UNIVERSITATEA ECOLOGICĂ DIN BUCUREȘTI
FACULTATEA DE DREPT
MASTER DREPTUL INTELIGENȚEI ARTIFICIALE

DREPTUL PENAL ȘI INTELIGENȚA ARTIFICIALĂ

Suport de Curs

Lect.univ.dr. Versavia Brutaru

2024-2025

Cuprins

- răspunderea pentru fapta ilicită comisă de o mașină/robot/ dirijat de inteligența artificială; mașinile autonome: responsabilitate.

Capitolul VI.

Impactul tehnologiei (inclusiv a Inteligenței Artificiale) asupra rețelelor de socializare; hărțuirea pe rețelele de socializare; Cyberbullying-ul.

Bibliografie

Capitolul I.
Introducere în conceptul de inteligență artificială. Emergența unui drept penal al inteligenței artificiale

Inteligența artificială (IA) este tot mai prezentă în lumea care ne înconjoară, fiind în egală măsură tehnologia cheie a procesului de digitalizare, dar și principalul factor disruptiv al lumii în care trăim[1].

Cu mașini inteligente care operează cu procese cognitive la înalt nivel, cum ar fi gândirea, perceperea, învățarea, rezolvarea problemelor și luarea deciziilor, la care se adaugă progresele obținute în colectarea și agregarea datelor, puterea de analiză și procesare computerizată, tehnologia senzorilor și robotica, IA prezintă oportunități de complementare și suplementare a inteligenței umane și de îmbunătățire semnificativă a modului în care oamenii trăiesc și lucrează.

Este evident că astfel de mașini nu numai că transformă sarcinile și deciziile de zi cu zi ale oamenilor, dar ridică o serie de întrebări importante pentru indivizi, societate, economie și guvernanță.

Problema sau întrebarea este: ce știm cu adevărat despre Inteligența Artificială? Ce știm despre modul în care această tehnologie ne afectează și ne va afecta viața? Ne va duce IA într-o situație mai bună sau va ajunge să fie o amenințare pentru bunăstarea noastră și poate chiar pentru supraviețuirea speciei umane?

Înțelegerea acestor probleme este importantă, mai ales când se iau în considerare aspecte ce pot conduce la situații de risc, deoarece, când vine vorba de Inteligența Artificială, aplicațiile practice și prevalența acestei tehnologii au crescut semnificativ în ultimii ani, iar IA se poate dovedi a fi una dintre cele mai de impact și disruptive tehnologii pe care omenirea le-a dezvoltat până acum.

Pentru a răspunde acestor provocări, mai multe state, instituții guvernamentale, asociații profesionale și alte organizații relevante, au exprimat poziții, în diverse forme, declarații de colaborare, elaborare de principii generale în privința folosirii sistemelor bazate pe IA care în

[1] *Analiza abordării europene și a inițiativelor din domeniul inteligenței artificiale la nivel internațional*, Autoritatea pentru digitalizarea României, Universitatea Tehnică din Cluj Napoca, Aceste rezultate au fost obținute prin finanțare în cadrul Programului Operațional Capacitate Administrativă 2014-2020 "Cadru strategic pentru adoptarea și utilizarea de tehnologii inovative în administrația publică 2021 – 2027 – soluții pentru eficientizarea activității", cod SIPOCA 704, Activitatea A6.1. Cadrul strategic național în domeniul inteligenței artificiale

esență propun asigurarea unui cadru juridic și etic adecvat, bazat pe drepturile și valorile fundamentale ale UE (în Europa), cu respectarea confidențialității protecția datelor cu caracter personal precum și principii precum transparența și responsabilitatea, dezvoltarea unui cadru etic pentru dezvoltarea și implementarea IA; Ghidarea tranziției digitale, astfel încât toată lumea să beneficieze de această revoluție tehnologică; Deschiderea unui forum național și internațional de discuții pentru a realiza în mod colectiv o dezvoltare echitabilă, favorabilă incluziunii și durabilă din punct de vedere ecologic a IA.

Problematica IA a devenit una centrală, iar evoluțiile din ultimii 3-4 ani, confirmând faptul că aceasta nu a rămas la nivel declarativ, ci s-a materializat într-o multitudine de acțiuni, strategii, investiții și reglementări.

Comisia Europeană a prezentat în aprilie 2018 o strategie europeană privind IA: "Inteligența artificială pentru Europa" COM(2018)237[2].

Obiectivele strategiei europene privind IA anunțate în această comunicare sunt:

- consolidarea capacității tehnologice și industriale a UE și adoptarea IA în întreaga economie, atât de către sectorul privat, cât și de către cel public;

- pregătirea pentru schimbările socio-economice generate de IA;

- asigurarea unui cadru etic și juridic adecvat.

Ulterior, în decembrie 2018, Comisia Europeană și statele membre au publicat un "Plan coordonat privind inteligența artificială", COM(2018)795, privind dezvoltarea IA în UE. Planul coordonat menționează rolul “AI Watch” de a monitoriza punerea sa în aplicare.

Carta albă privind inteligența artificială apărută în februarie 2020, este documentul care reflectă obiectivul Comisiei de a construi un ecosistem al excelenței și încrederii, fiind documentul care setează prioritățile din domeniul IA.

Comisia Europeană[3] a propus în anul 2021 noi norme și acțiuni menite să transforme Europa în centrul global pentru inteligența artificială de încredere (IA). Combinarea primului cadru juridic privind IA și a unui nou plan coordonat cu statele membre va garanta siguranța și drepturile fundamentale ale persoanelor și organizațiilor, consolidând în același timp adoptarea IA,

[2] https://ec.europa.eu/transparency/documents-register/api/files/COM(2018)237_0/de00000000142394?rendition=false

[3] Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682, Communication on Fostering a European approach to Artificial Intelligence https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence

investițiile și inovarea în întreaga UE. De altfel aceste norme și acțiuni sunt introduse sub deviza: „excelență și încredere”.

Margrethe Vestager, Executive Vice-President for a Europe fit for the Digital Age, a declarat: "În ceea ce privește inteligența artificială, încrederea este o necesitate (*MUST*), nu o opțiune dorită (*NICE TO HAVE*). Cu aceste norme de referință, UE este vârful de lance al dezvoltării de noi norme globale pentru a se asigura că IA poate fi de încredere. Prin stabilirea standardelor, putem deschide calea către tehnologia etică la nivel mondial și ne putem asigura că UE rămâne competitivă pe parcurs. Favorabile viitorului și favorabile inovării, normele noastre vor interveni acolo unde este strict necesar: atunci când siguranța și drepturile fundamentale ale cetățenilor UE sunt în joc."

Comisarul pentru piața internă, Thierry Breton, a declarat: "IA este un mijloc, nu un scop. IA a fost prezentă de zeci de ani, dar a ajuns acum la noi capabilități, alimentate de puterea de calcul. Acest lucru oferă un potențial imens în domenii atât de diverse precum sănătatea, transportul, energia, agricultura, turismul sau securitatea cibernetică. De asemenea, prezintă o serie de riscuri. Propunerile de astăzi vizează consolidarea poziției Europei ca centru global de excelență în IA de la laborator la piață, asigurarea faptului că IA în Europa respectă valorile și normele noastre și valorificarea potențialului IA pentru uz industrial."[4]

Provocările sunt proporționale cu mizele create și afirmate; dreptul nu poate constitui o frână în calea revoluției noilor tehnologii, iar depășirea capacității sale de manifestare corespunzătoare creează pericole grave pentru statutul individului și mersul societății în lipsa controlului social și democratic.

Experții în domeniu admit că inteligența artificială a ajuns deja la stadiul dezvoltării sale ce permite și chiar reclamă a ne imagina înțelegerea și exprimarea ei specifică de către drept. Aplicațiile acesteia se concretizează în ipostaze din ce în ce mai diverse și mai numeroase așa încât nu mai e nevoie de simple proiecții pentru a avea reprezentarea deopotrivă a progreselor de înregistrat și riscurilor de suportat. Consultanții anglo-saxoni specializați în informatică juridică au anunțat mai demult că profesiile juridice trebuie să integreze tehnologiile inteligenței artificiale (IA) în meseriile lor sub pericolul de a fi depășiți prin aplicațiile denumite legal tech și dezvoltate prin start-up-uri, care invadează domeniul și cel de a nu mai putea justifica suportarea remunerării

[4] *Analiza abordării europene și a inițiativelor din domeniul inteligenței artificiale la nivel internațional, op. cit*., p. 18

lor. În SUA s-a dezvoltat tehnologia cognitivă cu pretenția de a înlocui un tânăr avocat și a oferi multiple servicii conexe, ajungându-se până la o reprezentare (artificială) experimentală în instanță. La rândul lor avocații francezi și ordinele profesionale se îngrijorează deja de o atare problematică[5].

Sistemele de inteligență artificială prezintă astăzi numeroase riscuri pentru indivizi: ele pot aduce atingere vieții private ori se dovedesc discriminatorii, manipulatorii sau chiar pot produce prejudicii fizice, psihologice ori economice. De exemplu, ele pot perpetua prejudecăți sociale și discrimina persoanele la locul de muncă ori în contexte de prevenire a criminalității, tot așa pot manipula on-line comportamentele copiilor, persoanelor în vârstă ori altor categorii de consumatori, exploatându-le vulnerabilitățile grație analizelor de date avansate și forțându-le să ia decizii comerciale (ori chiar electorale, precum în scandalul Cambridge Analytica) nedorite și nerezonabile. În același timp, asemenea sisteme pot rămâne obscure și deci dificil de a fi abordate și puse în discuție.

Noua viziune se sprijină din ce în ce mai mult pe responsabilizarea întreprinderilor, pe măsurile tehnice și organizaționale vizând să atenueze riscurile generate de IA pentru oameni, o autoritate de supraveghere urmând să controleze conformarea.

În cadrul prefigurat (cel puțin în UE), anumite sisteme de IA sunt interzise: „dark patterns" (publicitatea on-line manipulatoare) care cauzează prejudicii neeconomice consumatorilor ori care exploatează vulnerabilitatea lor din cauza vârstei ori handicapului; sistemele de credit social, care produc efecte prejudiciabile disproporționate ori în afara contextului; sistemele de identificare biometrică utilizate de forțele de ordine în spațiul public (atunci când utilizarea lor nu este în mod strict necesară ori când riscul efectelor prejudiciabile e prea ridicat). Alte sisteme de IA sunt considerate de înalt risc, în special recunoașterea feței și a emoțiilor, cele utilizate în infrastructurile critice, în contextele educației ocupării, urgenței, azilului și frontierelor, în ajutorul social, pentru evaluarea creditului, de către forțele de ordine ori justiție. Furnizorii și utilizatorii lor vor trebui să efectueze gestionarea riscurilor, să notifice unei autorități de supraveghere, să obțină o certificare de conformitate ori să aplice acțiuni corective în cazuri de neconformare. În plus, va trebui să se pregătească un plan de guvernanță a datelor adecvat pentru întreținerea și validarea IA, acoperind eventualele „lacune", prevenind riscurile și contextualizând sistemul IA în cadrul social specific[6].

[5] Mircea Duțu, *Există un drept al inteligenței artificiale*?, www.juridice.ro
[6] Idem

Ajunsă în centrul cercetării, IA ridică provocarea dependenței tehnologice și a controlului proceselor de descoperire. Cine va fi responsabil de inovațiile produse de algoritmi? Cum garantăm că aceste avansuri servesc binelui comun? Formarea în domeniul instrumentelor de inteligență artificială devine inevitabilă și decisivă.

Despre imperativul reglementării IA. Emergența unei noi discipline.

De ce este nevoie de reglementare? Direcțiile adoptate de cercetători în domeniu, filosofi, juriști dar și programatori sau ingineri, sunt următoarele:

1. *Nevoia unei guvernanțe mondiale și a unei reglementări universale*. Raportul final al grupului de experți privind guvernanța IA desemnați de Secretarul general al ONU, publicat la 20 septembrie 2024, preciza că natura însăși a tehnologiei – transfrontalieră în structura și aplicarea sa – necesită o abordare mondială. Inteligența artificială e pe cale de a transforma lumea într-o multitudine de domenii, fie că e vorba de a deschise noi câmpuri de cercetare științifică, de a optimiza rețelele energetice, de a ameliora sănătatea publică și agricultura ori de a promova avansurile către atingerea celor 17 obiective ale dezvoltării durabile (ODD). În același timp, chiar dacă IA prezintă un enorm potențial benefic, avantajele sale, dacă nu se iau măsuri, ar putea să fie acaparate de un grup restrâns de state, întreprinderi și indivizi pionieri în materie, cu riscul de a agrava fracturile și inegalitățile digitale deja existente.

În scopul de a atenua riscurile asociate IA, documentul propune mai multe recomandări vizând stabilirea unui cadru mondial de guvernanță a inteligenței artificiale și exprimă imperativul unei reglementări adecvate. Și aceasta întrucât încep să apară deja disparități planetare; în termeni de reprezentare, regiuni întregi ale lumii au fost îndepărtate din discuțiile internaționale privind guvernanța IA. Numai șapte state (Canada, Franța, Germania, Italia, Japonia, Marea Britanie și SUA) sunt părți la șapte inițiative importante în materie de IA neguvernamentală, în timp ce altele 118, în principal din sud, nu participă la niciuna dintre ele. Echitatea cere ca un cât mai mare număr dintre ele să joace un rol semnificativ în cadrul deciziilor ce afectează buna guvernanță a tehnologiilor. Concentrarea luării deciziilor în sectorul aplicațiilor IA este nejustificată acum, cu atât mai mult cu cât, din punct de vedere istoric, numeroase comunități au fost total excluse de la dezbaterile privind guvernanța în materie care o privesc în mod direct, se concluzionează în raport. În același timp, îngrijorările iau formele unor amenințări reale și concrete.

Capacitatea inteligenței artificiale de a schimba lumea e de asemenea natură încât dezvoltarea și utilizarea sa nu pot fi lăsate numai la capriciile pieței, iar reglementarea lor e

indispensabilă. Din această perspectivă, grupul de experți formulează o serie de recomandări pentru reglementarea utilizării IA. Acestea includ crearea unui grup științific internațional independent privind IA, instituirea unui dialog politic interguvernamental și multipartit semestrial asupra guvernanței IA, care să permită partajarea celor mai bune practici și crearea unui fond mondial pentru IA destinat să reducă fractura digitală. Orice desfășurare a inteligenței artificiale în context militar trebuie să fie conformă dreptului internațional umanitar și normelor relative la drepturile omului și se recomandă statelor membre stabilirea de cadre juridice și mecanisme de supraveghere solide[7].

2. *Ce tip de reglementare pentru IA*? Independent de conținutul normelor și de obiectivele urmărite se pune, deopotrivă, cu aceeași acuitate și mai întâi problema tipului de reglementare care ar trebui să fie privilegiată[8]. Alegerea și promovarea unui dispozitiv normativ operează în funcție de criterii de adaptabilitate, de eficacitate, precum raportul dintre cost și randament, dar și de imperative economice, în particular voința de a nu împiedica ci, dimpotrivă, a stimula inovația în general, și în special sporirea acestui nou sector ca vector de creștere, cât și din considerații politice de legitimitate, cu precădere în raport cu responsabilitățile eminente ale puterilor publice în materie de protecție a drepturilor umane și de respectare a principiilor democrației și a exigențelor statului de drept. Se constată, la acest moment, avansul și coexistența a trei tendințe deopotrivă concurente și complementare: a) reglementarea zisă etică, de natură deontologică, care se exprimă mai ales prin adoptarea de carte și coduri de conduită; b) reglementarea juridică, tradusă nu numai prin legi și reglementări, ci rezultată și din jurisprudența provenită din aplicarea de către judecător la inteligența artificială a regulilor și principiilor ordinare de drept; c) reglementarea tehnică, care împrumută și promovează în domeniu vehiculul standardelor ori normelor tehnice regionale sau globale, de tip ISO ori IEEE. În timp, am asistat mai întâi la apariția de norme etice, care formulează principii și reguli generale deontologice pe care dezvoltatorii de IA și persoanele și instituțiile care le aplică ar trebui să le respecte. Aceste inițiative se înscriu în modelul general al responsabilității societale a întreprinderilor (RSE), cu adaptări și ajustări la particularitățile domeniului avut în vedere și împărtășesc o logică a autoregulării, vizând să evite o reglementare prematură ori intruzivă ce se presupune că ar împiedica dezvoltarea tehnologiilor și avansurile

[7] M. Duțu, *Dreptul inteligenței artificiale: imperativul reglementării și mizele unei noi discipline științifice*, www.juridice.ro
[8] Idem

economice preconizate. Ele s-au declinat, până acum, în special în instrumente precum: enunțarea de mari principii în acte declarative emanate, de exemplu, de la o organizație internațională, dar fără deschidere juridică obligatorie; coduri de conduită la nivel de întreprinderi, de sectoare etc.; obligații de raportare regulată subscrise de actorii respectivi; supravegherea mai mult sau mai puțin informală efectuată de auditori profesioniști ori de organizații ale societății civile. Printre aceste inițiative se citează, cu titlu exemplificativ, lucrările fundației AI for good instituită de Uniunea Internațională a Comunicațiilor (UIT), care fac parte din sistemul ONU, „liniile directoare" adoptate de Uniunea Europeană și Consiliul Europei ș.a. Dacă aceste demersuri și documente au jucat un rol precursor în formularea mizelor și a principiilor, dezvoltarea spectaculară și desfășurarea aplicațiilor în toate domeniile au condus la o tranziție rapidă spre norme juridice inclusiv în ipostaza asocierii cu cele tehnice mai ambițioase. În această ultimă perspectivă, se înregistrează o alunecare rapidă de reguli și instituții clasice (legi, regulamente, hotărâri judecătorești etc.) spre dispozitive preponderent tehnice (norme, certificări, etichete) la nivel european și global[9].

Prima reglementare juridică a fost Regulamentul (UE) 2016/679 al Parlamentului European și al Consiliului din 27 aprilie 2016 relativ la protecția persoanelor fizice față de tratamentul datelor cu caracter personal și libera circulație a acestor date (RGDP) care a stabilit reguli obligatorii în materie de colectare și tratare a datelor personale. Printre altele, acesta instituie pe seama persoanelor fizice „*dreptul de a nu face obiectul unei decizii fondate exclusiv pe un tratament automatizat, inclusiv profilajul, producând efecte juridice ce le privește ori afectează în mod semnificativ și similar*" (art. 22). Această interdicție este ridicată în situația în care persoana își dă consimțământul său explicit la tratament, atunci când e necesar pentru încheierea ori executarea unui contract sau atunci când tratamentul e autorizat de dreptul UE ori de dreptul intern al unui stat membru. Numărul și amploarea excepțiilor aproape că golește de conținut interdicția de principiu întrucât, pe de o parte, cea mai mare parte a relațiilor dintre consumatori și întreprinderi sunt de natură contractuală, iar, pe de alta, statele membre pot să autorizeze tot ceea ce vor, în special, dar nu numai, în ceea ce privește folosirea de către administrație de tratamente automatice. Tot articolul 22 cere, în același timp, atât întreprinderilor private, cât și colectivităților publice să stabilească garanții proporționale pentru a asigura respectarea drepturilor și libertăților persoanelor vizate, în special posibilitatea unei revizuiri umane. Din păcate, în reglementările

[9] Idem

subsecvente și conexe atât UE, cât și statele membre au deschis larg excepțiile tratamentului automatic. Astfel, pentru prima dată, Regulamentul (UE) 2022/2065 al Parlamentului European și Consiliului relativ la o piață unică a serviciilor digitale și de modificare a directivei 2000/31/CE (Digital Services Act – DSA) constrânge practic cele mai mari platforme să stabilească dispozitive algoritmice proactive de tratare preventivă a conținuturilor ilicite pe care le adăpostesc. Exemple asemănătoare se găsesc și în legislațiile naționale ale statelor membre, ca, de exemplu, Franța și Belgia. Desigur, actul legislativ cel mai complex adoptat până acum în materie e reprezentat de Regulamentul (UE) nr. 1689/2024 din 13 iunie 2024 (intrat în vigoare începând cu 1 august a.c.) de stabilire a unor norme armonizate privind inteligența artificială, care structurează un model de cadru juridic general, capabil a fi preluat, în multe privințe, ca exemplu și înalte regiuni și state terțe, în construirea reacției juridice la provocările IA.

Scurte clarificări conceptuale.

Tehnologiile bazate pe inteligență artificială au avut o dezvoltare rapidă și la scară largă în ultimii ani, arătându-și tot mai mult efectele în multe domenii ale vieții. Această dezvoltare a tehnologiilor inteligente a contribuit și contribuie în continuare atât la dezvoltarea societății dar și la schimbarea obiceiurilor acesteia. Termenul de ”*artificial*”, așa cum este definit de Webster Dictionary[10] este definit ca ”*făcut de oameni, adesea ca o copie a ceva natural*”. Așadar, această definiție nu cuprinde ipoteza în care un sistem de inteligență artificială ar crea alte sisteme de inteligență artificială.

O altă problemă ar fi definirea termenului ”*natural*”, în opoziție cu ”*artificial*”. Și descrierea noțiunii de ”inteligență” ar putea crea confuzii. Potrivit lui Mainzer[11] „*Un sistem poate fi definit ca inteligent dacă poate rezolva probleme în mod independent și eficient. Nivelul de inteligență depinde pe gradul de independență, gradul de complexitate al problemei, și gradul de eficiență al procesului de rezolvare a problemelor*”. În mod similar, dicționarul Cambridge[12] definește inteligența artificială ca fiind „*abilitatea de a înțelege, de a învăța și de a forma judecăți și opinii bazate pe rațiune*”.

