare și evaluare[41]. Recomandarea este utilă numai dacă are impact.

- Va fi important ca UNESCO să furnizeze instrumente de monitorizare orientate spre rezultate, pentru a continua punerea în aplicare a recomandării, precum și pentru a măsura progresul și impactul acesteia.
- Având în vedere evoluția rapidă a IA, este util să se reamintească statelor membre posibilitatea de a actualiza recomandarea și chestionarul utilizat de UNESCO pentru monitorizarea acestuia.

Scopuri urmărite[42]

- Recomandarea urmărește să ofere o bază pentru ca sistemele IA să funcționeze pentru binele umanității, indivizilor, societăților, mediului, ecosistemelor și pentru a preveni eventuale daune.
- În plus, față de normele etice existente în ceea ce privește IA în întreaga lume, recomandarea are ca scop să creeze un instrument normativ acceptat la nivel global. Acesta nu urmărește să se concentreze doar pe articularea valorilor și principiilor, pe realizarea lor practică, prin recomandări politice concrete, ci și pe aspectele legate de incluziune, de egalitatea de gen, de protecția mediului și a ecosistemelor.

Complexitatea problemelor etice din jurul IA necesită cooperarea mai multor părți interesate la diferite niveluri și sectoare ale comunităților internaționale, regionale și naționale. În acest sens, recomandarea are ca scop să permită părților interesate să își asume responsabilitatea comună pe baza unui dialog global și intercultural.

Obiective[43]

- Să ofere un cadru universal de valori, principii și acțiuni pentru a ghida statele în formularea legislației, a politicilor sau a altor

[41] *Ibidem*.

[42] *Ibidem*.

[43] *Ibidem*.

instrumente privind IA, fără a aduce atingere legislației internaționale existente.
- Să ghideze acțiunile indivizilor, grupurilor, comunităților, instituțiilor și companiilor din sectorul privat pentru a asigura încorporarea eticii în toate etapele ciclului de viață al sistemului IA.
- Să protejeze, să promoveze și să respecte drepturile omului și libertățile fundamentale, demnitatea umană și egalitatea, inclusiv egalitatea de gen; să protejeze interesele generațiilor prezente și viitoare; să păstreze mediul, biodiversitatea și ecosistemele; să respecte diversitatea culturală în toate etapele ciclului de viață al sistemului IA.
- Să promoveze dialogul multipartid, multidisciplinar și pluralist cu privire la aspectele etice legate de sistemele IA.
- Să promoveze accesul echitabil la evoluțiile și cunoștințele din domeniul IA și partajarea beneficiilor.

Valori[44]
- Respectarea, protecția și promovarea demnității umane, a drepturilor omului și a libertăților fundamentale
- Un mediu înconjurător și un ecosistem curat
- Asigurarea diversității și a incluziunii
- Să trăim în societăți pașnice, drepte și interconectate.

Principii[45]
- Proporționalitate și respectarea regulii „de a nu face rău”
- Siguranță și securitate
- Echitate și nediscriminare
- Durabilitate
- Confidențialitate și protecția datelor
- Supravegherea și determinarea umană
- Transparență și explicabilitate

[44] *Ibidem*.

[45] *Ibidem*.

- Responsabilitate
- Conștientizare și alfabetizare
- Guvernanța și colaborarea multipartită și adaptivă

Arii politice abordate[46]

- Evaluarea impactului etic
- Guvernanță și administrare etică
- Politica referitoare la date
- Dezvoltare și cooperare internațională
- Mediu și ecosisteme
- Gen
- Cultură
- Educație și cercetare
- Comunicare și informare
- Economie și muncă
- Sănătate și bunăstare socială

V. Concluzii

Ideea creării unor sisteme IA din ce în ce mai prezente în viața noastră, care depășesc inteligența umană, este uneori înfricoșătoare[47].

Problemele etice care vin odată cu adoptarea IA sunt foarte complexe. Soluția este să ținem cont de aceste aspecte etice pentru a analiza problemele generate de adoptarea IA[48].

Pentru a decide dacă sistemele IA sunt bune sau rele, se poate face o analiză din mai multe puncte de vedere, fără ca o teorie și/sau un punct de vedere să fie considerate determinante. Trebuie să continuăm să învățăm și să fim informați despre IA și ce presupune aceasta, pentru a lua decizii bune în viitor.

[46] *Ibidem*.

[47] Rodney Brooks, „The Seven Deadly Sins of Predicting the Future of AI", Rodney Brooks: Robots, AI, and Other Stuff, 2017.

[48] European Commission High-Level Expert Group on AI (26-06-2019), „Policy and investment recommendations for trustworthy Artificial Intelligence", Shaping Europe's digital future – Comisia Europeană. Arhivat pe baza originalului la 26,02,2020. Recuperat în 2019, 16,03,2020.

Inteligența artificială și drepturile omului: mult zgomot pentru nimic?

Iulia Motoc

Mi-am intitulat intervenția „Mult zgomot pentru nimic" pentru că în momentul de față cred că există foarte multă literatură consacrată problematicii avantajelor și riscurilor utilizării inteligenței artificiale, deci și posibilitatea de cercetare în domeniul intersecției drepturilor omului cu inteligența artificială. Trebuie să spun de la început că nu există o jurisprudență a drepturilor omului din perspectiva utilizării tehnologiilor care înglobează inteligența artificială. Cazuistica CEDO în domeniul drepturilor digitale, ca să spun așa, se rezumă la problemele de natură juridică ale supravegherii în masă (cazul Big Brother bine cunoscut) sau la cazuri individuale de supraveghere video la locul de muncă. Există chiar un caz, cazul românesc care a demarat întreaga jurisprudență în acest domeniu, Cauza „Bărbulescu împotriva României". Avem cazuistică privind legislația în domeniul supravegherii video, dar, mai departe de acest domeniu, în domeniul inteligenței artificiale încă nu avem jurisprudență, deoarece, așa cum știți, Curtea Europeană a Drepturilor Omului analizează cauzele numai după ce se epuizează căile de recurs interne, deci durează un timp de la introducerea cauzei pe rolul instanțelor naționale și, pentru că nu sunt nici măcar pe rol, nu putem avea un caz asupra căruia să ne pronunțăm.

Este interesant de arătat că nici la Curtea de Justiție a Uniunii Europene nu există o jurisprudență în acest domeniu. Încă nu știm foarte bine ceea ce trebuie să fie reglementat, dar totuși au fost enunțate niște principii ale protecției drepturilor omului, pentru o protecție preventivă, astfel încât să se prevină posibilele încălcări ale drepturilor omului în domeniul care ne interesează. De aceea spuneam câteodată, în special după pandemie, că numărul conferințelor, numărul reglementărilor din domeniu, numărul acțiunilor din domeniul inteligenței artificiale sau chiar al drepturilor digitale au crescut foarte mult. Chiar eu am organizat o conferință la CEDO, în urmă cu un an de zile, în acest domeniu, dar în ultimă instanță, dacă analizăm, poate că este încă mult zgomot pentru nimic, pentru că nu s-a ajuns încă la o reglementare, ci doar la enunțarea unor principii ale drepturilor omului aplicabile în acest domeniu, pe care am să încerc să le expun aici.

Ce este inteligența artificială? Vedem că nu există o definiție a inteligenței artificiale. IA este un sistem bazat pe mașini, pe mecanisme care ne influențează activitatea. Pentru utilizarea acestui sistem facem o serie de recomandări, predicții. Deci aici este problema: ce fel de predicții și ce fel de decizii, pentru că în acest domeniu trebuie să existe totuși un cadru clar, în care să fie utilizate aceste mașini. Trebuie să existe și un răspuns: cum trebuie reglementat inputul uman la utilizarea inteligenței artificiale. Să vedem care sunt eventualele consecințe ale acestor creionări, ale acestor tipuri de modele și care pot fi consecințele pentru sfera umanului, ceea ce presupune, din nou, recomandări, predicții și decizii. Deci suntem încă în domeniul predicțiilor. Există, fără îndoială, impactul inteligenței artificiale asupra drepturilor omului. Sunt de acord cu acest lucru. Impactul inteligenței artificiale asupra omului există și reprezintă, într-adevăr, unul dintre cele mai importante aspecte ale viitorului.

Vedem foarte bine cum tehnologia se dezvoltă în viața de zi cu zi, cum apar noi tehnologii extrem de performante. Aplicațiile pe care le avem în casă tip *smart home,* aplicațiile de social media care sunt utilizate pentru dezvoltarea capacităților personale, pentru modul

de a aloca resurse și alte tipuri de decizii pe care le luăm. De aceea, este important să existe o balanță între această tehnologie sofisticată și protecția drepturilor omului și, în acest context, este necesar, din perspectiva noastră, a celor care ne ocupăm cu protecția drepturilor omului, ca aceste obiective să fie ale drepturilor omului, să fie de la început subliniate, astfel încât balanța să nu se dezechilibreze, în sensul în care începem să fim dominați de problemele legate de tehnologie. Deja, un set de principii a fost dezvoltat la nivelul Consiliului, la nivelul Comisarului pentru drepturile omului de la Consiliul Europei. Este necesar să existe o preocupare pentru implementarea acestui sistem de inteligență artificială în sectorul privat și în sectorul public. Cât privește Consiliul Europei, dacă Curtea nu a avut încă ocazia să se pronunțe și nici măcar nu avem pe rol o astfel de cauză, nu poate fi luată în considerare o agendă specială în acest domeniu. Comisarul a adoptat acest set, a recomandat acest set de principii pentru a preveni posibilitatea unui dezechilibru de care discutam înainte.

Aceste principii încep cu recomandarea ca statele membre ale Consiliului Europei să stabilească un set de dispoziții legale pe care să le adoptăm, mă refer la autoritățile publice în ceea ce privește impactul inteligenței artificiale asupra drepturilor omului. Mă gândesc că ar fi util ca la nivelul parlamentelor să existe deja preocuparea de a reglementa un set legislativ de protecție a drepturilor omului. Ar fi necesar, cred, ca la nivelul statelor să existe un set de principii suficient de larg, pentru a acoperi toate consecințele utilizării sistemelor de inteligență artificială asupra drepturilor omului. Trebuie să avem posibilitatea să urmărim acest sistem al inteligenței artificiale cu expertiza care există în domeniu și, totodată, să pregătim instituțiile publice care protejează drepturile omului la nivel național pentru a interveni prin proceduri corespunzătoare. Mă refer aici în special la sistemul de tip Ombudsman, care să poată controla aceste decizii prin setul de principii care deja să fie adoptat de Parlament.

Autoritățile publice cu competență în materie trebuie să cunoască modelele sau algoritmii care se află în spatele funcționării sistemelor

de inteligență artificială. Vreau să spun că în mediul universitar din Occident a căpătat o amploare deosebită studierea inteligenței artificiale. Aș putea să adaug că, în marile universități europene, jumătate din numărul doctoranzilor și-au luat ca temă de cercetare inteligența artificială, ceea ce presupune înalte cunoștințe de matematică și informatică.

Deci deja vedem că nu juriștii sunt cei care au cel mai facil acces la acest sistem, ci cei care sunt educați în științele exacte și care apoi au diplome ce le atestă o înaltă calificare în domeniu, dar și o anumită instruire în materie de legislație. Deci o să avem foarte mulți matematicieni, foarte buni specialiști în științe tehnice, care se vor specializa și în drept și vor fi specialiștii noștri în inteligență artificială. Ei vor stăpâni modelele sau algoritmii utilizați în inteligența artificială și vor influența modul în care decidenții colectează aceste date și interpretează sistemele.

Dacă sistemele de inteligență artificială rămân sub controlul uman este una dintre principalele preocupări în acest domeniu. Cred că ați mai discutat despre acest lucru până acum. Circumstanțele și modul în care sunt revăzute aceste sisteme dau naștere la un risc real pentru protecția drepturilor omului. Se pune foarte mult accent pe mecanismele de protecție, pentru a diminua acest risc. Evident, un risc există. Cei care utilizează aceste mașini pot deja să confirme că există încălcări ale drepturilor omului sau posibile încălcări prin utilizarea inteligenței artificiale. În aceste cazuri, autoritățile publice trebuie să acționeze imediat și să remedieze acea încălcare, astfel încât să existe un mecanism de prevenție și riscul acestor încălcări să nu mai existe. Deci vedem că există un mecanism, care s-ar putea numi de *early warning*, de protecție a drepturilor omului, iar autoritățile publice sunt datoare să exercite un control regulat în acest domeniu.

Consultările publice sunt absolut necesare în ceea ce privește drepturile omului. Statele trebuie să utilizeze sistemele de bază ale inteligenței artificiale, iar, atunci când se stabilesc procedurile de protecție a drepturilor omului, este necesar să existe un acces liber la informații și o consultare publică la nivelul societății în legătură cu

sistemele de inteligență artificială. Consultările trebuie să includă pe toți cei care sunt interesați în conturarea unor soluții, de la actorii statali, la cei din mediul privat, ONG-uri, media, etc., pentru a se crea punți de legătură spre atingerea unui interes comun. Societatea civilă, autoritățile statului și mediul privat trebuie să conlucreze mult mai mult în interesul cetățeanului, ferindu-l de riscul utilizării greșite a inteligenței artificiale. După părerea mea, există obligația statelor de a facilita implementarea standardelor privind drepturile omului în sectorul privat. Și aici intrăm într-un domeniu care nu s-a discutat suficient în România.

Am atins un subiect destul de vechi în materie de protecție a drepturilor omului. El a început să fie discutat în cadrul ONU în anii 2000, la nivelul Subcomisiei drepturilor omului. Mai târziu, ONU a reușit să implementeze în 2016 principii directoare ale drepturilor omului și activităților economice în *business and human rights.* Aceste principii sunt astăzi foarte importante și vedem că ele sunt implementate sau se încearcă implementarea lor deja la o scară largă în societățile occidentale. Vă dau un singur exemplu. Modul în care unele firme utilizau munca copiilor în anumite state asiatice a fost sancționat și, în ultimă instanță, s-a renunțat la această practică, care la începutul anilor 2000 era destul de generalizată.

Deci vedem că intervențiile au avut o anumită eficiență. De aceea aceste principii privind protecția drepturilor omului sunt astăzi folosite larg. Avem o bază de la care să pornim ca societate și în domeniul protecției împotriva riscurilor folosirii inteligenței artificiale. Dar, pentru aceasta, statul și societatea civilă trebuie să lucreze direct și activ cu actorii privați. Așa cum știm foarte bine, normele referitoare la drepturile omului se aplică și actorilor privați. Apare însă o problemă în cazul în care este vorba de actori transnaționali, cărora este foarte complicat să le aplicăm normele care garantează drepturile omului în domeniul inteligenței artificiale. Complicațiile juridice ale aplicării unui set de reguli privind protecția drepturilor omului actorilor transnaționali care produc sau utilizează sisteme de inteligență artificială nu trebuie să ne descurajeze. Trebuie să avem o poziție fermă față de toți cei care sunt implicați, cei care produc

sau se ocupă cu aceste servicii, companiile care lucrează în domeniul inteligenței artificiale și care se află în jurisdicția unui stat. Și aici este o chestiune extrem de interesantă, cum analizăm jurisdicția în domeniul inteligenței artificiale în situația în care sunt mai multe state implicate?

Și pentru a fi în concordanță cu normele protecției drepturilor omului, așa cum le prevede CEDO și Convenția europeană a drepturilor omului, consider că trebuie să se ia măsurile necesare pentru protecția drepturilor omului de către acești actori pe întreaga durată de utilizare a sistemului de inteligență artificială. După părerea mea, statele trebuie să le ceară tuturor actorilor din domeniul inteligenței artificiale să protejeze drepturile omului. Aici intervine un principiu foarte important din drept, este vorba de *due diligence*, un principiu care se aplică prin excelență într-o zonă la fel de modernă, și anume cea legată de *climate changing,* și pe care îl putem invoca. Este vorba despre tema schimbărilor climatice, dar, înainte de impactul schimbărilor climatice asupra mediului înconjurător, temă care domină ultimii 30 de ani, în acest domeniu a existat standardul *due diligence*, care înseamnă că trebuie să se ia măsuri, deci este mai degrabă vorba de o obligație de mijloace decât de una de rezultat, cum am spune noi în drept, și trebuie să se ia măsuri pentru o protecție efectivă, pentru a preveni sau pentru a limita posibilitățile ca sistemele de inteligență artificială să genereze consecințe dăunătoare pentru om.

Deci vedem că, dincolo de această participare publică, dincolo de aplicarea principiilor directoare ale drepturilor omului și activităților economice pentru *business și human rights*, a principiului *due diligence*, care se aplică, cum spuneam, și în alte ramuri, inclusiv în jurisprudența noastră, există principiul informației. Cerința ca să existe o informare corectă și o transparență în acest domeniu este fundamentală, pentru că, în primul rând, înainte ca noi să adoptăm anumite norme, acestea trebuie să fie identificate. Cred că cei care se ocupă de acest sistem de inteligență artificială sunt datori să-l facă clar publicului pentru a fi accesat în siguranță. Este vorba, cu predilecție, de un domeniu încărcat de tehnicitate, care utilizează algoritmi dificil de înțeles pentru cei ce nu

au o educație matematică avansată, or omul obișnuit nu are cunoștințe științifice de specialitate. Vorbeam înainte de algoritm, vorbeam înainte de necesitatea cunoștințelor științifice și cunoștințelor de matematică avansate, or omul obișnuit nu realizează care sunt pericolele pentru el.

Pentru omul de rând trebuie să existe o informare corectă în acest domeniu. Inteligența artificială trebuie să fie utilizată astfel încât să existe o interacțiune între cei care au responsabilități în domeniul protecției medicale, domeniul justiției și domeniul, în general, al protecției sociale, pentru a se cunoaște foarte bine care este impactul inteligenței artificiale asupra drepturilor omului. Inteligența artificială nu trebuie să genereze creșterea și mai mare a disparităților dintr-o sferă socială, să ducă la lipsa accesului la educație, la sistemele de protecție medicală ș.a.m.d. Protecția drepturilor omului în domeniul sănătății trebuie să se păstreze. De aceea trebuie să existe un sistem al inteligenței artificiale cât mai transparent, un sistem în care publicul să cunoască, cum am spus înainte, riscurile la care se expune utilizând inteligența artificială, și să existe o posibilitate de audit efectiv.

Deci este foarte important ca niciun sistem de inteligență artificială să nu fie atât de complex încât omul să nu îl poată înțelege și să aibă posibilitatea de a-i evalua riscurile. Dar aici există o obligație a acestor factori, ale acestor firme private, în primul rând, care activează în domeniul licențelor, de a fi transparente și, mai mult decât atât, de a face accesibil omului sistemul lor de funcționare. Este necesară, de asemenea, existența unui mecanism independent care să analizeze ce se întâmplă în acest domeniu, așa cum este o cerință foarte, foarte avansată a comisarului.

Problema existenței unui mecanism independent de cadrul legislativ, a unui mecanism care să supravegheze, la rândul lui, din punct de vedere juridic, modul în care se autorizează accesul la supravegherea video în masă, a fost discutată în cadrul spețelor noastre Big Brother. Or, Curtea noastră nu a cerut existența unui mecanism independent judiciar pentru supravegherea în masă, pentru că s-a spus că nu

ar fi oportun să existe un astfel de organism independent, întrucât pot interveni situații în care, de pildă, servicii secrete să poată apela oricând la astfel de informații, care se obțin, inclusiv, pe tipuri de cablu transatlantice. Nu este posibil ca un judecător să ia o decizie într-un timp foarte scurt. De aceea s-a permis accesul pe scară largă la supravegheri în masă. Trebuie să spunem că, în ceea ce privește supravegherea în masă, există un alt aspect al tehnologiilor. Curtea noastră are un standard diferit de Curtea de la Luxemburg.

Curtea de la Luxemburg, în speța cuadraturii netului, cere un mecanism judiciar în acest domeniu. Standardele Curții de la Luxemburg sunt mai ridicate decât cele ale CEDO. Vedem cum este el aplicat în domeniul inteligenței artificiale de către comisar. În aceste principii se cere să existe un cadru legislativ.

Sunt necesare un mod de expertiză interdisciplinară și competențe care să supravegheze aceste mecanisme în mod independent. Deci vedem că nu se cere nici aici un mecanism judiciar. Se vorbește despre o combinație, o combinație de mecanisme administrative, judiciare, cvasijudiciare și parlamentare care să constituie acest grup care supraveghează și care trebuie să coopereze între ele. Aceste mecanisme interne trebuie să aibă o acțiune proactivă, să investigheze și să monitorizeze modul în care evoluează sistemul național

Un alt aspect foarte important, cu care aș dori să închei intervenția mea, este aspectul discriminării și egalității. Nu trebuie să existe vreo discriminare nici în acest domeniu. Trebuie să fim foarte atenți ca aceste mecanisme să nu genereze discriminare de niciun fel între grupuri sau discriminare legată de gen. Din nou intervin și cerințe legate de protecția datelor personale și de dreptul la viață privată. Problema drepturilor vieții private, a datelor biometrice, a datelor personale, a tuturor acestor drepturi impune să existe o protecție. Trebuie să existe însă și o limitare a lor, trebuie să fie proporționalitate. Acesta este un cuvânt magic în drept.

O perspectivă locală asupra unei viziuni europene

Lector dr. Ionuț Cristian Pistol,
Universitatea „Alexandru Ioan Cuza" din Iași

White Paper on Artificial Intelligence: a European approach to excellence and trust[1] (versiunea în limba română: *Cartea albă privind inteligența artificială – O abordare europeană axată pe excelență și încredere)* este un document publicat de Comisia Europeană, document ce a servit de la publicare (februarie 2020) ca o referință importantă, atât în literatura științifică de specialitate (peste 100 de citări conform Google Scholar), cât și ca suport tehnic, pentru diverse inițiative legislative sau dezbateri pe tema inteligenței artificiale. Documentul, deși acoperă principalele potențiale avantaje și dezavantaje aduse de tehnologiile moderne, precum și câteva propuneri rezonabile pentru facilitarea integrării acestor tehnologii în societate, include și câteva abordări care, din perspectivă personală, punctează câteva dintre viziunile nerezonabile asupra domeniului. În continuare voi indica, pe scurt, câteva puncte de vedere menționate în acel document, precum și o viziune diferită (cred, nu doar personală) asupra acestora.

Pentru început voi descrie pe scurt concepția actuală privind inteligența artificială. În documentul menționat, inteligența artificială este definită ca „un ansamblu de tehnologii care combină date, algoritmi și putere de calcul.". Această definiție explică multe dintre perspectivele conținute

[1] https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

în document, dar ea este imprecisă. Practic, orice tehnologie modernă include date, algoritmi și putere de calcul, mergând de la o mașină de spălat și până la fermele de calculatoare folosite pentru a „mina" criptomonede. În marea majoritate a cazurilor, inclusiv în extremele menționate, nu putem include acele tehnologii sub definiția de inteligență artificială.

Deși nu există o definiție unanim acceptată a acestui termen, în general este folosită o perspectivă computaționalistă asupra inteligenței, văzută ca o colecție de capabilități pe care, dacă un dispozitiv le posedă, poate fi numit inteligent. Plecând de la perspectiva lui Alan Turing, pentru care capabilitatea definitorie a unei inteligențe constă în folosirea limbajului natural la un nivel similar cu o ființă umană, în prezent capabilitatea definitorie este considerată în multe abordări ca fiind puterea de a lua decizii. Acest aspect este clar definit în antiteză cu urmarea unui algoritm care include toate deciziile pe care acel sistem ar putea fi pus vreodată să le ia, în funcționarea sa normală. De exemplu, un program capabil să joace șah nu ar putea rezonabil face acest lucru, urmând un algoritm determinist. Pe lângă dificultatea implementării unui astfel de algoritm, care ar trebui să aibă un răspuns la oricare dintre cele 10^{70} configurații posibile pe tabla de șah, un astfel de program ar fi și predictibil, deci relativ ușor de învins.

Pe lângă această capabilitate (de a lua decizii singur), un sistem inteligent mai poate avea și alte funcționalități asociate cu inteligența, cum ar fi folosirea limbajului natural, posibilitatea de a învăța, precum și multe alte capabilități specifice contextului în care va fi utilizat. Această abordare satisface marea majoritate a situațiilor în care dorim utilizarea unui sistem inteligent, dar nu pretinde substituirea unei inteligențe umane decât din perspectiva limitată a sarcinilor trasate acelui sistem, uneori nici măcar atât. Deficiența cea mai semnificativă a unei „inteligențe artificiale", mai ales din perspectiva acestui document, este imposibilitatea de a-și stabili singură obiective. Un astfel de sistem nu poate determina singur care dintre situațiile pe care le poate percepe sunt probleme pe care ar trebui să le rezolve. După stabilirea unui

obiectiv de către o inteligență umană, un sistem inteligent va putea urmări acel obiectiv chiar și identificând obiective intermediare, dar impulsul inițial este în afara capabilităților acestuia.

Plecând de la aceste clarificări, putem urmări acum câteva aspecte menționate în *Cartea albă privind inteligența artificială – O abordare europeană axată pe excelență și încredere*. Sunt trei aspecte principale ale inteligenței artificiale abordate în acest document: beneficiile potențiale, riscurile percepute și soluții propuse pentru ameliorarea acestora.

Ca beneficii, autorii punctează aspecte care ar putea fi îmbunătățite prin „dezvoltarea unui ecosistem de IA" din perspectiva cetățenilor, a întreprinderilor și a interesului public. O parte dintre beneficiile descrise sunt foarte vagi („servicii publice mai bune", „dezvoltarea unei noi generații de produse și servicii"), fără a include ceva din potențialul specific al inteligenței artificiale. Ca beneficii specifice ar fi putut fi menționate:

- accesul general la tehnologiile informaționale moderne (de exemplu prin posibilitatea interacțiunii folosind limbajul natural, cum e cazul sistemelor de tip Asistent Virtual);
- reducerea costurilor legate de utilizarea sau accesul la resurse (prin utilizarea unor sisteme-expert de consultanță/diagnostic);
- facilitarea comunicării și a accesului la informații (de exemplu prin utilizarea unor sisteme de traducere automată sau sisteme de suport pentru persoane cu dizabilități).

Riscurile menționate în documentul la care ne referim au în vedere, în special, drepturile fundamentale, inclusiv protecția datelor cu caracter personal și a vieții private și discriminarea. Documentul vorbește despre protejarea datelor personale (urmărirea activității, înregistrarea și valorificarea fără autorizație a datelor personale), prejudicii și discriminare (diferențe pe criterii ce țin de competențele tehnologice sau de vârstă, adaptarea la un public țintă) și limitarea interacțiunii și a explicațiilor disponibile. Aceste riscuri pot fi însă privite ca inerente procesului

de includere a tehnologiilor moderne în societate, nefiind specifice inteligenței artificiale. O parte dintre ele ar putea chiar să fie ameliorate prin includerea unor tehnologii specifice IA care să faciliteze accesul la informații (și să diminueze necesitatea unor competențe tehnologice sau fizice) sau care să descopere situații în care o categorie de persoane suferă prejudicii, prin descoperirea unor elemente specifice în interacțiunile avute cu aceste tehnologii.

Ca risc ce nu este precizat ca atare în document, putem menționa cazurile în care un sistem ce folosește tehnologii IA eșuează în interacțiunea cu un beneficiar, iar eșecul este perceput doar de acesta. Confirmarea, identificarea gradului și a cauzei posibile a eșecului ar trebui făcute de o persoană competentă atât în obiectul interacțiunii, cât și în tehnologiile implicate. În plus, o parte semnificativă dintre acestea sunt cel puțin parțial opace la o monitorizare detaliată a deciziilor luate, cum este, de exemplu, cazul sistemelor bazate pe învățare profundă (*deep learning*). Decizia luată de o inteligență artificială este, în unele cazuri, inexplicabilă, nu poate fi făcută o asociere directă între elemente din reprezentarea internă a datelor și din procesul de prelucrare a lor și decizia produsă de acel sistem. În acest caz nu doar descoperirea și corectarea unei erori vor fi dificile, dar apare și problema asumării unei responsabilități pentru posibilele pagube provocate. Așa cum am menționat însă în prima parte a acestei Perspective, inteligența artificială nu poate avea inițiativa stabilirii obiectivelor pentru atingerea cărora a luat acele decizii percepute ca erori. O soluție posibilă ar fi deci asumarea responsabilității de către persoana sau persoanele care au stabilit acele obiective pentru acel sistem. Legătura cauzală între acele persoane și erorile comise este stabilită prin faptul că dacă ele nu ar fi solicitat acelei inteligențe artificiale atingerea acelor obiective, aceasta nu ar fi luat deciziile considerate ca fiind cauza prejudiciilor. Spre deosebire de o persoană, care are opțiunea ignorării unei impuneri greșite, ilegale sau imorale (și posibilitatea determinării acestui lucru), sistemul inteligent este limitat prin natura sa la urmărirea atingerii obiectivelor stabilite prin luarea oricăror decizii determinate ca fiind optime.

În *Cartea albă privind inteligența artificială. – O abordare europeană axată pe excelență și încredere* sunt stabilite și niște soluții pentru facilitarea integrării tehnologiilor IA, date sub forma unor cerințe, ce ar trebui respectate de orice sistem inteligent. Aceste cerințe se referă la datele de antrenament, păstrarea datelor și a evidențelor, informațiile care trebuie furnizate, soliditatea și precizia, supravegherea umană și cerințe specifice pentru anumite aplicații specifice ale IA, cum ar fi cele utilizate în scopul identificării biometrice la distanță.

Legat de datele folosite pentru antrenarea unor sisteme inteligente, sunt propuse cerințe privind garantarea utilizării datelor cu caracter privat, a respectării standardelor de siguranță, a restricționării utilizării alternative ulterioare. Cred însă că impunerea unor restricții prea puternice în aceste privințe ar limita dezvoltarea sistemelor bazate pe învățare automată care ar respecta aceste restricții și ar păstra supremația unor sisteme existente, care fie au beneficiat de lipsa unor restricții în momentul dezvoltării lor, fie pot ignora aceste restricții prin valorificarea accesului la contribuții voluntare de date sau prin separarea juridică de zona Uniunii Europene. În plus, orice valorificare ilegală a unor informații sensibile incluse în aceste date poate fi sancționată deja, conform legislației curente. De exemplu, în prezent există angajați ai unor instituții publice sau private care au atras penalități instituției de care aparțin prin publicarea unor informații ce nu respectă General Data Protection Regulation (GDPR).

Pentru păstrarea datelor și a evidențelor se propune păstrarea datelor de antrenament folosite, a soluțiilor tehnice care au valorificat acele date, precum și „documentația privind metodologiile de programare și de antrenare, procesele și tehnicile utilizate pentru a elabora, testa și valida sistemele de IA". Aceste date ar urma să fie „puse la dispoziție la cerere, în special pentru a fi testate sau inspectate de către autoritățile competente."

Aceste soluții au potențialul de a limita interesul unor dezvoltatori care beneficiază de soluții tehnice sau resurse cu caracter restrictiv (protejate

de brevete sau cu caracter comercial). Pe de altă parte, posesia lor nu garantează, în multe cazuri, posibilitatea duplicării sau explicării comportamentului sistemului antrenat. Orice proces de învățare automată implică și o componentă aleatorie care poate altera rezultatul în anumite situații, în majoritatea cazurilor procesul de antrenare producând rezultate unice.

Cu privire la informațiile care trebuie comunicate public, se afirmă că ar fi importantă „asigurarea furnizării de informații clare în ceea ce privește capacitățile și limitările sistemelor de IA". Aceasta pare a fi o descriere vagă și inerent imprecisă. Nimeni nu poate ști precis și permanent care sunt capacitățile și limitările unui sistem, mai ales dacă este capabil de antrenare și interacțiunile sale nu sunt controlate complet. Mai mult decât o enumerare de capabilități aproximative, fără a detalia limitările în contextul acelor capabilități, nu cred că este posibil decât pentru aplicații simple din afara contextului IA. În privința transparenței funcționării acestor sisteme inteligente, cred că ar fi mai indicată comunicarea obiectivelor stabilite pentru acestea, precum și a instrumentelor puse la dispoziție acelui sistem în vederea atingerii acestor obiective.

Legat de soliditatea și precizia sistemelor inteligente, în documentul analizat se propun „cerințe care să asigure că sistemele de IA sunt solide și precise sau cel puțin reflectă în mod corect nivelul lor de precizie în toate etapele ciclului de viață; cerințe care să asigure reproductibilitatea rezultatelor; cerințe care să asigure că sistemele de IA pot face față în mod adecvat erorilor sau inconsecvențelor în toate etapele ciclului de viață; cerințe care să asigure reziliența sistemelor de IA atât împotriva atacurilor deschise, cât și a încercărilor mai subtile de manipulare a datelor sau chiar a algoritmilor, precum și faptul că în astfel de cazuri se iau măsuri de atenuare."

Termenii menționați (solide, precise, fac față erorilor, reziliente) sunt practic imposibil de formalizat, inclusiv în termeni juridici. Precizia unui sistem inteligent nu poate fi determinată decât într-un context controlat și static, rezultând valori puțin semnificative în alte contexte.

Reproductibilitatea rezultatelor nu poate fi asigurată decât în aceleași condiții. Reziliența la atacuri sau manipularea datelor este de dorit dar nu poate fi legiferată în condițiile în care realitatea confruntării cu imprevizibilitatea și imaginația utilizatorilor, precum și cu noile tehnologii și contexte de utilizare face imposibilă garantarea invulnerabilității oricărui sistem inteligent. Solicitări mai realiste ar fi facilitarea descoperirii eventualelor comportamente problematice ale sistemului sau ale utilizatorilor și obligativitatea îmbunătățirii comportamentului sistemului prin actualizări periodice.

În aceeași *Carte albă privind inteligența artificială – O abordare europeană axată pe excelență și încredere* sunt propuse soluții pentru supravegherea umană a sistemelor inteligente, mergând până la propuneri cum ar fi „rezultatul acțiunii sistemului de IA nu produce efecte decât dacă a fost revizuit și validat în prealabil de o persoană". Dacă există persoane capabile și dispuse să verifice și să certifice deciziile luate de un sistem inteligent, necesitatea existenței acelui sistem dispare.

Unul dintre motivele principale pentru care se simte necesitatea unor astfel de sisteme este tocmai apariția tot mai frecventă a unor contexte în care factorul uman este insuficient. Volume foarte mari de date sau solicitări foarte numeroase de acces exclud factorul uman în multe situații. Soluția ar trebui să fie descrisă în termeni de responsabilitate. Trebuie identificate persoanele dispuse să își asume deciziile luate de un sistem de IA, chiar și fără să fie capabile sau dispuse să valideze acele decizii.

În documentul la care ne referim, sunt clar sumarizate principalele deficiențe inerente în legiferarea tehnologiei și a comportamentului sistemelor inteligente: „Caracteristicile specifice ale multor tehnologii de IA, printre care se numără opacitatea („efectul de cutie neagră"), complexitatea, imprevizibilitatea și comportamentul parțial autonom, pot îngreuna verificarea conformității cu legislația existentă a UE, menită să protejeze drepturile fundamentale, și pot împiedica asigurarea respectării efective a acesteia. Autoritățile de aplicare a legii și

persoanele afectate ar putea să nu dispună de mijloacele necesare pentru a verifica modul în care a fost luată o anumită decizie cu implicarea IA și, prin urmare, dacă au fost respectate normele relevante. Persoanele fizice și entitățile juridice se pot confrunta cu dificultăți în ceea ce privește accesul efectiv la justiție în situațiile în care astfel de decizii îi pot afecta în mod negativ."

Inteligența artificială, în combinație cu dezvoltarea și expunerea tot mai mare a societății la tehnologiile informației, are potențialul de a aduce schimbări majore în societate. Riscul major legat de acest potențial este tocmai dispariția asocierii dintre interacțiunile în majoritatea contextelor socioeconomice și o altă persoană.

Tehnologiile neinteligente nu ridică semnificativ acest risc, ele au în general rolul de a facilita și automatiza procese simple, controlate în mare măsură de inițiator. Acest risc însă nu poate fi diminuat prin restrângerea capabilității sistemelor inteligente și nici prin impunerea unui control uman strict, problematic în special din punctul de vedere al limitărilor umane. Responsabilitatea pentru deciziile luate de acele sisteme poate fi însă asumată de beneficiarii direcți sau indirecți. Compania care decide să interacționeze cu publicul folosind un sistem inteligent de „Serviciu clienți" trebuie să își asume comportamentul acelui sistem. Utilizatorul care decide să își faciliteze activitatea proprie prin valorificarea unor tehnologii trebuie să fie responsabil pentru deciziile luate de acele sisteme în atingerea obiectivelor stabilite de către el.

Aspecte privind impactul noilor tehnologii asupra mediului juridic

Prof. univ. dr. Traian Briciu
Președinte al Uniunii Naționale a Barourilor
din România (U.N.B.R.)

Încep prin a afirma că legătura mea cu tehnologia este una destul de precară, dar am participat la destule întâlniri, conferinţe naţionale sau internaţionale care au adus în discuție această problemă a impactului noilor tehnologii asupra actului de justiţie.

Am observat că mulţi juriști îşi pun problema viitorului avocaturii sau a profilului pe care îl va avea avocatul în urma asaltului inteligenței artificiale. Am abordat o poziţie oarecum particulară. Profesiile în genere şi profesiile juridice în special vor exista atât timp cât va fi nevoie de ele în societate şi vor avea acea coloratură pe care o societate la un anumit moment o impune. În momentul în care nu mai au un rol social dispar de la sine. Este foarte greu de susținut existenţa unei profesii sau forma unei profesii de dragul profesiei respective, pentru că profesiile, în esenţă, au fost create pentru a servi oamenii, care au un anumit moment de dezvoltare istorică dorită sau nedorită. Prin urmare, am respins orice teză care pornește de la ideea că tehnologia/inteligenţa artificială ar reprezenta un pericol pentru unele profesii juridice sau pentru actul de justiţie în sine. Cu siguranţă că le va modela activitatea. Trebuie văzut însă cum şi care sunt limitele!

În abordarea impactului noilor tehnologii asupra juridicului ar trebui să distingem între mai multe aspecte în interiorul domeniului juridic.

În primul rând, trebuie distins între partea de prezentare comercială a inteligenţei artificiale, cu care lumea ia contact de obicei şi pe baza căreia își formează opiniile generale, și realitatea tehnică și juridică. Prezentarea comercială înseamnă ştiri, filme, în cel mai bun caz extrase din prezentări ştiinţifice sau transpuneri mai mult sau mai puțin riguroase ale unor informări cu conținut științific. Există o exuberanță a omului pentru orice idee de schimbare, de nou sau de înlăturare a unor limite. De cele mai multe ori însă nu este nimic nou, ci numai prezentarea situației creează aparența de noutate absolută, pentru a se putea vinde mai bine. Un exemplu: o știre anunță că *a apărut primul judecător robot*. În realitate el nu judecă, ci oferă nişte informaţii de specialitate unui judecător, iar decizia finală o ia judecătorul. Publicul va reține însă titlul, respectiv că a judecat robotul.

În al doilea rând, dincolo de prezentarea comercială a realizărilor tehnologiei, există o realitate care ne indică faptul că și la ora actuală inteligența artificială ocupă deja un loc semnificativ în activitatea juriștilor. Aceasta va creşte pe următoarele aspecte: *organizarea cabinetelor de avocatură* şi a *managementului judiciar*. Sunt contracte de sute de pagini. Un avocat sau un judecător trebuie să caute o informație în dosare care cuprind uneori zeci de volume. Sigur că în contextul ăsta, inteligența artificială care îți extrage documentul de care ai nevoie te scutește de mult timp alocat unei căutări. Acest timp este însă util pentru a reflecta cât mai profund asupra problemei de fond. Tehnologia ajută și managementul biroului de avocatură și al instanței de judecată. La bogăția de informații pe care o oferă legislația și jurisprudența este evident că fără o structurare specifică noilor tehnologii nu ne-am mai putea descurca în mod competitiv în domeniul juridic.

Putem lua în calcul și faptul că sub impactul inteligenţei artificiale o să dispară realmente o bună parte din serviciile juridice. Despre ce servicii juridice ar fi vorba? Vorbim de servicii juridice care ar reprezenta simple informări asupra unor cerințe legale, cum ar fi documentele

necesare pentru înființarea unei societăți. În cele mai multe cazuri, aceste informații nu reprezintă o provocare profesională pentru un avocat. Nu reprezintă un act de creație. De aspecte precum acesta se poate ocupa un robot, dar și atunci cineva va trebui să supervizeze rezultatul, măcar pentru a se asigura că nu decurge dintr-o eroare de procesare a datelor.

În fine, există și o zonă juridică în care credem că inteligența artificială nu are ce căuta, respectiv în partea decizională – fie că vorbim de adoptarea unei strategii de apărare (în cazul avocatului), fie că vorbim de o soluție asupra unei dispute (în cazul judecătorului).

Interdicția inteligenței artificiale în latura decizională are susținere în următoarele particularități, care ţin de drepturile şi libertăţile fundamentale :

– Este recunoscut dreptul de *a accede la o instanță*. Accesul la o instanţă implică faptul că aceasta operează cu concepte precum „echitatea". Or, noţiunea de echitate este una fundamental legată de fiinţa umană. Accesul la o instanță presupune ca aceasta să fie independentă. Ar putea fi total independent ceva creat sau totul este dependent de bagajul informaţional şi tehnologia utilizată?

– *Secretul profesional* reprezintă una dintre marile probleme ale lumii juridice. Acesta presupune încredinţarea anumitor informații unei persoane și asigurarea păstrării acestora, în logica garanțiilor morale și profesionale ale avocatului. În cazul utilizării tehnologiei, un secret profesional nu are garanția morală, fiind vorba numai de una tehnică – *cine poate să acceseze respectivul program*.

– *transparenţa procesului decizional*. Acceptabilitatea actului de justiţie se bazează pe înţelegerea mecanismului de gândire folosit de judecător. Or, este greu de acceptat soluția atunci când tipul de analiză nu este unul similar celui căruia i se adresează aceasta. Nu vorbim aici despre corectitudinea sau nu a raționamentului utilizat în stabilirea unei soluții, ci despre încrederea pe

care trebuie să o aibă destinatarii actului de justiție că problemele lor au fost analizate într-un mecanism de gândire compatibil cu cel utilizat de ei.

– *adâncirea crizei de încredere instituțională*. Dacă, într-o ipoteză de lucru, un stat, în pofida tuturor argumentelor prezentate, ar trece la o robotizare a judecăților, ne-am afla în fața unui aspect dificil, deoarece roboții sunt creați de anumite firme, iar alegerea unei tehnologii ar trebui să aibă la bază anumite criterii. Desigur, ar fi o licitație. În mod cert una cu un caiet de sarcini cu multe particularități și o comisie pe măsură. În final însă, nu ne putem gândi că nu vor fi critici cu privire la selecție. Și ce s-ar întâmpla atunci când aceste critici se vor uni cu miile de nemulțumiri față de o soluție sau alta? Va scădea profund încrederea în justiție, dacă nu cumva vor apărea revolte sociale. De altfel, și acum sunt critici cu privire la selecția judecătorilor, numai că în contextul actual există și speranța modelării lor în cursul activității profesionale!

Toate acestea și altele sunt aspecte care elimină din start ideea că inteligenţa artificială ar putea vreodată să joace rol de instanţă în sensul adevărat.

Totuși, aceasta nu înseamnă că inteligența artificială nu ar putea juca un anumit rol în luarea deciziei pe baza primelor elemente arătate, precum identificarea cu exactitate a informațiilor relevante, aducerea în lumină cu rapiditate a legilor aplicabile și a jurisprudenței ori a doctrinei relevante.

Atenția noastră însă ar trebui să se îndrepte spre o altă temă. Este clar că într-un fel sau altul inteligența artificială va juca un rol în justiție. Esența problemei este însă reglementarea cu rigurozitate a acestui rol și garantarea utilizării acesteia numai de către profesioniști, cel puțin pe anumite paliere care implică efectuarea de acte juridice.

Într-un registru foarte larg impactul inteligenței artificiale asupra mediului judiciar are anumite legături cu impactul pe care îl are asupra aceluiași mediu exacerbarea comunicărilor în mediul online. Fără

îndoială, sunt două fenomene distincte, dar care au în comun faptul că utilizarea temperată a acestora este benefică mediului judiciar, în timp ce utilizarea lor într-o manieră total liberă aduce grave prejudicii acestuia.

În legătură cu comunicările online și lipsa de responsabilitate pe care o cuprinde de multe ori exprimarea în spațiul virtual, s-a pierdut momentul ideal al reglementărilor. Desigur, acum se încearcă formule mai mult sau mai puțin eficiente de a diminua efectele negative ale formării de opinii în mediul virtual, dar sunt soluții care intervin *post factum*. Societatea este deja prea afectată de acest virus. Neîncrederea în orice valoare, în orice instituție, contestabilitatea generală a orice de către oricine este un fenomen specific apariției și dezvoltării fără vreo limită (până la un anumit moment) a dezbaterii în mediul virtual. Încetul cu încetul oamenii își raportează gândirea și comportamentul platformelor de comunicare. Chiar și judecata asupra a ceea ce este permis sau nu să fie spus/scris pe o anumită platformă de comunicare este cedat tot platformelor de comunicare, deoarece nu există un număr atât de mare de judecători pentru a primi și rezolva milioane de sesizări cu privire la încălcarea legii în comunicare în spațiul virtual. Cel mai probabil, restricționările se fac în prima fază automat, pe baza unui program, a unei tehnologii. Iată că o primă formă de judecător-robot a apărut în mod necesar, ca mod de răspuns la un flagel insuficient analizat atunci când societatea a îmbrățișat emoțional comunicarea în mediul virtual la scară globală.

În legătură cu utilizarea inteligenței artificiale în mediul judiciar, fie că vorbim de consultanța juridică, fie că vorbim de utilizarea acesteia în sfera instanțelor, reglementările sunt absolut necesare.

Înainte de a face şi aici să prolifereze neîncrederea în orice valoare şi orice instituţie, şi contestabilitatea generală a orice de către oricine, statele ar trebui să stabilească limitele utilizării acestor instrumente. Fără reglementări, pericolul este mai mare decât cel generat de transferul dezbaterilor în mediul virtual.

În esență, apreciem că problema reglementării modalității de utilizare a inteligenței artificiale în mediul judiciar ar trebui să formeze o prioritate.

În principiu, aceasta ar trebui manevrată exclusiv de profesioniști ai dreptului, la fel cum și aparatura medicală este utilizată numai de medici. Faptul că efectele utilizării într-un mod neprofesionist a informației juridice nu este unul cu efecte atât de abrupte ca cele ale utilizării unei tehnici medicale nu trebuie să ne inducă în eroare, deoarece, în final, rezultatele nocive vor fi similare, cu respectarea particularităților fiecărui domeniu. Numai o asigurare a modului profesionist de utilizare a rezultatelor primare indicate de evaluarea făcută pe baza inteligenței artificiale asupra unei chestiuni juridice poate asigura creșterea calității mediului judiciar și diminua pericolul unei profunde crize de încredere în justiție.

Inteligența artificială de încredere – între reglementare și valorificarea potențialului de creștere a competitivității economice

Alexandru Petrescu

Este foarte important, din perspectiva tematicii puse în dezbatere, să nu plecăm din această conferință cu impresia că Uniunea Europeană s-ar afla undeva în zona pionieratului în ceea ce privește elaborarea legislației privind inteligența artificială. Din punct de vedere geografic, Uniunea Europeană acoperă un spațiu cvasicontinental, cu tot ce interesează inteligența artificială, tehnologii avansate etc., situându-se după Asia și Statele Unite. Ca să vă dau o dimensiune a operaționalizării acestei tehnologii din perspectivă legislativă, deoarece s-a vorbit aici foarte mult despre tema legislației în domeniul inteligenței artificiale, Statele Unite au adoptat *National AI Act* undeva în 2020.

Există deci autoritate de reglementare. Agenția *Bloomberg Innovation* a considerat, în 2021, Coreea de Sud ca fiind cea mai inovativă țară. Din 2019 există reglementări, legislație, autoritate, certificări și toate celelalte. Calea este deci deschisă și nu trebuie să ne fie teamă nici să utilizăm fără reținere tehnologii bazate pe inteligența artificială, nici să înțelegem mecanismele psihologice ale raportărilor oamenilor – specialiști sau simpli utilizatori și beneficiari direcți ai unor instrumente sau proceduri însumând inteligență artificială.

Europa se află în faza de *whitepapers*, de *guidelines,* de discuții, unele chiar filozofice. Suntem în continuare departe de celelalte zone ale lumii, iar această discrepanță între noi, ca populație, respectiv, ca organizație europeană, și alte zone ale lumii, unde progresul cercetărilor în domeniul inteligenței artificiale reprezintă preocupări constante, la nivel statal, și în mod deosebit în domeniul privat, trebuie să ne dea de gândit. De aceea, mi se pare că astfel de întâlniri, cum este și această conferință, nu neapărat între oameni de știință implicați în crearea și utilizarea inteligenței artificiale, ci între cadre didactice din învățământul juridic, tehnic, psihologi, medici, juriști practicieni, membri ai Parlamentului sau reprezentanți ai societății civile, pot să dea un semnal că problematica generală a inteligenței artificiale ne privește pe fiecare dintre noi, în domeniile în care lucrăm sau ne desfășurăm activitatea. Concluziile unor astfel de simpozioane, mese rotunde sau conferințe au menirea să trezească și interesul legiuitorilor, pentru că, până la urmă, orice utilizare a tehnologiilor bazate pe inteligența artificială, cu beneficiile și riscurile sale, dă naștere unor raporturi sociale care, firește, trebuie reglementate într-un cadru normativ care să răspundă unor valori ocrotite de stat și de societate în ansamblul ei.

Vă anunț că România a semnat în septembrie 2019, sub mandatul meu de ministru, Declarația de aderare la principiile inteligenței artificiale. Cu toate acestea, nici astăzi nu avem o legislație corespunzătoare sau măcar o politică în domeniul IA care să fie în concordanță cu ceea ce se conturează la nivel internațional.

La nivelul Uniunii Europene, după cum dl Dragoș Tudorache a arătat deja, în aprilie 2021 a fost lansat proiectul de lege al Comisiei Europene și al Consiliului pe legislație europeană. Acesta este încă în discuție. La 27 iunie 2022 a fost lansat în Spania primul spațiu de testare numit *Regulatory Sunbox*. Thierry Breton, comisar european pentru Piața Internă, Roberto Viola, DG Connect, și dl Dragoș Tudorache au fost prezenți online la lansarea acestui important laborator de testare, care la rândul lui va furniza informații către decidenți. Aceștia, într-un viitor incert, vor ajunge să elaboreze o legislație în domeniu. Deci cam aici ne situăm și este important să cunoaștem acest lucru; de exemplu, în ceea ce privește tehnologia 5G, peste 65% dintre licențele de proprietate intelectuală se află în afara UE.

Vreau să vă spun că nici în ceea ce privește domeniul inteligenței artificiale nu stăm mai bine. La momentul la care vom avea o legislație și vom găsi un răspuns la toate întrebările pe care, pe bună dreptate, ni le punem, vom fi pe același loc unde ne-am consacrat și sper să rămânem, pentru că vine America de Sud foarte puternic din spate. Europa rămâne totuși „a treia geografie" la nivel mondial, în ceea ce privește reglementarea, adopția și, mai departe, valorificarea și, în ultimă instanță, monetizarea tehnologiilor electronice.

Conform PricewaterhouseCoopers, până în 2030 o să avem o creștere a PIB-ului cu 15,7 miliarde de dolari, doar din inteligență artificială. Nu e de mirare că, în acest context, China va atrage 20% din această plusvaloare, America de Nord, în principiu Statele Unite, 14% și Uniunea Europeană, undeva între 5% și 7%. Cam asta este estimarea în momentul de față. Se poate aprecia că 6,6% va veni din creșterea competitivității și a productivității, iar undeva pe la 9,1% va veni din siajul consumerismului creat de inteligența artificială.

În ceea ce privește România, conform EUROSTAT, în anul 2020 doar 5% dintre firmele cu mai mult de 10 angajați au folosit aplicații care încorporează inteligența artificială. La nivel european, media este de 7%, deci nu este atât de rău, iar țara care este premiată este Irlanda cu

23%, dar să nu uităm că Irlanda este *gateway*-ul pentru companiile de tehnologie mondiale, datorită unei fiscalități favorizante. Și pentru că am vorbit despre fiscalitate și am ascultat cu foarte mare interes ceea ce au spus domnul avocat Traian Briciu și ceilalți vorbitori, este clar că inteligența artificială are și în România o oarecare preponderență în multe sectoare de activitate: vorbim despre zona de conformare fiscală, vorbim despre zona de certificare, unde așa numitele *do-it-your-self audit* deja sunt pe scară largă întâlnite, deci e clar că inteligența artificială va interveni și în zona drepturilor fundamentale ale omului și în zona strict juridică. Este clar că în acest domeniu nivelul de interpretare a avantajelor și, totodată, al riscurilor recurgerii la inteligența artificială este foarte larg. Sunt sigur că colegii juriști sunt cei în măsură să aprecieze chiar necesitatea și riscurile utilizării inteligenței artificiale, iar în raport cu acestea, să stabilească și gradul de răspundere a celor care, prin utilizarea greșită a inteligenței artificiale, încalcă drepturi consacrate prin normă constituțională sau în legislația ordinară.