[10] https://dictionary.cambridge.org/de/worterbuch/englisch/artificial
[11] Klaus Mainzer, *Künstliche Intelligenz – Wann übernehmen die Maschinen*?, Berlin: Springer, 2016
[12] https://dictionary.cambridge.org/dictionary/english/intelligence

Unii autori au încercat definirea inteligenței artificiale în relație cu performanța umană, în timp ce alți autori preferă o definiție abstractă, formală a inteligenței numită raționalitate, adică ceva care face „lucrul corect", așadar fără nicio implicare emoțională[13]. Înțelegerea dar și definirea termenului de *Rațiune* în sine variază; unii cercetători consideră inteligența ca fiind o proprietate a proceselor de gândire interne și raționament specific uman (ființă biologică), în timp ce alții se concentrează pe comportamentul inteligent, așadar, o caracterizare externă. Pornind de la aceste două dimensiuni, *uman vs. rațional* și *gândire vs. comportament*, ar exista patru combinații posibile: gândirea umană, gândirea rațională, acțiunea umană și acțiunea rațională[14].

Așa cum am arătat mai sus, a defini termenul de "inteligență" poate fi destul de dificil. Până la urmă, ce fel de inteligență avem în vedere? Formele de inteligență pot fi clasificate ca inteligență emoțională, inteligență intelectuală, inteligență lingvistică, inteligență muzicală, coeficient de adversitate etc[15]. Dar, inteligența nu este singurul factor care definește o ființă biologică (omul). În luarea deciziilor, în formarea opiniilor, alături de inteligență și educație, pregătire școlară mai sunt implicați și alți factori, care nu pot fi cuantificați în mod obiectiv, cum ar fi, de exemplu conștiința, voința etc. Lawrence Solum[16], spre exemplu, a rezumat factorii necesari pentru existența personalității juridice a Inteligenței Artificiale în capitolul „Argumentul lipsește-ceva" (*missing something argument*) al cercetării sale, afirmând că există câteva criterii pe care încă nu le-am putut stabili, care caracterizează și diferențiază omul/ființa biologică de artificial: existența sufletului, conștiința, intenționalitatea și alte câteva elemente.

Potrivit lui Russel și Norvig[17], în primele decenii, "*agenții raționali*" (referindu-se la programele de computer) au fost construiți pe baze logice și a format planuri definite pentru a atinge obiective specifice. Mai târziu, metodele bazate pe teoria probabilității și învățarea automată au permis crearea de "agenți" (programe de inteligență artificială) care ar putea lua decizii în condiții de incertitudine pentru a obține cel mai bun rezultat așteptat.

---

[13] Stuart J. Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall Series în Artificial Intelligence (Upper Saddle River: Prentice Hall, 2010), p. 19 și urm.
[14] Stuart J. Russell, Peter Norvig, *op. cit*. p. 20
[15] Shubham Singh, *Attribution of Legal Personhood to Artificially Intelligence Beings*, Bharati Law Review July-Sept 2017 (2017): 195–96.
[16] Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, North Carolina Law Review 70, no. 4 (1992): 1262 ff.
[17] Stuart J. Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall Series in Artificial Intelligence, (Upper Saddle River: Prentice Hall, 2010), p. 19 și urm.

Definiția lui Helbig[18] este o explicație cuprinzătoare a inteligenței artificiale: „Efectuarea de operațiuni folosind sisteme tehnice sau cu ajutorul computerelor care îndeplinesc următoarele condiții: producerea lor necesită înțelegerea universală a inteligenței umane și nu există algoritmi special adaptați pentru realizarea lor" ("*Performing services using technical systems or with the help of computers that fulfill the following conditions: their production calls for universal understanding of human intelligence, and there are no specially adapted algorithms for their realization"*).

Kurzweil[19] (1990) a definit simplu inteligența artificială ca „*arta de a crea mașini care îndeplinesc funcții care necesită inteligență atunci când sunt realizate de oameni*". Tot Kurzweil ne aduce în atenție o întrebare care în opinia sa este hotărâtoare pentru înțelegerea lumii în care trăim: „Poate o inteligență să creeze o altă inteligență mai inteligentă decât ea însăși?[20]. Ori altfel spus, poate inteligența omului să creeze o altă inteligență, ca, de pildă, inteligența artificială (IA), mai inteligentă decât a lui?

Kaplan și Haenlein (2020)[21] au definit-o ca fiind „*capacitatea unui sistem de a interpreta corect datele externe, de a învăța din astfel de date și de a folosi acele învățăminte pentru a atinge obiective și sarcini specifice prin adaptare flexibilă*".

În afară de definițiile date de cercetători în lucrări academice, mai găsim și definiții legale, cuprinse în legi, cum ar fi Proiectul de lege nr. 511 al legislativului statului Nevada în anul 2011[22]: "*Utilizarea computerelor și a echipamentelor aferente pentru a permite unei mașini să duplice sau să imite comportamentul ființelor umane*", în acest sens arătându-se că sistemele care nu duplică sau nu imită comportamentele umane nu pot fi considerate sisteme tip inteligență artificială. În plus, definiția dată în cadrul UNCTAD (Conferința Națiunilor Unite pentru Comerț și Dezvoltare) este „... *Inteligența artificială este capacitatea mașinilor de a imita comportamentul uman inteligent. Aceasta poate implica îndeplinirea diferitelor sarcini cognitive, cum ar fi*

---

[18] Hermann Helbig, *Künstliche Intelligenz und automatische Wissensverarbeitung*, 2nd ed., Berlin: Verlag Technik Berlin, 1996, p. 11.
[19] Kurzweil, R. (1990.), *The Age of Intelligent Machines*, Cambridge MIT Press.
[20] Idem
[21] Kaplan, A., & Haenlein, M. (2020), *Rulers of the world, unite! The challenges and opportunities of artificial intelligence*, Journal of Business Ethics, 63(1).
[22] The Nevada State Legislature, Assembly Bill No. 511 (The Nevada State Legislature, 2011), http://search.leg.state.nv.us/isysquery/86bef777-54ca-45a5-9ee1-69561c80f3e/1/doc/AB511.pdf#xml=http://WebApp/isysquery/86bef777-54ca-45a5-9ee1-969561c80f3e/1/hilite/; *Ugo Pagallo, The Laws of Robots: Crimes, Contracts, and Torts*, 1st ed., Law, Governance and Technology Series, Springer, 2013, p. 5

detectarea, procesarea limbajului oral, raționamentul, învățarea, luarea deciziilor și demonstrarea abilității de a manipula obiectele în consecință"[23]. Dar, această definiție e considerată prea vagă pentru Inteligența Artificială. Cu toate acestea, definiția dată de principalul organism al Națiunilor Unite care se ocupă de problemele de comerț, investiții și dezvoltare, poate avea o funcție practică în determinarea politicilor de inteligență artificială, în special crearea unei foi de parcurs comune la scară globală. Comisia Europeană descrie IA ca fiind „*Mașini sau agenți capabile să observe mediul, să învețe și, pe baza cunoștințelor și experienței dobândite, să întreprindă acțiuni inteligente sau să ia decizii*"[24].

Pe lângă definiția inteligenței artificiale, există și diferite forme de interpretare din diferite perspective în funcție de nivelurile de dezvoltare ale AI. În acest sens, există o clasificare a inteligenței artificiale ca slabă, puternică și super AI (Artificial Narrow Intelligence [ANI], Artificial General Intelligence [AGI] și Artificial Super Intelligence [ASI]). ANI este inteligența artificială, care poate îndeplini doar o singură sarcină, cum ar fi traduceri, jocuri pe calculator, stabilirea prețurilor biletelor de avion și operarea de mașini autonome, de exemplu. AGI înseamnă că abilitățile inteligenței artificiale pot îndeplini sarcinile cognitive la nivel uman și ajunge la nivelul inteligenței umane[25]. Definiția ASI este „*Nivelul entității AI care este mult mai inteligentă decât cel mai inteligent creier uman în abilități practice și are creativitate, abilități sociale etc*." Cercetătorii presupun că entitățile AI pot atinge nivelul ASI după ce vor atinge nivelul AGI în viitor[26].

În anii 1950 s-au remarcat două viziuni referitoare la dezvoltarea IA. Prima a fost IA *simbolică* (care consta într-o reprezentare simbolică a lumii și a sistemelor care ar putea raționa

---

[23] United Nations Conference on Trade and Development UNCTAD (ed.), Digitalization, Trade and Development, Information Economy Report 2017 (New York Geneva: United Nations, 2017), p. 5, https://unctad.org/en/PublicationsLibrary/ier2017_en.pdf; Artificial Intelligence in Asia and the Pacific (United Nations ESCAP, November 2017), p. 1, https://www.unescap.org/sites/default/files/ESCAP_Artificial_Intelligence.pdf

[24] A European Perspective (Luxembourg: Publications Office of EU, 2018), 19, http://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf.

[25] Nicolas Miailhe/Cyrus Hodes, *The Third Age of Artificial Intelligence*, Field Actions Science Reports, no. Special Issue 17 (2017) p. 8; Singh, *Attribution of Legal Personhood to Artificially Intelligence Beings*, 196–97; Carolin Kemper, *Rechtspersönlichkeit für Künstliche Intelligenz*?, Cognitio 2018/1 (December 4, 2018), p. 2, https://doi.org/10.5281/ZEN ODO.1928171; Burcak Ünsal, Yapay Zeka, *Robotlar, Hukuki Düzenlemeler*, Istanbul Barosu Dergisi 93, no. 2019/4 (2019), p. 66; Tyler L. Jaynes, *Legal Personhood for Artificial Intelligence: Citizenship as the Exception to the Rule*, AI & SOCIETY, June 25, 2019, p. 2, https://doi.org/10.1007/s00146-019-00897-9; Lukas Staffler/Oliver Jany, *Künstliche Intelligenz und Strafrechtspflege – eine Orientierung*, Zeitschrift für Internationale Strafrechtsdogmatik, no. 4 (2020), p.165–66.

[26] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, 1st ed. (Oxford: Oxford University Press, 2014), p. 33; Singh, *Attribution of Legal Personhood to Artificially Intelligence Beings*, *op. cit*., p. 196.

despre lume) având ca susținători nume sonore din domeniu precum Marvin Minsky, Herbert A. Simon și Allen Newell). *Inteligența artificială simbolică* s-a fundamentat pe ”căutarea euristică”, aceasta revendicând cercetarea unui spațiu de posibilități pentru găsirea de răspunsuri. Cea de-a doua abordare privind dezvoltarea IA, denumită *conexionistă*, a fost susținută de Frank Rosenblatt un psiholog american însemnat în domeniul IA. El a pus bazele rețelelor neuronale artificiale motiv pentru care deseori a fost numit părintele învățării profunde. Viziunea conexionistă a încercat să obțină IA prin învățare, căutând să conecteze un algoritm de învățare supravegheată (perceptor) în moduri inspirate de legăturile neuronilor[27]. Kai-Fu Lee, vorbește că în perioada de pionierat a domeniului existau două tabere, mai precis două abordări: una pe bază de reguli și cealaltă fondată pe rețele neuronale[28]. Cu referire la prima dintre ele, cercetătorii „încercau să învețe computerele să gândească prin codarea unei serii de reguli logice: dacă X, atunci Y.”, cu precizarea că viziunea respectivă își îndeplinea rolul în cazul jocurilor simple și precise, dar nu se mai întâmpla același lucru atunci când „universul alegerilor sau mutărilor posibile se extindea. Pentru a face acest software mai aplicabil problemelor din lumea reală, tabăra bazată pe reguli (numite și „sisteme simbolice” sau „sisteme expert”) a încercat să intervieveze experți în domeniul problemelor vizate de inteligența artificială respectivă, după care să codeze înțelepciunea lor în procesul de luare a deciziilor integrat în program (de aici numele de „sisteme expert”)”. Cea de-a doua, în schimb, s-a orientat spre o altă manieră de lucru, respectiv au renunțat la a încerca „să învețe computerul regulile care fuseseră stăpânite de un creier omenesc” pentru a căuta „să reconstruiască însuși creierul omenesc”[29].

Apariția inteligenței artificiale (1941-1956)

Termenul „*inteligență artificială*” a fost folosit pentru prima dată în vara anului 1956, la o conferință organizată de Colegiul Dartmouth, în Hanover, New Hampshire, SUA, de către informaticianul și omul de știință cognitiv american John McCarthy[30], unul dintre fondatorii domeniului IA, alături de Alan Turing, Marvin Minsky, Allen Newell şi Herbert A. Simon[31]. Cu acest prilej McCarthy a utilizat termenul „Artificial Intelligence” pentru a descrie obiectivul de a

---

[27] Nicolae Sfetcu, *Inteligența de la originile naturale la frontierele artificiale: inteligența umană vs. inteligența artificială,* p. 49; https://en.wikipedia.org/wiki/Frank_Rosenblatt

[28] Ovidiu Predescu, *Unele considerații asupra conceptului de inteligență artificială,* www.juridice.ro

[29] Kai-Fu Lee, *Superputerile inteligenței artificiale: China, Silicon Valley și noua ordine mondială*, Ed. Corint Future, București, 2021, pp. 23, 24

[30] el este cel care a făcut cunoscute pe larg sistemele de partajare a timpului redenumite servere, cunoscute în prezent drept *cloud computing*

[31] O. Predescu, *op. cit.*, www.juridice.ro

crea mașini și programe care să fie capabile să simuleze gândirea și comportamentul inteligent. Aici a fost prezentat primul program de IA numit The Logic Theorist, „capabil să demonstreze teoreme în calculul propozițional”[32].

Cibernetica lui Norbert Wiener a descris controlul și stabilitatea în rețelele electrice. Teoria informației a lui Claude Shannon a descris semnalele digitale (binare). Teoria de calcul a lui Alan Turing a arătat că orice formă de calcul poate fi descrisă digital. Relația strânsă dintre aceste idei a sugerat că ar putea fi posibilă construirea unui „creier electronic”. În anii 1950, au apărut două viziuni despre dezvoltarea inteligenței artificiale. O abordare, sprijinită printre alții de Allen Newell, Herbert A. Simon și Marvin Minsky, a fost IA simbolică (GOFIA), apelând la o reprezentare simbolică a lumii și a sistemelor care ar putea raționa despre lume. Aceasta s-a bazat pe „căutarea euristică”, care necesită explorarea unui spațiu de posibilități pentru răspunsuri. A doua abordare, susținută de Frank Rosenblatt, este cea conexionistă, care a încercat să obțină IA prin învățare, încercând să conecteze perceptronul (un algoritm de învățare supravegheată) în moduri inspirate de conexiunile neuronilor. Abordările simbolice au dominat căutările pentru inteligența artificială în această perioadă, favorizată de tradițiile intelectuale ale lui Descartes, Boole, Gottlob Frege, Bertrand Russell și alții. Abordările conexioniste au renăscut în ultimele decenii[33].

1941: Alan Turing a publicat o lucrare despre inteligența mașinilor care ar putea fi cea mai veche lucrare din domeniul IA. 1943: Walter Pitts și Warren McCulloch au analizat rețele de neuroni artificiali idealizați, și au arătat cum aceștia ar putea îndeplini funcții logice simple[34].

1950: Alan Turing a publicat lucrarea „*Computing machinery and intelligence*”[35], întrebându-se: „*Pot mașinile să gândească*?”, speculând cu privire la posibilitatea de a crea mașini care gândesc, incluzând conceptul său cunoscut sub numele de ”testul Turing”: Dacă testul Turing implică o mașină care ar purta o conversație care nu se poate distinge de o conversație cu o ființă umană, atunci este rezonabil să spunem că mașina este inteligentă. Testul Turing a fost primul

[32] Idem; Adina Magda Florea, *Robustețe, siguranță și transparență pentru inteligența artificială de încredere*, în „*Caiet constituțional II: IA și drepturile omului*”, coordonatori Robert-Marius Cazanciuc, Cristian Pârvulescu, Ed. Monitorul Oficial, București, p. 22

[33] Nicolae Sfetcu, *Inteligența, de la originile naturale la frontierele artificiale*, Ed. MultiMedia Publishing, p. 51, https://www.telework.ro/ro/e-books/inteligenta-de-la-originile-naturale-la-frontierele-artificialeinteligenta-umana-vs-inteligenta-artificiala/

[34] McCulloch, Warren S., Walter Pitts, 1943, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, The Bulletin of Mathematical Biophysics 5 (4): 115–33, https://doi.org/10.1007/BF02478259

[35] Turing, A. M., *Computing Machinery and Intelligence*, Mind LIX (236): 433–60, 1950, https://doi.org/10.1093/mind/LIX.236.433

experiment propus să măsoare inteligența mașinilor. Au fost construiți primii roboți experimentali, precum țestoasele lui W. Gray Walter și Bestia Johns Hopkins, fiind controlați în întregime de circuite analogice[36].

1951: Christopher Strachey a scris un program de dame, iar Dietrich Prinz a scris unul pentru șah[37]. Marvin Minsky a construit prima mașină de rețea neuronală, SNARC[38] (McCorduck 2004, 102).

1952: Primul program care a demonstrat că computerele pot învăța și nu doar să realizeze ceea ce sunt programați pentru a face, a fost în jocul de dame.

1955: Allen Newell și Herbert A. Simon au creat „Logic Theorist", care a demonstrat 38 din primele 52 de teoreme din cartea Principia Mathematica[39] (Whitehead și Russell 1927) a lui Russell și Whitehead și a găsit demonstrații noi și mai elegante pentru altele. Au introdus concepte critice în inteligența artificială, cum ar fi euristica, procesarea listelor, „raționamentul ca și căutare" etc. Simon considera că au „rezolvat venerabila problemă minte/corp, explicând modul în care un sistem compus din materie poate avea proprietățile minții."[40]

1956: Atelierul de lucru de la Dartmouth a marcat începutul oficial al inteligenței artificiale ca disciplină academică. A fost organizat de Marvin Minsky, John McCarthy, Claude Shannon și Nathan Rochester, având ca obiectiv aprofundarea posibilităților de a crea mașini capabile să simuleze inteligența umană, testând afirmația potrivit căreia „fiecare aspect al învățării sau orice altă trăsătură a inteligenței poate fi descrisă atât de precis încât să poată fi făcută o mașină să o simuleze". Termenul „inteligență artificială" a fost introdus oficial de John McCarthy, devenind numele domeniului științific Declarația principală a conferinței a fost: „*Fiecare aspect al oricărei altei caracteristici de învățare sau inteligență ar trebui să fie descris cu acuratețe, astfel încât mașina să o poată simula*" (Russell și Norvig 2016). În toamna anului 1956 a avut loc o întâlnire a Grupului de interes special în teoria informației de la Institutul de Tehnologie din Massachusetts care a fost începutul „revoluției cognitive", care folosește instrumente conexe diverselor domenii

[36] McCorduck, Pamela, *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*, Taylor & Francis, 2004, p. 98.
[37] Copeland, Jack, *A Brief History of Computing*, 2000, https://www.alanturing.net/turing_archive/pages/Reference%20Articles/BriefHistofComp.html
[38] McCorduck, Pamela, *op. cit.*, p. 102
[39] Whitehead Alfred North, Bertrand Russell, *Principia Mathematica*, Cambridge University Press, 1927.
[40] Russell Stuart, Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th US ed." 2016

pentru a modela mintea. Abordarea cognitivă ia în considerare „obiecte mentale” (gânduri, planuri, scopuri, fapte sau amintiri), analizate folosind simboluri de nivel înalt în rețele funcționale.

Abordări ale IA

În primele decenii de după nașterea IA au existat multe programe de succes și noi direcții, precum:

• Raționamentul ca o căutare: Astfel de programe IA timpurii au mers pas cu pas spre obiectiv. Problema era numărul de căi posibile extrem de mare („explozie combinatorie”), reducerea acestora făcându-se folosind euristici sau „reguli generale”, cu riscul de a nu ajunge la o soluție.

• Rețele neuronale: Frank Rosenblatt a construit mașini perceptron (1957-1962) de până la patru straturi. Bernard Widrow și Ted Hoff au folosit ponderi reglabile[41]. Charles A. Rosen și Alfred E. (Ted) Brain, cu finanțare de la U.S. Army Signal Corps, au folosit, de asemenea, greutăți reglabile, putând clasifica simboluri pe hărțile armatei și recunoaște caracterele imprimate manual pe foile de codare Fortran. Sistemele construite în această perioadă foloseau hardware personalizat.

• Limbajul natural: Un obiectiv al cercetării IA a fost să permită computerelor să comunice în limbi naturale. În acest scop s-au folosit rețele semantice reprezentând concepte ca noduri, și relații între concepte ca legături între noduri. În acest fel a fost construit primul chatterbot, ELIZA al lui Joseph Weizenbaum, care putea desfășura conversații foarte realiste.

• Micro-lumi: La finele anilor '60, Marvin Minsky și Seymour Papert au propus ca cercetarea IA să se concentreze pe situații simple artificiale cunoscute sub numele de micro-lumi (modele simplificate).

• Automate: În Japonia a fost construit primul robot umanoid „inteligent” (android) la Universitatea Waseda (WABOT) în perioada 1967-1972.

Definiții propuse de doctrină.

Potrivit unui punct de vedere[42] IA reprezintă acel domeniu de cercetare, care conține mai multe subdomenii, și care „încearcă să imite inteligența umană la mașini”. Printre subdomeniile cuprinse în IA se numără ”sistemele bazate pe cunoaștere, sistemele expert, recunoașterea tiparelor, învățarea automată, înțelegerea limbajului natural, robotica și altele”. Apoi, dintr-un alt

[41] Widrow, B., M.A. Lehr, 1990, *30 years of adaptive neural networks: perceptron, Madaline, and backpropagation*, Proceedings of the IEEE 78 (9): 1415–42. https://doi.org/10.1109/5.58323
[42] Ray Kurzweil, *op. cit.*, pp. 438-439

punct de vedere[43], se relevă că IA este o „inteligență nebiologică" creată de om. Potrivit unei alte păreri[44] „Inteligența artificială este elucidarea procesului uman de învățare, cuantificarea procesului uman de gândire, explicația comportamentului uman și înțelegerea misterelor care fac posibilă inteligența".