Aceste elemente și niveluri largi de interpretare vin chiar din lipsa unei legislații măcar primare, secundare. În Europa, avem tendința, ca și în cazul altor tipuri de tehnologie, să depunem un efort fantastic de mare în zona omogenizării legislației și atingerii unei convergențe și mai puțin în zona creșterii nivelului de performanță și inovație. Cu alte cuvinte, este mai greu să gândim într-un numitor comun și avem mai puțină energie în zona efectiv de inovație. Nu facem decât să preluăm ceea ce se produce, se inovează, se înregistrează ca proprietate intelectuală în alte țări.

Aș dori, în continuare, să spun câteva cuvinte despre o altă legislație importantă, și anume, din zona NATO. Problema a căpătat o oarecare actualitate în ultima perioadă. În octombrie 2021, a fost adoptată strategia de implementare a inteligenței artificiale, alături de alte șapte componente de tehnologie (tehnologia supersonică, tehnologia cuantică, bioingineria sunt doar elemente ale unei serii de paliere tehnologice, adoptate la nivel de principii de către NATO). Este, apoi,

discuția identificare vs anihilare, unde, după cum toți ne așteptăm, în zona de anihilare, componenta umană rămâne primordială ca decizie. În continuare, sper că ceea ce ne va reveni din perspectivă practică, Regulatory Sandbox din Spania, ne va ajuta să ajungem cât mai repede la o variantă, la un proiect oarecum final de legislație, astfel încât să poată fi adoptat și diseminat la nivelul UE.

O ultimă idee si cu asta închei. Nu pot să nu-mi exprim un mare regret. România, undeva în anul 2019, era în *pole position* pentru a găzdui pe teritoriul nostru acest spațiu de testare. Din păcate, s-a pierdut puțin din preocuparea și concentrarea pe acest demers și, ca urmare, Spania a reușit să atragă pe teritoriul său această importantă unitate din perspectiva IA, ceea ce îi sporește securitatea juridică pe tot ceea ce înseamnă inovație și îi încurajează eforturile proprii pentru inovație și de a utiliza toate celelalte beneficii tangente.

Utilizarea inteligenței artificiale ca instrument de recrutare a personalului

Mihai Negroiu

În intervenția mea nu mă voi referi la problematica generală a inteligenței artificiale, ci mă voi rezuma să arăt modul în care o companie privată de consultare a înțeles să ofere servicii de specialitate, utilizând IA ca instrument de recrutare a personalului.

Cine suntem

Integrity Meter România este o companie specializată în furnizarea de servicii suport în procesele de recrutare și evaluare a angajaților. Practic, serviciile sunt destinate tuturor companiilor și organizațiilor din România care își doresc să îi identifice pe cei mai buni candidați pentru nevoile lor și să își formeze o imagine de ansamblu asupra angajaților și a activității. Astfel, sistemul oferă un ajutor considerabil în ceea ce privește procesul de luare a deciziilor, atât la nivelul departamentului de resurse umane, cât și la nivelul departamentului de securitate.

Cum a apărut

Compania are deja o experiență de peste 15 ani pe piețele internaționale, de când a fost lansată în Israel. Fondatorii ei sunt Gozlan

Menachem și Dotan Shavit. Pe de o parte, Gozlan este expert în evaluarea angajaților, în efectuarea testelor poligraf și în investigații. În prezent este membru al Asociației Americane a Examinatorilor Poligraf, precum și al Asociației Israeliene a Examinatorilor Poligraf. Acesta are, de asemenea, studii aprofundate de psihologie și fiziologie. Începând cu 1987, Menachem Gozlan a furnizat servicii de evaluare a integrității (evaluarea angajaților, intervievare, verificări de fond și testări poligraf) atât societăților din sectorul privat, precum și instanțelor. Dotan are experiență ca dezvoltator de software în cadrul unor companii de IT din Israel, dar și în securitatea informației. Construind pe baza expertizei acumulate de-a lungul carierei sale de către Gozlan, Dotan a transpus metodologia de lucru în platforma online, ceea ce a rezultat în crearea Integrity Meter.

Datorită îmbinării cunoștințelor celor doi, testele dezvoltate și implementate de Integrity Meter nu doar că oferă o radiografie a sistemului de valori al candidaților, dar respectă și cele mai înalte standarde în ceea ce privește securitatea informațiilor și protecția datelor cu caracter personal. Aplicația inițială a apărut ca răspuns pentru cerințe specifice venite din partea unor companii cu renume, precum McDonald's, Cellcom sau DHL. Mai exact, sistemul este bazat pe principii ale metodelor avansate de interogare, pe experiența acumulată în centre de evaluare și pe procesul analitic al verificării cu poligraf, dar este în totalitate online, nu necesită aparatură specializată și implică resurse mai puține și costuri mai mici.

Încă de la înființarea companiei, operațiunile s-au extins pe mai multe piețe la nivel internațional, precum Statele Unite ale Americii, Brazilia, Argentina, America Centrală și, mai recent, Franța, Italia, Germania sau Ungaria. De-a lungul celor 15 ani de activitate, serviciile au fost ajustate în funcție de clienți, de piețe și de schimbările care apar la nivelul societății și al economiei. Spre exemplu, lansarea serviciului în România a avut loc în 2018, numai după ce ne-am asigurat că am adaptat produsele noastre la piața românească, la particularitățile locale și, bineînțeles, la legislația națională și europeană.

Cum a fost validat

Validitatea testului Integrity Meter se bazează, în ceea ce privește caracterul impresionabil, pe informațiile acumulate în urma testării unor zeci de mii de candidați și pe experiența cu testul a clienților. Pentru un nivel suplimentar de validare, în decembrie 2009 a fost efectuată o cercetare asupra metodei de evaluare ocupațională Integrity Meter. Scopul acestei cercetări a fost să realizeze o examinare cantitativă a validității acestui instrument de testare și să investigheze relația dintre rezultatele Integrity Meter și rezultatele unei examinări cu poligraf. În acest studiu, examinarea poligraf a servit ca o bază pentru realizarea unei validări încrucișate. Rezultatele obținute de Integrity Meter au fost intersectate cu cele ale poligrafului și a fost examinată corelația dintre diversele decizii luate în cadrul ambelor metode. Studiul a concluzionat că Integrity Meter detectează un procent considerabil de candidați cu un trecut problematic în domeniilor cercetate din numărul de candidați care au fost semnalați astfel și de testul poligraf. Procentele de detecție, 76% pana la 100%, demonstrează că sistemul produce rezultate extrem de similare cu cele ale poligrafului, în ceea ce privește detectarea candidaților cu un trecut problematic.

În plus, a rezultat că, în comparație cu poligraful, sensibilitatea testului, mai exact abilitatea de a detecta și recunoaște candidații cu trecut problematic în domeniile studiate, este foarte ridicată, iar șansele ca un astfel de candidat să poată trece de evaluarea Integrity Meter sunt foarte mici. Rezultatele sunt extrem de semnificative pentru organizații și angajatori, în special în cazul celor pentru care exista un risc ridicat de a angaja un candidat cu un trecut problematic, de exemplu organizații de securitate și informații și organizații ce sunt în căutarea unui angajat într-o poziție de decizie.

În ceea ce privește candidații, gradul de specificitate al testului Integrity Meter este scăzut în comparație cu poligraful, ceea ce înseamnă că exista un număr considerabil de candidați ce nu au fost detectați de poligraf ca având un trecut problematic, dar au fost detectați de

Integrity Meter. Această disparitate provine din criteriile testului de definire a trecutului problematic, ce sunt mai cuprinzătoare decât cele ale poligrafului. Integrity Meter acoperă o varietate de subiecte mai largă decât cea a examinării cu poligraf și are un nivel mai mare de precizie în legătură cu orice alt subiect. De exemplu, chiar și o persoană care nu a fost implicată în activități infracționale (și, ca rezultat, acest lucru nu va fi detectat la un test poligraf) este posibil să fi comis la un moment dat o abatere minoră, iar acest lucru va apărea în rezultatul testului Integrity Meter. Studiul a descoperit un grad ridicat de compatibilitate între Integrity Meter și poligraf și, bazat pe cercetările anterioare ce au concluzionat că examinarea poligraf, atunci când este realizată conform unor protocoale bine cunoscute și studiate, este validă și de încredere, se poate afirma că validitatea testului Integrity Meter este susținută de examinarea poligraf și că abilitatea de a detecta candidați cu trecut problematic este crescută.

Structura soluției Integrity Meter

Testele sunt dezvoltate în funcție de nevoile clienților și sunt împărțite în două categorii: o serie de teste care vizează perioada pre-angajare, care se aplică celor care sunt implicați într-un proces de recrutare, și o serie de teste dezvoltate pentru evaluarea periodică a angajaților. Cele mai multe teste aplicate până în acest moment în România sunt cele pentru candidați. Platforma Integrity Meter măsoară atât caracteristici ale personalității unei persoane, cum ar fi onestitatea, integritatea sau autodisciplina, cât și factori de risc, precum dependențe, intenția de a prejudicia angajatorul sau implicarea în acțiuni ilegale. În plus, la cererea partenerului, se pot dezvolta subiecte conform nevoilor specifice.

În cazul testului periodic, subiectele sunt construite de o echipă formată din experți în examinări poligraf și în psihologie organizațională și furnizează cea mai completă imagine a angajatului. Sunt folosite ca evaluare a angajaților, aplicate la intervale stabilite de angajator, de exemplu evaluarea anuală. Pentru a asigura relevanța întrebărilor

și o îndeplinire cât mai completă a nevoilor, echipa Integrity Meter discută cu angajatorul spre a identifica ariile de interes pentru acesta, dar și detalii precum proceduri sau termeni specifici. În baza acestor informații este formulat un set de întrebări noi, care să cuprindă particularitățile companiei si pozițiilor de interes.

Toate testele pe care le oferim sunt personalizate în funcție de mai mulți factori, printre care domeniul de activitate al companiei care recrutează, poziția pentru care se recrutează sau atribuțiile specifice pe care persoana angajată va trebui să le îndeplinească. Abilitatea de a adapta testele la nevoile companiei face testele Integrity Meter să fie aplicabile la nivelul oricărui tip de angajator, indiferent de mărime sau industria în care activează.

În România, clienții noștri sunt prezenți în industria auto, FMCG, pariuri sportive, dezvoltare de software, securitate, tehnologie laser, rent-a-car sau leasing auto. În plus, lucrăm cu mai multe companii de recrutare, care utilizează serviciile Integrity Meter pentru a oferi un plus de valoare partenerilor lor. Până acum, au fost personalizate peste 50 de teste în România, pentru posturi vacante în mai multe tipuri de industrii, pentru poziții atât de tip *white collar*, cât și de tip *blue collar*. Mai mult decât atât, chiar dacă discutăm despre același tip de funcție – de exemplu, dacă vorbim despre o poziție de reprezentant de vânzări – testele diferă în funcție de nevoile și cerințele specifice ale organizațiilor care angajează. La nivel internațional, varietatea domeniilor de activitate a clienților Integrity Meter crește și mai mult. Pe lângă producători și furnizori de bunuri sau de servicii, printre clienți se află și angajatori din domeniul telecomunicațiilor, industria de apărare, industria petrolieră, dar și instituții de stat.

Metodologia de aplicare

Cei mai mulți dintre clienți preferă să aplice testele în fazele inițiale ale proceselor de recrutare, în principal datorită faptului că raportul este generat instant de către sistem și oferă informații concrete cu privire

la candidați. Pe baza acestora se pot lua decizii cu un grad mai ridicat de obiectivitate, în concordanță cu nevoile unei companii. Astfel, se pot identifica încă din primele etape ale procesului de recrutare candidații nepotriviți pentru pozițiile vacante, ceea ce are ca rezultat economisirea de resurse și de timp. Există, desigur, posibilitatea ca testele să fie folosite în orice moment al procesului de recrutare, în funcție de cerințele companiilor. În general, candidații sunt receptivi la teste, fiind asemănătoare testelor de personalitate folosite deja de mult timp de specialiștii în resurse umane și recrutare.

Algoritmii de verificare

Integrity Meter este un serviciu unic pe piața din România pentru că, în afară de informațiile pe care le furnizează și rapiditatea cu care oferă rezultatele, include o serie de algoritmi avansați care evaluează veridicitatea răspunsurilor oferite de candidați sau de angajați în chestionarele pe care le completează. Înțelegem că fiecare candidat va încerca să se prezinte într-o lumină favorabilă și va încerca să ascundă aspectele despre care crede că îi vor afecta șansele de a obține postul dorit. De exemplu, la o întrebare precum „Ai venit la muncă în ultimii doi ani sub influența alcoolului?", daca răspunsul este afirmativ, majoritatea candidaților vor ezita să răspundă onest.

Instrumentul, bazat pe tehnici profesionale de interogare și evaluare, are ca prim scop obținerea unor răspunsuri oneste și încurajarea recunoașterii unor situații. Dacă totuși cel testat alege să ascundă anumite aspecte, sistemul Integrity Meter, spre deosebire de alte testări computerizate, are abilitatea de a detecta acest lucru.

Calibrarea testului

În ciuda complexității sistemului, acesta este ușor de operat și furnizează un raport detaliat, imediat ce testul este finalizat. După ce persoana testată parcurge întrebările din cadrul testului, sistemul va analiza datele primite și va genera un raport detaliat și argumentat.

În funcție de răspunsurile date de persoana testată, raportul va fi încadrat într-o anumită categorie: „Respectă condițiile minime", „Necesită clarificări", „Nu îndeplinește condițiile minime". Aceste încadrări nu țin loc de decizie sau rezultat final, ci reprezintă o recomandare pentru angajator.

Chiar dacă doi candidați au parcurs același set de întrebări, este posibil că rezultatul testului să fie diferit. Acest lucru este cauzat de caracterul particular al fiecărei organizații și al fiecărei funcții și de adaptarea testului la acești factori. De exemplu, o companie poate pune accent pe un anumit aspect, pe care îl consideră cu o gravitate ridicată, în timp ce o altă companie îl poate considera relevant, dar nu grav. Un alt exemplu ar fi chiar între testele aceluiași angajator, în funcție de cerințele postului. Pentru o funcție cu drepturi mai multe, desigur că și riscurile sunt mai mari, iar angajatorul ar putea dori să crească sensibilitatea sistemului față de comportamente contraproductive. Acești parametri influențează algoritmii de calcul și încadrarea fiecărui raport într-una din cele trei categorii. Ei pot fi stabiliți încă dinaintea aplicării primului test de integritate sau pot fi calibrați pe parcursul activității, în funcție de analiza rezultatelor primite.

O altă funcție a platformei este calcularea de statistici pe baza datelor despre distribuirea rezultatelor. Cu alte cuvinte, se pot analiza procentul și numărul candidaților în funcție de rezultat (cei care îndeplinesc condițiile, cei ale căror rapoarte necesită clarificare, precum și cei care nu îndeplinesc aceste condiții).

Dacă la început distribuția se realizează în funcție de datele obținute în urma tuturor testelor realizate în România, după aplicarea unui număr de teste în cadrul unei organizații, sistemul poate calcula mai precis aceste medii și procente. Mai exact, statisticile vor fi bazate doar pe rezultatele testelor parcurse pentru organizația respectivă, oferind o imagine de ansamblu a nivelului de integritate din rândul candidaților sau angajaților. Mai mult decât atât, statisticile pot fi vizualizate și în funcție de alte filtre. De exemplu, pot fi analizate statisticile doar

pentru o anumită funcție testată sau doar pentru un anumit interval temporal, pentru a verifica dacă există un progres.

Așa cum a fost menționat anterior, sistemul de testare pe care îl furnizăm nu decide acceptarea sau respingerea unui candidat, ci oferă informații relevante în legătură cu activitatea pe care o va desfășura persoana respectivă și recomandări cu privire la nivelul de integritate al persoanei în cauză.

De exemplu, în cadrul unei companii din domeniul logistic, la finalul unui proces de recrutare, raportul a scos în evidență neconcordanțe majore între informațiile prezentate în CV de către candidatul preferat și cele oferite în timpul interviului, descalificându-l pe acesta. Un alt caz, din industria de asigurări, a avut un raport care a evidențiat faptul că persoana intervievată obișnuia să primească comisioane din partea furnizorilor, iar compania a decis să nu meargă mai departe cu angajarea pentru a preveni posibile pierderi. Pe de altă parte, în raportul unui candidat pentru funcția de investigator au fost surprinse atât răspunsuri afirmative cu privire la încălcări ale legii, cât și semne că acesta încerca să ascundă informații. În urma unor discuții suplimentare cu persoana testată, situația a fost clarificată și, în acest caz, angajatorul a decis că informațiile aflate nu reprezintă un risc pentru organizație.

Beneficiile testului

Integrity Meter le oferă companiilor o imagine completă a fiecărui candidat. Ca urmare, angajatorii pot lua decizii informate, astfel procentul candidaților potriviți pentru funcțiile și climatul organizațional crește.

O provocare pe care o întâlnesc angajatorii este că majoritatea candidaților se prezintă la interviuri cu CV-uri exemplare, făcute pentru pozițiile pentru care aplică. Acest lucru face dificilă filtrarea informațiilor și identificarea oricăror inconsecvențe în pregătirea candidaților sau experiența lor profesională. Iar acesta este un aspect pe care Integrity Meter îl acoperă, asigurând transparența totală a unui proces de recrutare și reducând dificultățile procesului.

Dacă un candidat potrivit aduce plus valoare organizației, un angajat necompatibil cu cerințele postului poate aduce pierderi însemnate. Un efect negativ cauzat de nepotrivirea unui nou angajat este productivitate scăzută a angajatului, din cauza nepotrivirii cu responsabilitățile poziției, din lipsa unor abilități necesare, sau din cauza incompatibilității cu restul echipei. Privind perspectiva opusă, aceasta incompatibilitate poate diminua și productivitatea celorlalți membri ai echipei.

Găsirea candidatului potrivit pentru o poziție deschisă din punctul de vedere al experienței și abilităților însă nu este suficientă, este necesară și potrivirea acestuia cu membrii echipei din care urmează să facă parte și cu sistemul de valori al organizației. Este esențial pentru succesul unei companii ca echipele să fie bine structurate și să existe acea chimie între angajați și între angajat și angajator.

Efectul pe termen lung al testului Integrity Meter este creșterea gradului de retenție a angajaților, ca rezultat direct al recrutării unor persoane care se pliază pe cultura organizațională și pe nevoile unei companii. Având în vedere costurile ridicate pe care le presupune angajarea unui nou angajat, instruirea acestuia și perioada de atingere a potențialului de performanță, cu cât fluctuația personalului este mai mică, cu atât riscul unor pierderi financiare nejustificate scade.

În plus, utilizarea serviciilor Integrity Meter conduce la reducerea pierderilor cauzate de angajați în contexte de tipul scurgerilor de informații, fraudelor sau furtului. Conform raportului realizat de Asociația Examinatorilor de Fraude Autorizați în 2022, în fiecare an, pierderi de 5% din profitul unei companii sunt cauzate de frauda ocupațională. Integrity Meter nu doar că poate ajuta la investigarea unui eveniment de securitate, odată ce acesta a avut loc, dar poate și preveni astfel de evenimente, prin angajarea unor oameni integri, care nu au intenția de a prejudicia angajatorul.

Acestea au fost câteva elemente care ilustrează pregnant, la nivelul unei companii, domeniile extrem de diferite în care inteligența artificială are un câmp larg de utilizare.

Limitele inteligenței artificiale și garantarea libertății. Actualitatea *Manifestului tehnorealist*

Prof. univ. dr. Sorin Bocancea
Rectorul Universității „Petre Andrei" din Iași

1. Limitele inteligenței artificiale

Pătrunderea fără precedent a inteligenței artificiale în toate domeniile vieții a generat discuții cu privire la modul de gestionare a acesteia și la implicațiile pe care le va avea în evoluția lumii civilizate și nu numai. În primul rând, s-a încercat definirea IA pentru a se putea trasa cadrele conceptuale ale fenomenului. Pe site-ul Parlamentului European, întâlnim o definiție simplă: „capacitatea unei mașini de a imita funcții umane, cum ar fi raționamentul, învățarea, planificarea și creativitatea"[1]. Într-o comunicare a Comisiei Europene către celelalte instituții fundamentale ale UE, apare o formulare mai complexă: „Inteligența artificială

[1] Cf. https://www.europarl.europa.eu/news/ro/headlines/society/20200827STO85804/ce-este-inteligenta-artificiala-si-cum-este-utilizata, accesat la 6 octombrie 2022.

(IA) se referă la sistemele care manifestă comportamente inteligente prin analizarea mediului lor înconjurător și care iau măsuri – cu un anumit grad de autonomie – pentru a atinge obiective specifice."[2] Între cele două definiții deja apar diferențe semnificative: prima vorbește de capacitatea mașinilor de a imita funcții umane, iar a doua de capacitatea lor de a lua măsuri, „cu un anumit grad de autonomie".

În urmă cu peste două decenii, Solomon Marcus semnala următoarea situație: dacă despre aparatele pe care ni le-a oferit până de curând progresul tehnic (radio, televizor și telefon) se poate spune că sunt „proteze", despre computer nu se mai poate spune la fel. „Programarea se sprijină pe anumiți algoritmi (succesiuni de instrucții și condiții logice) care se bazează pe anumite modele matematice ale proceselor pe care le avem în vedere. Sub raza de acțiune a acestor modele, intră, cu mai mult sau mai puțin succes, o parte a naturii înconjurătoare, a realității social-umane. […] Calculatorul devine un asociat organic al inteligenței noastre, spațiul de desfășurare a imaginației și curiozității noastre. Funcția de auxiliar al capacităților noastre cerebrale nu se realizează prin câteva manevre relativ simple, ca în cazul celorlalte proteze evocate mai sus. Gândirea algoritmică la care ne invită informatica amintește de cuiul lui Pepelea; inițial pe post de auxiliar, de proteză, această gândire se insinuează pe neobservate în comportamentul nostru, în logica noastră, în modul nostru de a vedea lumea."[3]

Într-adevăr, un computer iese din seria „protezelor" de până acum ale omului, dar poate el lua decizii în locul lui? Care este limita inteligenței artificiale? Cred că aceea dintre creator și creație. În primul rând, aceasta din urmă nu și-a putut decide propria apariție și funcționare. Iar, dacă un computer este folosit la crearea altui computer, nu el a decis oportunitatea și nici calea acestui act, ci omul.

[2] Comisia Europeană, *Inteligența artificială pentru Europa*, COM (2018) 237, Bruxelles, 25.04.2018, p. 1.

[3] Solomon Marcus, „Spațiul nostru comunicațional", în *Secolul 20*, nr. 4-9 (421-426)/2000, p. 71-79; p. 77.

Pentru a înțelege limitele inteligenței artificiale, consider că este utilă o întoarcere la Kant, pentru care cunoașterea umană „începe cu simțurile, înaintează de aici spre intelect și sfârșește cu rațiunea, deasupra căreia nu se găsește în noi nimic mai înalt pentru a prelucra materia intuiției și a o aduce sub cea mai înaltă unitate a gândirii"[4]. Toată cunoașterea configurată în intuițiile apriorice de spațiu și timp este ordonată de intelect, care operează cu categorii (precum cauză-efect, activ-pasiv, unitate-multiplicitate, posibil-imposibil, existent-non-existent, necesar-contingent etc.), configurând natura pe care o supune unei conștiințe de care el rămâne separat, pentru că intelectul rămâne tributar intuițiilor ce fac posibilă lumea fenomenală pe care o percepem în spațiu și timp. Spre deosebire de intelect, rațiunea nu mai este condiționată de lumea fenomenală, ea operând cu trei idei pure: Dumnezeu, libertate și nemurire. Între intelect și rațiune sunt plasate facultățile de judecare: cea estetică (ce distinge între frumos și urât) și cea teologică (ce distinge între bine și rău). Revenind la inteligența artificială, consider că ea este de domeniul intelectului, pentru că niciodată un computer nu va putea distinge între bine și rău sau între frumos și urât dacă el nu a fost programat de către om pentru astfel de categorii. De asemenea, un computer nu va opera cu ideile rațiunii, care sunt principii regulatoare ale vieții omului. În inteligența artificială se găsește mereu doar ceea ce a putut pune programatorul în ea și nimic mai mult. Un computer nu va putea lua decizii la originea cărora să stea raportarea la divinitate, libertate, frumos, urât, bine și rău. Inteligenței artificiale îi lipsește scopul a ceea ce face. Sub acest aspect, nu este niciun pericol ca viața omului să fie confiscată sau instrumentată de inteligența artificială,

Cu toate acestea, avertismentul lui Solomon Marcus trebuie luat în serios, pentru că inteligența artificială nu este o oarecare „proteză". Evoluția limbajului de programare a dus spre întâlnirea lui cu limbajul natural, ceea ce face ca inteligența artificială să se insinueze în viața

[4] Immanuel Kant, *Critica rațiunii pure*, traducere de Nicolae Bagdasar și Elena Moisuc, Editura Științifică, București, 1969, p. 283.

noastră mai ușor, pentru că nu mai poartă haina artificialului. „Depărtându-se inițial de limbajul natural, programarea la calculatorul electronic a demonstrat, prin progresele realizate, că această îndepărtare nu a avut alt scop decât acela de a se apropia chiar de ceea ce la început s-a depărtat... Artificialul și naturalul se cheamă unul pe altul pentru a constitui umanul."[5] Umanul însuși se constituie cu acest artificial ce ia chipul naturalului.

Care ar putea fi pericolul? Chiar și cei mai fervenți susținători ai inteligenței artificiale sunt de acord că există pericole ce pândesc utilizarea ei. În fond, orice instrument al omului e o sabie cu două tăișuri, decisiv fiind scopul în care îl folosește. Și ajungem din nou la scop. El rămâne mereu în mintea omului și, oricât de mult ar lua instrumentul „inteligență artificială" chipul naturii, el nu va putea să furnizeze scopul, rămânând mereu mijloc. În raport cu ce? În primul rând, s-ar putea spune că în raport cu intenția utilizatorului. Dar, dincolo de aceasta, care poate fi realizată mai bine sau mai rău, se află umanitatea, a cărei condiție de existență este libertatea. Și atunci când vorbim de realizarea libertății și de garantarea ei constituțională ajungem în fața nevoii de a reglementa.