Printre numeroasele definiții (de ordin tehnic) propuse pentru IA putem enumera câteva[45]. Una dintre acestea identifică IA cu un câmp interdisciplinar teoretic și practic care are ca obiect înțelegerea mecanismelor de cunoaștere și de reflecție, și imitarea lor printr-un dispozitiv material și care funcționează în baza unui program, a unui soft, având ca scopuri asistența sau substituirea activităților umane. O alta prezintă IA ca fiind cercetarea mijloacelor susceptibile de a dota sistemele informatice cu capacități intelectuale comparabile cu cele ale ființelor umane. O a treia definiție din această categorie înfățișează IA ca fiind o ramură a informaticii consacrată dezvoltării sistemelor de prelucrare a datelor care îndeplinește funcții în mod normal asociate inteligenței umane, cum ar fi raționamentul, învățarea ș.a. De asemenea, potrivit unei a patra definiții propuse, conceptul de IA desemnează acea capacitate a unei unități funcționale de a efectua funcții de regulă asociate inteligenței umane, precum raționamentul și învățarea. În fine, noțiunea de IA rezidă în punerea în aplicare a unui anumit număr de tehnici care au ca obiectiv să permită mașinilor să imite o formă de inteligență reală. În acest fel, IA se regăsește implementată într-un număr în creștere de zone de aplicare.

Dacă ne referim la IA la nivel uman, cunoscută și sub denumirea de IA generală (IAG), aceasta rezidă în „Capacitatea de a îndeplini orice sarcină cognitivă, cel puțin la fel de bine ca oamenii"[46]. Detaliind, inteligența artificială generală poate funcționa ca un sistem autonom, iar o entitate care deține IAG „poate raționa, poate folosi strategii, poate rezolva puzzle-uri (…) Mai mult, poate emite judecăți în condiții de incertitudine, luând chiar decizii pe marginea acestor judecăți"[47]; după cum, IA denumită super inteligență este cea care se situează „mult dincolo de nivelul uman". Și mai este IA îngustă (Narrow AI) ori IA aplicată care reprezintă stadiul actual al IA, adică toate înfăptuirile umane din domeniul IA de până acum. Aceasta „se ocupă cu rezolvarea

[43] Max Tegmark, *Viața 3.0. Omul în epoca inteligenței artificiale*, Ed. Humanitas, București, 2019, p. 49.
[44] Kai-Fu Lee, *Superputerile inteligenței artificiale: China, Silicon Valley și noua ordine mondială,* Ed. Corint Future, București, 2021, pp. 23.
[45] *ChatGPT dans le monde du droit. La cobotique juridique*, Éditions Bruylant, Bruxelles, 2024, p.17.
[46] Max Tegmark, *op. cit.*, p. 49
[47] Răzvan Boboc, *Ce este Inteligența Artificială Generală? Între inovație și amenințare pentru omenire*, https://www.euronews.ro/articole/ce-este-inteligenta-generala-artificiala-intre-inovatie-si-amenintare-pentru-omen

unei probleme predefinite, cum ar fi jocuri, identificarea imaginilor sau conducerea unei mașini. Narrow AI (s. orig.) este foarte util în sine și poate avea efecte mai mari asupra societății, făcând lucrătorii mai eficienți și ducând la automatizarea sarcinilor. Cu toate acestea nu presupune o inteligență pe deplin conștientă, de tipul celei umane". Într-o altă formulare, conceptul de IA desemnează „entitatea care are abilitatea de a percepe informații din mediul înconjurător, de a le interpreta, de a învăța din respectivele informații şi de a întreprinde acțiuni pentru a-şi creşte șansele de a-și îndeplini cu succes obiectivele".

Grupul de experți la nivel înalt privind inteligența artificială al Comisiei Europene (AI HLEG) a formulat următoarea definiție, care ulterior a făcut obiectul unor discuții suplimentare în cadrul grupului: „Inteligența artificială (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. Sistemele (…) IA se pot baza exclusiv pe software uri, acționând în lumea virtuală (de exemplu, asistenți vocali, software de analiză a imaginii, motoare de căutare, sisteme de recunoaștere vocală și facială), sau IA poate fi încorporată în dispozitive hardware (spre exemplu, roboți avansați, vehicule autonome, drone sau aplicații pentru internetul obiectelor)"[48].

Deopotrivă, s-a mai arătat că „Inteligența artificială (…) este aptitudinea unei mașini de a imita comportamentul uman, fiind programată să gândească și să acționeze asemenea unui om. Una dintre caracteristicile esențiale ale IA este învățarea continuă, pe baza stimulilor externi și a informațiilor colectate din mediul înconjurător IA observă realitatea înconjurătoare și acționează în consecință, fără a avea nevoie de ajutorul sau asistența umană". Totodată, s-a subliniat că această definiție cu toate că „este una actuală, trebuie să avem în vedere că standardul la care ne raportăm pentru a stabili ce este IA se află într-o evoluție continuă, asemenea tehnologiei.

Problema definirii Inteligenței Artificiale nu este încheiată nici azi, datorită evoluțiilor tehnologice din domeniu. Apreciem că o soluție în privința unei definiții cuprinzătoare a IA nu va fi găsită în curând, întrucât specialiștii în domeniu (ne referim aici la ingineri, în special) par să vină cu noi idei. Spre exemplu, "roboții" sunt incluși în diversele definiții ale IA; dar roboții, într-un sens restrâns, ar fi "corpul", în timp ce IA ar fi "creierul" . Ca să exemplificăm, robotul Da Vinci este menționat în lucrările de specialitate care discută despre IA; în principiu, robotul Da

[48] Agenția pentru Drepturi Fundamentale a Uniunii Europene, Raportul „*Înțelegerea viitorului. Inteligența artificială și drepturile fundamentale*" (rezumat), Luxemburg, Oficiul pentru Publicații al UE, 2021

Vinci este un sistem care permite doctorilor să facă intervenții chirurgicale de la distanță, fără a atinge ei înșiși pacientul. Dar, acest sistem chirurgical (Da Vinci) nu are autonomie în luarea deciziilor și nu este dotat cu IA, dar totuși este menționat în lucrările de specialitate.

Capitolul II.

1. Declarația Bletchley;

2. Declarația de la Montreal privind dezvoltarea responsabilă a inteligenței artificiale;

3. Principiile Asilomar pentru Inteligența Artificială; 4. Principiile Organizației pentru Cooperare și Dezvoltare Economică (OECD) pentru inteligența artificială (AI)

1. ***Declarația de la Bletchley a fost adoptată de țările participante la Summitul pentru Siguranța Inteligenței Artificiale desfășurat în perioada 1-2 noiembrie 2023.***

Inteligența artificială (IA) oferă oportunități globale enorme: are potențialul de a transforma și îmbunătăți bunăstarea umană, pacea și prosperitatea. Pentru a realiza acest lucru, afirmăm că, în beneficiul tuturor, IA ar trebui să fie concepută, dezvoltată, implementată și utilizată într-un mod sigur, centrat pe om, de încredere și responsabil. Salutăm eforturile comunității internaționale de până acum de a coopera pentru a promova o creștere economică incluzivă, o dezvoltare durabilă și inovarea, pentru a proteja drepturile omului și libertățile fundamentale, și pentru a încuraja încrederea publică în sistemele IA, astfel încât să le valorificăm pe deplin potențialul.

Sistemele IA sunt deja utilizate în multe domenii ale vieții cotidiene, inclusiv locuințe, ocuparea forței de muncă, transport, educație, sănătate, accesibilitate și justiție, iar utilizarea acestora este probabil să crească. Recunoaștem că acesta este un moment unic pentru a acționa și subliniem necesitatea dezvoltării sigure a IA și utilizarea oportunităților sale transformative într-un mod incluziv, în beneficiul tuturor, atât în țările noastre, cât și la nivel global. Aceasta include utilizarea în servicii publice precum sănătatea și educația, securitatea alimentară, știință, energie curată, biodiversitate și climă, pentru a sprijini drepturile omului și a contribui la atingerea Obiectivelor de Dezvoltare Durabilă ale Națiunilor Unite.

Riscurile și necesitatea acțiunii.

Pe lângă oportunități, IA prezintă și riscuri semnificative, inclusiv în aceste domenii ale vieții cotidiene. Salutăm eforturile internaționale relevante pentru a analiza și aborda impactul potențial al sistemelor IA și recunoaștem necesitatea protejării drepturilor omului, transparenței,

echității, responsabilității, eticii, siguranței, supravegherii umane, protecției datelor și intimității. Subliniem, de asemenea, riscurile legate de manipularea sau generarea de conținut înșelător.
Riscuri speciale apar la „frontiera” IA, reprezentată de modele generaliste extrem de capabile, inclusiv modele fundamentale care pot realiza o gamă largă de sarcini. Acestea pot amplifica riscuri precum dezinformarea și pot provoca daune grave, intenționate sau neintenționate. Datorită ritmului rapid de dezvoltare a IA, trebuie avută în vedere urgența înțelegerii riscurilor potențiale și luării de măsuri pentru a le aborda.

Cooperarea internațională și responsabilitatea.

Riscurile IA sunt în mod inerent internaționale și necesită cooperare internațională. Ne angajăm să lucrăm împreună pentru a asigura o IA centrată pe om, de încredere și sigură, prin forumuri internaționale existente și alte inițiative relevante. Recunoaștem, de asemenea, importanța guvernanței și reglementării proporționale care să maximizeze beneficiile IA, luând în considerare riscurile.

Actorii implicați în dezvoltarea IA de frontieră au o responsabilitate deosebită pentru siguranța acestor sisteme, inclusiv prin testare, evaluare și alte măsuri adecvate.

Direcții viitoare.

În cadrul cooperării noastre, agenda pentru abordarea riscurilor IA de frontieră se va concentra pe:

Identificarea riscurilor și construirea unei înțelegeri bazate pe știință.

Dezvoltarea de politici bazate pe riscuri în țările noastre, colaborând acolo unde este cazul.

Sprijinirea unei rețele internaționale de cercetare în siguranța IA, în beneficiul politicilor și al binelui public.

Recunoaștem potențialul pozitiv al IA și ne angajăm să menținem un dialog global incluziv care să contribuie la discuțiile internaționale mai ample și să asigure utilizarea responsabilă a tehnologiei pentru binele tuturor.

Țările reprezentate:

Australia, Brazilia, Canada, Chile, China, Uniunea Europeană, Franța, Germania, India, Indonezia, Irlanda, Israel, Italia, Japonia, Kenya, Arabia Saudită, Țările de Jos, Nigeria, Filipine, Coreea de Sud, Rwanda, Singapore, Spania, Elveția, Turcia, Ucraina, Emiratele Arabe Unite, Regatul Unit, Statele Unite ale Americii.

Contextul Declarației

Momentul ales are în vedere ritmul accelerat al dezvoltării IA: Ritmul rapid al progresului în domeniul IA a generat atât entuziasm, cât și preocupări legate de riscuri potențiale, precum pierderea controlului, amplificarea dezinformării, sau utilizarea malițioasă a tehnologiei.

Frontiera IA: Modelele de frontieră (de exemplu, modele generative precum GPT-4 sau mai avansate) prezintă riscuri unice, inclusiv abilitatea de a produce conținut fals, manipulativ sau de a impacta infrastructuri critice.

Impactul global: IA este deja integrată în domenii critice (sănătate, educație, justiție etc.) și influențează viața cotidiană. Se anticipează o extindere rapidă a utilizării, ceea ce crește necesitatea reglementării.

Semnificația Bletchley Park

Locul summitului are o semnificație istorică: Bletchley Park a fost centrul criptografic al Regatului Unit în timpul celui de-al Doilea Război Mondial, unde au fost descifrate codurile Enigma. Alegerea acestui loc simbolizează importanța tehnologiei în securitate și colaborare globală.

Obiectivele principale ale summitului: Crearea unui cadru pentru cooperare internațională privind siguranța IA. Abordarea riscurilor asociate cu modelele IA extrem de capabile. Echilibrarea necesității de reglementare cu promovarea inovației.

Puncte-cheie din Declarație. Oportunități și beneficii

IA are potențialul de a transforma sectoare precum sănătatea, educația, energia curată și biodiversitatea. IA poate sprijini realizarea Obiectivelor de Dezvoltare Durabilă ale ONU, reducând decalajele de dezvoltare și promovând creșterea economică incluzivă.

Riscuri identificate

Manipularea și dezinformarea: Generarea de conținut fals sau înșelător. *Risc de securitate*: Utilizarea IA în atacuri cibernetice sau în biotehnologie. *Pierderea controlului*: Modelele avansate pot deveni dificil de înțeles sau gestionat. *Impact asupra drepturilor omului*: Potențial de discriminare, amplificarea inechităților sau invazii ale intimității.

a) Manipularea și dezinformarea

Modelele generative IA, cum ar fi cele capabile să creeze imagini, texte sau videoclipuri, pot produce conținut fals foarte convingător. Implicațiile includ amplificarea propagandei, manipularea electorală și subminarea încrederii în informațiile publice.

b) Securitatea cibernetică.

Frontier AI (modelele avansate) ar putea fi folosite pentru a genera atacuri cibernetice sofisticate. Exemple includ: Crearea de malware avansat. Exploatarea vulnerabilităților în infrastructurile critice, cum ar fi rețelele energetice sau financiare.

c) *Controlul și pierderea intenției umane.*

Modelele IA avansate, în special cele care nu sunt complet înțelese, ar putea acționa într-un mod neintenționat sau nesigur, punând în pericol utilizatorii sau societatea în ansamblu.

d) Impact asupra pieței muncii.

Înlocuirea locurilor de muncă de către IA reprezintă un risc major. Domeniile vulnerabile includ transportul (vehicule autonome), customer support, dar și profesii creative (datorită IA generativă).

Responsabilitatea actorilor-cheie.

Dezvoltatorii de modele IA avansate trebuie să implementeze măsuri stricte de testare și monitorizare. Guvernele și organizațiile internaționale trebuie să colaboreze pentru a reglementa IA într-un mod proporțional și bazat pe risc. Este necesară implicarea societății civile și a mediului academic pentru a asigura o IA incluzivă și etică.

Acțiuni planificate.

Dezvoltarea unor rețele internaționale pentru cercetare în siguranța IA. Crearea unui cadru de politici bazate pe evaluarea riscurilor specifice. Organizarea viitoarelor Summituri pentru Siguranța IA, începând cu anul 2024.

Implicații pentru reglementare și guvernanță.

a) *Necesitatea clasificării riscurilor.*

Declarația subliniază că reglementarea IA trebuie să fie proporțională cu riscurile implicate. Exemple: Modelele utilizate în sectorul sănătății ar necesita testări mai riguroase decât cele din divertisment. Dezvoltatorii de IA de frontieră ar trebui să respecte protocoale stricte de siguranță, cum ar fi transparența privind utilizările posibile.

b) *Rolul cadrelor juridice internaționale.*

Declarația sugerează extinderea reglementărilor internaționale, posibil prin cooperare cu organizații existente precum ONU, OECD sau G20. Este recunoscut faptul că reglementările trebuie adaptate la contexte naționale, dar standardele globale sunt esențiale pentru a preveni utilizările abuzive ale IA.

Provocări în implementarea Declarației.

a) Divergențe între țări.

Țările au abordări diferite privind guvernanța IA: UE se concentrează pe reglementare strictă (Legea IA din UE). SUA promovează inovația, favorizând autoreglementarea. China are un control centralizat, dar și ambiții mari în dezvoltarea IA.

b) Lipsa unor mecanisme de implementare imediate.

Declarația este un document de principii, fără măsuri obligatorii. Pentru ca obiectivele să devină realitate, va fi nevoie de acorduri ulterioare concrete.

c) Resurse necesare.

Țările în curs de dezvoltare ar putea întâmpina dificultăți în respectarea standardelor globale fără sprijin financiar și tehnologic.

Semnificația globală a Declarației.

Colaborare internațională. Declarația subliniază că multe riscuri legate de IA sunt transfrontaliere și necesită soluții globale. Implicarea unor țări precum SUA, China și Uniunea Europeană arată dorința unui consens global, chiar dacă abordările naționale pot varia. Echilibru între reglementare și inovare. Este recunoscută importanța reglementării pentru a preveni riscurile, dar fără a frâna inovația. Documentul sugerează abordări flexibile, adaptate contextelor naționale și juridice. *Rolul frontierelor IA*. Modelele generative și cele generaliste sunt văzute ca având cel mai mare potențial disruptiv. Prin urmare, ele sunt în centrul preocupărilor de siguranță. Declarația reflectă un moment rar de cooperare internațională între țări cu agende și priorități foarte diferite, cum ar fi: SUA și China: Două superputeri tehnologice care, deși sunt în competiție strategică, au participat la discuții. Este un semn că siguranța IA este recunoscută ca o problemă globală ce depășește rivalitățile tradiționale.

Statele din Sudul Global: Țări precum Nigeria, Kenya, Rwanda sau Filipine au fost incluse, subliniind importanța asigurării accesului echitabil la beneficii și tehnologii IA, precum și nevoia de sprijin pentru dezvoltarea capacităților lor tehnologice.

Această abordare incluzivă este esențială pentru a evita adâncirea decalajului digital și pentru a preveni inechități în utilizarea IA.

Critici și provocări. Lipsa unor măsuri concrete imediate. Declarația este mai degrabă un document de principii decât un plan de acțiune specific. *Tensiuni geopolitice*: Implicarea unor țări cu perspective diferite asupra reglementării IA (ex. China și SUA) poate face dificilă adoptarea

unor standarde globale uniforme. *Provocări de implementare*: Punerea în practică a unor măsuri precum testarea standardizată a siguranței IA sau prevenirea dezinformării necesită resurse considerabile și cooperare extinsă.

2. Declarația de la Montreal privind dezvoltarea responsabilă a inteligenței artificiale

Preambul: Pentru prima dată în istoria omenirii, este posibil să se creeze sisteme autonome capabile să îndeplinească sarcini complexe pe care se considera că doar inteligența naturală le poate realiza: procesarea unor cantități mari de informații, calcularea și prezicerea, învățarea și adaptarea răspunsurilor la situații în schimbare și recunoașterea și clasificarea obiectelor. Având în vedere natura imaterială a acestor sarcini și prin analogie cu inteligența umană, desemnăm aceste sisteme extinse sub denumirea generală de inteligență artificială. Inteligența artificială constituie o formă majoră de progres științific și tehnologic, care poate genera beneficii sociale considerabile prin îmbunătățirea condițiilor de viață și a sănătății, facilitarea justiției, crearea de bogăție, consolidarea siguranței publice și atenuarea impactului activităților umane asupra mediului și climatului. Mașinile inteligente nu sunt limitate doar la realizarea unor calcule mai bune decât oamenii; ele pot interacționa și cu ființele sensibile, le pot ține companie și le pot îngriji. Cu toate acestea, dezvoltarea inteligenței artificiale ridică provocări etice majore și riscuri sociale. De fapt, mașinile inteligente pot restricționa alegerile indivizilor și grupurilor, pot reduce nivelul de trai, pot perturba organizarea muncii și piața muncii, pot influența politica, pot intra în conflict cu drepturile fundamentale, pot exacerba inegalitățile sociale și economice și pot afecta ecosistemele, clima și mediul înconjurător. Deși progresul științific și viața într-o societate presupun întotdeauna un risc, este responsabilitatea cetățenilor să determine scopurile morale și politice care dau sens riscurilor întâlnite într-o lume incertă. Cu cât riscurile implementării sale sunt mai mici, cu atât beneficiile inteligenței artificiale vor fi mai mari. Primul pericol al dezvoltării inteligenței artificiale constă în a crea iluzia că putem stăpâni viitorul prin calcule. Reducerea societății la o serie de numere și guvernarea acesteia prin proceduri algoritmice este un vechi vis iluzoriu care încă motivează ambițiile umane. Dar, în ceea ce privește afacerile umane, ziua de mâine rareori seamănă cu cea de azi, iar numerele nu pot stabili ce are valoare morală, nici ce este de dorit din punct de vedere social.

Principiile din prezenta declarație sunt ca niște puncte pe o busolă morală, care vor ajuta la ghidarea dezvoltării inteligenței artificiale spre scopuri moral și social dorite. Ele oferă, de asemenea, un cadru etic care promovează drepturile fundamentale ale omului, recunoscute internațional, în domeniile afectate de implementarea inteligenței artificiale. Privite în ansamblu, principiile exprimate pun bazele cultivării încrederii sociale în sistemele inteligente artificiale.

Principiile din prezenta declarație se bazează pe convingerea comună că ființele umane doresc să crească ca ființe sociale înzestrate cu senzații, gânduri și sentimente, și tind să-și îndeplinească potențialul prin exercițiul liber al capacităților lor emoționale, morale și intelectuale. Este responsabilitatea diferiților actori publici și privați, precum și a factorilor de decizie de la nivel local, național și internațional, să se asigure că dezvoltarea și implementarea inteligenței artificiale sunt compatibile cu protejarea capacităților și obiectivelor fundamentale ale oamenilor și contribuie la realizarea acestora într-o măsură mai mare. Având acest scop în vedere, principiile propuse trebuie interpretate într-o manieră coerentă, ținând cont de contextul social, cultural, politic și legal specific al aplicării lor.

Adoptată în 2018, Declarația a fost inițiată de Institutul de Etică Aplicată (Institut d'éthique appliquée) din cadrul Universității din Montreal. Este rezultatul unui proces colaborativ care a implicat experți din diverse domenii (filosofie, tehnologie, drept, economie), dar și publicul larg. Scopul său principal este de a oferi principii pentru guvernarea dezvoltării și utilizării inteligenței artificiale, având în vedere impactul social, economic și etic al acesteia.

Principii fundamentale ale Declarației

Declarația de la Montreal este structurată în jurul mai multor principii esențiale:

a) Binele social.

Sistemele de inteligență artificială (AIS) trebuie să ajute indivizii să își îmbunătățească condițiile de viață, sănătatea și condițiile de muncă. AIS trebuie să permită indivizilor să își urmeze preferințele, atâta timp cât acestea nu cauzează daune altor ființe sensibile. AIS trebuie să permită oamenilor să își exercite capacitățile mentale și fizice. AIS nu trebuie să devină o sursă de suferință, decât dacă permite obținerea unei bunăstări superioare față de ceea ce s-ar putea obține altfel. Utilizarea AIS nu trebuie să contribuie la creșterea stresului, anxietății sau la sentimentul de a fi hărțuit de mediul digital.

IA ar trebui să fie folosită pentru a îmbunătăți calitatea vieții și bunăstarea generală a oamenilor, contribuind la reducerea inegalităților sociale și economice. În ceea ce privește

implicarea practică a acestui principiu, notăm: Politicile publice: Implementarea IA în sănătate, educație și transport pentru a îmbunătăți accesul și calitatea serviciilor. Exemple: Algoritmi pentru diagnosticare timpurie a bolilor. Platforme educaționale personalizate pentru elevi în zone defavorizate. Reducerea inegalităților: IA poate analiza date pentru a identifica grupuri vulnerabile și pentru a îmbunătăți accesul lor la resurse esențiale, cum ar fi locuințele sau locurile de muncă.

b) Respectul pentru autonomie.

Sistemele de inteligență artificială (AIS) trebuie să permită indivizilor să își îndeplinească propriile obiective morale și propria concepție despre o viață demnă de trăit. AIS nu trebuie să fie dezvoltate sau utilizate pentru a impune un anumit stil de viață asupra indivizilor, fie direct, fie indirect, prin implementarea unor mecanisme de supraveghere și evaluare opresive sau prin mecanisme de stimulente. Instituțiile publice nu trebuie să utilizeze AIS pentru a promova sau discredita o anumită concepție despre viața bună. Este esențial să se încurajeze cetățenii să dobândească cunoștințele necesare în ceea ce privește tehnologiile digitale, promovând învățarea abilităților fundamentale (alfabetizare digitală și media) și dezvoltarea gândirii critice. AIS nu trebuie să fie dezvoltate pentru a răspândi informații de încredere scăzută, minciuni sau propagandă și trebuie să fie concepute cu scopul de a limita răspândirea acestora. Dezvoltarea AIS trebuie să evite crearea de dependențe prin tehnici de captare a atenției sau prin imitarea caracteristicilor umane (aspect, voce etc.) în moduri care ar putea cauza confuzie între AIS și oameni.

Tehnologia nu trebuie să limiteze libertatea de alegere a indivizilor. IA trebuie proiectată astfel încât să sprijine și să respecte deciziile umane, fără a manipula sau constrânge, manipularea subtilă fiind una din provocările căreia trebuie să-i facă față orice utilizator.

c*) Protecția vieții private.*

Spațiile personale în care oamenii nu sunt supuși supravegherii sau evaluării digitale trebuie protejate de intruziunea AIS și a sistemelor de achiziție și arhivare a datelor (DAAS). Intimitatea gândurilor și emoțiilor trebuie protejată strict de utilizările AIS și DAAS care ar putea cauza daune, în special acele utilizări care impun judecăți morale asupra oamenilor sau alegerilor lor de stil de viață. Oamenii trebuie să aibă întotdeauna dreptul la deconectare digitală în viața lor privată, iar AIS ar trebui să ofere explicit opțiunea de a se deconecta la intervale regulate, fără a încuraja oamenii să rămână conectați. Oamenii trebuie să aibă control extins asupra informațiilor privind preferințele lor. AIS nu trebuie să creeze profiluri individuale de preferințe pentru a

influența comportamentul indivizilor fără consimțământul lor liber și informat. DAAS trebuie să garanteze confidențialitatea datelor și anonimitatea profilului personal. Fiecare persoană trebuie să poată exercita control extins asupra datelor sale personale, în special în ceea ce privește colectarea, utilizarea și diseminarea acestora. Accesul la AIS și la serviciile digitale de către indivizi nu trebuie să fie condiționat de abandonarea controlului sau a proprietății asupra datelor lor personale. Indivizii ar trebui să fie liberi să își doneze datele personale organizațiilor de cercetare pentru a contribui la avansarea cunoașterii. Integritatea identității personale trebuie garantată. AIS nu trebuie să fie utilizate pentru a imita sau modifica aspectul, vocea sau alte caracteristici individuale ale unei persoane în scopul de a-i dăuna reputației sau de a manipula alte persoane.

IA trebuie să respecte confidențialitatea și să protejeze datele personale ale utilizatorilor. Este important să existe transparență în modul în care sunt colectate, stocate și utilizate datele.

d) Solidaritate.

Sistemele de inteligență artificială (AIS) nu trebuie să amenințe păstrarea relațiilor umane morale și emoționale satisfăcătoare și trebuie dezvoltate cu scopul de a sprijini aceste relații și de a reduce vulnerabilitatea și izolarea oamenilor. AIS trebuie dezvoltate cu scopul de a colabora cu oamenii în sarcini complexe și trebuie să sprijine munca colaborativă între oameni. AIS nu trebuie implementate pentru a înlocui oamenii în îndatoririle care necesită relații umane de calitate, ci trebuie dezvoltate pentru a facilita aceste relații. Sistemele de sănătate care folosesc AIS trebuie să ia în considerare importanța relațiilor pacientului cu familia și cu personalul medical. Dezvoltarea AIS nu trebuie să încurajeze comportamentele crude față de roboții care sunt concepuți să semene cu ființele umane sau cu animale non-umane în aspectul sau comportamentul lor. AIS trebuie să ajute la îmbunătățirea gestionării riscurilor și să sprijine condițiile pentru o societate cu o distribuție mai echitabilă și mutuală a riscurilor individuale și colective.

e) *Sistemul IA trebuie să îndeplinească criteriile de inteligibilitate, justificabilitate și accesibilitate și trebuie supus controlului, dezbaterii și scrutinului democratic. Principiul participării democratice*: Procesele AIS care iau decizii ce afectează viața unei persoane, calitatea vieții sau reputația acesteia trebuie să fie inteligibile pentru creatorii lor.

Deciziile luate de AIS care afectează viața unei persoane, calitatea vieții sau reputația acesteia ar trebui să fie întotdeauna justificabile într-un limbaj înțeles de persoanele care le utilizează sau care sunt supuse consecințelor utilizării lor. Justificarea constă în a face transparente

cele mai importante factori și parametri care formează decizia, iar aceasta ar trebui să ia aceeași formă ca justificarea pe care am solicita-o unui om care ia același tip de decizie.

Codul pentru algoritmi, fie că sunt publici sau privați, trebuie să fie întotdeauna accesibil autorităților publice și părților interesate relevante pentru scopuri de verificare și control.

Descoperirea erorilor de funcționare ale AIS, a efectelor neașteptate sau nedorite, a breșelor de securitate și a scurgerilor de date trebuie să fie raportată imperativ autorităților publice relevante, părților interesate și celor afectați de situație.

Conform cerinței de transparență pentru deciziile publice, codul algoritmilor de luare a deciziilor folosiți de autoritățile publice trebuie să fie accesibil tuturor, cu excepția algoritmilor care prezintă un risc mare de pericol grav în cazul unei utilizări abuzive.

Pentru AIS publice care au un impact semnificativ asupra vieții cetățenilor, cetățenii ar trebui să aibă oportunitatea și abilitățile necesare pentru a delibera asupra parametrilor sociali ai acestor AIS, obiectivelor lor și limitelor utilizării lor.

Trebuie să putem verifica în orice moment dacă AIS fac ceea ce au fost programate să facă și ceea ce sunt folosite să facă. Oricine folosește un serviciu ar trebui să știe dacă o decizie care îi privește sau care îi afectează a fost luată de un AIS.

Oricare utilizator al unui serviciu care utilizează chatboți ar trebui să poată identifica cu ușurință dacă interacționează cu un AIS sau cu o persoană reală.

Cercetarea în domeniul inteligenței artificiale ar trebui să rămână deschisă și accesibilă tuturor.

f) Echitate.

AIS trebuie să fie proiectate și instruite astfel încât să nu creeze, să nu întărească și să nu reproducă discriminarea pe baza diferențelor sociale, sexuale, etnice, culturale sau religioase, printre altele. Dezvoltarea AIS trebuie să ajute la eliminarea relațiilor de dominație între grupuri și oameni bazate pe diferențe de putere, bogăție sau cunoaștere.

Dezvoltarea AIS trebuie să genereze beneficii sociale și economice pentru toți, reducând inegalitățile sociale și vulnerabilitățile. Dezvoltarea industrială a AIS trebuie să fie compatibilă cu condițiile de muncă acceptabile în fiecare etapă a ciclului lor de viață, de la extragerea resurselor naturale până la reciclare, inclusiv procesarea datelor.

Activitatea digitală a utilizatorilor de AIS și de servicii digitale trebuie recunoscută ca muncă ce contribuie la funcționarea algoritmilor și creează valoare. Accesul la resurse fundamentale, cunoștințe și instrumente digitale trebuie garantat pentru toți.

Ar trebui să susținem dezvoltarea algoritmilor de tip commons — și a datelor deschise necesare pentru a-i instrui — și extinderea utilizării lor ca obiectiv social echitabil.

Dezvoltarea IA trebuie să promoveze echitatea și să evite discriminarea. Algoritmii trebuie să fie proiectați astfel încât să nu reproducă sau să amplifice prejudecăți existente.

g) *Diversitate și incluziune*.

AIS nu trebuie să conducă la omogenizarea societății prin standardizarea comportamentului și opiniilor. În momentul în care sunt concepute, dezvoltarea și implementarea AIS trebuie să ia în considerare multitudinea de expresii ale diversității sociale și culturale prezente în societate. Mediile de dezvoltare a AIS, fie în cercetare, fie în industrie, trebuie să fie incluzive și să reflecte diversitatea indivizilor și grupurilor din societate. AIS nu trebuie să folosească datele obținute pentru a închide indivizii într-un profil de utilizator, pentru a fixa identitatea personală sau pentru a-i închide într-o bulă de filtrare care le-ar restrânge și limita posibilitățile de dezvoltare personală — mai ales în domenii precum educația, justiția sau afacerea. AIS nu trebuie să fie dezvoltate sau utilizate cu scopul de a limita exprimarea liberă a ideilor sau oportunitatea de a auzi opinii diverse, ambele fiind condiții esențiale ale unei societăți democratice. Pentru fiecare categorie de servicii, oferta AIS trebuie diversificată pentru a preveni formarea de monopoluri de facto care ar submina libertățile individuale.

h) Prudență.

Este necesar să se dezvolte mecanisme care să ia în considerare potențialul de utilizare dublă — benefică și dăunătoare — a cercetării în domeniul AI și a dezvoltării AIS (fie publice, fie private) pentru a limita utilizările dăunătoare. Atunci când utilizarea greșită a unui AIS pune în pericol sănătatea publică sau siguranța și are o probabilitate ridicată de a se întâmpla, este prudent să se restricționeze accesul deschis și difuzarea publică a algoritmului acestuia.

Înainte de a fi plasate pe piață, indiferent dacă sunt oferite contra cost sau gratuit, AIS trebuie să îndeplinească cerințe stricte de fiabilitate, securitate și integritate și trebuie să fie supuse unor teste care să nu pună în pericol viața oamenilor, să nu le afecteze calitatea vieții sau să le dăuneze reputației ori integrității psihologice. Aceste teste trebuie să fie deschise autorităților publice relevante și părților interesate.

Dezvoltarea AIS trebuie să prevină riscurile de utilizare abuzivă a datelor utilizatorilor și să protejeze integritatea și confidențialitatea datelor personale. Erorile și deficiențele descoperite în AIS și SAAD (Sisteme de Achiziție și Arhivare de Date) ar trebui să fie împărtășite public, la nivel global, de către instituțiile publice și întreprinderile din sectoarele care prezintă un pericol semnificativ pentru integritatea personală și organizarea socială.

i) Responsabilitate.

Numai ființele umane pot fi trase la răspundere pentru deciziile care provin din recomandările făcute de AIS și pentru acțiunile care urmează din acestea.

În toate domeniile în care trebuie luată o decizie ce afectează viața unei persoane, calitatea vieții sau reputația acesteia, atunci când timpul și circumstanțele permit, decizia finală trebuie să fie luată de o ființă umană și acea decizie trebuie să fie liber aleasă și informată. Decizia de a ucide trebuie să fie întotdeauna luată de ființe umane, iar responsabilitatea pentru această decizie nu trebuie transferată unui AIS.

Persoanele care autorizează un AIS să comită o crimă sau o infracțiune, sau care demonstrează neglijență permițând unui AIS să le comită, sunt responsabile pentru această crimă sau infracțiune. Acest principiu afirmă că, în ciuda recomandărilor oferite de un AIS, oamenii trebuie să rămână responsabili pentru deciziile finale și pentru acțiunile care urmează acestor recomandări.

În toate domeniile în care trebuie luată o decizie care afectează viața unei persoane, calitatea vieții sau reputația sa, atunci când timpul și circumstanțele permit, decizia finală trebuie să fie luată de o ființă umană și acea decizie trebuie să fie liber aleasă și informată. Se subliniază că, în cazuri importante care afectează viața unui individ, deciziile finale trebuie să rămână în mâinile oamenilor și să fie bine informate, fără a fi complet automatizate de către AIS.

Acest principiu face referire la interdicția de a delega responsabilitatea unei decizii extreme, cum ar fi aceea de a lua viața cuiva, unui AIS. În mod clar, deciziile legate de viață și moarte trebuie să fie controlate de oameni.

Persoanele care autorizează un AIS să comită o crimă sau o infracțiune, sau care demonstrează neglijență prin permiterea unui AIS să le comită, sunt responsabile pentru această crimă sau infracțiune. Aici se afirmă că persoanele care dau permisiunea unui AIS de a comite acte ilegale, sau care nu previn aceste acte, vor fi considerate responsabile pentru faptele respective.

Când daunele sau răul sunt cauzate de un AIS, și dacă AIS este dovedit a fi fiabil și a fost utilizat conform scopului său, nu este rezonabil să punem vina pe persoanele implicate în dezvoltarea sau utilizarea acestuia. Acest principiu se referă la faptul că, dacă un AIS a fost folosit corect și a produs un rezultat dăunător, responsabilitatea nu ar trebui să fie atribuită dezvoltatorilor sau utilizatorilor AIS, în cazul în care aceștia nu au contribuit direct la daune.

j) Dezvoltarea sustenabilă.

Hardware-ul AIS, infrastructura sa digitală și obiectele relevante pe care se bazează, cum ar fi centrele de date, trebuie să urmărească obținerea celei mai mari eficiențe energetice și să atenueze emisiile de gaze cu efect de seră pe întreaga durată de viață a acestora. Hardware-ul AIS, infrastructura sa digitală și obiectele relevante pe care se bazează trebuie să urmărească generarea celor mai puține deșeuri electrice și electronice și să asigure proceduri de întreținere, reparare și reciclare conform principiilor economiei circulare. Hardware-ul AIS, infrastructura sa digitală și obiectele relevante pe care se bazează trebuie să minimizeze impactul nostru asupra ecosistemelor și biodiversității în fiecare etapă a ciclului său de viață, în special în ceea ce privește extracția resurselor și eliminarea finală a echipamentului atunci când acesta ajunge la sfârșitul duratei sale de viață utilă.

Actorii publici și privați trebuie să sprijine dezvoltarea responsabilă din punct de vedere ecologic a AIS pentru a combate risipa de resurse naturale și bunuri produse, pentru a construi lanțuri de aprovizionare și comerț durabile și pentru a reduce poluarea globală.

Obiective specifice ale Declarației:

Prevenirea utilizării abuzive a IA. Declarația subliniază riscurile utilizării IA în scopuri care pot pune în pericol drepturile omului, cum ar fi supravegherea în masă, războiul autonom sau manipularea politică. *Reducerea decalajului digital*. Este recunoscut faptul că IA ar putea adânci inegalitățile între țările dezvoltate și cele în curs de dezvoltare, iar acest lucru trebuie prevenit prin politici incluzive. *Promovarea unei reglementări etice*. Declarația propune implicarea guvernelor și a comunităților în stabilirea unui cadru normativ pentru utilizarea IA, cu accent pe protecția drepturilor fundamentale.

Comparație cu Declarația de la Bletchley. Deși ambele declarații împărtășesc preocupări similare, există câteva diferențe notabile: Montreal se concentrează pe principii etice generale, valabile în toate etapele dezvoltării IA. Bletchley are o abordare mai specifică, axată pe riscurile IA de frontieră și pe cooperarea internațională pentru siguranță.

Exemple concrete de aplicare.

Justiție: Sistemele IA sunt utilizate pentru a evalua riscurile asociate eliberării condiționate. Aplicarea Declarației ar presupune: Testarea acestor sisteme pentru a evita discriminarea pe bază de rasă sau statut socio-economic. Oferirea unor explicații clare pentru deciziile IA către judecători și părțile implicate.

Educație: IA poate personaliza procesul de învățare, dar: Datele elevilor trebuie protejate pentru a evita utilizarea lor în scopuri comerciale. Algoritmii trebuie verificați pentru a evita favorizarea unor grupuri sau stiluri de învățare.

Recrutare: Sistemele IA sunt folosite pentru screeningul candidaților, dar: este esențial să fie auditate pentru a preveni excluderea unor categorii (ex: femei în domenii tehnice). Companiile ar trebui să ofere candidaților acces la motivele deciziei automate.

3. Principiile Asilomar pentru Inteligența Artificială: Un set de 23 de principii elaborate de experți în domeniu, care subliniază importanța siguranței, securității și beneficiilor pe termen lung ale IA.

Probleme referitoare la cercetare.

1. Scopul Cercetării: Scopul cercetării în IA ar trebui să fie crearea nu a inteligenței neîndreptate, ci a inteligenței benefice.
2. Finanțarea Cercetării: Investițiile în IA ar trebui să fie însoțite de finanțare pentru cercetarea privind asigurarea utilizării benefice a acesteia, inclusiv întrebări spinoase din informatică, economie, drept, etică și studii sociale, precum: Cum putem face viitoarele sisteme de IA extrem de robuste, astfel încât să facă ceea ce dorim fără să funcționeze defectuos sau să fie piratate? Cum putem crește prosperitatea prin automatizare, menținând în același timp resursele și scopul oamenilor? Cum putem actualiza sistemele noastre juridice pentru a fi mai echitabile și eficiente, pentru a ține pasul cu IA și pentru a gestiona riscurile asociate cu IA? Cu ce set de valori ar trebui să fie aliniată IA și ce statut juridic și etic ar trebui să aibă?
3. Legătura Știință-Politică: Ar trebui să existe un schimb constructiv și sănătos între cercetătorii în IA și factorii de decizie politică.
4. Cultura Cercetării: Ar trebui să fie promovată o cultură de cooperare, încredere și transparență între cercetătorii și dezvoltatorii de IA.

5. Evitarea Competiției Nesănătoase: Echipele care dezvoltă sisteme de IA ar trebui să coopereze activ pentru a evita reducerea standardelor de siguranță.

Etică și Valori.

6. Siguranță: Sistemele de IA ar trebui să fie sigure și securizate pe tot parcursul duratei lor de funcționare și verificabile, acolo unde este posibil și fezabil.

7. Transparența Eșecului: Dacă un sistem de IA provoacă daune, ar trebui să fie posibil să se afle motivul.

8. Transparența Judiciară: Orice implicare a unui sistem autonom în luarea deciziilor judiciare ar trebui să ofere o explicație satisfăcătoare, auditabilă de către o autoritate umană competentă.

9. Responsabilitate: Proiectanții și constructorii de sisteme AI avansate sunt părți interesate în implicațiile morale ale utilizării, utilizării greșite și acțiunilor acestora, având responsabilitatea și oportunitatea de a modela aceste implicații.

10. Alinierea Valorilor: Sistemele AI foarte autonome ar trebui să fie proiectate astfel încât obiectivele și comportamentul lor să poată fi garantate că se aliniază cu valorile umane pe tot parcursul funcționării lor.

11. Valorile Umane: Sistemele de IA ar trebui să fie proiectate și operate astfel încât să fie compatibile cu idealurile demnității umane, drepturilor, libertăților și diversității culturale.

12. Confidențialitatea Personală: Oamenii ar trebui să aibă dreptul de a accesa, gestiona și controla datele pe care le generează, având în vedere puterea sistemelor de IA de a analiza și utiliza aceste date.

13. Libertate și Confidențialitate: Aplicarea IA asupra datelor personale nu trebuie să limiteze în mod nejustificat libertatea reală sau percepută a oamenilor.

14. Beneficiu Comun: Tehnologiile AI ar trebui să beneficieze și să împuternicească cât mai mulți oameni posibil.

15. Prosperitate Împărtășită: Prosperitatea economică creată de IA ar trebui să fie împărtășită pe scară largă, pentru a beneficia întreaga umanitate.

16. Controlul Uman: Oamenii ar trebui să aleagă cum și dacă să delege deciziile sistemelor de IA, pentru a îndeplini obiectivele alese de oameni.