2. Nevoia de reglementare

Pe 8 februarie 1996, John Perry Barlow lansa la Davos *A Declaration of The Independence of Cyberspace*, în care le cerea guvernelor lumii industrializate să-i lase pe cei din viitor în pace, pentru că nu au suveranitate acolo unde se adună ei. Iată câteva extrase din acest document: „Noi nu avem niciun guvern ales și nici nu este posibil să avem unul, așa că mă adresez cu o autoritate mai mare decât cea cu care vorbește întotdeauna libertatea însăși. Declar că spațiul social global pe care îl construim este independent în mod natural de tiraniile pe care doriți să ni le impuneți. Nu aveți niciun drept moral să ne conduceți și nici nu aveți vreo cale de constrângere de care

[5] Solomon Marcus, *art. cit.*, p. 75.

să avem motive să ne temem. […] Ciberspațiul nu se află în granițele voastre. Să nu credeți că-l puteți construi ca pe un proiect public. Nu se poate. Este un proiect al naturii și se dezvoltă prin acțiunile noastre colective. […] Nu cunoașteți cultura noastră, etica noastră sau codurile noastre nescrise care oferă deja societății noastre mai multă ordine decât ar putea fi obținută prin oricare dintre impunerea voastră. Pretindeți că avem probleme pe care trebuie să le rezolvați. Folosiți această afirmație ca pe o scuză pentru a invada incinta noastră. […] Creăm o lume în care toți pot intra fără privilegiile acordate de rasă, putere economică, forță militară sau statut social. Creăm o lume unde fiecare om își poate exprima oriunde dorește convingerile, oricât de ieșite din comun ar fi, fără teama de a fi constrâns la tăcere sau conformism. Conceptele voastre juridice de proprietate, expresie, identitate, mișcare și context nu se aplică în cazul nostru. Toate se bazează pe materie și nu există nicio materie aici. […] În Statele Unite ați creat azi o lege, *Telecomunication Reform Act*, care vă respinge propria constituție și insultă visele lui Jefferson, Washington, Mill, Madison, De Toqueville și Brandeis. Aceste vise trebuie să renască în noi. […] În lumea noastră, orice poate crea mintea umană poate fi reprodus și distribuit la infinit fără costuri. Transmiterea globală a gândirii nu mai necesită fabricile voastre pentru a se realiza. […] Trebuie să ne declarăm eul virtual imun la suveranitatea voastră, chiar dacă continuăm să fim de acord cu stăpânirea voastră asupra trupurilor noastre. Ne vom răspândi pe planetă astfel încât nimeni să nu ne poată opri gândurile. Vom crea o civilizație a Minții în spațiul cibernetic. Fie ca aceasta să fie mai umană și mai corectă decât lumea pe care guvernele voastre au făcut-o înainte."[6]

Un fan al acestui manifest plin de entuziasm și de anarhism, Edward Snowden, își amintește ce gândea la începutul anilor 2000: „îmi dădeam seama că tehnologia comunicațiilor avea șansa de a reuși acolo unde tehnologia violenței dăduse greș. Niciodată democrația

[6] John Perry Barlow, *A Declaration of The Independence of Cyberspace*, https://www.eff.org/cyberspace-independence, accesat la 25 iunie 2022.

nu va putea fi impusă cu sau sub amenințarea armelor; ea poate fi sădită doar prin răspândirea siliconului și a fibrei. La începutul anilor 2000, internetul abia dacă ieșise din găoace și... chiar și atunci oferea o întruchipare mai autentică și mai completă a idealurilor americane decât America însăși. Un loc unde toți oamenii sunt egali? Bifat. Un loc dedicat vieții, libertății și căutării fericirii. Bifat, bifat, bifat"[7].

Tânărul informatician avea să descopere mai târziu că și acel nou spațiu care-l entuziasma prin oportunitățile deschise s-a transformat din unul al libertății în unul al ingerinței și al supravegherii. Și el a fost unul dintre agenții acestei transformări, ajungând astăzi să fie cetățeanul Federației Ruse, unde libertatea nu este un lucru „bifat".

Pe data de 12 martie 1998, pornind de la un concept al lui Andrew Shapiro, David Shenk și Steven Johnson, un grup de 12 scriitori preocupați de implicațiile tehnologiei asupra indivizilor[8] au elaborat un document pe care l-au publicat sub titlul *Manifestul tehnorealist*[9], în care au formulat răspunsuri la următoarea întrebare: schimbările tehnologice rapide „sunt bune sau rele, trebuie întâmpinate cu bucurie sau cu teamă"? Răspunsul simplu a fost: „și una și alta. Tehnica ne face viața mai comodă și mai plăcută și pe mulți dintre noi mai sănătoși, mai bogați și mai înțelepți. Dar, în același timp, afectează imprevizibil munca, familia și economia, generând noi forme de tensiune și confuzie și aducând noi amenințări coeziunii comunităților noastre fizice"[10]. Însă acest răspuns a fost urmat de varianta detaliată.

Constatând că față de progresul tehnic s-au conturat două poziții extreme, semnatarii documentului arată că s-a conturat și un anume

[7] Edward Snowden, *Dosar permanent*, Nemira, București, 2019, p. 186.

[8] Aceștia au fost: David Bennahum, Brooke Shelby, Paulina Borsook, Marisa Bowe, Simson Garfinkel, Steven Johnson, Douglas Rushkoff, Andrew Shapiro, David Shenk, Steve Silberman, Mark Stahlman și Stefanie Syman.

[9] Varianta originală poate fi lecturată accesând https://www.technorealism.org/. În România, documentul a apărut tradus de Ana Tăbîrcă în revista *Secolul 20,* nr. 4-9 (421-426)/2000, p. 307-311, fiind disponibil la https://secolul21.ro/arhive/621.

[10] *Manifestul tehnorealist*, loc. cit., p. 307.

consens cu privire la subiect, pe care ei îl numesc „tehnorealism": „A fi tehnorealist înseamnă a gândi critic rolul pe care uneltele și interfețele îl joacă în evoluția omului și în viața cotidiană. [...] Ca tehnorealiști, căutăm să extindem terenul fertil aflat la mijloc între tehno-utopism și neo-ludism. Suntem «critici» ai tehnologiei așa cum și din aceleași motive pentru care alții sunt critici în gastronomie, critici de artă sau critici literari. Putem fi extraordinar de optimiști în legătură cu unele tehnologii, sceptici și disprețuitori în legătură cu altele. Totuși, scopul nostru nu este să supraapreciem, nici să desființăm tehnologia, ci mai degrabă să o înțelegem și să o aplicăm într-un mod corespunzător valorilor umane fundamentale."[11]

Mai este de actualitate acest document astăzi, după 24 de ani de la lansarea lui? Cred că da și vom vedea aceast lucru urmărind cele opt principii ale tehnorealismului pe care le propune.

Primul principiu este „tehnologiile nu sunt neutre". Având scopuri economice, sociale și politice, acestea au un impact direct asupra indivizilor, oferindu-le oportunități de rezolvare a problemelor, dar prezentând și pericole. Tehnologiile provin dintr-un tip de raționalitate, anume aceea științifică, și contribuie la menținerea acesteia. În opinia autorilor manifestului, „este important ca fiecare dintre noi să ia în considerare părtinirile diferitelor tehnologii și să le căutăm pe acelea care ne reflectă valorile și aspirațiile"[12].

Așa cum am arătat și mai sus, tehnologia nu ne lasă indiferenți pentru că ne structurează acțiunile. Și atunci când vorbim de aparate care nu doar imită acțiunile umane, ci sunt construite pentru a realiza acțiuni în locul omului, nu se mai poate vorbi de „neutralitate" a lor. Făcând pasul de la operațiunile computerelor la consecințele mesajelor lansate în ciberspațiu, constatăm că lipsa de neutralitate a tehnologiei este evidentă. Așa cum nici tiparul și nici emițătorul radio nu au fost neutre, nici computerele și nici ciberspațiul pe care l-au făcut posibil

[11] *Ibidem*, p. 308.

[12] *Ibidem*, p. 309.

nu sunt neutre. Dacă la început utilizarea lor presupunea competența unei minorități, astăzi orice individ alfabetizat devine lesne utilizatorul unui computer și al unui *Iphone* cu care navighează în ciberspațiu, acolo unde este bombardat cu informații fără a fi adesea pregătit pentru a discerne între cele viabile și cele manipulatorii sau cu intenții antisociale.

Al doilea principiu: „Internetul este revoluționar, dar nu utopic." Oferind oportunități uriașe de comunicare și documentare, Internetul a devenit o altă realitate. „Pe măsură ce ciberspațiul devine mai populat, seamănă tot mai mult cu societatea în general, în toată complexitatea acesteia."

Într-un eseu din anul 2000, Ștefan Augustin Doinaș scria: „În spațiul virtual încep să se manifeste, cu vehemență, absolut toate infracțiunile pe care omul – singurul animal cu vocația delincvenței – le-a săvârșit, de mii de ani până acum, în spațiul real. Imaginația nu face decât să adauge noi piese la dosarul criminalității noastre: băncile vor fi jefuite, de pildă, nu cu pistolul în mână, ci doar cu degetele pe taste. Internetul va face din cel mai mediocru cap-pătrat un erudit prin simplul acces la banca de date științifice, sedentarul care nu-și va da osânza din fotoliul său de acasă va întreprinde cele mai palpitante călătorii, revoluțiile vor putea fi provocate oricând și de către oricine prin simpla difuzare de informație în ciberspațiu ș.a.m.d."[13] Este suficient să urmărim astăzi știrile despre infracțiunile cibernetice și revoluțiile de pe rețelele de interacțiune socială pentru a avea confirmarea celor spuse de scriitor în urmă cu peste două decenii.

Recent, în perioada pandemiei, o parte importantă a activității profesionale a fost mutată în ciberspațiu, fapt ce a creat atât avantaje (de natură economică), cât și dezavantaje (apariția unor psihoze). Spațiul virtual în care indivizii și-au desfășurat activitatea nu este unul ireal, ci face parte din aceeași viață a lor. Universul virtual nu este unul ireal

[13] Ștefan Augustin Doinaș, „Confesiunea unui analfabet cibernetic", în *Secolul 20*, nr. 4-9 (421-426)/2000, p. 8-10; p. 8.

și nici paralel, ci un univers real care îmbogățește și ameliorează, pe de o parte, și complică și alienează, pe de altă parte, viața individului. În consecință, trebuie tratat ca atare.

Al treilea principiu stabilește că „Guvernul trebuie să joace un rol important la frontiera electronică". Principiul vine în continuarea celui de-al doilea, pentru că ciberspațiul este tot în această lume, chiar dacă există pretenția că ar fi un univers paralel, așa cum îl prezintă Barlow. Tot ce se întâmplă în ciberspațiu afectează viața reală a individului, motiv pentru care, „ca reprezentant al oamenilor și paznic al valorilor democratice, statul are dreptul și responsabilitatea de a ajuta ciberspațiul și societatea tradițională să se completeze reciproc. Standardele tehnologice și problemele puse de protecția vieții private, de exemplu, sunt prea importante pentru a fi încredințate spre rezolvare pieței libere. Firmele de software concurente au un interes redus în păstrarea standardelor deschise, esențiale pentru o rețea interactivă, deplin funcțională. Piețele încurajează inovația, dar nu respectă în mod necesar și interesul public"[14].

Aici am un amendament. La fel ca în viața reală, guvernul trebuie să-i apere pe cetățenii de abuzurile semenilor, dar trebuie să-i apere și de propriile abuzuri. Și mă refer la cazul lui Edward Snowden, cel care a devoalat faptul că guvernul american a utilizat ciberspațiul pentru o supraveghere în masă. În UE, este deja o realitate cu care ne-am obișnuit, aceea creată de aplicarea *General Data Protection Regulation (GDPR)*[15], cu implicații și în ciberspațiu.

Al patrulea principiu: „Informația nu este cunoaștere." Este o veche dispută pe această temă. Filosoful Heraclit din Efes spunea că „mulțimea cunoștințelor nu te face înțelept". Este lesne de observat astăzi că multitudinea de informații cu care sunt bombardați zilnic

[14] *Ibidem*.

[15] Regulamentul (UE) 2016/679 al Parlamentului European și al Consiliului din 27 aprilie 2016 privind protecția persoanelor fizice în ceea ce privește prelucrarea datelor cu caracter personal și privind libera circulație a acestor date și de abrogare a Directivei 95/46/CE, cunoscut și ca GDPR.

oamenii în mediul virtual nu-i face pe aceștia mai buni cunoscători. Autorii manifestului arată că, „indiferent cât de performante devin computerele, nu trebuie să le folosim niciodată ca pe un substitut pentru aptitudinile noastre cognitive de bază ca percepția, conștiința, raționamentul și judecata"[16].

Al cincilea principiu: „conectarea în rețea nu va salva școlile. [...] Arta de a preda nu poate fi reprodusă de computere, de Internet sau de «învățământul la distanță». Aceste instrumente pot, desigur, augmenta o experiență educațională de o înaltă calitate. Dar a te baza pe ele ca pe un soi de panaceu ar fi o greșeală usturătoare"[17].

Încă nu este încheiată disputa pe marginea introducerii învățământului online, odată cu pandemia de COVID-19, apărând pe această temă opinii divergente. Tradiționaliștii susțin că școala nu poate fi înlocuită de noile tehnologii; moderniștii susțin că noile tehnologii sunt viitorul. Consider că adevărul este la mijloc. Varianta clasică de învățământ poate fi îmbogățită cu varianta online.

Al șaselea principiu stabilește că „informația vrea să fie protejată". Pentru că evoluțiile tehnologiei vulnerabilizează drepturile de autor, „trebuie să aducem la zi vechile legi și interpretări, astfel încât informația să fie protejată la fel ca în contextul vechilor mijloace de comunicare. Scopul este același: să le dea autorilor suficient control asupra muncii lor astfel încât să fie stimulați să creeze, în timp ce este menținut dreptul publicului de a utiliza cinstit acea informație".

Cel puțin în România, ne confruntăm cu fenomenul plagiatului, ajuns la cele mai înalte niveluri ale societății. Problema este că tinerii vin de pe băncile școlii cu ideea că pot lua de pe internet informații fără să indice sursa, că tot ce este pe rețele e „la liber", așa cum afirma Barlow. Au apărut site-uri, precum „referate.ro". În 2013, semnalam faptul că la un examen de bacalaureat circula o cărticică ce putea

[16] *Ibidem*, p. 310.

[17] *Ibidem.*

fi ascunsă în palmă, intitulată *Fițuici. Kit de prim ajutor pentru elevi*. Produsul putea fi achiziționat de pe site-ul firmei. Firma îl asigura pe consumator că are în față un „material elaborat conform cerințelor de bacalaureat 2013". „Și, preluând în notă ironică atenționările de pe pachetele de țigări, mai notează că «Folosirea fițuicilor este o fraudă. Nu recomandăm utilizarea lor în timpul examenelor». Păi, dacă nu în examen, atunci când?"[18]. Bineînțeles că în cazul acestui kit nu apare vreun autor care să ceară să i se respecte drepturile. Dar astfel de produse vin să completeze nerespectarea drepturilor de autor în online și să răspândească fraudarea examenelor.

Al șaptelea principiu stabilește că: „Publicul deține frecvențele – publicul trebuie să beneficieze de pe urma utilizării lor. [...] Cetățenii ar trebui să beneficieze și să profite de pe urma utilizării frecvențelor publice și ar fi normal să rețină o parte a spectrului pentru a-l folosi în scopuri educaționale, culturale și de acces public. Ar fi cazul să cerem mai mult de la utilizarea privată a proprietății publice."[19]

În ultimii ani, tot mai multe autorități publice locale au făcut investiții în abonamente la internet, astfel încât, treptat, accesul la ciberspațiu se va generaliza, fiind suportat din bani publici.

În fine, *al optulea principiu* stabilește că „înțelegerea tehnologiei ar trebui să fie o componentă esențială a calității de cetățean al planetei. Într-o lume mânată de fluxul informațional, interfețele și codul care le stă la bază ce fac informația vizibilă devin forțe sociale extrem de puternice. [...] Aceste unelte ne afectează viețile la fel de mult ca legile și ar trebui să le supunem unui scrutin democratic similar"[20].

Lumea occidentală, și nu numai ea, a trecut prin criza pandemiei, când în mediul virtual s-a declanșat un adevărat război astfel încât Comisia

[18] Sorin Bocancea, „Industria copiatului la examene", în *Adevărul,* 11 iunie 2013, disponibil la https://adevarul.ro/blogurile-adevarul/industria-copiatului-la-examene-1455061.html, accesat la 5 octombrie 2022.

[19] *Manifestul tehnorealist.*, loc. cit., p. 310.

[20] *Ibidem.*

Europeană a decretat chiar apariția unei „infodemii", a unui val de dezinformări menite să zădărnicească măsurile de gestionare a crizei și să destabilizeze UE. A urmat agresiunea Federației Ruse asupra Ucrainei, fapt ce a dus la un alt val de dezinformări, motiv pentru care s-a pus problema controlării fluxurilor de dezinformări fără a afecta libertatea de exprimare. Situația este departe de a fi rezolvată, cea mai singură cale de a combate dezinformarea fiind buna informare a utilizatorilor, care, așa cum am văzut, sunt asaltați de informații mai mult decât au timp și dispoziție pentru a discerne.

După cum se poate constata, principiile *Manifestului* rămân valabile, realitatea din ultimele două decenii confirmându-le. Un asemenea document poate fi luat ca reper în elaborarea unei legislații care să reglementeze prezența inteligenței artificiale în viața individului.

Ca stat membru al UE, România are deja un reper în cazul în care ar demara un proces de reformă constituțională în care să introducă și prevederi care să vizeze inteligența artificială. Reperul este *Cartea albă privind inteligența artificială – O abordare europeană axată pe excelență și încredere*[21], iar astfel de prevederi pot să apară la capitolul „Drepturile, libertățile și îndatoririle fundamentale" al Constituției.

[21] Comisia Europeană, *Cartea albă privind inteligența artificială – O abordare europeană axată pe excelență și încredere*, COM(2020) 65 final, Bruxelles, 19.02.2020.

Este necesară o legislație națională pentru domeniul inteligenței artificiale?

Prof. univ. emerit dr. ing. Doru Adrian Pănescu

Sintagma „inteligență artificială" este una des folosită în prezent. La o căutare pe internet, vom găsi aproape o jumătate de milion de răspunsuri, dacă folosim expresia în limba română, și de circa trei ori mai multe răspunsuri în limba engleză. Diferența este normală, inteligența artificială (o prescurtez IA) este nativă de limba engleză. S-a născut în anii '50 ai secolului trecut și a generat mereu multe controverse, iar întrebarea din titlu este justificată, deoarece, așa cum voi explica, IA produce în prezent un nivel serios de îngrijorare, față de care un grad corespunzător de reglementare din partea statului poate fi singura soluție.

1. Definirea inteligenței artificiale și a potențialelor pericole

IA înseamnă studierea și dezvoltarea sistemelor tehnice care manifestă o comportare inteligentă, caracteristici apropiate de cele ale inteligenței umane. Evoluția IA a fost de la IA slabă (în urmă cu 20, 30 de ani), la IA puternică (din limba engleză, unde termenii folosiți sunt *Weak Artificial Intelligence, Strong Artificial Intelligence*). Atributele slab/puternic sugerează gradul de dezvoltare a domeniului, fără a fi

însă etichetele cele mai potrivite[1]. Diferențierea se face mai bine din punctul de vedere al modului de funcționare. IA slabă se referă la sistemele care acționează ca și cum ar fi inteligente (mimează inteligența). Acestea au fost primele sisteme de IA care au apărut. O denumire mai potrivită ar fi însă *IA restrânsă* (în limba engleză există termenul *Narrow Artificial Intelligence*), deoarece este vorba despre sisteme ce rezolvă o sarcină specifică, pentru un domeniu bine individualizat[2].

Chiar dacă este folosit termenul de *IA slabă*, nu înseamnă că performanțele sale sunt reduse; dimpotrivă, ele sunt, de fapt, egale sau adesea superioare operatorului uman din acel domeniu. Exemple în acest sens sunt IBM Deep Blue, sistemul de IA ce joacă șah la cel mai înalt nivel – este specializat strict pentru acest domeniu, doar în privința șahului se manifestă inteligența, sau asistenții software inteligenți, ce au o componentă performantă de IA pentru procesarea limbajului natural, precum sistemele software Apple Siri, Amazon Alexa, Google Assistant. Cealaltă categorie, IA puternică, se referă la sistemele care iau decizii în mod „conștient", nu doar simulează gândirea. În acest caz, este vorba despre rezolvarea de sarcini din diferite domenii, cu capabilități cognitive apropiate de cele umane: percepție, înțelegerea contextului și efectuarea a diferite tipuri de raționament, învățare generalizată, creativitate. Astfel, o denumire mai adecvată ar fi aceea de *IA generală* (în engleză, *Artificial General Intelligence*[3]), deoarece diferența importantă față de *IA slabă* se referă la gradul de generalitate, diversitatea sarcinilor ce pot fi rezolvate.

Asemenea sisteme sunt încă în faza de cercetare. Un exemplu la granița *IA slabă – IA puternică* ar fi robotul Pillo Health, un sistem care poate îndeplini diferite sarcini de asistență sanitară[4]. Se discută și despre o a treia

[1] S. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Pearson, 2016; https://www.professional-ai.com/types-of-ai.html, 2020.

[2] https://www.professional-ai.com/types-of-ai.html, 2020

[3] https://www.professional-ai.com/types-of-ai.html, 2020

[4] C. Recchuto et al., „Cloud Services for Culture Aware Conversation: Socially Assistive Robots and Virtual Assistants", 2020 International Conference on Ubiquitous Robots, 2020, p. 270-277, doi: 10.1109/UR49135.2020.9144750.

categorie, cea care ar ridica problemele cele mai dificile de reglementare, și anume *super IA* (în engleză, *Artificial Super Intelligence*). Aceasta se referă la sistemele care ar depăși în mod clar capabilitățile umane actuale, atât în planul activităților fizice, cât și intelectuale.

Deși pentru această categorie nu există încă realizări concrete, un număr de cercetători sunt convinși că *super IA* va fi atinsă, ceea ce este mai greu de prezis fiind orizontul de timp pentru momentul respectiv. Tot la nivel de speculație se vorbește și de momentul atingerii „singularității", aceasta însemnând momentul în care IA va avea acele capabilități și respectiv autonomie care să-i permită să autogenereze sisteme de IA din ce în ce mai performante[5].

În consecință, temerile formulate încă de pe acum, cu atât mai mult cu cât unele dintre ele fac referire și la pericole deja existente, conform nivelului atins în prezent de IA, sunt de luat în considerare (pentru o tratare completă, trebuie să observăm că există și păreri conform cărora IA nici nu există și, cu atât mai mult, o variantă care să depășească inteligența umană nu se pune în discuție[6]).

În opinia mea, există câteva aspecte-cheie (sunt metode folosite pentru realizarea sistemelor de IA) care au accentuat îngrijorarea față de IA. Acestea sunt: *big data, machine learning, deep learning*, rețele neuronale; am păstrat și denumiri în limba engleză, acestea fiind folosite, în prezent, aproape exclusiv.

Trebuie să înțelegem în ce măsură aceste metode au schimbat abordarea în IA. Fără a intra în detalii, toate se referă la sisteme capabile să învețe din exemple și să generalizeze. Spre ilustrare, rezolvarea unor probleme de clasificare (de exemplu, să fie determinate și separate piesele de diferite forme într-un proces de fabricație sau să fie

[5] P. Boucher, „Artificial intelligence: How does it work, why does it matter, and what can we do about it?", Study Panel for the Future Science and Technology, European Parliamentary Research Service, iunie 2020.

[6] E. Larson, *Mitul inteligenţei artificiale*, Polirom, 2022.

distinși pietonii, cicliștii și autoturismele din trafic pentru un vehicul autonom) se poate face prin metode de tip *machine learning* și rețele neuronale. Este de remarcat felul în care utilizarea tot mai frecventă, în prezent, a metodelor respective este legată de câțiva factori: creșterea puterii de procesare și stocare a informațiilor, dar și folosirea noilor tehnologii bazate pe internet (folosirea cloudului) și creșterea vitezei de comunicare.

Bazat pe toate acestea, așa-numitele seturi de antrenare (datele folosite în faza de învățare, cea care precedă funcționarea propriu-zisă a sistemului de IA) au ajuns la dimensiuni inimaginabile cu unul, două decenii în urmă: texte de trilioane de cuvinte, seturi de miliarde de imagini, milioane de ore de înregistrări audio și video, cantități enorme de date privind genomuri, cantități imense de date din rețelele sociale[7]. Folosirea unor asemenea volume de date pentru învățare a avut un efect favorabil, în privința performanțelor la care au ajuns sistemele de IA, justificând trecerea spre ceea ce am numit anterior IA puternică.

Au apărut însă și efecte negative. Pentru a le înțelege, trebuie precizat un aspect care deosebește cele două nivele – *IA slabă* și, respectiv, *IA puternică*. În cazul *IA slabă*, rezultatul furnizat de sistem este întotdeauna facil de explicat. De exemplu, un sistem expert destinat diagnozei maladiilor infecțioase – Mycin[8], caz de IA tipic pentru nivelul *Weak Artificial Intelligence*, furniza rezultate ce se puteau ușor explica, conform elementelor de cunoaștere preluate de la medici, de cei care au dezvoltat sistemul. La modul principial, într-un asemenea sistem software, concluziile se deduc prin execuția unei succesiuni de reguli (instrucțiuni) de forma *„Dacă condiția... este satisfăcută, atunci deducem concluzia...".*

[7] C. Recchuto et al., art. cit.; Artificial Intelligence Index Report 2021, Stanford University, Human-Centered Artificial Intelligence, 2021; „How to assess Artificial Intelligence Startups", Part II, BlueBull Tech Financial Advisory, 2021.

[8] E.H. Shortliffe, *Computer-Based Medical Consultations: Mycin*, Elsevier, New York, 1976.