17. Non-Subversiune: Puterea conferită de controlul sistemelor AI foarte avansate ar trebui să respecte și să îmbunătățească, mai degrabă decât să submineze, procesele sociale și civice de care depinde sănătatea societății.

18. Cursa înarmării: Ar trebui evitată o cursă de înarmare cu arme autonome letale.

Probleme pe termen lung

19. Prudență în privința capacităților: În absența unui consens, ar trebui să evităm presupunerile puternice cu privire la limitele superioare ale capacităților viitoare ale IA.

20. Importanță: IA avansată ar putea reprezenta o schimbare profundă în istoria vieții pe Pământ și ar trebui planificată și gestionată cu o atenție și resurse corespunzătoare.

21. Riscuri: Riscurile prezentate de sistemele de IA, în special riscurile catastrofale sau existențiale, trebuie să fie supuse eforturilor de planificare și atenuare proporționale cu impactul lor așteptat.

22. Îmbunătățirea Recursivă de Sine: Sistemele de IA proiectate să se îmbunătățească recursiv de sine sau să se autoreproducă într-un mod care ar putea duce la o creștere rapidă a calității sau cantității trebuie să fie supuse unor măsuri stricte de siguranță și control.

23. Binele comun: Superinteligența ar trebui dezvoltată doar în serviciul unor idealuri etice larg împărtășite și în beneficiul întregii umanități, mai degrabă decât a unui singur stat sau organizație.

4. Principiile Organizației pentru Cooperare și Dezvoltare Economică (OECD) pentru inteligența artificială (AI) sunt orientări importante pentru guverne și organizații în promovarea dezvoltării și utilizării AI într-un mod responsabil și benefic pentru societate. Aceste principii au fost adoptate de către 42 de țări membre ale OECD în 2019 și vizează asigurarea unui echilibru între inovație, responsabilitate și protecția drepturilor fundamentale.

Principiul 1.1. Creştere incluzivă, dezvoltare durabilă şi bunăstare.

Acest principiu recunoaște că orientarea dezvoltării și utilizării IA spre prosperitate și rezultate benefice pentru oameni și planetă este o prioritate. IA de încredere poate juca un rol important în promovarea creșterii inclusive, a dezvoltării durabile, a bunăstării și a obiectivelor de dezvoltare globală. Într-adevăr, IA poate fi utilizată pentru binele social și poate contribui substanțial la realizarea Obiectivelor de Dezvoltare Durabilă (SDGs) în domenii precum educația, sănătatea, transportul, agricultura, mediul și orașele durabile, printre altele.

Acest rol de administrare responsabilă ar trebui să urmărească abordarea preocupărilor legate de inegalitate și riscul ca disparitățile în accesul la tehnologie să crească diviziunile existente în cadrul și între țările dezvoltate și în curs de dezvoltare. Cadrul OECD pentru Acțiuni Politice

privind Creșterea Incluzivă oferă un punct de referință util. Acest cadru este destinat să orienteze acțiunile politice care „aduc pe toată lumea împreună” către un viitor mai robust și încrezător.

Acest principiu recunoaște, de asemenea, că sistemele de IA ar putea perpetua prejudecățile existente și ar putea avea un impact disproporționat asupra populațiilor vulnerabile și subreprezentate, cum ar fi minoritățile etnice, femeile, copiii, vârstnicii și persoanele mai puțin educate sau cu calificări scăzute. Impactul disproporționat este un risc deosebit în țările cu venituri mici și medii. Acest principiu subliniază că IA poate fi, de asemenea, și ar trebui să fie, utilizată pentru a împuternici toți membrii societății și pentru a ajuta la reducerea prejudecăților.

Administrarea responsabilă este, în plus, o recunoaștere a faptului că, pe tot parcursul ciclului de viață al sistemului de IA, actorii și părțile interesate în domeniul IA pot și ar trebui să încurajeze dezvoltarea și implementarea IA pentru rezultate benefice, cu garanții adecvate. Definirea acestor rezultate benefice și a celor mai bune modalități de a le realiza va beneficia de colaborarea multidisciplinară și multi-stakeholder și de dialogul social. În plus, un dialog public semnificativ, bine informat și iterativ, care să includă toate părțile interesate, poate spori încrederea publicului în IA și înțelegerea acesteia.

Principiul 1.2. Respectarea statului de drept, a drepturilor omului și a valorilor democratice, inclusiv a echității și a vieții private.

IA ar trebui dezvoltată în concordanță cu valorile centrate pe om, cum ar fi libertățile fundamentale, egalitatea, echitatea, statul de drept, justiția socială, protecția datelor și viața privată, precum și drepturile consumatorilor și corectitudinea comercială.

Unele aplicații sau utilizări ale sistemelor de IA au implicații asupra drepturilor omului, inclusiv riscuri ca drepturile omului (așa cum sunt definite în Declarația Universală a Drepturilor Omului) și valorile centrate pe om să fie încălcate în mod deliberat sau accidental. Prin urmare, este important să promovăm „alinierea la valori” în sistemele de IA (adică proiectarea lor cu măsuri de siguranță adecvate), inclusiv capacitatea de intervenție și supraveghere umană, după cum este adecvat contextului. Această aliniere poate ajuta la asigurarea faptului că comportamentele sistemelor de IA protejează și promovează drepturile omului și se aliniază cu valorile centrate pe om pe tot parcursul funcționării lor. Rămânerea fidelă valorilor democratice comune va ajuta la consolidarea încrederii publice în IA și la susținerea utilizării IA pentru protejarea drepturilor omului și reducerea discriminării sau a altor rezultate nedrepte și/sau inegale.

Acest principiu recunoaște, de asemenea, rolul unor măsuri precum evaluările de impact asupra drepturilor omului (HRIAs) și diligența datorată în ceea ce privește drepturile omului, determinarea umană (adică un „om în buclă”), codurile de conduită etică sau etichetele și certificările de calitate destinate promovării valorilor centrate pe om și echității.

Principiul 1.3. Transparență și explicabilitate.

Termenul de transparență are mai multe semnificații. În contextul acestui Principiu, accentul se pune, în primul rând, pe dezvăluirea momentului în care se utilizează IA (într-o predicție, recomandare sau decizie, sau că utilizatorul interacționează direct cu un agent de Inteligență Artificială, cum ar fi un chatbot). Dezvăluirea ar trebui făcută proporțional cu importanța interacțiunii. Ubicuitatea crescândă a aplicațiilor de IA poate influența dorința, eficacitatea sau fezabilitatea dezvăluirii în unele cazuri.

În plus, transparența înseamnă a permite oamenilor să înțeleagă modul în care un sistem de IA este dezvoltat, antrenat, operează și este implementat în domeniul de aplicare relevant, astfel încât consumatorii, de exemplu, să poată face alegeri mai informate. Transparența se referă, de asemenea, la capacitatea de a furniza informații semnificative și claritate cu privire la informațiile furnizate și motivul pentru care sunt furnizate. Astfel, transparența nu se extinde, în general, la dezvăluirea codului sursă sau a altui cod proprietar sau la partajarea seturilor de date proprietar, toate acestea putând fi prea complexe din punct de vedere tehnic pentru a fi fezabile sau utile pentru înțelegerea unui rezultat. Codul sursă și seturile de date pot fi, de asemenea, supuse proprietății intelectuale, inclusiv secretelor comerciale.

Un alt aspect al transparenței privește facilitarea discursului public, multi-stakeholder și înființarea de entități dedicate, după cum este necesar, pentru a promova conștientizarea generală și înțelegerea sistemelor de IA și pentru a crește acceptarea și încrederea.

Explicabilitatea (capacitatea unui sistem de AI de a explica modul în care ajunge la o anumită decizie sau recomandare) înseamnă a permite oamenilor afectați de rezultatul unui sistem de IA să înțeleagă cum a fost obținut. Aceasta implică furnizarea de informații ușor de înțeles oamenilor afectați de rezultatul unui sistem de IA, care să le permită celor afectați să conteste rezultatul, în special - în măsura în care este posibil - factorii și logica care au condus la un anumit rezultat. Cu toate acestea, explicabilitatea poate fi realizată în diferite moduri, în funcție de context (cum ar fi semnificația rezultatelor). De exemplu, pentru anumite tipuri de sisteme de IA, solicitarea explicabilității poate afecta negativ acuratețea și performanța sistemului (deoarece

poate necesita reducerea variabilelor soluției la un set suficient de mic pentru a fi înțeles de oameni, ceea ce ar putea fi suboptimal în probleme complexe, de înaltă dimensiune) sau confidențialitatea și securitatea. De asemenea, poate crește complexitatea și costurile, potențial punând actorii AI care sunt IMM-uri într-un dezavantaj disproporționat.

Prin urmare, atunci când actorii din domeniul IA oferă o explicație a unui rezultat, ei pot lua în considerare furnizarea - în termeni clari și simpli și, după cum este adecvat contextului - a principalilor factori într-o decizie, a factorilor determinanți, a datelor, logicii sau algoritmului din spatele rezultatului specific sau explicarea motivului pentru care circumstanțe similare au generat un rezultat diferit. Acest lucru ar trebui făcut într-un mod care să permită indivizilor să înțeleagă și să conteste rezultatul, respectând în același timp obligațiile privind protecția datelor cu caracter personal, dacă este cazul.

Principiul 1.4. Robustețe, securitate și siguranță.

Abordarea provocărilor de siguranță și securitate ale sistemelor complexe de IA este crucială pentru a promova încrederea în IA. În acest context, robustețea semnifică capacitatea de a rezista sau de a depăși condițiile adverse, inclusiv riscurile de securitate digitală. Acest principiu afirmă, de asemenea, că sistemele de IA nu ar trebui să prezinte riscuri de siguranță nejustificate, inclusiv pentru securitatea fizică, în condiții de utilizare sau utilizare greșită normale sau previzibile pe parcursul ciclului lor de viață. Legile și reglementările existente în domenii precum protecția consumatorilor identifică deja ce constituie riscuri de siguranță nejustificate. Guvernele, în consultare cu părțile interesate, trebuie să determine în ce măsură se aplică sistemelor de IA.

Actorii din domeniul IA pot utiliza o abordare de gestionare a riscurilor pentru a identifica și proteja împotriva utilizării necorespunzătoare previzibile, precum și împotriva riscurilor asociate cu utilizarea sistemelor de IA în scopuri diferite de cele pentru care au fost inițial proiectate. Problemele de robustețe, securitate și siguranță a IA sunt interconectate. De exemplu, securitatea digitală poate afecta siguranța produselor conectate, cum ar fi automobilele și aparatele electrocasnice, dacă riscurile nu sunt gestionate în mod corespunzător.

Recomandarea evidențiază două modalități de menținere a sistemelor de IA robuste, sigure și securizate:

a) trasabilitatea și analiza și ancheta ulterioară (Trasabilitatea (traceability) în contextul inteligenței artificiale (IA) se referă la capacitatea de a urmări și documenta fiecare etapă a dezvoltării,

implementării și utilizării unui sistem de IA. Este o caracteristică esențială pentru asigurarea responsabilității, transparenței și conformității cu reglementările și normele etice).
b) aplicarea unei abordări de gestionare a riscurilor.

La fel ca explicabilitatea (vezi 1.3), trasabilitatea poate ajuta la analiza și investigarea rezultatelor unui sistem de IA și este o modalitate de promovare a responsabilității. Trasabilitatea diferă de explicabilitate prin faptul că accentul se pune pe menținerea înregistrărilor caracteristicilor datelor, cum ar fi metadatele, sursele de date și curățarea datelor, dar nu neapărat pe datele în sine. În acest sens, trasabilitatea poate ajuta la înțelegerea rezultatelor, la prevenirea erorilor viitoare și la îmbunătățirea fiabilității sistemului de IA.

Abordare de gestionare a riscurilor: Recomandarea recunoaște riscurile potențiale pe care sistemele de IA le prezintă pentru drepturile omului, integritatea fizică, confidențialitatea, echitatea, egalitatea și robustețe. De asemenea, recunoaște costurile protejării împotriva acestor riscuri, inclusiv prin construirea transparenței, responsabilității, siguranței și securității în sistemele de IA. De asemenea, recunoaște că diferite utilizări ale IA prezintă riscuri diferite și că unele riscuri necesită un standard mai ridicat de prevenire sau atenuare decât altele.

O abordare de gestionare a riscurilor, aplicată pe tot parcursul ciclului de viață al sistemului de IA, poate ajuta la identificarea, evaluarea, prioritizarea și atenuarea riscurilor potențiale care pot afecta negativ comportamentul și rezultatele unui sistem. Alte standarde OECD privind gestionarea riscurilor, de exemplu în contextul gestionării riscurilor de securitate digitală și al diligenței datorate bazate pe risc în cadrul Ghidurilor pentru întreprinderile multinaționale și Ghidului OECD privind diligența datorată pentru o conduită responsabilă a afacerilor, pot oferi îndrumări utile1. Documentarea deciziilor de gestionare a riscurilor luate la fiecare fază a ciclului de viață poate contribui la implementarea celorlalte principii de transparență (1.3) și responsabilitate (1.5).

Principiul 1.5. Responsabilitate.

Termenii de responsabilitate, răspundere și răspundere juridică sunt strâns legați, dar diferiți, și au, de asemenea, semnificații diferite în culturi și limbi diferite. În general, „responsabilitatea” implică o așteptare etică, morală sau de altă natură (de exemplu, așa cum este prevăzută în practicile de management sau codurile de conduită) care ghidează acțiunile sau conduita indivizilor sau organizațiilor și le permite să explice motivele pentru care au fost luate deciziile și acțiunile. În cazul unui rezultat negativ, implică, de asemenea, luarea de măsuri pentru

a asigura un rezultat mai bun în viitor. „Răspunderea juridică” se referă, în general, la implicații juridice adverse care decurg din acțiunile sau inacțiunea unei persoane (sau a unei organizații). „Răspunderea” poate avea, de asemenea, așteptări etice sau morale și poate fi utilizată atât în contexte juridice, cât și non-juridice pentru a se referi la o legătură cauzală între un actor și un rezultat.

Având în vedere aceste semnificații, termenul „responsabilitate” surprinde cel mai bine esența acestui principiu. În acest context, „responsabilitatea” se referă la așteptarea ca organizațiile sau indivizii să asigure funcționarea corectă, pe tot parcursul ciclului de viață, a sistemelor de IA pe care le proiectează, dezvoltă, operează sau implementează, în conformitate cu rolurile lor și cadrele de reglementare aplicabile, și pentru a demonstra acest lucru prin acțiunile și procesul lor decizional (de exemplu, prin furnizarea de documentație privind deciziile cheie pe tot parcursul ciclului de viață al sistemului de IA sau prin efectuarea sau permisiunea auditului, acolo unde este justificat).

Principiul 2.1. Investiții în cercetare și dezvoltare în domeniul inteligenței artificiale.

Progresele științifice permise de IA ar putea ajuta la rezolvarea provocărilor societale și la crearea de industrii complet noi. Aceste posibilități subliniază importanța cercetării fundamentale și a luării în considerare a unor orizonturi de timp îndelungate în politica de cercetare. În timp ce sectorul privat a preluat conducerea în investițiile în cercetare și dezvoltare în domeniul IA aplicată în ultimii ani, guvernele, uneori completate de fundații axate pe binele public, au un rol important de jucat în furnizarea de investiții susținute în cercetarea publică cu orizonturi pe termen lung. Acest tip de investiție este esențial pentru a conduce și modela inovația AI de încredere și pentru a asigura rezultate benefice pentru toți, în special în domeniile insuficient acoperite de investițiile orientate spre piață. Cercetarea finanțată public poate ajuta la abordarea problemelor tehnologice dificile care afectează o gamă largă de actori și părți interesate în domeniul IA.

În plus, deoarece IA are o acoperire largă și implicații omniprezente asupra mai multor fațete ale vieții, această recomandare solicită investiții în cercetare interdisciplinară, privind implicațiile sociale, juridice și etice ale IA care sunt relevante pentru politica publică.

Datorită importanței datelor pentru ciclul de viață al sistemului de IA, un element cheie în asigurarea unei cercetări și dezvoltări IA mai bune și mai bune este disponibilitatea unor seturi de date deschise, accesibile și reprezentative care nu compromit confidențialitatea și protecția datelor personale și a consumatorilor, drepturile de proprietate intelectuală și alte drepturi importante. În

special, deși poate fi imposibil să se realizeze un mediu complet „fără prejudecăți”, prin furnizarea (și oferirea de stimulente pentru) seturi de date reprezentative care sunt disponibile publicului, guvernele pot contribui la atenuarea riscurilor de prejudecăți necorespunzătoare în sistemele de IA. De exemplu, sistemele de IA care utilizează seturi de date care nu sunt suficient de reprezentative în conformitate cu utilizarea intenționată a sistemului, chiar și fără intenția de a discrimina.

Această recomandare completează recomandarea privind promovarea unui ecosistem digital pentru IA (2.2), deoarece investițiile pe termen lung în tehnologii și infrastructură digitală și mecanismele de partajare a cunoștințelor privind IA sunt mijloace de promovare a acestui ecosistem digital. În special, investiția în seturi de date deschise, accesibile și reprezentative facilitează partajarea cunoștințelor privind IA.

Principiul 2.2. Promovarea unui ecosistem AI incluziv și favorabil.

Dezvoltarea unei IA de încredere necesită un ecosistem favorabil. Această recomandare solicită guvernelor - angajându-se cu sectorul privat, după cum este cazul - să lucreze pentru a furniza sau a promova furnizarea infrastructurii și tehnologiilor digitale pentru IA și a mecanismelor de partajare a cunoștințelor privind IA, ținând cont de cadrele lor naționale.

Tehnologiile și infrastructura digitală necesare includ accesul la rețele și servicii broadband de mare viteză și la prețuri accesibile, putere de calcul și stocare de date, precum și tehnologii de generare a datelor, cum ar fi Internetul lucrurilor (IoT). De exemplu, progresele recente în domeniul IA pot fi atribuite, în parte, creșterii exponențiale a vitezei de calcul, inclusiv cu resurse de unitate de procesare grafică. Mecanismele adecvate pentru partajarea cunoștințelor privind IA, inclusiv date, cod, algoritmi, modele, cercetare și know-how, sunt, de asemenea, necesare pentru a înțelege și a participa la ciclul de viață al sistemului de IA. Astfel de mecanisme trebuie să respecte confidențialitatea, proprietatea intelectuală și alte drepturi.

Instrumentele open-source și seturile de date de înaltă calitate pentru gestionarea și utilizarea IA, care permit difuzarea tehnologiei IA și soluțiile de crowdsourcing pentru erorile software, joacă un rol cheie în dezvoltarea IA.

Atunci când dezvoltă mijloace de partajare a datelor, cum ar fi trusturile de date sau terțe părți de încredere, guvernele ar trebui să acorde atenție riscurilor legate de accesul la date și partajarea acestora: riscurile pentru persoane (inclusiv consumatori), organizații și țări ale partajării datelor pot include încălcări ale confidențialității și vieții private, riscuri pentru drepturile

de proprietate intelectuală, protecția datelor, concurența și interesele comerciale, precum și potențiale riscuri de securitate națională și digitală. În ceea ce privește seturile de date în sine și în conjuncție cu recomandarea 2.1, guvernele sunt invitate să promoveze și să utilizeze seturi de date cât mai inclusive, diverse și reprezentative.

Recomandările pentru factorii de decizie politică acordă o atenție deosebită politicilor pentru întreprinderile mici și mijlocii (IMM-uri): pentru a facilita accesul IMM-urilor la date, tehnologii IA și infrastructură relevantă (cum ar fi conectivitatea, capacitățile de calcul și platformele cloud) pentru a promova antreprenoriatul digital, concurența și inovația prin adoptarea IA.

Principiul 2.3. Modelarea și facilitarea unui mediu de guvernanță și politică interoperabil pentru IA.

Această recomandare se concentrează pe mediul politic care permite inovația în domeniul IA, adică cadrul instituțional, politic și juridic. Astfel, completează recomandarea 2.2 privind infrastructura fizică și tehnologică necesară. Ca și în recomandarea 2.2, guvernele sunt invitate să acorde o atenție deosebită IMM-urilor.

Având în vedere ritmul rapid al dezvoltărilor în domeniul IA, stabilirea unui mediu politic suficient de flexibil pentru a ține pasul cu evoluțiile și a promova inovația, rămânând în același timp sigur și oferind certitudine juridică, este o provocare semnificativă. Această recomandare urmărește să abordeze această provocare prin identificarea mijloacelor de îmbunătățire a adaptabilității, reactivității, versatilității și aplicării instrumentelor de politică, pentru a accelera în mod responsabil tranziția de la dezvoltare la implementare și, acolo unde este cazul, comercializare. În promovarea utilizării IA, ar trebui adoptată o abordare centrată pe om.

Recomandarea subliniază rolul experimentării ca mijloc de a oferi medii controlate și transparente în care sistemele de IA pot fi testate și în care modelele de afaceri bazate pe IA, care ar putea promova soluții la provocările globale, pot prospera. Experimentele de politică pot funcționa în „modul startup", în care experimentele sunt implementate, evaluate și modificate, apoi extinse sau abandonate, în funcție de rezultatele testelor.

În cele din urmă, recomandarea recunoaște importanța mecanismelor de supraveghere și evaluare pentru a completa cadrele de politică și experimentare. În acest sens, poate exista spațiu pentru încurajarea actorilor din domeniul IA să dezvolte mecanisme de autoreglementare, cum ar fi codurile de conduită, standardele voluntare și cele mai bune practici. Împreună cu Ghidurile

OECD pentru întreprinderile multinaționale (MNE), astfel de inițiative pot ajuta la îndrumarea actorilor din domeniul IA pe parcursul ciclului de viață al IA, inclusiv pentru monitorizare, raportare, evaluare și abordarea efectelor negative sau a utilizării necorespunzătoare a sistemelor de IA. În măsura posibilului, aceste mecanisme ar trebui să fie transparente și publice.