Regulile sunt incluse în sistem conform cunoașterii preluate de la experții umani (medici în cazul Mycin) și, atunci când un rezultat obținut nu este corect, se poate ușor determina cauza funcționării greșite: fie a existat o eroare de programare (caz puțin probabil, de obicei eliminat în faza de testare), fie piesa de cunoaștere achiziționată de la expertul uman nu a fost corectă. Astfel, se afirmă despre sistemele de tip *IA slabă* că sunt transparente, se bucură de proprietatea de explicabilitate: atunci când acestea furnizează un rezultat, elementele determinante pot fi ușor de identificat (funcționarea sistemului este explicabilă, relația de cauzalitate se poate pune ușor în evidență, în detaliu). În contrast, la un sistem de tip *IA puternică*, proprietatea de explicabilitate nu mai este satisfăcută; când o rețea neuronală artificială sau un sistem bazat pe *machine learning* a folosit un set de date imens, devine practic imposibil să găsești explicația producerii unui anume rezultat.

În plus, mecanismele folosite în IA puternică pot implica metode de aproximare, o parte statistică. Cu alte cuvinte, un asemenea sistem de IA poate greși. Iar într-o asemenea situație, așa cum am precizat, cauza și respectiv remedierea deficienței devin greu de obținut.

Pentru a avea o imagine completă a potențialelor pericole legate de IA, trebuie introduși în tablou și roboții, cei care includ frecvent componente de IA. Aceștia și-au găsit cea mai semnificativă arie de aplicare în industrie încă din anii '60 ai secolului trecut, scopul inițial fiind acela de a înlocui operatorul uman în activități fizice, solicitante. În pasul următor, ne-am dorit să fim suplinițI și în ceea ce privește partea decizională, deliberativă. Pe de o parte, sunt sarcini decizionale solicitante, din punct de vedere al timpului și al volumului mare de informații de prelucrat, pentru care, dacă folosim doar experții umani, soluția va fi găsită după un interval lung de timp, iar pe de altă parte, fără folosirea sistemului software, soluția s-ar putea să nu fie cea optimă. Rezultă concluzia necesității asistării de către calculator. Astfel, componente decizionale de tip IA au fost îmbarcate pe roboți, dar au fost folosite și de sine

stătător. În prima fază (cea a *IA slabe*), controlul asupra deciziei se păstra în cvasitotalitatea lui la nivel uman.

Odată cu varianta *IA puternică*, controlul decizional este, într-o anumită măsură sau complet, trecut la nivelul sistemului de IA; de exemplu, așa se întâmplă în unele sisteme folosite în mediul financiar, care decid asupra creditării. Un exemplu semnificativ este și cel al vehiculelor autonome, unde decizii care pot avea consecințe asupra vieții oamenilor sunt luate de sistemele de IA. Iar această schimbare (transferul controlului de la om la „mașină") s-a produs pentru că așa am dorit noi. Deci nu este vorba despre literatura SF, în care IA face în mod voit rău omenirii, ci este vorba despre rezultatul modului în care au fost proiectate și al felului în care funcționează sistemele de IA actuale. Evident, o zonă încă mai sensibilă din punctul de vedere al consecințelor este aceea a folosirii IA în domeniul militar. Acest aspect este în sine o problemă distinctă și nu o voi trata aici. În concluzie, felul în care unii sunt îngrijorați de evoluția IA este explicabil.

2. Cât de mare este pericolul inteligenței artificiale

A avea niște sisteme decizionale de tip IA implicate în tratarea unor evenimente critice, așa cum pot apărea la folosirea IA în zona militară, dar și în medicină, vehicule autonome, finanțe (și lista poate continua, practic, într-un viitor apropiat, toate domeniile vor avea partea decizională cuplată cu componente software, multe folosind metode din IA), a căror funcționare nu este în totalitate explicabilă, este un potențial pericol. Și nu sunt puțini cei care și-au exprimat temeri semnificative în acest sens. Iată câteva citate, conform cărora IA este un pericol pentru omenire sau chiar cel mai mare pericol:

Elon Musk: ***„With artificial intelligence we are summoning the demon"*** – „Cu inteligența artificială, noi îl chemăm pe diavol"; aceasta a fost formularea la o conferință din 2014, preluată de *The Washington Post*.

Profesorul Stephen Hawking: ***„The development of artificial intelligence could spell the end of the human race"*** – „Inteligența artificială ar putea însemna sfârșitul rasei umane"; aceasta a fost formularea utilizată într-un interviu acordat BBC în 2014.

Sir Clive Sinclair: ***„Once you start to make machines that are rivaling and surpassing humans with intelligence, it's going to be very difficult for us to survive"***– „Odată ce vom avea sisteme tehnice care vor rivaliza cu inteligența oamenilor și o vor depăși, ne va fi foarte greu să supraviețuim"; aceasta a fost formularea utilizată de Sir Clive Sinclair, cunoscutul inventator britanic, de asemenea la un interviu acordat BBC în anul 2015.

Trebuie făcute câteva observații pe marginea afirmațiilor de mai sus. Dacă citim comentariile complete ale celor trei personalități, lucrurile nu mai apar așa de sumbre. De exemplu, Sir Clive Sinclair remarca faptul că nu are o temere imediată, că momentul în care IA ar putea fi un mare pericol nu este unul iminent. Apoi, mai ales Elon Musk, dar și o mare parte a comunității științifice care a luat în discuție subiectul, au previzionat și mecanismul de îndepărtare a pericolului: o reglementare corespunzătoare a domeniului IA.

Să remarcăm și faptul că afirmațiile de mai sus au fost formulate cu ceva ani în urmă, iar, între timp, viitorul IA nu mai este văzut așa de negativ, exact din motivul enunțat: modul în care IA va evolua depinde de noi, de felul în care vom avea reglementări, o legislație care să nu permită efecte distructive. Aș face o comparație cu domeniul auto. La momentul apariției primelor autovehicule, acestea au părut ceva extravagant, dar și-au câștigat destul de repede mulți adepți, iar apoi progresul a fost unul spectaculos. Odată cu perfecționarea autovehiculelor, acestea au adus cu ele și pericole, iar consecința a fost aceea a necesității introducerii unei legislații adecvate; ce ar însemna azi să avem o circulație fără nicio regulă! Desigur, cu toată legislația, există

încă prea multe accidente de circulație cu consecințe grave; cu toate acestea, nu ne mai putem lipsi de mijloacele de transport.

Putem previziona că lucrurile sunt, într-o anumită măsură, similare și în cazul IA: chiar dacă acest domeniu aduce cu el și pericole, mult mai importante vor fi avantajele, IA poate determina schimbări în bine pentru umanitate. De exemplu, chiar pentru domeniul auto, se consideră că IA poate pune la dispoziție mecanisme importante pentru reducerea pericolelor în circulație. O părere clară despre viitorul favorabil pentru IA este cea exprimată de Eric Horowitz (unul dintre cei mai importanți cercetători de la Microsoft), care crede că putem fi proactivi în privința dezvoltării IA și acest domeniu va determina beneficii incredibile pentru umanitate (ideea a fost exprimată într-un interviu acordat Forbes, în 2015), iar proactivitatea trebuie să cuprindă reglementări pentru IA.

3. Primul palier de intervenție față de potențialele pericole determinate de inteligența artificială – etica

Conform celor spuse, odată cu creșterea performanțelor sistemelor de IA (ca parte bună), s-a ajuns la sisteme a căror funcționare nu mai este în totalitate predictibilă, explicabilă (ca neajuns). Consecința este aceea că trebuie intervenit, iar măsurile care pot reduce riscurile trebuie avute în vedere în toate etapele – proiectarea sistemelor de IA, implementarea și utilizarea acestora. Primul nivel pe care s-a acționat a fost cel etic. Fără să exagerez, există sute de lucrări publicate în ultimii ani pe acest subiect – etica și IA. La modul rezumativ, pe de o parte se poate remarca existența unei relații complexe între IA, etică și lege, drept, iar pe de altă parte, stabilirea unor principii etice ale IA poate fi un element de suport pentru legislația domeniului IA, un punct de plecare, fără însă a elimina necesitatea părții legislative. Dintre multiplele lucrări științifice, recomandări, memorandumuri privind etica și IA am ales câteva exemple.

The UNI Global Union, o organizație reprezentând lucrători din peste 150 de țări[9], cu sediul în Elveția, a adoptat zece principii etice pentru IA: 1) sistemele de IA trebuie să fie transparente, 2) este necesară o echipare a sistemelor de IA cu o „cutie neagră etică", 3) IA trebuie să servească oamenii și planeta, 4) abordarea de adoptat pentru IA este cea în care controlul este la nivel uman, 5) IA nu trebuie să determine vreo discriminare de gen sau de altă natură, 6) beneficiile IA trebuie să fie accesibile tuturor, 7) schimbările produse de IA nu trebuie să afecteze drepturile și libertățile fundamentale ale omului, 8) este necesară stabilirea de mecanisme globale pentru o guvernanță etică a IA, 9) trebuie interzisă atribuirea de responsabilități roboților și 10) trebuie interzisă o cursă a înarmării bazată pe IA.

Ca elemente deosebite, specifice, putem observa ceea ce am subliniat deja – ideea că rezultatele pe care le produc sistemele de IA trebuie să fie controlabile atât *ante*, cât și *post factum*; aceasta se poate obține prin aplicarea principiilor notate cu 1), 2) și 4) mai sus.

Un alt exemplu privind importanța eticii pentru IA este ***Carta etică europeană privind utilizarea inteligenței artificiale în sistemele judiciare*** (*European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*) din anul 2018. Și în cazul acestui document, care se referă la aplicarea IA în domeniul justiției, ca și în general pentru utilizarea IA, putem observa cele două fațete. Pe de o parte, este clar că IA poate crește calitatea rezultatelor din domeniul respectiv, calitatea actului de justiție. Astfel, sistemele de IA pot manipula volume mari de date, iar procedurile de efectuare a raționamentelor, inclusiv în condiții de incertitudine, s-au perfecționat, astfel că asistența pe care aceste sisteme o pot furniza celor care lucrează în justiție mi se pare că va avea un rol din ce în ce mai mare în viitor. Pe de altă parte, realizarea și utilizarea mecanismelor respective

[9] „Top 10 Principles for Ethical Artificial Intelligence", UNI Global Union, Nyon, http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/

trebuie să ia în considerare și să evite orice posibilitate ca un rezultat furnizat de sistemul de IA să fie părtinitor; de exemplu, asta se poate întâmpla când, pentru faza de învățare într-un sistem de tip *machine learning*, se folosesc seturi de date neechilibrate. De aceea, mi se pare semnificativ primul capitol din documentul citat mai sus, cel dedicat respectării drepturilor fundamentale ale omului. Se subliniază în acest capitol că obținerea protecției drepturilor omului față de sistemele de IA presupune ca acest aspect să fie luat în considerare de la momentul proiectării (formulările în limba engleză sunt: *ethical-by-design* și *human-rights-by-design*)[10].

Un ultim exemplu pe care îl comentez este ***Orientări în materie de etică pentru o inteligență artificială fiabilă***, document redactat de Grupul independent de experți la nivel înalt privind inteligența artificială, instituit de Comisia Europeană în iunie 2018 (după publicarea unei forme preliminare, varianta finală a apărut în 2019[11]). Preluând documentul în limba română, așa cum apare în mod oficial, remarc folosirea sintagmei *IA fiabilă*. De fapt, formularea în limba engleză este *trustworthy artificial intelligence* și mi se pare că mai inspirată ar fi varianta *IA de încredere.* Trecând peste această nuanță, documentul formulează niște recomandări privind aplicarea unor principii etice în cazul sistemelor de IA. De remarcat că se vorbește despre un subdomeniu al eticii aplicate, și anume etica inteligenței artificiale; este corectă ideea că aplicarea eticii pentru IA implică aspecte specifice, este o arie care a evoluat și a ajuns acum să aibă un contur de sine stătător. Dintre mecanismele care pot contribui la obținerea unei IA de încredere, și acest document vorbește despre respectarea drepturilor fundamentale ale omului, despre supravegherea umană, despre transparență și explicabilitate.

[10] *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment,* European Commission for the Efficiency of Justice (CEPEJ), Strasbourg, 2018.

[11] *Orientări în materie de etică pentru o inteligenţă artificială fiabilă*, Grupul de experţi la nivel înalt privind inteligenţa artificială, Comisia Europeană, Bruxelles, 2019.

Departe de a fi o prezentare exhaustivă, mai remarc că preocuparea privind stabilirea unor principii etice pentru IA a fost și este una globală. De exemplu, pentru zona Asia putem aminti documentul intitulat *Beijing Artificial Intelligence Principles*[12], iar la nivelul Statelor Unite s-au emis mai multe documente, dintre care amintesc *Principles of Artificial Intelligence Ethics for the Intelligence Community*, elaborat de Office of the Director of National Intelligence, în 2020[13].

Chiar dacă există acest număr considerabil de formulări, se pot observa multe elemente comune: accentul pus pe respectarea drepturilor și libertăților fundamentale ale omului, importanța păstrării unui grad important al controlului la nivel uman (de exemplu, în documentul *Orientări în materie de etică pentru o inteligență artificială fiabilă* elaborat de Grupul independent de experți la nivel înalt privind inteligența artificială sunt puse în evidență trei variante în funcție de implicarea factorului uman: *human-in-the-loop*, care presupune o intervenție umană în fiecare ciclu decizional, *human-on-the-loop*, ce presupune doar o monitorizare a funcționării sistemului de IA, și *human-in-command*, când omul decide când și cum se folosește sistemul de IA), păstrarea transparenței și responsabilității privind funcționarea sistemului de IA, asigurarea siguranței și securității, inclusiv controlul riscurilor, respectarea intimității, reflectarea diversității.

Este ușor de remarcat felul în care apar două tendințe contradictorii: aplicând cerințe etice stricte se reduce autonomia sistemelor de IA (și s-ar putea reduce chiar performanțele în general), respectiv, cu cât acestea sunt mai autonome, cu atât respectarea tuturor criteriilor etice devine mai dificilă. În fine, există o părere cvasiunanimă că, deși palierul etic este necesar și important, acesta nu este suficient, astfel că ajungem la partea legislativă.

[12] *Beijing AI Principles*, Datenschutz und Datensicherheit 43, 656, 2019.

[13] *Principles of Artificial Intelligence Ethics for the Intelligence Community*, Office of the Director of National Intelligence, iulie, 2020.

4. Cum abordăm legislația pentru inteligența artificială și care ar putea fi rolul României

Etica și legea nu sunt sinonime, prin niște principii etice stabilite pentru IA nu se poate substitui legislația. A ne opri la nivelul etic ar fi o variantă incompletă. Printre cei care observă acest lucru îl putem aminti pe Ben Wagner, director al AI Futures Lab de la Universitatea Tehnică Delft[14]. Acesta spune că etica este o opțiune „ușoară" sau „soft", care poate ajuta, da sens inițiativelor de autoreglementare existente, dar nu ne putem baza doar pe autoreglare. Simplificând și rezumând explicația, bazându-ne doar pe principii etice ni se pare că am rezolvat problema, dar, de fapt, nu va exista o obligație clară, vreo responsabilitate în a respecta aspectele de etică. În plus, dacă pe partea de etică se ajunge mai ușor la soluții, la consens, pe partea legislativă este mai greu, apar dificultăți. Aici se poate face o comparație – IA versus domeniul nuclear, plecând de la unul dintre cele mai sensibile aspecte asupra căruia IA poate avea o influență majoră: armele letale autonome, cele care, în fapt, sunt sisteme de IA; de altfel, într-un discurs din 2018, Elon Musk vedea IA pe zona militară mai periculoasă decât armele nucleare. Dacă avem în vedere dificultățile, lipsa de consens în privința domeniului nuclear (s-a născut ușor legislația privind domeniul nuclear, este o chestiune simplă, este unanim respectată azi?), atunci, mutatis mutandis, ne putem aștepta ca legislația privind IA să fie dificil de conceput și respectat.

Comparația între IA și cazul folosirii energiei nucleare merită încă o remarcă. O problemă este aceea a accesului la tehnologiile de IA, la fel cum s-a pus problema accesului la tehnologiile nucleare. Pe de o parte, fiecare stat își dorește să-și protejeze produsele, tehnologiile de IA, din motive economice, politice, de securitate (spre comparație, a se vedea adevăratul război UE, SUA versus China privind tehnologiile 5G); pe de altă parte, legislația privind IA ar trebui să asigure accesul cât mai larg la noile tehnologii, în anumite condiții. Aici, ca model am

[14] B. Wagner, „Ethics As an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?", în M. Hildebrandt, (Ed.), *Being profiled. Cogitas ergo sum*, Amsterdam University Press, 2018.

putea avea Tratatul pentru neproliferarea armelor nucleare, în care, pe lângă partea militară, se precizează că toate părțile trebuie să participe la „un schimb larg de echipamente, materiale și informații tehnologice și științifice privind folosirea pașnică a energiei nucleare"[15]; la fel ar trebui să se întâmple și în privința accesului la tehnologiile de IA.

Cu toate dificultățile, există totuși, mai ales în ultima perioadă, preocupări clare pentru legiferarea în IA. Trebuie remarcat faptul că legislația pentru IA nu este creată de la zero. Există mai multe puncte de plecare, dintre care putem aminti (pe lângă posibila inspirație din domeniul energiei nucleare, cea remarcată mai sus): legislația privind protecția datelor cu caracter personal; legislația privind protecția drepturilor de autor; legislația privind siguranța produselor; legislația privind protecția muncii; legislația privind relațiile dintre companii și beneficiari. Chiar dacă este în funcțiune această legislație, trebuie să observăm că ea nu acoperă toate aspectele, toate riscurile determinate de IA.

O întrebare la care trebuie dat un răspuns este și aceea cu privire la caracterul național sau internațional al legislației privind IA. În acest sens, este clar că IA are un caracter global. Nu putem sau este foarte greu să restrângem sfera de acțiune a IA; drept comparație, chiar dacă nu e imposibil, a îngrădi utilizarea IA ar fi ca și cum am restrânge folosirea internetului. Datorită caracterului global, rolul legislației internaționale pentru domeniul IA va fi pregnant, primordial. Iar în acest sens, la nivelul Uniunii Europene se întreprind acțiuni (a se vedea, de exemplu, *The Artificial Intelligence Act and emerging EU digital acquis*[16]). Lucrurile sunt în derulare și, de exemplu, în privința documentului amintit, există o previziune că un acord ar putea fi atins într-un viitor apropiat. Totodată, se remarcă o dependență de fiecare membru al UE, adică, citez: „Un acord pare posibil până la jumătatea anului 2023, dar acest lucru va depinde de convergența co-legislatorilor asupra unor aspecte cheie,

[15] Treaty on the Non-Proliferation of Nuclear Weapons, United Nations Office for Disarmament Affairs, https://www.un.org/disarmament/wmd/nuclear/npt/text/.

[16] A. Bogucki, A. Engler, C. Perarnaud, A. Renda, „The AI Act and emerging EU digital acquis", ceps.eu, 2022.

cum ar fi definiția IA, clasificarea riscurilor și măsurile de reglementare asociate, măsurile administrative și normele de aplicare". Rezultă implicit importanța stabilirii unor reglementări la nivel național.

Altfel spus, chiar dacă va exista o legislație internațională, asta nu înseamnă că nu este posibilă și o abordare de jos în sus, și, oricum, va fi necesară o legislație națională. Se mai pot aduce câteva argumente de susținere a importanței legislației pentru IA la nivel național. Cum am încercat să sugerez, există și un scenariu cu un progres lent al legislației de nivel internațional, și atunci o abordare plecând de la nivel național poate fi utilă și importantă. De asemenea, în mecanismele de IA bazate pe date (IA puternică) se lucrează cu seturi de date specifice, din fiecare țară; apar astfel aspecte particulare, existente în România, de care se poate ține seama doar printr-o legislație națională. În fine, din anumite puncte de vedere, domeniul IT și în particular cercetarea și folosirea IA la noi în țară sunt la un nivel avansat. Cum România are printre cei mai buni programatori, ar fi firesc să aibă și printre cele mai bune sisteme legislative pentru acest domeniu. În consecință, țara noastră ar putea fi un model în privința legislației pentru IA, iar avantajele ar fi multiple; menționez doar unul: încrederea în produsele IT ale României ar fi întărită de o asemenea legislație; idei de susținere în acest sens apar în documentele unor organisme ale UE, de exemplu[17]. Ca o concluzie, legislația națională nu poate să lipsească și este important unde se va situa România.

Un punct de vedere bine argumentat privind importanța legislației cu privire la IA, ce înseamnă partea de etică și posibilii pași pentru stabilirea legislației privind IA este prezentat de M. Robles Carrillo în studiul „Artificial intelligence: From ethics to law"[18]. Este de remarcat una dintre concluziile studiului respectiv, și anume, necesitatea de a crea un organism internațional dedicat legiferării în domeniul IA. Iar previziunea

[17] *Orientări în materie de etică pentru o inteligenţă artificială fiabilă*, Grupul de experţi la nivel înalt privind inteligenţa artificială, Comisia Europeană, Bruxelles, 2019.

[18] M. Robles Carrillo, „Artificial intelligence: From ethics to law", Telecommunications Policy, 44, 2020, doi: 10.1016/j.telpol.2020.101937.

că un asemenea organism va fi creat într-un viitor nu prea îndepărtat pare suficient de realistă. Ideea pe care o lansez este aceea ca, la momentul oportun, România să fie printre inițiatorii creării acestui organism și, de ce nu, să fie avansată oferta ca acesta să fie situat în România. Fără a intra în detalii, chiar dacă niște ținte pentru găzduirea unor organisme internaționale în țara noastră au fost ratate în trecut, aceasta nu înseamnă că, *a priori,* un asemenea demers este inutil, lipsit de șansă. Astfel, sunt unele puncte de susținere în acest sens: dezvoltarea domeniului IA în Romania, faptul că țara noastră este membră a UE, dar nu este membră OECD (e doar parteneră), iar între cele două organisme există și divergențe pe subiectul legislației privind IA, astfel că România ar putea fi considerată un partener de negociere, de arbitrare a controverselor.

Evident, pentru atingerea unui asemenea obiectiv trebuie făcute pregătiri. Menționez câteva acțiuni, legate de mediul universitar, cel pe care îl cunosc mai bine; desigur, va fi necesar și un sprijin la nivelul Guvernului și al Parlamentului. Realizarea legislației privind IA este o problemă complexă, interdisciplinară. În consecință, un atu important ar fi acela al existenței unui grup de experți pentru realizarea legislației în domeniul IA. Chiar dacă am afirmat că pentru IA există specialiști în România (cel puțin în marile centre universitare, atât pe plan didactic, cât și în privința cercetării în IA există preocupări, o tradiție), aceasta nu este suficient. Este necesară o specializare mai focalizată pe subiectul legislației pentru IA. Iar pasul care se poate face în acest sens este acela al organizării unui masterat interdisciplinar pe domeniul legislației privind IA. Nu ar fi ceva ieșit din comun.

Există asemenea masterate, dintre care îl menționez pe cel intitulat *European Master in Law, Data and AI*, organizat de un consorțiu format din Dublin City University, Avignon Université, Universidad de Léon, Università di Pisa, care are printre obiective instruirea de specialiști pentru legislația în materie de IA[19]. Dacă există acest masterat și are audiență,

[19] https://www.dcu.ie/courses/postgraduate/school-law-and-government/european-master-law-data-and-artificial-intelligence

înseamnă că s-a identificat o asemenea necesitate și putem presupune că și acum, dar mai ales în viitorul apropiat, în domeniul legislației privind IA vor exista locuri de muncă. Este de interes să remarcăm cine sunt candidații la acest masterat, care sunt cerințele impuse la admitere: sunt două ramuri (în engleză, *stream*): una pentru absolvenții cursurilor de licență (*undergraduate*) pe domeniul Calculatoare, Informatică, iar a doua ramură pentru absolvenții facultăților de drept sau o altă calificare ce include o componentă semnificativă în domeniul juridic (Științe politice, Relații internaționale).

Devine retorică întrebarea dacă un asemenea masterat ar avea căutare în România; până la urmă, nu este vorba doar de legislația în domeniul IA, ci de felul în care absolvenții unui asemenea masterat vor putea trata aspecte legate de securitatea și managementul datelor, GDPR, confidențialitate în domeniul sistemelor software. Câteva cursuri din programul de masterat european menționat sunt[20]: *European and International Human Rights Law*, *Artificial Intelligence*, *Information and Information Seeking*, *EU Data Protection Law*, *Contemporary Legal Issues in Commercial Practice*, și, bineînțeles, un curs general, intitulat *Introduction to Law, Data and AI*. Este clar, conform acestei enumerări, că ar fi destule companii în România interesate să angajeze absolvenți care au urmat asemenea cursuri. Alte activități ce pot contribui la formarea specialiștilor necesari realizării legislației privind IA pot fi organizarea de școli de vară și cursuri opționale în cadrul unor programe de licență și masterat existente, din zona ingineriei, informaticii sau a dreptului.

5. Exemple care justifică necesitatea legislației în inteligența artificială

Pentru a nu rămâne la un nivel abstract, teoretic, este util să punem în discuție situații concrete, cu care fie ne confruntăm astăzi, fie sunt foarte aproape de a pătrunde în viața de zi cu zi și pentru care legislația

[20] https://www.dcu.ie/courses/postgraduate/school-law-and-government/european-master-law-data-and-artificial-intelligence

în domeniul IA este deja necesară. Exemplul cel mai des citat este cel al vehiculelor autonome. Pentru IA, acestea sunt o provocare, deoarece reprezintă primii agenți inteligenți (sisteme software) ale căror decizii pot avea consecințe privind viața oamenilor.

Problema des abordată în literatura de specialitate este așa-numita „dilemă a troleului"[21]. Reformulată, într-o variantă mai ușor de înțeles și comentat, problema ar fi cum să acționeze un vehicul autonom în situația iminenței unui accident: a) să salveze viața pasagerului din mașina respectivă, situație care poate conduce la decesul mai multor persoane din celelalte mașini implicate; b) să salveze cât mai multe persoane, chiar cu riscul ca pasagerul din mașina respectivă să-și piardă viața. Aceasta este încă o problemă deschisă, soluția implicând tehnici de optimizare și sisteme decizionale care să lucreze în timp real, fiind totodată și o problemă din punct de vedere juridic.

Exemple privind necesitatea legislației pentru IA se pot da și în domeniul roboticii, în ceea ce privește aplicarea IA în robotică. Aceste două arii – robotica și IA – au fost și, mai ales în prezent, sunt corelate. Astfel, cum am mai remarcat, primii roboți au fost folosiți în aplicații industriale, efectuând sarcini repetitive, solicitante din punctul de vedere al efortului fizic. Ulterior, s-a dorit ca roboții să poată lucra cât mai mult autonom, să se poată adapta la schimbările din mediu, ceea ce a presupus îmbarcarea IA pe roboți. Dacă în privința IA am remarcat o evoluție atât din punct de vedere cronologic, cât și în privința performanțelor – *IA slabă* și *IA puternică* –, tot așa, și la nivelul roboților putem vorbi în prezent de două generații: mai întâi au fost roboții industriali tradiționali, iar acum este vremea coboților; aceste două variante sunt prezentate succint în continuare.

[21] A., Wolkenstein, „What has the Trolley Dilemma ever done for us (and what will it do in the future)? On some recent debates about the ethics of self-driving cars", *Ethics and Information Technology*, 20, p. 163-173, 2018.

Fig. 1. Robot industrial clasic în laboratorul Facultății de Automatică și Calculatoare a Universității Tehnice „Gheorghe Asachi" din Iași

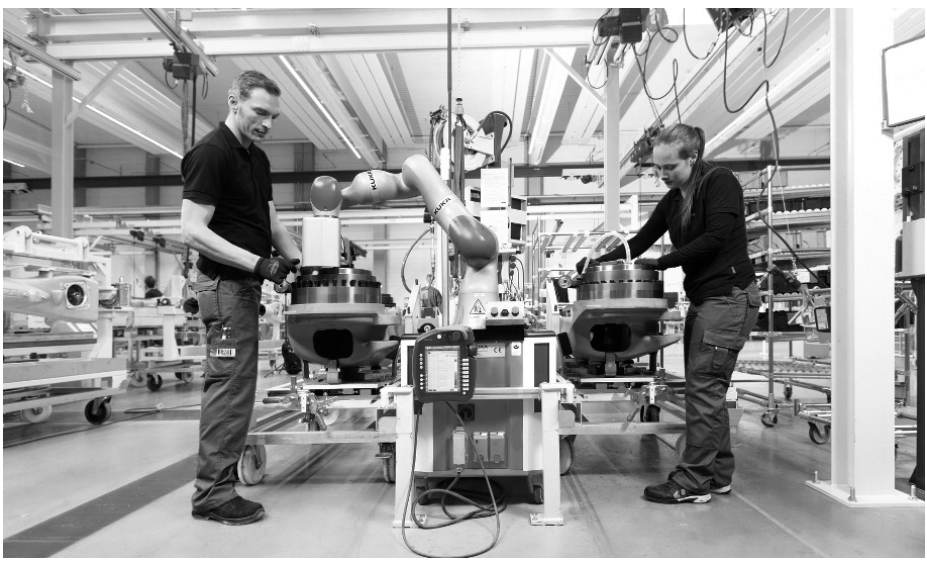

Fig. 2. Cobot lucrând alături de operatorii umani într-o aplicație de robotică colaborativă

În Fig. 1 apare un robot industrial tradițional; nota bene, cel din imagine este primul robot industrial ABB (produs de compania suedeză ABB) importat de România la sfârșitul anilor '90 și care se găsește într-unul dintre laboratoarele Facultății de Automatică și Calculatoare de la Universitatea Tehnică „Gheorghe Asachi" din Iași. Un asemenea robot cântărește peste 200 kg și, în deplasarea brațului robotului cu mare viteză, dacă ar lovi un operator uman în timpul ciclului de lucru, accidentul ar putea fi unul serios. În consecință, la acești roboți, conform regulilor de protecție a muncii, operatorul nu trebuie să intre în zona de operare a robotului, în timpul ciclului de lucru. De exemplu, în situația prezentată în Fig. 1 există niște garduri despărțitoare, pe care studenții ce folosesc robotul nu trebuie să le depășească, atunci când acesta execută o sarcină de lucru. Situația este total schimbată în cazul noilor roboți, adică în cazul coboților. Denumirea provine de la *collaborative robots*, aceștia fiind destinați unor activități în care operatorii să poată lucra alături de roboți, pentru a colabora. Prin perfecționarea părții mecanice și prin cuplarea cu componente de IA (sisteme performante decizionale, de control și planificare a traiectoriei, senzori avansați, inclusiv sisteme de vedere artificială), coboții pot lucra în vecinătatea oamenilor, fără să mai reprezinte un pericol, ca în exemplul prezentat în Fig. 2. Astfel, anumite sarcini tehnologice pot fi rezolvate mai eficient, dacă operatorii colaborează cu coboții.

În cazul lucrului cu coboți, accidentele sunt evitate datorită elementelor mai sus menționate – senzori și sisteme de control avansate. Aceasta nu înseamnă însă că accidentele sunt excluse; în anul 2016, conform BBC[22], un cobot a accidentat un copil. A fost un caz izolat, implicând un cobot folosit într-o aplicație neindustrială; totuși, asemenea incidente ar putea apărea în diferite aplicații ale coboților. Problema este aceea că, la producerea unui asemenea eveniment, trebuie găsită cauza, vinovatul. În cazul roboților industriali clasici, nu pot fi elemente de dubiu: singura situație în care apare un accident este cea în care operatorul

[22] https://www.bbc.com/news/technology-36793790

nu a respectat regulile de protecție. În contrast, în cazul unor activități cu coboți lucrurile sunt sensibile, mai greu de detectat, deoarece pot fi trei surse care să conducă la un incident: sau există o eroare în sistemul software cu care a venit cobotul de la producător, sau a greșit cel care a scris aplicația software pentru situația particulară în care este implicat cobotul, conform problemei de rezolvat și mediului în care a fost acesta integrat, sau poate fi de vină operatorul uman cu care colaborează cobotul, pentru că nu a respectat întru totul protocolul de lucru. Este clar că, pe de o parte, este necesară o legislație care să țină cont de aceste trei situații posibile, iar, pe de altă parte, sistemul software trebuie să asigure transparența pentru găsirea vinovatului. Legislația privind roboții dotați cu IA are, în prezent, și unele aspecte distincte, mai degrabă cu un caracter exotic. Astfel, în 2017, robotul Sophia a primit cetățenie în Arabia Saudită, eveniment ce a stârnit controverse[23]. Deocamdată nu există pericole din partea roboților, de tipul celor ce apar în literatura SF, cu privire la revendicări ale acestora. Linia roșie se va depăși atunci când vom avea hibrizi robot-om, adică proteze computaționale conectate la creierul uman.

6. Concluzii

Fără a avea pretenția unei prezentări exhaustive a subiectului legislației privind IA, am identificat câteva aspecte importante pentru evoluția acestuia; în fapt, conform preocupărilor actuale, este clar că reglementarea domeniului IA se va înfăptui odată cu lărgirea sferei de aplicare a IA și cu creșterea impactului acesteia asupra vieții noastre. În opinia mea, cele mai importante aspecte privind legiferarea în domeniul IA (în ordinea priorităților) ar fi:

a) apărarea drepturilor fundamentale ale omului (intervenția poate fi inclusiv la nivel constituțional);

[23] https://www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics

b) stabilirea statutului legal al sistemelor de IA (cui îi revine responsabilitatea în cazul unui incident);

c) relația om – IA.

Doar o abordare interdisciplinară poate rezolva aceste probleme. Ceea ce vreau să subliniez este faptul că nu toată responsabilitatea este în domeniul juridic. Sunt destule aspecte tehnice ce trebuie rezolvate. Proprietatea de transparență, explicabilitatea acțiunilor, a rezultatelor pe care le furnizează un sistem de IA, asigurarea controlabilității la nivel uman sunt caracteristici care trebuie standardizate și reglementate de tehnicienii care proiectează, implementează și utilizează sistemele de IA. Putem remarca felul în care acest proces a fost demarat; argumente în acest sens sunt noile discipline de studiu apărute, precum Agenți inteligenți morali (*Artificial moral agents*) și Robotica responsabilă. Iar marea majoritate a cercetătorilor din domeniul IA urmăresc acele abordări care să maximizeze beneficiile, ceea ce se poate întâmpla prin colaborarea om – IA, om – robot. Un termen recent apărut în acest sens în literatura de specialitate este acela de inteligență artificială complementară (în engleză, *complementary artificial intelligence*[24]), ceea ce înseamnă că IA nu este într-o competiție cu noi, ci trebuie luate în considerare mecanismele prin care capabilitățile umane și ale IA să fie folosite în mod complementar.

Despre roboți, despre IA s-au scris multe povești. Într-o lume de poveste, există cele două tipuri de personaje – pozitive și negative. Într-un fel, cei care lucrează în domeniul IA scriu acum „o poveste" și trebuie să decidem ce personifică inteligența artificială: este zâna bună (conform beneficiilor pe care le aduce) sau vrăjitoarea cea rea (prin pericolele pe care le creează)? O legislație adecvată poate elimina vrăjitoarea din poveste!

[24] J. Sourati, J. Evans, „Complementary artificial intelligence designed to augment human discovery", arXiv preprint arXiv:2207.00902, 2022.

Premisele fundamentale ale constituționalismului și provocările inteligenței artificiale

Conf. dr. Marius Bălan,
Facultatea de Drept, Universitatea „Alexandru Ioan Cuza" din Iași

Reprezintă oare inteligența artificială un pericol pentru ordinea constituțională și pentru libertatea individuală? Fără îndoială că da. Nu pentru că ar constitui un rău intrinsec, ci pentru că libertatea este în permanență amenințată. În celebra formulare a lui Ronald Reagan, din discursul inaugural rostit la preluarea funcției de guvernator al statului California, „libertatea este un lucru fragil și nu este niciodată la mai mult de o generație distanță de propria extincție".[1] Nu mă voi pronunța aici în privința aspectelor tehnice ale inteligenței artificiale – în privința cărora, ca jurist, sunt aproape cu totul ignorant –, ci asupra consecințelor practice, foarte vizibile ale acesteia în privința raporturilor de putere.

Cum se raportează individul uman la inteligența artificială? Dar la cunoaștere în general și, mai ales, la consecințele foarte palpabile ale utilizării acesteia în sfera raporturilor de putere? Din perspectiva

[1] „Freedom is a fragile thing and it's never more than one generation away from extinction." Ronald Reagan, *Inaugural Address (Public Ceremony)*, 5 ianuarie, 1967, text accesibil la: https://www.reaganlibrary.gov/archives/speech/january-5-1967-inaugural-address-public-ceremony.

dreptului constituțional, orice raporturi de putere – inclusiv cele care derivă din instrumentalizarea cunoașterii – sunt tratate prin prisma drepturilor și libertăților fundamentale. Aici există un consens nominal – ne înțelegem în privința cuvintelor, dar nu întotdeauna și în privința conținutului acestora. Se pune mai întâi problema, cum anume ne raportăm la cunoaștere și la puterea dobândită prin cunoaștere. Am să încep evocând o dimensiune mitică a acestei probleme. Două teme au marcat mitologia europeană (și nu numai) a secolelor trecute în această privință: cea a pietrei filosofale și cea a lui „homunculus". Piatra filosofală, în diversele variante ale mitului, este o substanță – extrem de greu sau aproape imposibil de sintetizat – care oferă celui care o obține, nu fără eforturi inimaginabile, tinerețe veșnică, viață veșnică, bogăție și putere.

În variantele mai „materialiste", aceasta transformă mercurul (sau alte substanțe) în aur. Homunculus reprezintă o repetare, la scară redusă, a Creației. Omul ajuns la nivelul de cunoaștere absolută poate face orice, poate produce chiar și miracole, și pentru că încununarea Creației o reprezintă însăși facerea omului, atunci poate crea și el ființe umane. Modestia ne îndeamnă să credem că această creație derivată produce un om de dimensiuni mai mici: homunculus. Mitul acesta, al producerii unui om artificial, a fost pe larg tematizat în literatura europeană: în romanul *Frankenstein* al lui Mary Shelley, în partea a doua a lui *Faust* sau în mitul Golemului, relansat în secolul XX de Gustav Meyrink. Cunoașterea absolută face deci posibilă crearea unei ființe (cvasi) umane. Această dimensiune mitică ne determină în mare măsură perceperea inteligenței artificiale, după cum este vizibil în filmografia contemporană. În „Matrix" sau în „Star Trek", omenirea este amenințată de „borgi" sau de rețele cibernetice scăpate de sub control, care au dobândit autonomie și voință proprie. Nu putem concepe răul decât ca pe o ființă autonomă, dispunând de voință și liber arbitru, sau, în orice caz, doar o asemenea imagine este de natură să impresioneze publicul, să facă o narațiune atractivă și să asigure succesul unei producții cinematografice.

Deja s-a stabilit aici, de către antevorbitori, că inteligența artificială oferă cunoaștere, știință și – prin aceasta – putere (pentru cei care o controlează), dar în niciun caz voință, liber arbitru și putere decizională

autonomă. Pericolul decurge din reaua voință a utilizatorilor. Drepturile omului sunt în pericol, dar nu din partea inteligenței artificiale, ci, mai degrabă, din partea prostiei naturale, a înclinației umane, perenă și omniprezentă, spre abuz a utilizatorilor săi. Modul în care operăm cu cunoașterea este problema.

Câteva probleme relativ recente: un rabin conservator american, Denis Prager, cunoscut pentru o serie de conferințe în care popularizează valorile iudaismului tradițional, a fost cenzurat pe youtube, în sensul că cel puțin unul dintre clipurile destinate procesului didactic a fost oprit de la distribuție în rețelele instituțiilor de învățământ americane. În acel clip, dedicat popularizării celor zece porunci, el a ilustrat prima poruncă („Eu sunt Domnul Dumnezeul tău: să nu ai alți dumnezei afară de Mine") cu o imagine reprezentând figuri stilizate ale unor oameni, purtând semnul zvasticii. Era evidentă trimiterea la o religie politică, la idolatrizarea unei ideologii, la consecințele posibile ale ideii de a ne închina altor valori decât lui Dumnezeu. Motivul restricției – operate în baza unui algoritm netransparent – era, evident, utilizarea unui simbol nazist. Cel puțin la prima vedere, algoritmul a fost atât de stupid, încât nu a putut face distincția între utilizarea unui simbol în scop de propagandă național-socialistă efectivă și prezentarea aceluiași simbol într-un context negativ.[2] Aceasta cu atât mai mult cu cât în toate filmele documentare sau artistice consacrate epocii național-socialiste sau Celui de-Al Doilea Război Mondial, simbolurile naziste sunt prezentate supraabundent (și fără obiecții din partea publicului larg).

Un alt aspect: platformele de socializare – facebook, în primul rând – beneficiază de privilegiile unor simple platforme, ale unor instrumente de comunicare pur tehnice, neutre față de conținutul lor, pentru care nu răspund, spre deosebire de o publicație online, un ziar sau o revistă. Acestea trebuie să-și asume răspunderea pentru conținutul lor, pentru ceea ce publică, chiar dacă răspunde în primul rând autorul. În subsidiar

[2] A se vedea audierile privind cenzura în rețelele sociale, din 19 iulie 2019, în fața Subcomisiei pentru Justiție (*Judiciary Subcommittee*) a Senatului SUA, cu depoziția lui Denis Prager. Înregistrarea video este accesibilă la https://www.judiciary.senate.gov/meetings/google-and-censorship-though-search-engines, precum și la https://www.youtube.com/watch?v=nO2hOe61yeM.

va răspunde și editorul sau proprietarul publicației. Proprietarii platformelor social media sunt scutiți de orice răspundere în privința actelor incitate sau provocate de către diverși utilizatori, dar au dreptul să cenzureze conținutul, în sensul că pot limita accesul și bloca sau șterge anumite postări ale utilizatorilor platformei. Spre deosebire de situația din presă și audiovizual, aici deținătorii unor puternice surse de informare, dezinformare și modelare a opiniei publice nu răspund pentru conținutul mesajelor postate, pe care nu și le asumă, dar își rezervă dreptul de a le cenzura pe baza unor criterii stabilite de ei înșiși.[3] De unde vine această putere? Avem de-a face cu raporturile clasice de putere, unde o mare corporație își creează un lobby și-și poate permite să controleze sau să influențeze în destul de mare măsură legiuitorul.

Rezultatele nedorite sau chiar periculoase ale unor asemenea evoluții nu sunt datorate însă inteligenței artificiale ca atare și nici neapărat răutății, lăcomiei sau setei de putere a unor actori umani (fie ei și sub forma unor corporații lipsite de scrupule). Consecințele negative sunt de multe ori determinate de imperativele unor optimizări tehnologice – care au, evident, și dimensiuni economice și politice – care impun anumite soluții, în folosul unor obiective punctuale dorite de guvernanți, cu prețul bulversării relațiilor interumane care constituie premisa respectării demnității umane și a libertății individuale. Sacrificiile pe seama acestor valori (privite uneori ca inactuale sau depășite) pot fi ușor justificate retoric ca fiind prețul plătit de omenire pentru progresul tehnic și științific.

[3] Conform prevederilor legislației federale americane, niciun operator de internet (furnizor și utilizator de „servicii interactive de computer") nu va fi tratat ca „publicist" (*publisher*) sau autor al discursului (speaker) în privința oricărei informații furnizate de un alt furnizor de conținut. Pe de altă parte, cei dintâi nu răspund pentru nicio acțiune adoptată în mod deliberat și cu bună-credință pentru a restrânge, bloca accesul sau a restrânge accesibilitatea materialului pe care aceștia îl consideră obscen, lasciv, senzual, excesiv de violent, hărțuitor sau în alt mod repugnant, indiferent dacă acesta este sau nu protejat constituțional. Vezi „*§230. Protection for private blocking and screening of offensive material*", litera c) (47 U.S.C.), accesibil online la: https://uscode.house.gov/view.xhtml?req=(title:47%20section:230%20edition:prelim).

Aș face o scurtă referință la un roman citit cu decenii în urmă, *Bombardierul*, de Len Deighton.[4] Acțiunea se petrece în 1943, la data fictivă de 31 iunie, și se concentrează în jurul unei misiuni de bombardament masiv a aviației britanice, care, cu mari pierderi, distruge orașul, de asemenea fictiv, Altenberg, situat în zona industrială a Văii Ruhr-ului și confundat cu obiectivul real al misiunii, orașul Krefeld. Cu alte cuvinte, orașul fusese nimicit în mod inutil. Personajul principal, pilotul Sam Lambert îi atrage atenția unui prieten, mecanicul Worthington, că „e caraghios […] să vopsești pe avion numărul de misiuni făcute. Niciunul dintre numerele însemnate pe avionul nostru nu coincide cu numărul de misiuni în care a zburat echipajul respectiv, ci doar zborurile făcute de aparatul propriu-zis […] Am uneori impresia că mașinile nemțești se bat cu cele englezești, și atât."[5] Într-o discuție aparent banală și într-o formă ușor accesibilă este evidențiată subordonarea acțiunii umane față de imperativele născute din considerente de ordin tehnic. Acest proces are cumva un caracter demonic sau titanic.

Tema a fost aprofundată într-un eseu din 1932 al lui Ernst Jünger, *Der Arbeiter* (Muncitorul)[6]. În imaginarul colectiv, zeii, ființe luminoase, senine, spirituale sunt înlocuiți cu titanii, ființe supraumane, gigantice dispunând de o forță imensă și de cunoaștere, dar care au ceva inerent malefic. Acest caracter se vede în dependența noastră față de tehnică. În prima conflagrație mondială, războiul submarin a fost deosebit de crud și de devastator, nu din cauza educației greșite sau a cruzimii înnăscute a germanilor, ci datorită posibilităților tehnice pe care le oferea un submarin de a întrerupe sau de a suprima comunicațiile adversarului, precum și constrângerilor – tot de ordin tehnic – care făceau imposibilă salvarea vieții naufragiaților, pe spațiul absolut limitat al submarinului. În același fel, bombardamentele devastatoare din Al Doilea Război Mondial nu sunt datorate unei cruzimi inerente

[4] Len Deighton, *Bombardierul*, traducere de Daniela Parau-Staicu, Meridiane, București, 1982.

[5] *Ibidem*, vol. 1, p. 129.

[6] Ernst Jünger, *Der Arbeiter. Herrschaft und Gestalt*, Klett-Cotta, Stuttgart, 2013 (prima ediție Hanseatische Verlagsanstalt, Hamburg, 1932).

a germanilor, britanicilor sau americanilor, ci sunt explicabile prin posibilitățile tehnice și oportunitatea strategică a utilizării unor astfel de mijloace de luptă. Acest lucru este valabil cu atât mai mult în privința bombei atomice. Caracterul „demonic" se manifestă prin aceea că puterea care apare și se manifestă va fi inevitabil utilizată în cele din urmă de către cineva, ca în regula nescrisă a scenariilor din teatru: dacă în actul întâi apare pe scenă o armă de foc, în mod sigur, în actul trei, cineva o va folosi. Problema este dependența noastră de tehnică, conformarea treptată a acțiunii colective și a conduitei individuale unor criterii de optimizare artificiale și unilaterale, și nu neapărat pericolul acesta mitic, imaginar, al unei inteligențe artificiale personificate, care ar prelua controlul asupra omenirii și asupra lumii.

Un alt aspect care pune paradigma de gândire a drepturilor omului constă în obedienta idolatră față de știință și tehnică. Sunt de invocat aici două referințe, cunoscute mai mult sau mai puțin, în funcție de vârstă. Cea cunoscută de către toți este mitul anilor 2000. În copilăria generației mele, televiziunea era saturată de emisiuni și cântece optimiste despre ceea ce vom face, „noi în anul 2000". Mă mai gândesc la filmul lui Stanley Kubrick, „2001: Odiseea spațială", an care urma să apară, din perspectiva de atunci, ca absolut miraculos. Erau avute în vedere întâlnirea cu o civilizație extraterestră, performanțele inteligenței artificiale, dezvoltarea explozivă a științei, capabilă de a ne rezolva toate problemele. Apare și aici tema preluării controlului unei misiuni spațiale de către un computer, care elimină – la nevoie chiar fizic – din procesul decizional ființa umană.

A doua referință, azi mai puțin amintită, privește Congresul al XXII-lea al PCUS din 17-30 noiembrie 1961.[7] Pentru ultima dată, partidul a promis „solemn" că „actuala generație de oameni sovietici va trăi în comunism". Pornind (și) de la succesele eclatante ale primilor ani ai programului spațial sovietic – primul satelit artificial din 1957, fotografierea părții

[7] A se vedea documentele congresului: *The Road to Communism: Documents of the 22nd Congress of the Communist Party of the Soviet Union*, 17-31 octombrie, 1961, Foreign Languages Publishing House, Moscova, în special p. 187 și urm.

invizibile a lunii, lansarea primului om în cosmos –, liderii comuniști prevedeau că în anul 1970 Statele Unite urmau să fie depășite din punct de vedere economic, în 1980 toți sau aproape toți cetățenii urmau să dispună de automobile, pentru a nu mai aminti ce miracole urmau să se producă în anul 2000. Acest optimism de factură științifico-fantastică – alimentat și de o campanie intensă în mass-media, literatură, cinematografie, cultură și învățământ – era foarte strâns legat de optimismul canonic și imperativ al doctrinei și propagandei comuniste. Ne aflăm la puțină vreme după Sputnik, după Gagarin, cosmosul părea să fie accesibil. Secolul XX a fost jalonat de o succesiune de asemenea descoperiri științifice și tehnice majore: submarinele, automobilul, aviația, îngrășămintele chimice, insecticidele, precum DDT-ul, care urma să elibereze omenirea de o serie de boli și de spectrul foametei, penicilina, rachetele, zborurile cosmice și, nu în ultimul rând, fisiunea nucleară și bomba atomică. De fiecare dată ne-am simțit amenințați de pericole enorme, declanșate de aceste forțe dezlănțuite, scăpate de sub control, dar am văzut, totodată, „viitorul luminos" ca fiind foarte aproape.

Acest tip de discurs și acest tip de așteptări, dar și de temeri ne fac să uităm, să ignorăm sau să relativizăm paradigma fundamentală a drepturilor omului. Fără a intra în detalii tehnice – față de care sunt prea puțin apt a mă pronunța – legate de inteligența artificială, se poate observa că aceasta oferă guvernanților, dar și celor ce dețin putere financiară și economică, un instrument redutabil de a-și optimiza acțiunile. Fiind vorba de raporturi de putere, dreptul constituțional devine relevant. În discuție este, și în acest caz, echilibrul dintre sfera de libertate individuală a cetățeanului și sfera acțiunii legitime a statului. În cazul actorilor non-statali, un anumit echilibru între acțiunea acestora și libertatea individuală este de asemenea necesar, fiind conceptualizat prin noțiunea efectului orizontal al drepturilor omului.

În privința unor astfel de probleme ridicate de inteligența artificială – și cu asta închei – există o decizie mai veche, dar foarte pertinentă a Curții Constituționale Federale a Germaniei, din decembrie 1983, deci dintr-o epocă în care se lucra încă cu cartele perforate, referitoare

la așa-zisul drept de „autodeterminare informatică" (*informationelle Selbstbestimmung*).[8] Cu ocazia recensământului, li se cerea cetățenilor să furnizeze o listă destul de lungă de date, pe care birocrația ministerială federală le dorea cu ardoare, mergând de la situația locativă, dimensiunile locuințelor deținute, nivelul veniturilor, al studiilor, până la informațiile referitoare la situația profesională, la bibliotecile consultate, cluburile și asociațiile frecventate de cei recenzați. Era vorba de date utile pentru a forma o imagine economică și sociologică cât mai exactă, în vederea calibrării adecvate a actelor legislative și a măsurilor administrative de către autorități. Interesul guvernanților de a avea o imagine clară, precisă și detaliată a societății și a fiecărui individ în parte intra în conflict cu libertatea individuală. Aceste date ofereau posibilitatea de a contura o hartă a profilului social și cultural sau chiar a preferințelor politice, la nivelul întregii societăți. Care era problema? În anii '70 existau temeri foarte mari legate de angajabilitatea absolvenților în serviciul public.[9] Cei care aveau simpatii de stânga, în special cei care, în siajul mișcărilor studențești din 1968 și din anii următori, au manifestat sub portretele lui Lenin, Troțki, Mao Zedong, Ho Și Min, Fidel Castro sau Che Guevara ori au propagat idei ale acestora se temeau că, în baza unor acțiuni din trecut (de care între timp, poate, se distanțaseră), vor fi în pericol să-și piardă locul de muncă din serviciul public sau din învățământ ori să întâmpine dificultăți la angajare. Serviciul de

[8] Decizia din 16 decembrie 1983 a celui de-al doilea colegiu a Curții: BVerfGE 65, 1 (conform uzanțelor academice germane, deciziile Curții Constituționale Federale sunt citate prin indicarea numărului volumului culegerii de decizii, urmat de numărul primei pagini a textului deciziei respective). Textul deciziei este accesibil la https://www.servat.unibe.ch/dfr/bv065001.html. Pentru o discuție detaliată, vezi Erhard Denninger, „Das Recht auf inforrnationelle Selbstbestimmung und Innere Sicherheit", în Andreas von Schoeler (ed.), *Informationsgesellschaft oder Überwachungsstaat?: Strategien zur Wahrung der Freiheitsrechte im Computerzeitalter*, Westdeutscher Verlag, Opladen, 1986, p. 107-160.