Principiul 2.4.Clădirea capacității umane și pregătirea pentru transformarea pieței muncii.

Se preconizează pe scară largă că inteligența artificială (IA) va transforma natura multor aspecte ale vieții, pe măsură ce aceasta se răspândește în diverse sectoare. Acest lucru este valabil în mod special în contextul muncii, al angajării și al locului de muncă, unde IA va completa activitatea umană în unele sarcini, o va înlocui în altele și va genera noi tipuri de locuri de muncă și organizații de lucru. Dacă aceste transformări ale pieței muncii nu sunt gestionate corespunzător, ele ar putea genera costuri economice și sociale semnificative. Pentru a gestiona aceste tranziții și a se asigura că ele sunt echitabile, factorii de decizie politică – împreună cu părțile interesate, cum ar fi partenerii sociali, organizațiile patronale și sindicatele – vor trebui să ia în considerare aspecte legate de protecția socială, programele educaționale, dezvoltarea competențelor, reglementarea pieței muncii, serviciile publice de ocupare a forței de muncă, politica industrială și fiscalitatea, precum și finanțarea tranzițiilor.

Gestionarea unor tranziții echitabile necesită politici pentru învățarea pe tot parcursul vieții, dezvoltarea competențelor și formarea profesională, care să le permită oamenilor – în special lucrătorilor, în diverse contexte contractuale – să interacționeze cu sistemele IA, să se adapteze la schimbările generate de IA și să acceseze noi oportunități pe piața muncii. Aceasta include competențele necesare practicienilor din domeniul IA (care sunt în prezent insuficiente) și cele de care au nevoie alți lucrători (cum ar fi medicii sau avocații) pentru a valorifica IA în domeniile lor de expertiză, astfel încât IA să amplifice capacitățile umane. În paralel, politicile de dezvoltare a competențelor vor trebui să se concentreze asupra aspectelor distinct umane necesare pentru a completa sistemele IA, cum ar fi judecata, gândirea creativă și critică, precum și comunicarea interpersonală.

În plus, schimbările generate de IA pe piața muncii pot necesita adaptarea sau adoptarea unor standarde de muncă și acorduri între management și lucrători, acolo unde este cazul, pentru a reflecta aceste schimbări, a aborda eventualele provocări legate de egalitate, diversitate și echitate (de exemplu, cele generate de colectarea și prelucrarea datelor) și a consolida locurile de muncă

fiabile, sigure și productive. Acest lucru poate fi realizat printr-o combinație de reglementări, dialog social și negocieri colective. Este important să se permită flexibilitatea la locul de muncă, protejând în același timp autonomia și calitatea muncii lucrătorilor.

Principiul 2.5. Cooperare internațională pentru o IA de încredere.

Această recomandare solicită cooperare internațională între guverne și cu părțile interesate, pentru a aborda oportunitățile și provocările globale ale inteligenței artificiale (IA). O astfel de cooperare include promovarea implementării și diseminării acestor principii și politici în țările OCDE, țările partenere, inclusiv în țările în curs de dezvoltare și cele mai puțin dezvoltate, precum și cu alte părți interesate.

Cooperarea internațională poate valorifica forumurile internaționale și regionale pentru a împărtăși cunoștințe despre IA, cu scopul de a construi expertiză pe termen lung în acest domeniu; pentru a dezvolta standarde tehnice care să asigure interoperabilitatea și încrederea în IA; și pentru a crea, disemina și utiliza metrici care să evalueze performanța sistemelor IA, precum acuratețea, eficiența, contribuția la obiectivele societale, echitatea și robustețea. De asemenea, poate contribui la fluxurile transfrontaliere de date bazate pe încredere, care protejează securitatea, confidențialitatea, proprietatea intelectuală, drepturile omului și valorile democratice, elemente esențiale pentru inovația în domeniul IA.

**Raportul Comitetului selectat de inteligență artificială a Camerei Lorzilor din Regatul Unit, publicat în aprilie 2018*, include cinci principii generale pentru un cod al inteligenței artificiale. Aceste principii sunt esențiale pentru ghidarea dezvoltării și utilizării responsabile a inteligenței artificiale (IA) în Regatul Unit, având ca scop protejarea valorilor democratice, asigurarea echității și prevenirea riscurilor etice.

1. Inteligența artificială trebuie să fie dezvoltată pentru a fi un beneficiu al umanității. Toate sistemele de IA ar trebui să fie concepute pentru a servi binelui comun, îmbunătățind viețile oamenilor și promovând progresul social.
2. Inteligența artificială trebuie să funcționeze pe baza principiilor de transparență. Deciziile luate de sistemele de IA trebuie să fie inteligibile și explicabile. Utilizatorii ar trebui să aibă o înțelegere clară asupra modului în care funcționează aceste sisteme și cum ajung la concluzii.

3. Sistemele de inteligență artificială nu trebuie să fie utilizate pentru a diminua sau eroda drepturile individuale. Drepturile fundamentale ale omului, cum ar fi confidențialitatea, libertatea de exprimare și egalitatea, trebuie protejate în toate aplicațiile de IA.

4. Toți cetățenii ar trebui să aibă acces la beneficiile inteligenței artificiale. Sistemele de IA ar trebui să fie dezvoltate astfel încât să evite exacerbarea inegalităților sociale sau economice și să promoveze incluziunea.

5. Inteligența artificială nu ar trebui să fie concepută în mod autonom pentru a pune viața umană în pericol sau pentru a provoca prejudicii. Dezvoltarea IA trebuie să fie reglementată pentru a preveni utilizarea sa în scopuri dăunătoare, precum arme autonome sau alte aplicații care pot afecta viața umană.

Aceste principii sunt o parte esențială a strategiei Regatului Unit pentru a asigura o utilizare etică a inteligenței artificiale și pentru a contribui la dezvoltarea de politici care să protejeze atât indivizii, cât și societatea în ansamblu.

**Alte seturi de principii adoptate*:

- Principiile Parteneriatului pentru IA (Partnership on AI 2024) publicat în colaborare cu cadre universitare, cercetători, organizații societății civile, companii care construiesc și utilizează tehnologia IA și alte grupuri.
- China și-a lansat propriile principii IA, denumite Principiile Beijing IA.
- Principiile Google IA (Google 2024) se concentrează pe construirea unei inteligențe artificiale benefice din punct de vedere social.

Au fost formate multe comitete etice IA, inclusiv Institutul Stanford pentru IA centrată pe om (HIA), Institutul Alan Turing, Parteneriatul IA, IA Now, IEEE și altele.

Capitolul III.

Reglementarea IA în cadrul internațional (ONU, Consiliul Europei, SUA) preambulul unei viitoare reglementări naționale a IA.

a) AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (*COMITETUL).

b) UE-AI Act.

c) Consiliul UE – Framework Convention on AI (primul tratat internațional obligatoriu).

d) ONU 2024 – Governing AI for Humanity (cadrul global pentru guvernanța IA).

e) SUA- Blueprint for an AI Bill of Rights.

a) ***Comitetul ad-hoc pentru inteligență artificială (*COMITETUL)*** și-a îndeplinit mandatul (2019-2021) și a fost succedat de Comitetul pentru Inteligența Artificială (CAI).

Comitetul a examinat fezabilitatea și potențialele elemente, pe baza unor ample consultări cu părțile interesate, ale unui cadru juridic pentru dezvoltarea, proiectarea și aplicarea inteligenței artificiale, bazat pe standardele Consiliului Europei privind drepturile omului, democrația și statul de drept.

https://rm.coe.int/*Comitetul-2021-09rev-elements/1680a6d90d

Acest document conține rezultatele muncii Comitetului ad-hoc al Consiliului Europei pentru Inteligența Artificială (*COMITETUL) privind potențialele elemente ale unui cadru juridic pentru dezvoltarea, proiectarea și aplicarea inteligenței artificiale, bazat pe standardele Consiliului Europei privind drepturile omului, democrația și statul de drept.

Observații generale.

*COMITETUL observă că aplicarea sistemelor de inteligență artificială (AI) are potențialul de a promova prosperitatea umană și bunăstarea individuală și socială, prin îmbunătățirea progresului și inovației, însă, în același timp, anumite aplicații ale sistemelor de AI generează îngrijorare, deoarece pot pune în pericol drepturile omului, democrația și statul de drept.

Pentru a preveni și/sau atenua în mod eficient aceste riscuri, se consideră că un cadru juridic adecvat pentru AI, bazat pe standardele Consiliului Europei privind drepturile omului, democrația și statul de drept, ar trebui să ia forma unui instrument transversal juridic obligatoriu. Se notează că - pe lângă instrumentul transversal juridic obligatoriu propus, care stabilește principii generale și norme juridice specifice - pot fi necesare instrumente juridice obligatorii și/sau neobligatorii existente sau viitoare la nivel sectorial, pentru a oferi orientări mai detaliate privind asigurarea

faptului că proiectarea, dezvoltarea și aplicarea AI se realizează în conformitate cu drepturile omului, democrația și statul de drept în domenii specifice.

Instrumentul transversal juridic obligatoriu ar trebui să se concentreze pe prevenirea și/sau atenuarea riscurilor provenite din aplicațiile sistemelor de AI cu potențialul de a interfera cu exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept, promovând în același timp aplicațiile AI benefice din punct de vedere social. Ar trebui să fie susținut de o abordare bazată pe risc: cerințele legale privind proiectarea, dezvoltarea și utilizarea sistemelor de AI ar trebui să fie proporționale cu natura riscului pe care îl prezintă pentru drepturile omului, democrația și statul de drept. Principiile de bază care permit determinarea unui astfel de risc (de exemplu, cerințele de transparență) ar trebui să fie aplicabile tuturor sistemelor de AI.

În conformitate cu articolul 1 d din Statutul Consiliului Europei, chestiunile legate de apărarea națională nu intră în domeniul de aplicare al unui cadru juridic al Consiliului Europei și, prin urmare, nu sunt incluse în domeniul de aplicare al unui instrument juridic (sau nejuridic) obligatoriu al Consiliului Europei. *COMITETUL este de părere că problema dacă acest domeniu de aplicare ar putea include „dubla utilizare” și securitatea națională ar trebui examinată în continuare în contextul elaborării unui cadru juridic al Consiliului Europei privind AI, ținând cont de posibilele dificultăți în această privință.

Diverse probleme juridice.

Diversele probleme juridice ridicate de aplicarea sistemelor de AI nu sunt specifice statelor membre ale Consiliului Europei, ci sunt, datorită numeroșilor actori globali implicați și efectelor globale pe care le generează, de natură transnațională. Prin urmare, se recomandă ca un instrument transversal juridic obligatoriu al Consiliului Europei, deși evident bazat pe standardele Consiliului Europei, să fie redactat într-un mod care să faciliteze aderarea statelor din afara regiunii care împărtășesc standardele menționate. Nu numai că acest lucru va crește semnificativ impactul și eficiența instrumentului propus, dar, în plus, va oferi un nivel mult-necesar de egalitate a șanselor pentru actorii relevanți, inclusiv industria și cercetătorii în domeniul AI, care adesea operează peste frontierele naționale și regiunile lumii. Standardele Consiliului Europei privind drepturile omului, democrația și statul de drept sunt suficient de universale pentru a face acest lucru o opțiune realistă. Există mai multe precedente de tratate ale Consiliului Europei aplicate dincolo de regiunea europeană, cf. în special Convenția de la Budapesta (Cybercrime) și Convenția pentru protecția persoanelor cu privire la prelucrarea automată a datelor cu caracter personal (CETS nr. 108), care

în prezent au 66 și, respectiv, 55 de părți, multe dintre acestea nefiind state membre ale Consiliului Europei.

Se recomandă, de asemenea, ca, pentru a asigura atât coerența juridică globală, cât și regională, un instrument transversal juridic obligatoriu al Consiliului Europei să țină cont de cadrele juridice și de reglementare existente și viitoare ale altor foruri internaționale și regionale, în special sistemul Națiunilor Unite (inclusiv UNESCO), Uniunea Europeană și Organizația pentru Cooperare Economică și Dezvoltare - toate acestea fiind în prezent implicate în dezvoltarea diferitelor forme de standarde legate de sistemele de AI.

Se notează că scopul unui cadru juridic internațional nu ar trebui să fie stabilirea unor parametri tehnici detaliați pentru proiectarea, dezvoltarea și aplicarea sistemelor de AI, ci stabilirea anumitor principii și norme de bază care guvernează dezvoltarea, proiectarea și aplicarea sistemelor de AI și reglementarea, într-un mod consecvent și deliberat, dacă și în ce condiții sistemele de AI care pot prezenta riscuri pentru exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept pot fi dezvoltate, proiectate și aplicate de toate tipurile de organizații, inclusiv actori publici și privați deopotrivă.

Partea a II-a: Elemente pentru un instrument transversal juridic obligatoriu.

III. Elemente referitoare la obiect și scop, domeniu de aplicare și definiții.

În ceea ce privește obiectul și scopul instrumentului transversal juridic obligatoriu, se recomandă să se precizeze, în special, că scopul instrumentului este de a asigura deplina coerență cu respectarea drepturilor omului, funcționarea democrației și respectarea statului de drept în dezvoltarea, proiectarea și aplicarea sistemelor de IA, indiferent dacă aceste activități sunt întreprinse de actori privați sau publici. În continuare, ar trebui să se precizeze că instrumentul va facilita cooperarea în acest sens între Părțile sale, atât la nivel internațional, cât și la nivel național, și că vor fi stabilite mecanismele de urmărire necesare. În final, obiectul și scopul ar trebui să sublinieze necesitatea de a stabili un cadru juridic comun care să conțină anumite standarde minime pentru dezvoltarea, proiectarea și aplicarea IA în raport cu drepturile omului, democrația și statul de drept.

*COMITETUL consideră că instrumentul transversal juridic obligatoriu ar trebui să conțină o dispoziție care să definească domeniul său de aplicare. Această dispoziție ar trebui să clarifice că instrumentul se va aplica dezvoltării, proiectării și aplicării sistemelor de IA, indiferent dacă aceste activități sunt întreprinse de actori publici sau privați, cu un accent deosebit pe astfel

de sisteme care sunt evaluate ca prezentând riscuri potențiale pentru exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept. După cum este necesar, ar trebui abordate și potențialele excepții de la domeniul de aplicare.

În ceea ce privește definițiile, se consideră că, la minimum, ar trebui incluse următoarele definiții într-un instrument transversal juridic obligatoriu: „Sistem de inteligență artificială"; „ciclu de viață"; „furnizor de IA"; „utilizator de IA"; „subiect AI"; „vătămare ilegală". Se recomandă ca toate definițiile utilizate să fie, pe cât posibil, compatibile cu definițiile similare utilizate în alte instrumente relevante privind IA. Mai mult, definițiile ar trebui redactate cu atenție pentru a asigura, pe de o parte, precizia juridică, iar, pe de altă parte, să fie suficient de abstracte pentru a rămâne valabile în ciuda viitoarelor evoluții tehnologice privind sistemele de IA.

IV. *Elemente referitoare la principiile fundamentale de protecție a demnității umane și respectarea drepturilor omului, a democrației și a statului de drept*.

Se consideră că este necesar ca un instrument transversal juridic obligatoriu să conțină anumite principii fundamentale de protecție a demnității umane și respectarea drepturilor omului, a democrației și a statului de drept, care ar trebui să se aplice tuturor dezvoltărilor, proiectărilor și aplicărilor sistemelor de IA, indiferent dacă actorul este public sau privat.

În același timp, recunoscând riscurile de duplicare sau chiar fragmentare a standardelor generale existente ale dreptului internațional, inclusiv dreptul umanitar, se recomandă ca aceste principii fundamentale să fie redactate într-un mod care să minimizeze în mod corespunzător riscurile de duplicare sau fragmentare nejustificate. Acest lucru implică, printre altele, adaptarea în continuare a drepturilor și obligațiilor legate de drepturile omului, democrație și statul de drept în scopul acestui instrument numai atunci când, după o examinare atentă, se ajunge la concluzia că standardele existente în forma lor actuală nu pot oferi o protecție suficientă a drepturilor indivizilor în contextul specific al dezvoltării, proiectării și aplicării sistemelor de IA.

În ceea ce privește conceptul de „demnitate umană", se notează că demnitatea persoanei umane este universal recunoscută ca fiind baza reală a drepturilor omului, cf. și importanța acordată conceptului în preambulul Declarației Universale a Drepturilor Omului din 1948. Este deosebit de util să se utilizeze acest concept într-un instrument transversal juridic obligatoriu privind potențialele impacturi negative asupra drepturilor fundamentale ale indivizilor cauzate de dezvoltarea, proiectarea și aplicarea sistemelor de IA.

Se mai notează că unele dintre dispozițiile legate de aceste elemente particulare pot fi formulate ca drepturi directe pozitive ale indivizilor sau, alternativ, ca obligații pentru Părți de a asigura introducerea în legislația și practica internă a măsurilor vizând protejarea drepturilor indivizilor în raport cu sistemele de IA. Pe baza deliberărilor sale, s-ar tinde, acolo unde este fezabil și necesar, să favorizeze o combinație atât a stabilirii anumitor drepturi directe, concrete și pozitive ale indivizilor în raport cu dezvoltarea, proiectarea și aplicarea sistemelor de IA, cât și a stabilirii anumitor obligații pentru Părți, pentru a asigura o aplicare mai uniformă a instrumentului transversal juridic obligatoriu între Părți.

Elemente referitoare la clasificarea riscului sistemelor de inteligență artificială și aplicațiile interzise ale inteligenței artificiale.

Se recomandă ca un instrument transversal juridic obligatoriu să prevadă stabilirea unei metodologii de clasificare a riscurilor sistemelor de IA, cu accent pe drepturile omului, democrație și statul de drept. Criteriile utilizate pentru evaluarea impactului aplicării sistemelor de IA în această privință ar trebui să fie concrete, clare și cu o bază obiectivă, iar evaluarea în sine să se facă într-un mod echilibrat, asigurând astfel atât certitudinea juridică, cât și nuanța.

Se consideră că clasificarea riscurilor ar trebui să includă o serie de categorii (de exemplu, „risc scăzut", „risc ridicat", „risc inacceptabil"), bazate pe o evaluare a riscului în raport cu exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept. Clasificarea riscului se va baza pe o revizuire inițială pentru a determina dacă este necesară o evaluare completă a *impactului asupra drepturilor omului, democrației și statului de drept (HUDERIA)* (cf. capitolul XII), la fel cum însăși evaluarea de impact poate avea un impact asupra menținerii sau modificării clasificării inițiale a riscului sistemului de IA în cauză. Această evaluare de impact este considerată ca un element al cadrului juridic general privind sistemele de IA propus de *COMITET. Cu toate acestea, modelul HUDERIA specific nu trebuie neapărat să facă parte dintr-un posibil instrument juridic obligatoriu.

În ceea ce privește criteriile care ar putea fi luate în considerare în scopul evaluării riscurilor, se face referire la elementele enumerate la punctul 51 din capitolul XII de mai jos. Unele dintre aceste criterii ar putea trebui să fie consacrate în instrumentul juridic obligatoriu, pentru a se asigura că sunt luate în considerare și aplicate în mod consecvent.

În ceea ce privește aplicațiile interzise ale AI (așa-numitele „linii roșii" sau „risc inacceptabil"), se consideră că un instrument transversal juridic obligatoriu ar trebui să prevadă

posibilitatea de a impune un moratoriu total sau parțial sau o interdicție asupra aplicării sistemelor de IA, care, în conformitate cu clasificarea riscului menționată mai sus, sunt considerate a prezenta un risc inacceptabil de interferență cu exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept. O astfel de posibilitate ar trebui, de asemenea, luată în considerare pentru cercetarea și dezvoltarea anumitor sisteme de IA care prezintă un risc inacceptabil. Se atrage atenția, de exemplu, asupra unor sisteme de IA care utilizează biometria pentru a identifica, categorisi sau deduce caracteristici sau emoții ale indivizilor, în special dacă acestea conduc la supravegherea de masă, și sisteme de IA utilizate pentru scorul social, pentru a determina accesul la servicii esențiale, ca aplicații care pot necesita o atenție deosebită, ținând cont de posibilele excepții legitime. Cu toate acestea, un moratoriu sau o interdicție ar trebui luate în considerare numai atunci când, pe o bază obiectivă, a fost identificat un risc inacceptabil pentru drepturile omului, democrația sau statul de drept și, după o examinare atentă, nu există alte măsuri fezabile și la fel de eficiente disponibile pentru atenuarea acestui risc și având în vedere sfera specifică de aplicare. Ar trebui puse în aplicare proceduri de revizuire pentru a permite revocarea unei interdicții sau a unui moratoriu dacă riscurile sunt reduse suficient sau dacă măsuri de atenuare adecvate devin disponibile, pe o bază obiectivă, pentru a nu mai prezenta un risc inacceptabil.

Elemente referitoare la dezvoltarea, proiectarea și aplicarea sistemelor de inteligență artificială în general.

Se recomandă ca un instrument transversal juridic obligatoriu să includă o serie de dispoziții aplicabile tuturor dezvoltărilor, proiectărilor și aplicărilor sistemelor de IA, pentru a permite clasificarea adecvată a acestora în ceea ce privește riscul potențial pentru exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept și pentru a asigura conformitatea acestora prin stabilirea unor garanții minime. Acestea pot include, de exemplu, dispoziții privind transparența sistemelor de IA. În conformitate cu abordarea bazată pe risc menționată mai sus, ar trebui aplicate dispoziții suplimentare sistemelor de IA pe baza și proporțional cu clasificarea riscului acestora, pentru a se asigura că riscurile pe care le prezintă pentru drepturile omului, democrația și statul de drept sunt atenuate corespunzător.