[9] A se vedea, în acest sens, Martin Kriele, „Verfassungsfeinde im öffentlicehn Dienst – Ein unlösbares Problem?", în idem, *Legitimitätsprobleme der Bundesrpublik*, C.H. Beck, München, 1977, p. 146-161, unde autorul combate ideea, foarte răspândită în epocă, a existenței unui „decret privind radicalii" (Radikalenerlaß), în baza căruia personalul din serviciul public, inclusiv personalul didactic din învățământ, era sau urma a fi supus unei epurări sub pretextul condiției legale a loialității față de Constituție, cu interdicția angajării sau respectiv îndepărtarea din serviciu a celor care în timpul studenției au manifestat simpatii față de stânga radicală sau au exprimat opinii interpretabile în acest sens.

informații german – Bundesamt für Verfassungsschutz (Serviciul Federal pentru Protecția Constituției), a cărui misiune principală constă în protecția „ordinii liberale și democratice" (*freiheitliche demokratische Grundordnung*), realiza, și atunci, o monitorizare atentă a mișcărilor politice suspecte sau periculoase pentru această ordine. Unele vor fi în cele din urmă declarate neconstituționale, în cazul partidelor politice, de către Curtea Constituțională Federală.[10] Dată fiind raritatea situațiilor în care o formațiune politică posibil extremistă este scoasă în afara legii, există un număr destul de mare de asociații care pot intra în atenția Serviciului Federal pentru Protecția Constituției. Cei ce sunt în atenția acestui serviciu se pot simți destul de inconfortabil. Indiferent de declararea neconstituționalității activității politice a unei formațiuni politice, apare problema dubiului în privința loialității față de „ordinea liberală și democratică". În plus, asociații și formațiuni politice care astăzi nu sunt în atenția Serviciului s-ar putea ca ulterior să fie declarate ca problematice, pe seama unor evenimente survenite între timp sau a reconsiderării unor evaluări anterioare. Temerea cetățeanului este că, dacă astăzi se înscrie într-o asociație care este legală sau participă la întrunirile sau evenimentele organizate de aceasta, este posibil ca, mâine, autoritățile să se răzgândească.

Pe calea unei evaluări retroactive și pe baza unor date stocate la un moment în care acestea nu implicau niciun pericol sau o consecință nefavorabilă pentru cel ce le furniza, cetățeanul recenzat putea avea de suferit. În opinia Curții, în prelucrarea datelor, protecția individului contra preluării, stocării, utilizării și transmiterii nelimitate a datelor sale personale intră sub incidența dispozițiilor din art. 2 alin. 1 (libertatea individuală) și art. 1 alin. 1 (intangibilitatea demnității umane) din Legea

[10] „Privilegiul partidelor" de a fi scoase în afara legii doar prin declararea neconstituționalității de către Curtea Constituțională Federală este prevăzut în art. 21 din Legea fundamentală, iar procedura este stabilită prin §§ 43 și urm. din legea Curții. La soluția declarării neconstituționalității unui partid politic se ajunge extrem de rar. Curtea a pronunțat doar două decizii în acest sens: în 1952 privind „Partidul Socialist al Reichului" (BVerfGE 2,1) și în 1956 referitor la Partidul Comunist (BVerfGE 5, 85). În privința unei alte formațiuni extremiste, Partidul Național-Democrat al Germaniei (Nationaldemokratische Partei Deutschlands – NPD), au fost inițiate două proceduri de declarare a neconstituționalității, în 2001 și în 2013, ambele cu rezultat negativ (BVerfGE 107, 339 și respectiv BVerfGE 144, 20).

fundamentală. Acest drept fundamental garantează prerogativa individului de a decide, în principiu, el însuși asupra divulgării și utilizării datelor sale personale, iar restrângerile acestui drept de „autodeterminare informațională" sunt admisibile doar în cazul în care interesul general este, în mod clar, precumpănitor. Pentru aceasta este necesară o reglementare prin lege care să îndeplinească cerințele, specifice statului de drept, ale clarității normei și proporționalității, asigurând, totodată, cadrul organizatoric și procedural necesar.[11]

În esență, Curtea afirmă implicit că este mai bine ca statul să nu știe aceste lucruri, pentru ca nu cumva eu, cetățeanul, să ratez opțiunea de a face astăzi ceva legal, din cauza riscului că mâine acțiunea mea ar putea deveni ilegală sau doar suspectă. S-ar crea un culoar de conformism social și ideologic, pe baza evaluărilor precaute și deseori inhibate ale individului în privința a ceea ce este acceptabil de către majoritate sau de către guvernanți, spre care ar fi dirijată conduita tuturor. Cred că ar trebui să fie menținută o zonă de ignoranță a statului în privința acțiunilor individului, chiar și cu riscul ca acesta să comită acte reprobabile sau criticabile.

Jaloanele fundamentale ale conceptualizării problemelor inteligenței artificiale existau deja într-o epocă în care inteligența artificială nu era încă înțeleasă cum este astăzi. Nu înseamnă că astăzi nu sunt provocări mai mari. Riscurile sunt mult mai mari, reacția trebuie să fie mult mai promptă, iar legislația trebuie să ia în considerare toate aceste aspecte. Volumul și impactul mult mai mari ale comunicațiilor prin social media și timpul de reacție foarte redus impun cu mare probabilitate soluții care pun în pericol dreptul fundamental la liberă exprimare și chiar dreptul la informare sau accesul la cultură. Din perspectiva dreptului constituțional, reperul fundamental rămâne intangibilitatea demnității umane,[12] premisă esențială a garantării efective a libertăților fundamentale, inclusiv a dreptului la liberă exprimare.

[11] BVerfGE 65,1.

[12] A se vedea Josef Isensee, „Menschenwürde: die säkulare Gesellschaft auf der Suche nach dem Absoluten", în *Archiv des öffentlichen Rechts* 131. Bd., Heft 2 (2006), p. 173-218; Marius Bălan, „Valențe constituționale ale demnității umane", în *Dreptul*, nr. 2/2021, p. 142-160.

Interferența inteligenței artificiale cu drepturile omului în cazul investigării infracțiunilor contra mediului

Dr. Remus Jurj
Profesor asociat, Universitatea Maritimă din Constanța

1. Categoria infracțiunilor contra mediului

În introducere la intervenția mea, aș dori să fac unele precizări referitoare la sfera infracțiunilor contra mediului. Acestea acoperă o gamă largă de fapte penale, care pot fi împărțite în mai multe categorii, în funcție de subiecte și priorități specifice, astfel: (i) defrișarea ilegală a pădurilor și comerțul ilegal cu cherestea (infracțiuni contra fondului forestier); (ii) depozitarea, eliminarea și comerțul ilegal cu deșeuri și chimicale; (iii) poluarea apei, solului, subsolului și aerului; (iv) comerțul ilegal cu substanțe care duc la epuizarea stratului de ozon; (v) mineritul ilegal și comerțul cu metale prețioase și minerale; (vi) comerțul ilegal și braconajul cu animale sălbatice și al plantelor (infracțiuni contra fondului cinegetic); (vii) pescuitul ilegal, nedeclarat și nereglementat (infracțiuni contra fondului piscicol).

Potrivit Interpol și Programului Națiunilor Unite pentru Mediu, criminalitatea împotriva mediului se află în primele patru dintre cele mai mari activități infracționale din lume, după traficul de droguri, traficul de persoane și contrafacerea, crescând cu o rată între 5% și 7% pe an, de două până la trei ori, ritmul creșterii economice globale. Astfel, comerțul ilegal și braconajul cu flora și fauna sălbatică generează un profit anual între 7 și 23 miliarde USD; infracțiunile forestiere (inclusiv infracțiuni corporative și exploatare ilegală) între 50,7 și 152 miliarde USD, pescuitul ilegal între 11 și 23,5 miliarde USD, infracțiunile privind traficul ilegal și depozitarea de deșeuri între 10 și 12 miliarde USD, iar mineritul ilegal între 12 și 48 miliarde USD.

2. Definirea inteligenței artificiale

Deși termenul de inteligență artificială datează din 1955, el a început să fie utilizat cu precădere din 2010[1] , ca urmare a evoluției internetului prin așa-numita „revoluție digitală". Inteligența artificială este definită[2] ca fiind capacitatea unei mașini de a imita funcții umane, cum ar fi raționamentul, învățarea, planificarea și creativitatea și care permite sistemelor tehnice să perceapă mediul în care funcționează, să prelucreze această percepție și să rezolve probleme, acționând pentru a atinge un anumit obiectiv. Altfel spus, calculatorul primește datele (deja pregătite sau colectate prin intermediul propriilor senzori, cum ar fi o cameră video), le prelucrează și reacționează, fiind capabil să își adapteze, într-o anumită măsură, comportamentul, analizând efectele acțiunilor anterioare și funcționând autonom.

Inteligența artificială poate fi de tipul software, cum ar fi asistenți virtuali, programe informatice de analiză a imaginilor, motoare de căutare,

[1] A se vedea ghidul Interpol și Unicri *Artificial Intelligence and Robotics for Law Enforcement*, 2019 (https://unicri.it/artificial-intelligence-and-robotics-law-enforcement). Cu privire la evoluția termenului de inteligență artificială a se vedea C. Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs* (https://www.ojp.gov/pdffiles1/nij/252038.pdf).

[2] A se vedea definiția postată pe site-ul https://www.europarl.europa.eu/news/ro/headlines/society/20200827STO85804/ce-este-inteligenta-artificiala-si-cum-este-utilizata.

sisteme de recunoaștere vocală și facială sau încorporată, cum ar fi roboți, automobile autonome, drone, internetul obiectelor[3]. În același sens, Organizația pentru Cooperare și Dezvoltare Economică (OECD)[4] definește un sistem de inteligență artificială (IA) ca sistem bazat pe mașină, care poate, pentru un set dat de obiective definite de om, să facă predicții, recomandări sau să ia decizii care influenţează mediile reale sau virtuale. În propunerea de Regulament (Legea privind inteligența artifcială)[5], „sistemul de inteligență artificială" (sistem de IA) este definit ca un software care este dezvoltat prin una sau mai multe tehnici de optimizare matematică și care, pentru un anumit set de obiective definite de om, poate genera rezultate precum conținuturi, previziuni, recomandări sau decizii, care influențează mediile cu care interacționează.

În sens tehnic, inteligența artificială sau „sistemele algoritmice" sunt înțelese ca aplicații care, folosind adesea tehnici de optimizare matematică, efectuează una sau mai multe sarcini, cum ar fi colectarea, combinarea, curățarea, sortarea, clasificarea și deducerea datelor, precum și selecția, prioritizarea, realizarea recomandării și luarea deciziilor[6]. Deși *modus operandi* al inteligenței artificiale exclude orice intervenție umană, ea este creată de oameni și, în acest sens, implică un anumit spațiu de eroare. Toate seturile de date introduse în algoritmii IA pentru a genera rezultate sunt date umane, ceea ce înseamnă că conțin deja părtinire umană, care sunt apoi transmise în rezultatele IA.

[3] *Ibidem*

[4] https://www.oecd.org/digital/artificial-intelligence/

[5] Propunere de Regulament al Parlamentului European și al Consiliului de stabilire a unor norme armonizate privind inteligența artificială (Legea privind inteligența artificială) și de modificare a anumitor acte legislative ale Uniunii – COM(2021) 206 final (https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0023.02/DOC_1&format=PDF).

[6] Recomandarea CM/Rec(2020)1 a Comitetului de Miniștri pentru statele membre cu privire la impactul sistemelor algoritmice asupra drepturilor omului, adoptată de Comitetul de Miniștri la 8 aprilie 2020 la a 1373-a ședință a adjuncților miniștrilor (https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154).

Bazele de date utilizate de forțele de ordine sunt, de fapt companii private, cum ar fi Clearview[7], cea mai mare companie de rețea facială din lume creată pentru aplicarea legii. Deși Clearview este legată prin contract de guverne, ea implică un transfer parțial al anumitor funcții ale statului către companii private, ceea ce ar putea duce la alte rezultate negative, cum ar fi o bază de date „otrăvită" sau pirateria cibernetică, care ar încălca principiile drepturilor la confidențialitate a sute de mii de oameni.

3. Utilizarea generală a IA de către autoritățile publice responsabile de aplicarea legii

În contextul aplicării legii au fost enumerate[8] patru categorii principale în care IA și robotica pot interacționa cu spațiul cyber-fizic: (i) predicție și analiză; (ii) recunoaștere; (iii) explorare; (iv) comunicare.

Prin intermediul analizei video și al imaginilor, inteligența artificială este utilizată pentru a îmbunătăți rezultatele aplicării legii, reducând sarcinile consumatoare de timp și erorile umane. Abilitățile de recunoaștere facială IA pot stabili identitatea și locul unde se află un individ, îmbunătățind considerabil rezultatele supravegherii mulțimii. Recunoașterea facială IA evaluează îmbrăcămintea, structura scheletului și mișcările corpului, pentru a detecta comportamentul anormal sau suspect în rândul maselor, cum ar fi hoții sau șoferii periculoși care încalcă legile rutiere. De asemenea, ajută la identificarea vehiculelor, deoarece programele IA sunt învățate să descifreze plăcuțele de înmatriculare chiar și cu rezoluție slabă sau lumină ambientală scăzută.

Inteligența artificială poate fi, de asemenea, foarte utilă în detectarea accidentelor de trafic prin supravegherea televiziunii cu circuit închis (CCTV) și a infracțiunilor online, inclusiv a traficului de persoane, spălării banilor, fraudei și abuzului sexual.

[7] https://www.clearview.ai/

[8] Ghidul Interpol și Unicri, *Artificial Intelligence and Robotics for Law Enforcement,* 2019, sus-citat.

IA are o utilizare semnificativă și în instanțele de judecată. Prin rezolvarea crimelor și din punct de vedere științific, IA îmbunătățește activitatea laboratoarelor criminalistice și a muncii anchetatorilor în testarea și analiza ADN-ului prin procesarea probelor ADN de nivel scăzut sau degradate, care nu ar fi putut fi folosite cu un deceniu în urmă. În plus, cazuri vechi de zeci de ani au fost redeschise pentru a depune dovezi privind cazurile de agresiune sexuală și omucidere pentru identificarea făptuitorilor. O astfel de utilizare a IA reduce criminalitatea nerezolvată, ceea ce întărește sentimentul de încredere al civililor în justiție. O utilizare importantă a IA se referă și la semnalarea postărilor suspecte pe rețelele sociale, deoarece postările pe rețelele sociale sunt în mod inerent digitale și conțin date text, date de imagine, timpi de postare etc.

4. Aplicarea inteligenței artificiale în cazul infracțiunilor contra fondului forestier

Inteligența artificială poate ajuta la monitorizarea ecosistemelor și a faunei sălbatice și a interacțiunilor acestora. Vitezele sale rapide de procesare pot oferi date din satelit aproape în timp real pentru a urmări tăierile ilegale din păduri. Aceasta se face prin analiza fotografiilor de înaltă rezoluție efectuate fiecărui arbore din anumite parcele. Folosind informațiile de la sol din aceste parcele specifice, IA își poate da seama cum arată diferitele specii de copaci de sus în imaginile de zbor. Aceste informații sunt folosite pentru a extrapola într-o zonă mai mare[9].

Înțelegerea modului în care distribuția și compoziția pădurilor se modifică ca răspuns la tăierile ilegale sau la doborâturile de vânt (uragane) este importantă, deoarece atunci când pădurile sunt deteriorate, vegetația se descompune și emite mai mult CO2 în atmosferă. Pe măsură ce copacii cresc din nou, deoarece sunt mai mici, stochează mai puțin carbon. Tăierile ilegale sau schimbările climatice

[9] A se vedea R. Cho, „Artificial Intelligence — A Game Changer for Climate Change and the Environment", 2018 (https://news.climate.columbia.edu/2018/06/05/artificial-intelligence-climate-environment/).

au ca rezultat faptul că unele păduri nu se vor recupera, cu consecința stocării unui volum mai mic de carbon, iar prin faptul că și mai mult carbon va rămâne în atmosferă se exacerbează încălzirea globală[10].

În România, specialiștii au realizat o aplicație – Sistem integrat de monitorizare și management al fondului forestier[11] – prin care se calculează volumul bușteanului pe baza fotografiilor, se marchează arborii cu cipuri și, cu ajutorul unui cititor, se calculează volumul de masă lemnoasă și APV-ul care stă la baza exploatării. Tot o aplicație a specialiștilor, dar din cadrul Ministerului Mediului, Apelor și Pădurilor este și SUMAL 2.0, prin care se poate urmări practic tot procesul de exploatare a lemnului, de la marcare până la transport[12].

5. Aplicarea inteligenței artificiale în cazul infracțiunilor contra faunei sălbatice

Inteligența artificială poate fi folosită atât la monitorizarea ecosistemelor, cât și la detectarea împușcăturilor în cazul braconajului. Descoperirea semnăturilor de model în analiza împușcăturilor oferă un alt domeniu în care se pot folosi algoritmi IA. Dacă în orașe este mai ușor de utilizat prin aplicarea unor senzori în zonele aglomerate[13], identificarea științifică a unei împușcături în pădure este o provocare pentru autorități.

În cadrul unui proiect, National Institute of Justice (NIJ) a finanțat Cadre Research Labs, LLC pentru a analiza fișierele audio cu împușcături de pe smartphone-uri și dispozitive inteligente „pe baza observației că conținutul și calitatea înregistrărilor cu împușcături sunt influențate de tipul armei de foc și al muniției, de geometria scenei și de calitatea înregistrării de pe dispozitivul utilizat".

[10] *Ibidem*.

[11] https://www.forestdesign.ro/index.php/ro/servicii/28-management-silvic

[12] http://www.mmediu.ro/categorie/sumal-2-0/321

[13] În SUA se folosește IA pentru identificarea împușcăturilor în 80 de orașe (a se vedea B. Dupont, Y. Stevens, H. Westermann, M. Joyce, *Artificial Intelligence in the Context of Crime and Criminal Justice*, raport pentru The Korean Institute of Criminology, p. 83 (https://www.researchgate.net/publication/337402608_Artificial_Intelligence_in_the_Context_of_Crime_and_Criminal_Justice).

Folosind un model matematic bine definit, oamenii de știință de la Cadre[14] lucrează pentru a dezvolta algoritmi care să detecteze împușcături, să diferențieze exploziile botului de undele de șoc, să determine timpul de la împușcare la împușcătură, să determine numărul de arme de foc prezente, să atribuie împușcături specifice armelor de foc și să estimeze probabilități de clasă și calibru — toate acestea putând ajuta forțele de ordine în investigații.

Referitor la obținerea unor informații despre traficanții de animale sălbatice protejate, Environmental Investigation Agency, un ONG[15], a creat platforma Global Environmental Crime Tracker care colaționează datele despre arestări, sechestrări și condamnări în contextul comerțului ilegal de animale sălbatice. Aplicația elaborează hărți cu rutele de transport și legăturile cu grupările infracționale organizate. Prin analiza cu IA a rețelelor sociale ale unor braconieri cunoscuți, se poate stabili predictibilitatea unor rute de braconaj și a riscurilor capturării unor animale sălbatice. În opinia noastră, având în vedere multitudinea de specii protejate CITES, autoritățile de aplicare a legii au nevoie de o aplicație pentru identificarea speciilor CITES (indexul are 1514 pagini), precum și de aplicații pentru detectarea plantelor protejate, a habitatelor pierdute ori a speciilor invazive.

6. Aplicarea inteligenței artificiale în cazul infracțiunilor contra aerului, apei, solului și subsolului

Inteligența artificială poate ajuta la monitorizarea calității aerului, a captării de carbon, la detectarea sursei poluării și la monitorizarea calității apei potabile, poate gestiona utilizarea apei rezidențiale, poate detecta scurgerile subterane în sistemele de alimentare cu apă potabilă și poate prezice când plantele de apă au nevoie de întreținere. De asemenea, IA poate ajuta la identificarea unor fisuri în digurile și barajele de acumulare. În agricultură IA poate conduce la îmbunătățirea producției agricole (de exemplu, fertilizarea și irigarea automate)

[14] https://www.cadreforensics.com/

[15] https://eia-international.org/global-environmental-crime-tracker/

și la bunăstarea animalelor, precum și la reducerea riscurilor de boli, de dăunători sau de amenințări meteorologice[16].

Pentru toate aplicațiile, beneficiile potențiale trebuie să fie echilibrate cu impactul asupra mediului în întregul ciclu de producție al IA și al tehnologiilor informaționale (IT). Aceasta include minerit pentru elemente rare din pământuri și alte materii prime, energia necesară pentru a produce și alimenta mașinile, precum și deșeurile generate în timpul producției și la sfârșitul ciclurilor de viață[17].

Folosirea IA fiind în creștere, este probabil să se adauge la preocupările cu privire la volumul tot mai mare de deșeuri electronice și presiunea asupra materiilor prime rare generate de industria de calcul. Pe lângă mediu și sănătate, impactul deșeurilor electronice are implicații socio-politice importante, în special legate de exportul către țările în curs de dezvoltare și populațiile vulnerabile. Nu în ultimul rând, folosirea IA conduce la urmărirea traseului unor deșeuri periculoase și, în cele din urmă, la identificarea grupărilor infracționale organizate implicate în depozitarea ilegală a deșeurilor.

7. Interferența inteligenței artificiale cu drepturile omului

În opinia noastră aplicarea IA în protecția penală a mediului poate încălca dreptul la respectarea vieții private și de familie; dreptul la confidențialitate și protecția datelor; dreptul la căi de atac efective; prezumția de nevinovăție; dreptul la un proces echitabil (inclusiv egalitatea armelor în procedurile judiciare, dreptul de a interoga martorii, care au introdus datele în soft, dreptul de a efectua o contraexpertiză); principiul nediscriminării și egalității, iar principiul legalității (adică *lex certa*) estompează standardele existente de probă. Rapoartele de cercetare independente arată că utilizarea inteligenței artificiale poate duce la oprirea și percheziția mai frecventă a anumitor grupuri de

[16] „Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence", 2019 (https://unesdoc.unesco.org/ark:/48223/pf0000367422).

[17] *Ibidem*.

oameni de către forțele de ordine decât altele, de exemplu, privând cetățeanul de principiile echității și egalității în drepturi.

De exemplu, supravegherea prin IA a „hotspoturilor" criminale poate crește discriminarea geografică, deoarece acele zone sunt mai controlate de poliție decât alte zone, ceea ce duce la arestări mai mari în astfel de zone monitorizate de IA. Anumiți algoritmi de IA, atunci când sunt exploatați pentru anticiparea recidivelor infracționale, pot prezenta opinii părtinitoare bazate pe gen și rasă, demonstrând o probabilitate diferită de recidivă pentru femei față de bărbați sau pentru cetățenii naționali față de cetățenii străini (de exemplu, par a fi mai mulți vânători bărbați decât femei[18]).

Normele UE privind protecția datelor interzic[19], în principiu, prelucrarea datelor biometrice în scopul identificării unice a unei persoane fizice, cu excepția unor condiții specifice . Mai precis, în temeiul RGPD, o astfel de prelucrare nu poate avea loc decât într-un număr limitat de situații, în principal, din motive de interes public major. În acest caz, prelucrarea trebuie să aibă loc în temeiul dreptului UE sau al dreptului intern, sub rezerva cerințelor de proporționalitate, a respectării esenței dreptului la protecția datelor și a unor garanții adecvate.

În temeiul Directivei privind protecția datelor în materie de asigurare a respectării legii[20], prelucrarea trebuie să fie strict necesară și trebuie să

[18] European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, 2019.

[19] Art. 9 alin. 1 din Regulamentul (UE) 2016/679 al Parlamentului European și al Consiliului din 27 aprilie 2016 privind protecția persoanelor fizice în ceea ce privește prelucrarea datelor cu caracter personal și privind libera circulație a acestor date și de abrogare a Directivei 95/46/CE (Regulamentul general privind protecția datelor), publicat în JO L119/4.05.2016 prevede următoarele:
„Se interzice prelucrarea de date cu caracter personal care dezvăluie originea rasială sau etnică, opiniile politice, confesiunea religioasă sau convingerile filozofice sau apartenența la sindicate și prelucrarea de date genetice, de date biometrice pentru identificarea unică a unei persoane fizice, de date privind sănătatea sau de date privind viața sexuală sau orientarea sexuală ale unei persoane fizice."

[20] Directiva (UE) 2016/680 a Parlamentului European și a Consiliului din 27 aprilie 2016 privind protecția persoanelor fizice referitor la prelucrarea datelor cu caracter personal de către autoritățile competente în scopul prevenirii, depistării, investigării sau urmăririi penale a infracțiunilor sau al

existe, în principiu, o autorizație acordată în temeiul dreptului UE sau al dreptului intern, precum și garanții adecvate. Deoarece orice prelucrare a datelor biometrice în scopul identificării unice a unei persoane fizice ar fi o excepție de la o interdicție prevăzută în dreptul UE, aceasta ar intra sub incidența *Cartei drepturilor fundamentale a UE*.

8. Inteligența artificială în practica CJUE

CJUE, Hotărârea Curții (Marea Cameră) din 6 octombrie 2020 – Cauzele conexate C-511/18, C-512/18 și C-520/18)[21]

Dispozitivul:

1) Articolul 15 alineatul (1) din Directiva 2002/58/CE a Parlamentului European și a Consiliului din 12 iulie 2002 privind prelucrarea datelor personale și protejarea confidențialității în sectorul comunicațiilor publice (Directiva asupra confidențialității și comunicațiilor electronice), astfel cum a fost modificată prin Directiva 2009/136/CE a Parlamentului European și a Consiliului din 25 noiembrie 2009, citit în lumina articolelor 7, 8 și 11, precum și a articolului 52 alineatul (1) din *Carta drepturilor fundamentale a Uniunii Europene*, **trebuie să fie interpretat** în sensul că **se opune** unor măsuri legislative care prevăd, în scopurile prevăzute la acest articol 15 alineatul (1), cu titlu preventiv, o stocare generalizată și nediferențiată a datelor de transfer și a datelor de localizare.