Un instrument transversal juridic obligatoriu ar trebui, în general, să stabilească că, sub rezerva anumitor limitări, dezvoltarea și proiectarea, precum și cercetarea în domeniul sistemelor de IA ar trebui să se efectueze liber, ținând seama de siguranță și securitate și în deplină conformitate cu standardele Consiliului Europei privind drepturile omului.

În plus, se recomandă includerea unei dispoziții care să încurajeze Părțile să stabilească „spații de reglementare” pentru a stimula inovarea responsabilă în sistemele de IA, permițând testarea sistemelor de IA sub supravegherea autorității naționale competente, asigurând în același timp conformitatea cu standardele stabilite în Convenția pentru protecția persoanelor cu privire la prelucrarea automată a datelor cu caracter personal (CETS nr. 108) și Protocolul său modificator (CETS nr. 223), precum și cu standardele stabilite în acest instrument transversal juridic obligatoriu privind proiectarea, dezvoltarea și aplicarea IA și orice alte standarde aplicabile.

Pentru a promova o abordare multi-actor și pentru a crește gradul de conștientizare în societate cu privire la impactul dezvoltării, proiectării și aplicării sistemelor de IA, se consideră util să se includă o dispoziție care să îndemne Părțile să promoveze deliberări publice bazate pe dovezi și implicarea incluzivă în acest subiect. Inspirația pentru formularea unei astfel de dispoziții poate fi găsită în articolul 28 din Convenția pentru protecția drepturilor omului și a demnității ființei umane în ceea ce privește aplicarea biologiei și medicinei (CETS nr. 164).

Se propune astfel includerea unei dispoziții privind prevenirea prejudiciului ilegal care poate rezulta din dezvoltarea, proiectarea și aplicarea sistemelor de IA, inclusiv clarificarea conceptului de „prejudiciu ilegal” în scopul instrumentului transversal privind IA, drepturile omului, democrația și statul de drept.

Se propune, de asemenea, să includă o dispoziție privind respectarea egalității de tratament și nediscriminarea indivizilor în raport cu dezvoltarea, proiectarea și aplicarea sistemelor de IA pentru a evita integrarea unor prejudecăți nejustificate în sistemele de IA și utilizarea sistemelor de IA care conduc la efecte discriminatorii.

Din aceleași motive, un instrument transversal juridic obligatoriu ar trebui să conțină dispoziții privind asigurarea respectării egalității de gen și a drepturilor legate de grupurile vulnerabile și persoanele aflate în situații vulnerabile, inclusiv copiii, pe tot parcursul ciclului de viață al sistemelor de inteligență artificială. Se consideră, de asemenea, că este prudent să se includă o dispoziție referitoare la guvernanța datelor pentru sistemele IA, în conformitate cu și bazată pe Convenția Consiliului Europei pentru Protecția Persoanelor cu Privire la Prelucrarea Automată a Datelor cu Caracter Personal (CETS nr. 108) și Protocolul său de modificare (CETS nr. 223). Aceasta poate include cerința de a stabili mecanisme de guvernanță a datelor pentru a evalua și asigura acuratețea, integritatea, securitatea și reprezentativitatea datelor, într-un mod adecvat scopului sistemului și proporțional.

Se recomandă, de asemenea, introducerea de dispoziții privind robustețea, siguranța și securitatea cibernetică, transparența, explicabilitatea, auditabilitatea și responsabilitatea de-a lungul ciclului de viață al acestor sisteme; noțiunile de „transparență”, „explicabilitate” și „responsabilitate” sunt de o importanță majoră pentru protecția drepturilor indivizilor în contextul sistemelor IA. În plus, se recomandă ca aspectul durabilității sistemelor IA de-a lungul ciclului lor de viață să fie luat în considerare într-un mod adecvat. Nu în ultimul rând, un instrument transversal cu caracter obligatoriu ar trebui să includă o prevedere care să asigure un nivel necesar de supraveghere umană asupra sistemelor IA și efectelor acestora pe parcursul întregului lor ciclu de viață.

Elemente legate de dezvoltarea, proiectarea și aplicarea sistemelor de inteligență artificială în sectorul public.

Dezvoltarea, proiectarea și aplicarea sistemelor de inteligență artificială în sectorul public generează anumite preocupări privind modul de a asigura respectarea drepturilor omului, a democrației și a statului de drept atunci când sistemele IA sunt utilizate pentru a lua sau a influența decizii care afectează drepturile și obligațiile persoanelor fizice și juridice. Cu toate acestea, *COMITETUL subliniază că nu toate aplicațiile IA din sectorul public reprezintă riscuri pentru exercitarea drepturilor omului, pentru funcționarea democrației sau pentru respectarea statului de drept. Astfel, este important să se examineze cu atenție riscul potențial al unei aplicații IA anume, de la caz la caz. În consecință, ar trebui să se facă o distincție între, pe de o parte, sistemele IA care pot interfera cu drepturile omului, democrația sau statul de drept și, pe de altă parte, sistemele IA care, deși sunt operate de aceleași autorități publice, nu prezintă astfel de riscuri.

Pornind de la premisa că un instrument transversal cu caracter obligatoriu ar trebui să fie de natură generală, se recomandă ca acest instrument să se concentreze asupra riscurilor potențiale care apar din dezvoltarea, proiectarea și aplicarea sistemelor de inteligență artificială în scopuri legate de aplicarea legii, administrarea justiției și administrația publică. În ceea ce privește „administrația publică”, se subliniază că un astfel de instrument juridic obligatoriu nu ar trebui să abordeze multitudinea de activități administrative specifice întreprinse de autoritățile publice, cum ar fi asistența medicală, educația, beneficiile sociale etc., ci să se limiteze la dispoziții generale privind utilizarea responsabilă a sistemelor IA în administrația publică. Problemele legate de diversele sectoare ale administrației publice pot fi, dacă este necesar, abordate prin instrumente sectoriale adecvate. Se consideră că un instrument transversal cu caracter obligatoriu, atunci când

abordează dezvoltarea, proiectarea și aplicarea sistemelor de IA în sectorul public, ar trebui să includă, cel puțin, dispoziții privind accesul la o cale de atac eficientă, un drept obligatoriu la revizuire umană a deciziilor luate sau influențate de un sistem IA, cu excepția cazurilor în care există motive legitime și predominante care exclud acest drept. De asemenea, autoritățile publice ar trebui să fie obligate să implementeze o revizuire umană adecvată pentru procesele influențate sau susținute de sisteme IA și să ofere persoanelor fizice sau juridice informații relevante privind rolul sistemelor IA în luarea sau influențarea deciziilor referitoare la acestea, cu excepția cazurilor în care motive legitime și predominante exclud sau limitează această revizuire sau divulgare.

Mai mult, părțile ar trebui să fie obligate să se asigure că legislația lor internă oferă garanții adecvate și eficiente împotriva practicilor arbitrare și abuzive cauzate de aplicarea unui sistem IA în sectorul public. Părțile ar trebui să asigure conformitatea cu standardele referitoare la sistemele de IA în ceea ce privește drepturile omului, democrația și statul de drept, în măsura în care actorii privați care acționează în numele lor sunt implicați.

Elemente legate de democrație și guvernanța democratică.

Deși *COMITETUL recunoaște că IA poate avea un rol pozitiv în funcționarea democrației și guvernanței democratice, facilitând procese incluzive și participative, există și îngrijorări cu privire la potențiala utilizare a IA pentru a interveni în mod ilegal sau nejustificat în procesele democratice. Formarea opiniei publice prin IA, precum și potențialele efecte de descurajare generate de utilizarea IA ar trebui, prin urmare, analizate în contextul unui posibil instrument juridic obligatoriu. Totodată, problemele mai specifice legate de manipularea alegerilor, cum ar fi micro-targetarea, profilarea și manipularea conținutului (inclusiv „deep fakes"), ar putea fi abordate prin instrumente sectoriale specifice.

Rolul entităților private, de exemplu, al platformelor online care contribuie la formarea sferei publice, ar trebui, de asemenea, avut în vedere în acest context, în măsura în care concentrarea tot mai mare a puterii economice și a datelor ar putea submina procesele democratice. În acest context, se subliniază necesitatea respectării dreptului la libertatea de exprimare, inclusiv libertatea de a forma și de a păstra opinii și de a primi și transmite informații și idei politice, precum și dreptul la libertatea de întrunire și asociere, cu scopul de a asigura că toate părțile și grupurile de interes au acces egal la procesele democratice și că se poate asigura un spațiu liber pentru dezbaterea publică.

Elemente legate de măsurile de protecție.

Se recomandă ca un instrument transversal cu caracter obligatoriu să includă o serie de dispoziții privind măsurile legale de protecție care trebuie aplicate tuturor aplicațiilor sistemelor de IA utilizate în scopul luării sau informării deciziilor care afectează drepturile legale și alte interese semnificative ale persoanelor fizice și juridice. Aceste protecții ar trebui să asigure un echilibru între utilizarea tehnologică și drepturile fundamentale ale individului, precum și transparența și responsabilitatea în procesele decizionale automate.

Aceste măsuri de protecție ar trebui să includă cel puțin următoarele drepturi: dreptul la o cale de atac eficientă în fața unei autorități naționale (inclusiv autorități judiciare) împotriva deciziilor luate de un sistem IA; dreptul de a fi informat despre aplicarea unui sistem IA în procesul decizional; dreptul de a alege interacțiunea cu un om în plus față de sau în locul unui sistem IA; și dreptul de a ști că se interacționează cu un sistem IA și nu cu o persoană umană. Alte măsuri de protecție pot fi relevante în funcție de specificul sistemelor IA utilizate. Modalitățile de exercitare a acestor drepturi ar trebui prevăzute de legislația națională. Excepțiile legitime de la aceste drepturi pot fi prevăzute de lege, acolo unde este necesar și proporționat într-o societate democratică.

În final, se consideră că un instrument transversal cu caracter obligatoriu ar trebui să includă și o prevedere privind protecția avertizorilor de integritate (whistle-blowers) în legătură cu dezvoltarea, proiectarea și aplicarea sistemelor IA care ar putea afecta negativ exercitarea drepturilor omului, funcționarea democrației și respectarea statului de drept. Această prevedere ar trebui să respecte limitele legale legitime privind divulgarea informațiilor. Astfel, protejarea celor care semnalează posibile abuzuri sau riscuri legate de IA este esențială pentru a preveni încălcările drepturilor fundamentale și a asigura transparența în aplicarea tehnologiilor.

Elemente legate de răspunderea civilă.

Deși Comitetul recunoaște că problemele legate de răspunderea civilă și dezvoltarea, proiectarea și aplicarea sistemelor IA ar fi, în general, reglementate de legislația internă existentă a părților unui eventual instrument juridic obligatoriu, totuși consideră utilă examinarea acestei probleme în detaliu, pentru a explora necesitatea asigurării unei abordări comune fundamentale privind răspunderea civilă în raport cu IA.

Acest punct subliniază importanța clarificării modului în care răspunderea civilă ar trebui aplicată în cazurile în care sistemele IA cauzează daune sau prejudicii indivizilor sau entităților

legale, într-un context global. În condițiile în care tehnologia IA poate opera autonom și poate lua decizii care afectează direct drepturile fundamentale, stabilirea unui cadru comun pentru răspunderea civilă poate contribui la asigurarea transparenței și la protejarea drepturilor persoanelor afectate. De asemenea, acest cadru ar putea facilita cooperarea internațională și standardizarea reglementărilor, prevenind astfel conflictele legale și încurajând responsabilitatea tehnologică.

Elemente legate de autoritățile de supraveghere, conformitate și cooperare.

Se consideră că un instrument transversal cu caracter obligatoriu ar trebui să includă dispoziții care să oblige părțile să ia toate măsurile necesare și adecvate pentru a asigura conformitatea eficientă cu instrumentul respectiv, în special prin stabilirea de mecanisme și standarde de conformitate. În plus, ar trebui să fie prevăzute dispoziții referitoare la stabilirea sau desemnarea autorităților naționale de supraveghere, definirea atribuțiilor și funcționării acestora, asigurarea expertizei, independenței și imparțialității acestora în îndeplinirea funcțiilor lor, precum și alocarea de resurse și personal suficient.

De asemenea, instrumentul juridic obligatoriu ar trebui să conțină dispoziții care reglementează cooperarea între părți și asistența mutuală juridică și de alt tip, inclusiv schimbul de date și alte forme de informații, asigurând totodată coerența cu alte instrumente deja aplicabile ale Consiliului Europei în domeniul asistenței juridice internaționale.

Un instrument juridic obligatoriu ar trebui, de asemenea, să conțină dispoziții privind înființarea unui „comitet al părților” pentru a sprijini implementarea instrumentului. În această privință, se face referire la dispozițiile standard utilizate în alte instrumente juridice obligatorii ale Consiliului Europei, care, dacă și pe măsură ce este necesar, pot fi modificate pentru a se potrivi mai bine scopurilor acestui instrument juridic obligatoriu.

Aceste măsuri sunt esențiale pentru asigurarea unui cadru de reglementare eficient și pentru sprijinirea aplicării uniforme a normelor privind IA în întreaga regiune, promovând în același timp transparența, responsabilitatea și cooperarea internațională.

Partea a III-a: Elemente pentru posibile instrumente juridice suplimentare.

Evaluarea impactului asupra drepturilor omului, democrației și statului de drept.

Se consideră util și necesar să completeze un instrument transversal juridic obligatoriu cu un model non-obligatoriu pentru evaluarea impactului sistemelor de IA asupra exercitării drepturilor omului, funcționării democrației și respectării statului de drept.

O evaluare bine realizată a impactului asupra drepturilor omului, democrației și statului de drept poate sprijini analiza modului în care implementarea sistemelor de IA poate influența drepturile fundamentale, democrația și statul de drept. Totuși, trebuie menționat că acest tip de evaluare nu este destinat să echilibreze impacturile negative și pozitive, ceva care ar putea depinde de specificitățile sistemului juridic din jurisdicția în care se intenționează aplicarea sistemului IA. Într-o etapă ulterioară, se poate examina dacă și cum riscurile identificate prin HUDERIA (Evaluarea impactului asupra drepturilor omului, democrației și statului de drept) pot fi atenuate și dacă și cum un interes legitim poate justifica utilizarea sistemului în ciuda interferenței cu drepturile omului, democrația și standardele statului de drept, atunci când astfel de limitări sunt prescrise de lege, proporționale și necesare într-o societate democratică.

Într-adevăr, un HUDERIA (*Evaluarea impactului asupra drepturilor omului, democrației și statului de drept*) nu ar trebui să funcționeze izolat, ci să fie completat la nivelul legislației interne sau internaționale prin alte mecanisme de conformitate, precum certificarea și etichetarea calității, audituri, "sandboxuri" reglementare și monitorizare continuă, așa cum a fost subliniat în Studiul de Fezabilitate. Este esențial ca evaluarea impactului să fie aliniată cu aceste mecanisme de conformitate, deoarece ar fi costisitor și inutil să se ceară evaluări separate ale impactului asupra drepturilor omului, democrației și statului de drept, care să nu corespundă abordărilor publice de supraveghere sau reglementare stabilite în legislația internă. În plus față de mecanismele de conformitate, trebuie asigurată disponibilitatea unor remedii eficiente pentru persoanele care ar putea fi afectate negativ de implementarea sistemelor IA.

În ceea ce privește resursele necesare și timpul pentru efectuarea unei astfel de evaluări și pentru a proteja proporționalitatea unei abordări bazate pe risc, Comitetul consideră că, în mod normal, o evaluare formalizată și extinsă a impactului asupra drepturilor omului, democrației și statului de drept ar trebui solicitată doar dacă există indicii clare și obiective ale unor riscuri relevante provenite din aplicarea unui sistem IA. Acest lucru presupune ca toate sistemele IA să treacă printr-o revizuire inițială pentru a determina dacă trebuie sau nu să fie supuse unei astfel de evaluări formalizate. Se recomandă dezvoltarea în continuare a unor indicații pentru necesitatea unei evaluări mai detaliate. De asemenea, ar trebui luat în considerare faptul că utilizarea unui sistem IA într-un context nou sau diferit, pentru un scop nou sau diferit, sau alte modificări relevante, ar necesita o reevaluare.

Comitetul subliniază că adoptarea unei abordări bazate pe risc presupune ca orice impact relevant al aplicării unui sistem IA asupra drepturilor omului, funcționării democrației și respectării statului de drept să fie evaluat și revizuit în mod sistematic și regulat, în scopul identificării măsurilor de atenuare adaptate riscurilor respective. În cazul în care aceste măsuri de atenuare nu sunt considerate suficiente, trebuie aplicate măsuri de interzicere, acolo unde este necesar. Mai mult, având în vedere necesitatea unui proces de evaluare iterativ, astfel de evaluări ar trebui realizate din nou ori de câte ori un sistem IA suferă modificări substanțiale. Se recomandă ca, cel puțin, următorii pași principali să fie incluși într-o evaluare a impactului asupra drepturilor omului, democrației și statului de drept, sub rezerva realizării unei revizuiri inițiale și a implicării părților interesate, acolo unde este relevant.

Identificarea riscurilor: Identificarea riscurilor relevante pentru drepturile omului, democrație și statul de drept. Evaluarea impactului: Evaluarea impactului, luând în considerare probabilitatea și severitatea efectelor asupra acestor drepturi și principii.

Evaluarea guvernanței: Evaluarea rolurilor și responsabilităților celor care trebuie să apere drepturile, titularii drepturilor și părțile interesate în implementarea și guvernarea mecanismelor pentru atenuarea impactului.

Mitigare și evaluare: Identificarea măsurilor adecvate de atenuare și asigurarea unei evaluări continue a eficienței acestora.

În ceea ce privește pasul de evaluare a impactului, se recomandă ca evaluarea unui sistem IA să includă cel puțin următoarele elemente: Evaluarea contextului și scopului sistemului IA: Analiza utilizării sistemului în contextul și pentru scopurile pentru care a fost creat.

Nivelul de autonomie al sistemului IA: Determinarea autonomiei sistemului în deciziile sale, adică cât de mult depinde de intervenția umană. Tehnologia de bază a sistemului IA: Evaluarea tehnologiilor pe care se bazează sistemul (ex. rețele neuronale, algoritmi de învățare automată etc.).

Utilizarea sistemului IA: Atât utilizarea intenționată, cât și posibilele utilizări neintenționate ale sistemului. Complexitatea sistemului IA: Evaluarea complexității, de exemplu, dacă este un sistem care utilizează multiple rețele neuronale adânci sau se bazează pe alte sisteme IA. Transparența și explicabilitatea: În ce măsură utilizatorii pot înțelege procesul decizional al IA. Scopul geografic și temporal: Limitările geografice și temporale ale utilizării sistemului IA.

Evaluarea probabilității și severității daunelor potențiale: Riscurile posibile și măsura în care acestea ar putea afecta drepturile fundamentale. Reversibilitatea daunelor: Posibilitatea de a remedia efectele negative ale unui sistem IA.

Aplicația „linie roșie”: Dacă sistemul IA este utilizat într-un domeniu reglementat de lege, unde utilizarea acestuia ar putea fi considerată neacceptabilă din punct de vedere legal (în conformitate cu legislația internă sau internațională).

Supraveghere umană și mecanisme de control: Măsura în care există mecanisme de control și supraveghere din partea furnizorului IA și utilizatorului.

Calitatea datelor: Asigurarea că datele utilizate de IA sunt corecte și de înaltă calitate. Robustețea și securitatea sistemului: Capacitatea sistemului IA de a rezista atacurilor cibernetice și erorilor de operare. Implicarea grupurilor vulnerabile: Dacă anumite grupuri vulnerabile (de exemplu, minori sau persoane cu dizabilități) sunt afectate de sistemul IA. Scalabilitatea: Evaluarea măsurii în care sistemul IA poate fi extins pentru a afecta un număr mare de persoane sau situații.

Mai mult, se observă că, în timp ce evaluarea impactului sistemelor de IA este relativ simplă în ceea ce privește drepturile omului, datorită existenței unor obligații clare și universale în acest domeniu, evaluarea impactului sistemelor de IA asupra democrației și statului de drept poate fi mai dificilă în unele cazuri. Cu toate acestea, având în vedere puternica interconexiune dintre drepturile omului, pe de o parte, și democrație și statul de drept, pe de altă parte, în unele situații, un impact negativ asupra primului poate oferi, de asemenea, o indicație a unui impact negativ asupra celui din urmă. De exemplu, atunci când dreptul la libertatea de întrunire și asociere sau dreptul la alegeri libere este îngrădit, acesta împiedică funcționarea democrației. În același sens, o interferență cu dreptul la un proces echitabil afectează negativ statul de drept. În plus, pot fi luate în considerare și alte elemente, cum ar fi scopul și funcția sistemului într-o societate democratică, domeniul său de aplicare (cu o atenție deosebită acordată utilizării sistemelor de IA în sectorul public sau în sfera publică) și modul în care poate împiedica anumite principii democratice și ale statului de drept (cum ar fi principiul legalității, prevenirea abuzului de putere sau imparțialitatea și independența judiciară).

În final, se consideră că ar trebui asigurată implicarea părților interesate în evaluarea impactului. Cu cât impactul este considerat a fi mai sever sau cu cât este mai mare amploarea acestuia, cu atât mai extinsă ar trebui să fie implicarea părților interesate. În această privință, ar

trebui acordată o atenție deosebită implicării părților interesate externe și a membrilor societății (adică celor care nu sunt incluși în categoriile de „furnizori de IA” și „utilizatori de IA”, așa cum sunt enumerate în Capitolul III) care ar putea fi afectați negativ de implementarea sistemului de IA.