Articolul 15 alineatul (1) din Directiva 2002/58, astfel cum a fost modificată prin Directiva 2009/136, citit în lumina articolelor 7, 8 și 11, precum și a articolului 52 alineatul (1) din *Carta drepturilor fundamentale* **nu se opune** unor măsuri legislative:

– care permit, în scopul protejării **securității naționale**, impunerea unei obligații furnizorilor de servicii de comunicații electronice de a efectua

executării pedepselor și privind libera circulație a acestor date și de abrogare a Deciziei-cadru 2008/977/JAI a Consiliului, publicată în JO L119/4.05.2016.

[21] CJUE, Hotărârea Curții (Marea Cameră) din 6 octombrie 2020 în cauzele conexate C511/18, C512/18 și C520/18.

o stocare generalizată și nediferențiată a datelor de transfer și a datelor de localizare, în situații în care statul membru în cauză se confruntă cu o amenințare gravă la adresa securității naționale, care se dovedește reală și actuală sau previzibilă, în condițiile în care decizia care prevede această obligație poate face obiectul unui control efectiv fie de către o instanță, fie de către o entitate administrativă independentă, a cărei decizie are efect obligatoriu, prin care se urmărește să se verifice existența uneia dintre aceste situații, precum și respectarea condițiilor și a garanțiilor care trebuie să fie prevăzute, iar obligația menționată nu poate fi impusă decât pentru o perioadă limitată în timp la strictul necesar, dar care poate fi reînnoită în cazul menținerii acestei amenințări;

– care prevăd, în scopul protejării securității naționale, al **combaterii infracționalității** grave și al prevenirii amenințărilor grave la adresa siguranței publice, o stocare direcționată a datelor de transfer și a datelor de localizare care să fie delimitată, pe baza unor elemente obiective și nediscriminatorii, în funcție de categoriile de persoane vizate sau prin intermediul unui criteriu geografic, pentru o **perioadă limitată în timp** la strictul necesar, dar care poate fi reînnoită;

– care prevăd, în scopul protejării securității naționale, al combaterii infracționalității grave și al prevenirii amenințărilor grave la adresa siguranței publice, o stocare generalizată și nediferențiată a adreselor IP atribuite sursei unei conexiuni, pentru o perioadă limitată în timp la strictul necesar;

– care prevăd, în scopul protejării securității naționale, al **combaterii infracționalității** și al protejării siguranței publice, o stocare generalizată și nediferențiată a datelor referitoare la identitatea civilă a utilizatorilor de mijloace de comunicații electronice; și

– care permit, în scopul **combaterii infracționalității grave** și, *a fortiori*, al protejării securității naționale, impunerea unei obligații furnizorilor de servicii de comunicații electronice, prin intermediul unei decizii a autorității competente, supuse unui control jurisdicțional efectiv, de a realiza, pentru o perioadă determinată, conservarea rapidă a datelor

de transfer și a datelor de localizare de care dispun acești furnizori de servicii. Aceste măsuri garantează, prin norme clare și precise, că stocarea datelor în discuție este condiționată de respectarea condițiilor materiale și procedurale aferente acestora și că persoanele în cauză dispun de garanții efective împotriva riscurilor de abuz.

2) Articolul 15 alineatul (1) din Directiva 2002/58, astfel cum a fost modificată prin Directiva 2009/136, citit în lumina articolelor 7, 8 și 11, precum și a articolului 52 alineatul (1) din *Carta drepturilor fundamentale*, trebuie să fie interpretat în sensul că **nu se opune** unei reglementări naționale care impune furnizorilor de servicii de comunicații electronice să recurgă, pe de o parte, la **analiza automatizată**, precum și la colectarea în timp real, printre altele, a datelor de transfer și a datelor de localizare și, pe de altă parte, la colectarea în timp real a datelor tehnice referitoare la localizarea echipamentelor terminale utilizate, atunci când:

– recurgerea la analiza automatizată **se limitează** la situațiile în care un stat membru se confruntă cu o amenințare gravă la adresa securității naționale, care se dovedește reală și actuală sau previzibilă, în condițiile în care recurgerea la această analiză poate face obiectul unui control efectiv fie de către o instanță, fie de către o entitate administrativă independentă, a cărei decizie are efect obligatoriu, prin care se urmărește să se verifice existența unei situații care justifică măsura menționată, precum și respectarea condițiilor și a garanțiilor care trebuie să fie prevăzute, iar

– recurgerea la colectarea în timp real a datelor de transfer și a datelor de localizare **este limitată** la persoanele în privința cărora există un motiv valabil pentru a suspecta că sunt implicate într-un mod sau altul în activități de terorism și este supusă unui control prealabil efectuat fie de o instanță, fie de o entitate administrativă independentă, a cărei decizie are efect obligatoriu, pentru a asigura că o astfel de colectare în timp real nu este autorizată decât în limita a ceea ce este strict necesar. În caz de urgență justificată corespunzător, controlul trebuie să aibă loc în termen scurt.

3) Directiva 2000/31/CE a Parlamentului European și a Consiliului din 8 iunie 2000 privind anumite aspecte juridice ale serviciilor societății informaționale, în special ale comerțului electronic, pe piața internă (Directiva privind comerțul electronic) trebuie să fie interpretată în sensul că nu se aplică în materie de protecție a confidențialității comunicațiilor și a persoanelor fizice în raport cu prelucrarea datelor cu caracter personal în cadrul serviciilor societății informaționale, această protecție fiind, după caz, reglementată de Directiva 2002/58, astfel cum a fost modificată prin Directiva 2009/136, sau de Regulamentul (UE) 2016/679 al Parlamentului European și al Consiliului din 27 aprilie 2016 privind protecția persoanelor fizice în ceea ce privește prelucrarea datelor cu caracter personal și privind libera circulație a acestor date și de abrogare a Directivei 95/46. Articolul 23 alineatul (1) din Regulamentul 2016/679, citit în lumina articolelor 7, 8 și 11, precum și a articolului 52 alineatul (1) din *Carta drepturilor fundamentale* trebuie să fie interpretat în sensul că se opune unei reglementări naționale care impune furnizorilor de acces la servicii de comunicații publice online și furnizorilor de servicii de stocare – hosting stocarea generalizată și nediferențiată printre altele a datelor cu caracter personal aferente acestor servicii.

4) O instanță națională nu poate aplica o dispoziție a dreptului său național care îi permite să limiteze în timp efectele unei declarații de nelegalitate pe care trebuie să o facă, în temeiul acestui drept, în privința unei legislații naționale care impune furnizorilor de servicii de comunicații electronice, în vederea, printre altele, a protejării securității naționale și a combaterii infracționalității, o stocare generalizată și nediferențiată a datelor de transfer și a datelor de localizare, incompatibilă cu articolul 15 alineatul (1) din Directiva 2002/58, astfel cum a fost modificată prin Directiva 2009/136, interpretat în lumina articolelor 7, 8 și 11, precum și a articolului 52 alineatul (1) din *Carta drepturilor fundamentale*. Acest articol 15 alineatul (1), citit în lumina principiului efectivității, impune instanței penale naționale să înlăture informațiile și elementele de probă care au fost obținute printr-o stocare generalizată și nediferențiată a datelor de transfer și a datelor de localizare,

incompatibilă cu dreptul Uniunii, în cadrul unei proceduri penale inițiate împotriva unor persoane suspectate de săvârșirea unor infracțiuni, în cazul în care persoanele respective nu sunt în măsură să prezinte în mod eficient observații cu privire la aceste informații și elemente de probă, care provin dintr-un domeniu care nu este cunoscut de judecători și care pot influența în mod preponderent aprecierea faptelor.

9. Inteligența artificială în practica CEDO, Cauza Big Brother Watch și alții împotriva Regatului Unit[22], Hotărârea din 25 mai 2021

Dispozitivul:
CURTEA
1. hotărăște, în unanimitate, că a fost încălcat art. 8 din Convenție în ceea ce privește regimul prevăzut la art. 8 alin. (4) din RIPA;

2. hotărăște, în unanimitate, că a fost încălcat art. 8 din Convenție în ceea ce privește regimul prevăzut la Capitolul II din RIPA;

3. hotărăște, cu douăsprezece voturi pentru și cinci împotrivă, că nu a fost încălcat art. 8 din Convenție în ceea ce privește primirea de informații de la servicii de informații străine;

4. hotărăște, în unanimitate, că, în măsura în care a fost invocat de părțile reclamante din cea de-a doua din cauzele conexate, a fost încălcat art. 10 din Convenție în ceea ce privește regimul prevăzut la art. 8 alin. (4) și regimul prevăzut la Capitolul II din RIPA;

5. hotărăște, cu douăsprezece voturi pentru și cinci împotrivă, că nu a fost încălcat art. 10 din Convenție în ceea ce privește primirea de informații de la servicii de informații străine.

În cauză au fost dezbătute:
Art. 8 Viață privată, respectiv:

[22] CEDO, Cauza Big Brother Watch și alții împotriva Regatului Unit, Marea Cameră nr. 58170/13, 62322/14 și 24960/15, hotărârea din 25 mai 2021, tradusă în limba română (http://ier.gov.ro/wp-content/uploads/2022/01/Big-Brother-Watch-s.a.-impotriva-Regatului-Unit.pdf).

– conformitatea cu Convenția a unui regim de supraveghere secretă care include interceptarea în masă a comunicațiilor și schimbul de informații;

– necesitatea dezvoltării jurisprudenței în lumina diferențelor importante dintre interceptarea țintită și interceptarea în masă;

– criteriu adaptat pentru examinarea regimurilor de interceptare în masă printr-o evaluare globală;

– accentul pus pe „garanții de la un capăt la altul" pentru a ține seama de gradul crescut de intruziune în dreptul la viață privată, pe măsură ce procesul de interceptare în masă trece prin diferite etape;

– deficiențe fundamentale prezente în regimul de interceptare în masă, ca urmare a lipsei unei autorizații independente, neincluderea unor categorii de selectori în cererea de emitere a unui mandat și neincluderea unor selectori legați de o persoană înainte de autorizarea prealabilă internă;

– previzibilitate și garanții suficiente în regimul primirii de informații de la servicii de informații străine;

– regim de obținere a datelor de comunicații de la furnizorii de servicii de comunicații care nu este „prevăzut de lege".

Art. 10 Libertatea de exprimare, respectiv protecția insuficientă a materialelor confidențiale ale jurnaliștilor în cadrul schemelor de supraveghere electronică.

10. Reglementarea inteligenței artificiale la nivelul Uniunii Europene

Până în prezent, la nivelul Uniunii Europene s-a elaborat în 2021 Propunerea de Regulament de stabilire a unor norme armonizate privind inteligența artificială (Legea privind inteligența artificială)[23], iar

[23] https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0023.02/DOC_1&format=PDF

recent, la 28 septembie 2022, Comisia Europeană a emis o Propunere de Directivă privind adaptarea normelor în materie de răspundere civilă extracontractuală la inteligența artificială (Directiva privind răspunderea în materie de IA)[24] prin care se introduce o prezumție legală relativă, astfel că pârâtul ar trebui să o poată răsturna, în special demonstrând că culpa sa nu ar fi putut cauza prejudiciul, și prin care se stabilesc norme comune cu privire la: (i) divulgarea elementelor de probă privind sistemele de inteligență artificială (IA) cu grad ridicat de risc, pentru a permite unui reclamant să justifice o acțiune în despăgubire de drept civil extracontractuală bazată pe culpă; (ii) sarcina probei în cazul acțiunilor de drept civil extracontractuale bazate pe culpă introduse în fața instanțelor naționale pentru prejudiciile cauzate de un sistem de IA.

Directiva menționată nu se aplică răspunderii penale, ci doar acțiunilor în despăgubire de drept civil extracontractuale bazate pe culpă, în cazurile în care prejudiciile cauzate de un sistem de IA se produc ulterior. Anterior a fost emis Raportul Parlamentului European privind inteligența artificială în dreptul penal și utilizarea sa de către autoritățile polițienești și judiciare în procedurile penale, 2020/2016 (INI)[25], în care s-a propus emiterea unei Rezoluții a Parlamentului European, subliniindu-se că unele țări, inclusiv unele state membre, utilizează mai mult aplicațiile sau sistemele bazate pe IA integrate în aplicarea legii și în sistemul judiciar decât altele, ceea ce este în parte consecința lipsei de reglementări și a diferențelor de reglementare care permit sau interzic utilizarea IA în anumite scopuri. De asemenea, s-a subliniat că utilizarea tot mai mare a IA în domeniul dreptului penal se bazează mai ales pe promisiunile că aceasta ar reduce incidența anumitor tipuri de infracțiuni și ar conduce la luarea de decizii mai obiective; cu toate acestea, aceste promisiuni nu se confirmă întotdeauna.

Totodată, se constată că IA este utilizată de autoritățile de aplicare a legii în aplicații cum ar fi tehnologiile de recunoaștere facială, de

[24] http://www.cdep.ro/afaceri_europene/CE/2022/COM(2022)496_RO_ACT_part1_v2.pdf

[25] https://www.europarl.europa.eu/doceo/document/A-9-2021-0232_RO.html

exemplu, căutarea în baze de date de suspecți și identificarea victimelor traficului de persoane sau ale exploatării și abuzurilor sexuale asupra copiilor, recunoașterea automată a numărului de înmatriculare, identificarea vorbitorului, identificarea vocală, tehnologii de lectură labială, supraveghere sonoră (adică algoritmi de detectare a focurilor de armă), cercetare și analiză autonomă a bazelor de date identificate, prognoză (activități polițienești predictive și de analiză a focarelor infracționale), instrumente de detectare a comportamentului, instrumente avansate de autopsie virtuală care să contribuie la stabilirea cauzei decesului, instrumente autonome de identificare a fraudelor financiare și a finanțării terorismului, monitorizarea platformelor de comunicare socială (extragere și recoltare de date în vederea explorării conexiunilor) și sisteme automate de supraveghere care includ diferite capacități de detectare (cum ar fi detectarea bătăilor inimii și camerele cu termoviziune). Aplicațiile menționate mai sus, alături de alte aplicații potențiale sau viitoare ale tehnologiei IA în domeniul aplicării legii, pot prezenta grade diferite de fiabilitate și acuratețe și afectează protejarea drepturilor fundamentale și dinamica sistemului justiției penale; multe dintre aceste instrumente sunt utilizate în țări din afara UE, dar ar fi ilegale în baza cadrului legislativ și al jurisprudenței Uniunii privind protecția datelor. Utilizarea de rutină a algoritmilor, chiar și cu o rată mică de rezultate fals pozitive, poate face ca numărul de alerte corecte să fie cu mult depășit de numărul celor false.

11. Concluzii

Inteligența artificială este o familie de tehnologii care se dezvoltă rapid și care necesită forme noi de supraveghere normativă și un spațiu sigur pentru experimentare, asigurând, în același timp, inovarea responsabilă și integrarea unor garanții adecvate și a unor măsuri de atenuare a riscurilor, astfel că, în acord, cu Propunerea de Regulament de stabilire a unor norme armonizate privind inteligența artificială, apreciem că noțiunea de sistem de IA ar trebui definită în mod clar pentru a asigura securitatea juridică, oferind, în același timp, flexibilitatea necesară

pentru a ține seama de evoluțiile tehnologice viitoare. Definiția ar trebui să se bazeze pe caracteristicile funcționale esențiale ale software-ului, în special pe capacitatea, pentru un anumit set de obiective definite de om, de a genera rezultate cum ar fi conținutul, previziunile, recomandările sau deciziile care influențează mediul cu care interacționează sistemul.

În dreptul penal, probele obținute în mod ilegal sunt inadmisibile la proces. Dacă partea împotriva căreia sunt introduse probe în timpul unui proces nu poate contesta acuratețea și fiabilitatea acestora, atunci se pune întrebarea dacă probele colectate printr-un sistem IA nu sunt supuse criticii, deoarece inaccesibilitatea codului sursă sau a altor caracteristici ale software-ului este permisă legal[26].

Este necesar să se asigure, prin cadre legislative, de reglementare și de supraveghere adecvate legate de sisteme algoritmice, că actorii din sectorul privat care s-au implicat în proiectarea, dezvoltarea și continua implementare a unor astfel de sisteme respectă legile aplicabile și își îndeplinesc responsabilitățile în materie de respectare a drepturilor omului[27].

Totodată, ținând seama de elemente precum complexitatea și opacitatea multor sisteme de IA și de posibilele dificultăți aferente de a verifica în mod eficace conformitatea cu normele aplicabile și de a asigura respectarea acestora, sunt necesare cerințe legate de păstrarea evidențelor în ceea ce privește programarea algoritmului, datele utilizate pentru antrenarea sistemelor de IA cu risc ridicat și, în anumite cazuri, păstrarea datelor însele[28].

[26] A se vedea și https://www.oecd-ilibrary.org/sites/ba682899-en/1/3/3/index.html?itemId=/content/publication/ba682899-en&_csp_=02d27ef0d7308d76a010fd2a9882228f&itemIGO=oecd&itemContentType=book.

[27] Recomandarea CM/Rec(2020)1 a Comitetului de Miniștri pentru statele membre cu privire la impactul sistemelor algoritmice asupra drepturilor omului, adoptată de Comitetul de Miniștri la 8 aprilie 2020 la a 1373-a ședință a adjuncților miniștrilor.

[28] Comisia Europeană, *Cartea albă privind inteligența artificială – O abordare europeană axată pe excelență și încredere,* 2020.

Este necesară folosirea unor surse certificate și a unor date intangibile cu modele elaborate într-o manieră multidisciplinară, într-un mediu tehnologic securizat, respectiv datele introduse într-un software care implementează un algoritm de învățare automată ar trebui să provină din surse certificate și nu trebuie modificate până când nu au fost efectiv utilizate de mecanismul de învățare. Prin urmare, întregul proces trebuie să fie trasabil pentru a se asigura că nu a intervenit nicio modificare care să schimbe conținutul sau sensul deciziei în curs de procesare.

De asemenea, modelele și algoritmii creați trebuie să poată fi stocate și executate în medii securizate, astfel încât să asigure integritatea sistemului și intangibilitate.

Profesioniștii din sistemul de justiție ar trebui, în orice moment, să poată cere revizuirea deciziilor judecătorești și a datelor utilizate pentru a produce un rezultat pentru a nu fi obligatoriu legați de acesta în lumina caracteristicilor specifice ale fiecărui caz în parte. Nu în ultimul rând, utilizatorul trebuie să fie informat într-un limbaj clar și ușor de înțeles dacă soluțiile oferite de instrumentele de inteligență artificială sunt obligatorii, cu privire la diferitele opțiuni disponibile, precum și dacă are sau nu dreptul la consiliere juridică și dreptul de acces la tribunal. El/ea trebuie, de asemenea, să fie informat în mod clar cu privire la orice prelucrare prealabilă a unui caz de către inteligența artificială înainte sau în timpul unui proces judiciar și are dreptul de a se opune, astfel încât cazul său să poată fi audiat direct de instanță în sensul articolului 6 din CEDO[29].

[29] European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, 2019*.

Inteligența artificială – oportunitate, rezistență la schimbare și drepturile omului

Cătălin Luca
Psiholog clinician principal
și cadru didactic asociat la
Universitatea „Alexandru Ioan Cuza" din Iași

Conform informațiilor postate pe site-ul[1] Parlamentului European, inteligența artificială este „capacitatea unei mașini de a imita funcții umane, cum ar fi raționamentul, învățarea, planificarea și creativitatea." Respectiv, aceasta ar putea permite sistemelor tehnice să perceapă mediul în care funcționează, să prelucreze această percepție și să rezolve probleme, acționând pentru a atinge un anumit obiectiv. Sistemele de inteligență artificială pot avea capacitatea să își adapteze, într-o anumită măsură, comportamentul, analizând efectele acțiunilor anterioare și funcționând autonom.

Sunt lideri de opinie care susțin că „La acest moment inteligența artificială nu este inteligență. Este capacitatea de a acumula și prelucra date statistice prin intermediul formulelor matematice, algoritmilor care,

[1] https://www.europarl.europa.eu/news/ro/headlines/society/20200827STO85804/ce-este-inteligenta-artificiala-si-cum-este-utilizata, accesat la 4.10.2022.

datorită dezvoltării internetului, motoarelor de căutare și rețelelor sociale, permite controlul vieții private". Asociația Psihologilor Americani definește inteligența ca fiind „capacitatea de a obține informații, de a învăța din experiență, de a se adapta la mediu, de a înțelege și de a utiliza corect gândirea și rațiunea[2]".

Analizând cele două definiții prezentate anterior, am putea spune că nu găsim mari diferențe între ele, cu excepția unor nuanțări pe care aleg să nu le prezint aici pentru a evita o digresiune. În general, când vorbim despre inteligență, vorbim despre putere care poate însemna și politică. Dar, când vorbim despre politică nu ne raportăm la inteligența artificială, care pare să fie logico-matematică, ci mai curând la inteligență socială și emoțională.

Oricum, cred că cei mai mulți dintre noi am trăit sentimente contradictorii, precum curiozitate, frică, neliniște etc. atunci când ne uitam la primele filme științifico-fantastice și vedeam roboți răi care voiau să distrugă comunități și personaje pozitive, gândindu-ne că într-o zi toate acestea s-ar putea întâmpla și că oamenii nu vor avea niciun control asupra unei astfel de stări de lucruri. Prin urmare, dominația roboților inteligenți și răi riscă să fie peste puterea umană.

Ceea ce în anii '80-'90 mă înfiora în filmele SF astăzi pare că nu mă mai neliniștește atât de mult; utilizarea tehnologiilor care folosesc inteligența artificială în medicină, comunicare, transporturi etc., inclusiv sistemele inteligente pe care le utilizez în fiecare zi, precum telefonul mobil, aplicații etc. par să simplifice și să ușureze existența, însă cu riscurile de rigoare (viruși pe internet, hackeri etc.), iar, când văd ușurința copiilor care se adaptează și utilizează gadgeturi bazate pe inteligența artificială, tind să mă gândesc din ce în ce mai serios la rezistența la schimbare, definită ca fiind „un mecanism interior de apărare cu rolul de semnal de alarmă pe care mintea noastră îl trage pentru a ne apăra".

Am experimentat personal această stare de lucruri, dar mi-au mărturisit și cunoscuți situații trăite de rezistență la schimbare care, acum,

[2] https://dictionary.apa.org/intelligence, accesat la 4.10.2022.

par ridicole. Bunica mea, în tinerețea ei, a avut primul aparat de radio din comuna Dobrovăț și persoanele mai în vârstă din acea comunitate spuneau că într-o cutie de lemn din care se aud voci și cineva cântă nu poate fi decât diavolul. Când au apărut telefoanele mobile mama mea spunea că este imposibil să vorbești cu cineva mergând pe stradă, pentru că trebuie să fii cablat și nu poți târâi atâta fir după tine. Un prieten de-al meu proaspăt hirotonisit preot și cu parohie primită cu eforturi într-un sat a trebuit să se mute, deoarece soția lui, preoteasa, avea semnal la telefon doar la fântână, iar babele din sat au văzut-o „vorbind singură, dând din mâini și învârtindu-se în jurul fântânii", ceea ce însemna că face vrăji la ape. Eu însumi când mi-am cumpărat primul telefon mobil – abia apăruse – am vrut să-l duc înapoi la magazin pentru că nu avea ton.

Timpul de utilizare a tehnologiilor, respectiv al ecranelor în general, părea, înainte de pandemia de COVID-19, o temă nelipsită din discuțiile cu părinții care solicitau consultații pentru abilități parentale și modalități de gestionare a copiilor catalogați drept dependenți de calculator, telefon, tabletă etc. Erau cercetări și, prin urmare, existau recomandări clare, nu mai mult de șase ore pe săptămână, pentru că altfel pot apărea următoarele efecte negative...

Când inclusiv copiii de grădiniță au început să facă cursuri online, situația s-a schimbat; nici în momentul actual nu avem încă un răspuns la întrebarea cât timp este sănătos pentru copii să stea în fața ecranelor. Dacă o persoană se spăla des pe mâini și îi era frică de microbi, diagnosticul era clar, însă în timpul pandemiei paradigma s-a schimbat, erai catalogat ca fiind pe puțin irațional rezistent dacă nu te dădeai cât de des se putea pe mâini cu gel antimicrobian, nu te fereai de ceilalți, nu dezinfectai etc. Contextul creează regulile și impune conduita, important este să utilizăm inteligența și să ne adaptăm la schimbare.

În ediția din 12 octombrie 2022 a publicației *The Guardian* a apărut un articol care mi-a atras atenția și care se numește „Dead-eyed AI robot

AI-Da sets the bar high for Truss and Kwarteng"[3]. Membrii Comisiei digitale și de comunicare din Camera Lorzilor, în sesiunea din 11 octombrie 2022, au adus robotul cu figură umană AI-Da, care s-a adresat acestora critic și ironic cu privire la politici, șefi de instituții, buget etc. În ciuda aspectului de tânără energică și activistă, AI-Da a avut un discurs programat, o privire inexpresivă și răspunsuri preprogramate la întrebări care au fost pregătite. Autorul articolului a concluzionat că AI-Da a oferit răspunsuri destul de plictisitoare, dar care au respectat sintaxa și punctuația în construcția propozițiilor.

Într-un articol[4] postat pe site-ul Parlamentului European se afirmă că luarea automată a deciziilor de către sistemele care folosesc inteligența artificială pentru a îmbunătăți performanța este încă lipsită de „perspectivă umană" și de flexibilitate pentru a se adapta la nuanțele particulare ale cazurilor specifice, probabil pentru că lipsește viclenia de a ascunde părtinirile. În conținutul acestuia, se mai susține că sistemele automate ar lua decizii mai corecte decât o fac oamenii, atunci când aceste date se bazează pe informații care au fost selectate, organizate și prezentate folosind cunoștințe profesionale sau de specialitate.

Asigurarea că drepturile omului sunt respectate și nu subminate de inteligența artificială este un factor foarte important care va continua să preocupe lumea în care trăim. Este evident că tehnologia bazată pe inteligența artificială a intrat și va continua să intre în mai multe aspecte ale vieții personale și sociale. Găsirea unui echilibru corect între dezvoltarea tehnologică și protecția drepturilor omului este o chestiune importantă și urgentă, care preocupă și va preocupa atât specialiștii, cât și pe cetățeni în general.

[3] https://www.theguardian.com/politics/2022/oct/11/dead-eyed-ai-robot-ai-da-sets-the-bar-high-for-truss-and-kwarteng

[4] https://www.europarl.europa.eu/RegData/etudes/ATAG/2022/729501/EPRS_ATA(2022)729501_EN.pdf