Elemente complementare privind inteligența artificială în sectorul public.

Așa cum este stabilit în Capitolul VII, dezvoltarea, proiectarea și aplicarea sistemelor IA în sectorul public ar trebui să fie reglementate printr-un instrument juridic obligatoriu transversal, care să acopere cele mai importante drepturi și obligații transversale ce trebuie respectate în acest domeniu. În plus, Comitetul consideră că, având în vedere specificitatea riscurilor generate de IA în sectorul public, în lumina rolului său specific în societate, un astfel de cadru transversal ar putea fi completat cu instrumente suplimentare, juridic obligatorii sau neobligatorii, la nivel sectorial. Aceste instrumente ar putea, de exemplu, să elaboreze în continuare principii și cerințe specifice pentru serviciile publice, referitoare la transparență, corectitudine, responsabilitate, responsabilizare, explicabilitate și remediere, pentru a asigura utilizarea responsabilă a IA. Se recomandă ca utilizarea, proiectarea, achiziționarea, dezvoltarea și implementarea sistemelor IA în sectorul public să fie supuse unor mecanisme adecvate de supraveghere, pentru a proteja respectarea drepturilor omului, principiilor democratice și statului de drept, și pentru a consolida încrederea publicului prin garantarea unui utilizării de încredere a sistemelor IA, adică inteligibile, urmărite și auditabile.

În plus, având în vedere că distincția dintre implicarea sectorului public și cel privat este adesea ambiguă și având în vedere problemele de răspundere legate de externalizarea serviciilor publice către actori privați, orice dispoziții aplicabile la proiectarea, dezvoltarea și aplicarea IA în sectorul public ar trebui să se aplice și actorilor privați care acționează în numele sectorului public. „*Se consideră că următoarele elemente referitoare la proiectarea, achiziționarea, dezvoltarea și implementarea unui sistem de IA de către o entitate publică ar putea, în plus față de elementele deja descrise în Capitolul VII, să fie abordate ca parte a unui instrument legal sau non-legal referitor la IA în sectorul public*:”

În faza de proiectare a sistemului, Comitetul este de părere că un instrument legal sau non-legal ar putea aborda modul în care ar trebui să se acorde o atenție deosebită analizei problemei pe care entitatea publică intenționează să o rezolve, pentru a evalua dacă un sistem de IA este soluția adecvată pentru problemă și, dacă da, ce caracteristici ar trebui să aibă. Un instrument legal sau

non-legal ar putea aborda, de asemenea, următoarele aspecte: seturile de date care urmează a fi utilizate pentru sistemul de IA ar trebui să fie clar identificate, iar protecția acestor date și proveniența lor să fie respectate. Alegerile de design ale sistemului ar trebui să fie explicite și documentate. Utilizatorii vizați ai sistemului, atât funcționarii publici, cât și publicul, precum și cei care ar putea fi afectați de sistem, ar trebui să fie implicați din timp, iar capacitățile lor de a utiliza sistemul de IA în cauză ar trebui luate în considerare. Ar trebui favorizată o abordare deschisă și transparentă de co-proiectare. În final, ar trebui realizată o evaluare a impactului asupra drepturilor omului, democrației și statului de drept pentru a anticipa, preveni și atenua riscurile potențiale. Aceasta presupune, de asemenea, implementarea unor cadre de management și atenuare a riscurilor, relevante pe parcursul tuturor fazelor.

În faza de achiziție, ar trebui realizată o revizuire amănunțită a legislației aplicabile și a măsurilor politice în vigoare. Acestea, dacă este necesar, ar trebui adaptate, iar ghiduri pentru achizițiile publice de IA ar trebui adoptate, pentru a asigura că sistemele de IA achiziționate respectă standardele privind drepturile omului, democrația și statul de drept. Ar trebui asigurată o abordare multidisciplinară și multi-părți interesate pentru a implica diverse perspective, inclusiv cele ale grupurilor vulnerabile. Deoarece entitățile publice sunt responsabile pentru sistemele pe care le adoptă și aplică, trebuie acordată o atenție deosebită impactului potențial asupra responsabilității publice.

În timpul fazei de dezvoltare, care este deosebit de sensibilă, procesele de documentare și înregistrare ar trebui păstrate cu meticulozitate pentru a asigura transparența și trasabilitatea sistemului. Ar trebui implementate procese adecvate de testare și validare, precum și mecanisme de guvernanță a datelor. Printre alte riscuri, riscul de acces sau tratament inegal, diverse forme de părtinire și discriminare, precum și impactul asupra egalității de gen ar trebui evaluate.

Modelele de gestionare a riscurilor și de reducere a acestora stabilite în etapele anterioare ar trebui evaluate, adaptate și menținute în timpul fazei de implementare. Având în vedere natura riscului, implicarea umană poate fi necesară pentru a asigura o supraveghere adecvată asupra sistemului. Dacă este cazul, sistemul de IA ar trebui să fie auditabil inițial și periodic de un actor independent, iar rezultatele să fie făcute publice pentru a sprijini încrederea publicului. În acest sens, se consideră că ar trebui abordată în cadrul unui instrument juridic sau non-juridic privind IA în sectorul public înființarea unor registre publice care să listeze sistemele de IA utilizate în sectorul public, conținând informații esențiale despre sistem, precum scopul acestuia, actorii

implicați în dezvoltarea și implementarea sa, informații de bază despre model și indicatorii de performanță, acolo unde este cazul, precum și rezultatul unui HUDERIA. În plus, acest instrument ar putea aborda înființarea unui mecanism de feedback pentru a colecta opinii despre cum poate fi îmbunătățit sistemul direct de la utilizatorii săi și de la cei care ar putea fi afectați de acesta. Mai mult, instrumentul ar putea aborda necesitatea ca sistemul de IA să fie supus unei evaluări și actualizări periodice, inclusiv prin luarea în considerare a feedback-ului.

Transparența și comunicarea către utilizatori și cetățeni ar trebui, de asemenea, abordate, precum și posibilitatea accesului la mecanisme de responsabilizare și de compensare individuală și colectivă. Nu în ultimul rând, instrumentul ar trebui să abordeze dreptul publicului de a fi informat despre faptul că interacționează cu un sistem de IA și nu cu o persoană umană, precum și dreptul de a interacționa cu o persoană umană în loc de a interacționa doar cu un sistem de IA, în special atunci când drepturile și interesele indivizilor sau persoanelor juridice ar putea fi afectate negativ. Excepțiile legitime de la aceste drepturi pot fi prevăzute de lege, atunci când sunt necesare și proporționale într-o societate democratică.

În cele din urmă, un instrument legal sau non-legal obligatoriu privind IA în sectorul public ar putea aborda măsuri pentru a spori alfabetizarea digitală și abilitățile atât ale funcționarilor publici, cât și ale publicului larg, în special prin investiții în dezvoltarea capacității (formare inițială și continuă și educație) a oficialilor publici și creșterea conștientizării cu privire la beneficiile, riscurile, capabilitățile și limitările sistemelor IA, precum și prin sprijinirea cercetării în interes public. Aceste abilități ar trebui să includă atât cunoștințe teoretice, cât și practice despre interacțiunea dintre proiectarea, dezvoltarea și aplicarea sistemelor IA, pe de o parte, și drepturile omului, democrația și statul de drept, pe de altă parte. Mai mult, instrumentul menționat mai sus ar putea aborda și modul în care aceste sisteme ar trebui să fie supravegheate și cum ar trebui gestionate riscurile care decurg din utilizarea acestora.

*b) **UE-AI Act[49]. REGULAMENTUL (UE) 2024/1689 AL PARLAMENTULUI EUROPEAN ȘI AL CONSILIULUI***

La 1 august 2024, a intrat în vigoare Regulamentul UE privind inteligența artificială (IA). Acesta urmărește să promoveze dezvoltarea și implementarea responsabilă a inteligenței artificiale în UE.

[49] https://eur-lex.europa.eu/legal-content/RO/TXT/PDF/?uri=OJ:L_202401689#page=44.71

Propus de Comisie în aprilie 2021 și aprobat de Parlamentul European și de Consiliu în decembrie 2023, Regulamentul privind IA abordează riscurile potențiale la adresa sănătății, siguranței și drepturilor fundamentale ale cetățenilor. Acesta prevede cerințe și obligații clare pentru dezvoltatori și operatori în ceea ce privește utilizările specifice ale IA, reducând, în același timp, sarcinile administrative și financiare pentru companii.

Regulamentul privind IA introduce un cadru uniform în toate țările UE, bazat pe o definiție prospectivă a IA și pe o abordare bazată pe ***riscuri.***

Risc minim sau fără risc: majoritatea sistemelor de IA, cum ar fi filtrele antispam și jocurile video bazate pe IA, nu au nicio obligație în temeiul Regulamentului privind IA, dar companiile pot adopta în mod voluntar coduri de conduită suplimentare.

Riscurile limitate se referă la riscurile asociate lipsei de transparență în utilizarea IA. Legea privind IA introduce obligații specifice de transparență pentru a se asigura că oamenii sunt informați atunci când este necesar, stimulând încrederea. De exemplu, atunci când utilizează sisteme de IA, cum ar fi roboții de chat, oamenii ar trebui să fie informați că interacționează cu o mașină, astfel încât să poată lua o decizie în cunoștință de cauză de a continua sau de a face un pas înapoi. Furnizorii trebuie, de asemenea, să se asigure că conținutul generat de IA este identificabil. În plus, textul generat de IA publicat cu scopul de a informa publicul cu privire la chestiuni de interes public trebuie etichetat ca fiind generat artificial. Acest lucru este valabil și pentru conținutul audio și video care constituie deepfake-uri.

Risc specific în materie de transparență: sistemele precum roboții de chat trebuie să îi informeze în mod clar pe utilizatori că interacționează cu o mașină, iar anumite conținuturi generate de IA trebuie etichetate ca atare.

Risc ridicat: sistemele de IA cu grad ridicat de risc, cum ar fi software-ul medical bazat pe IA sau sistemele de IA utilizate pentru recrutare, trebuie să respecte cerințe stricte – sisteme de atenuare a riscurilor, seturi de date de înaltă calitate, informații clare pentru utilizatori, supraveghere umană etc.

Risc inacceptabil: de exemplu, sistemele de IA care permit acordarea unui „punctaj social” de către guverne sau companii sunt considerate o amenințare clară la adresa drepturilor fundamentale ale cetățenilor și, prin urmare, sunt interzise.

Recent, Comisia a lansat o consultare privind un cod de bune practici pentru furnizorii de modele de inteligență artificială de uz general. Acest cod, prevăzut de Regulamentul privind IA,

va aborda domenii critice, cum ar fi transparența, normele legate de drepturile de autor și gestionarea riscurilor. Furnizorii de modele de inteligență artificială de uz general care desfășoară activități în UE, companiile, reprezentanții societății civile, titularii de drepturi și experții universitari sunt invitați să își prezinte opiniile și constatările. Comisia va ține cont de acestea în elaborarea viitorului proiect de cod de bune practici pentru furnizorii de modele de inteligență artificială de uz general.

Dispozițiile privind modelele de inteligență artificială de uz general vor intra în vigoare în termen de 12 luni. Comisia își propune să finalizeze codul de bune practici până în aprilie 2025. În plus, observațiile primite în urma consultării vor fi utilizate și de Oficiul pentru IA, care va supraveghea punerea în aplicare și respectarea dispozițiilor Regulamentului privind IA în relație cu modelele de inteligență artificială de uz general.

Pentru a asigura un nivel consecvent și ridicat de protecție a intereselor publice în ceea ce privește sănătatea, siguranța și drepturile fundamentale, ar trebui să fie stabilite norme comune pentru sistemele de IA cu grad ridicat de risc. Normele respective ar trebui să fie în concordanță cu carta și ar trebui să fie nediscriminatorii și în conformitate cu angajamentele comerciale internaționale ale Uniunii. De asemenea, ele ar trebui să țină seama de Declarația europeană privind drepturile și principiile digitale pentru deceniul digital și de Orientările în materie de etică pentru o IA fiabilă ale Grupului independent de experți la nivel înalt privind inteligența artificială.

Noțiunea de „sistem de IA” din regulament ar trebui să fie definită în mod clar și ar trebui să fie aliniată îndeaproape la lucrările organizațiilor internaționale care își desfășoară activitatea în domeniul IA, pentru a asigura securitatea juridică și a facilita convergența internațională și acceptarea pe scară largă, oferind în același timp flexibilitatea necesară pentru a ține seama de evoluțiile tehnologice rapide din acest domeniu. În plus, definiția ar trebui să se bazeze pe caracteristicile esențiale ale sistemelor de IA, care le diferențiază de sistemele software sau abordările de programare tradiționale mai simple, și nu ar trebui să acopere sistemele care se bazează pe reguli definite exclusiv de persoane fizice pentru a executa în mod automat anumite operațiuni. *O caracteristică esențială a sistemelor de IA este capabilitatea lor de a realiza deducții*. Această capabilitate se referă la procesul de obținere a rezultatelor, cum ar fi previziuni, conținut, recomandări sau decizii, care pot influența mediile fizice și virtuale, precum și la capabilitatea sistemelor de IA de a deriva modele sau algoritmi, sau ambele, din date sau din datele de intrare. Printre tehnicile folosite în timpul conceperii unui sistem de IA care fac posibilă

realizarea de deducții se numără abordările bazate pe învățarea automată, care presupun a învăța, pe baza datelor, cum pot fi atinse anumite obiective, și abordările bazate pe logică și pe cunoaștere, care presupun realizarea de deducții pornind de la cunoștințe codificate sau de la reprezentarea simbolică a sarcinii de rezolvat. Capacitatea unui sistem de IA de a realiza deducții depășește simpla prelucrare de date, făcând posibile învățarea, raționamentul sau modelarea.

Termenul „bazat pe calculator" se referă la faptul că sistemele de IA rulează pe calculatoare. Trimiterea la obiective explicite sau implicite subliniază faptul că sistemele de IA pot funcționa în conformitate cu obiective definite explicite sau cu obiective implicite. Obiectivele sistemului de IA pot fi diferite de scopul preconizat al sistemului de IA într-un context specific. În sensul prezentului regulament, mediile ar trebui să fie înțelese ca fiind contextele în care funcționează sistemele de IA, în timp ce rezultatele generate de sistemele de IA reflectă diferitele funcții îndeplinite de acestea și includ previziuni, conținut, recomandări sau decizii. Sistemele de IA sunt concepute pentru a funcționa cu diferite niveluri de autonomie, ceea ce înseamnă că au un anumit grad de independență a acțiunilor față de implicarea umană și anumite capabilități de a funcționa fără intervenție umană. Adaptabilitatea de care un sistem de IA ar putea da dovadă după implementare se referă la capabilitățile de autoînvățare, care permit sistemului să se modifice în timpul utilizării. Sistemele de IA pot fi utilizate în mod independent sau ca o componentă a unui produs, indiferent dacă sistemul este integrat fizic în produs (încorporat) sau dacă servește funcționalității produsului fără a fi integrat în acesta (neîncorporat).

c) Convenția-cadru a Consiliului Europei privind inteligența artificială și drepturile omului, democrația și statul de drept, din 17 mai 2024.

La 14 martie 2024 Comitetul privind inteligența artificială (CAI) al Consiliului Europei (CE) a finalizat proiectul său de Convenție-cadru asupra inteligenței artificiale, drepturilor omului, democrației și a statului de drept, care va fi *primul tratat internațional în materie*. Această „realizare extraordinară" (după caracterizarea Marijei Pejčinović Burić, Secretar general al CE) acoperă sistemele de inteligență artificială de-a lungul ciclului lor de viață și va garanta că avântul fenomenului respectă normele juridice ale organizației europene în materie de drepturi umane fundamentale, democrație și stat de drept. La 20 martie a.c. Comitetul Miniștrilor al CE a decis să transmită textul Adunării Parlamentare a Consiliului Europei pentru un aviz, așteptat cu ocazia sesiunii sale plenare din aprilie. După examinarea sa de către organismul parlamentar, proiectul va

trebui să fie adoptat cu ocazia reuniunii Comitetului Miniștrilor din 17 mai a.c., care se va ține la nivelul miniștrilor de externe ai celor 46 de statele membre (dintre care 27 membre ale UE) și care va marca cea de-a 75 aniversare a organizației interstatale continentale.

Convenția-cadru va fi deschisă ratificării la nivel internațional în scopul de a permite țărilor din întreaga lume să adere și să respecte normele etico-juridice astfel stabilite. Înscriindu-se în viziunea unei abordări globale a problematicii IA prefigurată și de rezoluția Adunării Generale a ONU din 21 martie a.c. Sesizarea posibilităților oferite de sistemele de inteligență artificială sigure, securizate și demne de încredere pentru dezvoltarea durabilă, ineditul tratat deschide calea instituirii unui cadru juridic internațional general, cu o vocație mondială și pune fundamentele dreptului internațional al inteligenței artificiale (DIA). În același timp, alături de Regulamentul de stabilire a unor norme armonizate privind inteligența artificială (legea privind inteligența artificială) al UE, adoptat la 13 martie a.c., Convenția-cadru a CE contribuie la afirmarea influenței normative internaționale a Europei și a capacității organizațiilor sale – CE și UE – de a sesiza provocarea globală majoră a IA.

Consiliul Europei a reacționat relativ timpuriu la problematica IA, în special în raport cu impactul său asupra drepturilor omului, democrației și a statului de drept. Astfel, în 2019, sub egida sa, s-a publicat Recomandarea Decodarea IA: 10 măsuri pentru a proteja drepturile umane, prin care se oferea statelor membre o serie de orientări generale referitor la marile principii de urmat spre a preveni ori atenua efectele negative ale IA asupra drepturilor fundamentale. Documentul are în vedere mai ales riscurile specifice care fac să apese din partea inteligenței artificiale asupra dreptului la discriminare și la egalitate, dreptului la protecția datelor și a vieții private, libertății de expresie, de reuniune și dreptului la muncă. În plus, sunt evidențiate șapte domenii cheie cărora se impune a li se acorda o atenție particulară în acest sens: necesitatea de a se realiza studii de impact asupra drepturilor înainte de achiziția, dezvoltarea și/ori desfășurarea unui sistem de IA; respectarea normelor în materie de drepturi umane în sectorul privat; informare și transparență; necesitatea de consultări publice pe deplin pertinente; promovarea educației privind IA; nevoia unui control independent și acces la căi de recurs aferente. Statele membre au luat măsuri în unele dintre sectoarele identificate în Recomandarea din 2019, dar fără o coerență globală. Astfel, în multe privințe reglementarea sistemelor de IA centrată pe drepturi umane e încă absentă și cel mai adesea autoritățile publice intervin prea târziu și avansează prea lent pentru ca a lor contribuție să fie cu adevărat reală. Normele și garanțiile în materie de drepturi umane, chiar

dacă neutre din punct de vedere tehnologic și aplicabile în toate contextele, inclusiv în cele care operează sistemele de IA, sunt arareori aplicabile și controalele rămân sporadice. Documentul din 2019 a fost urmat în 2023 de Recomandarea: Drepturile umane ale concepției IA – o protecție durabilă a drepturilor umane în era inteligenței artificiale, care examinează provocările cu care se confruntă statele membre în aplicarea recomandărilor anterioare, în special în ceea ce privește pertinența evaluării riscurilor și studiul de impact asupra drepturilor umane, aplicarea de garanții de transparență mai solide și necesitatea unui control independent.

După Preambul (în 17 puncte), cele 36 articole ale textului final al proiectului de *Convenție-cadru* sunt structurate în opt capitole: dispoziții generale, obligații generale, principii relative la activități desfășurate în cadrul ciclului de viață al sistemelor de inteligență artificială, recurs, evaluare și atenuare a riscurilor și a impacturilor negative, punerea în aplicare a Convenției, mecanisme de urmărire și cooperare, clauze finale. Considerentele Preambulului evocă, în primul rând, dubla perspectivă, în privința efectelor antrenate ale activităților desfășurate în cadrul ciclului de viață al sistemelor de inteligență artificială și, în consecință, nevoia de reglementare a acestora în consens cu valorile organizației de cooperare politică interstatală europeană. Pe de o parte, potențialul de a promova prosperitatea umană și bunăstarea indivizilor și a societății, dezvoltarea durabilă, oportunitățile fără precedent pentru protecția și promovarea drepturilor omului, democrației și statului de drept (pct. 3 și 4). Pe de alta, faptul că anumite activități de acest gen „pot aduce atingerea demnității umane și autonomiei persoanelor, drepturilor omului, democrației și statului" (pct. 5), riscuri de discriminare în cadrele digitale, în particular cele implicând sistemele de IA și prin aceasta efectul potențial de creare ori agravare de inegalități (pct. 6) ori utilizarea abuzivă a sistemelor respective, în scopuri represive, cu violarea dreptului internațional al drepturilor omului, în special prin practici de supraveghere și de cenzură arbitrare ori ilegale ce aduc atingere vieții private și autonomiei individuale" (pct. 7). În același timp se recunoaște că documentul „are valoare de cadru și că poate să fie completată prin alte instrumente destinate să trateze chestiuni specifice legate de activitățile desfășurate în cadrul ciclului de viață al sistemelor de inteligență artificială" (pct. 11). În mod declarat, preconizata convenție-cadru se alătură celorlalte instrumente internaționale aplicabile în materie de drepturi ale omului, începând cu Declarația universală din 1948 și are ca preocupare definitorie afirmarea „angajamentului Părților de a proteja drepturile omului, democrația și statul de drept și a favoriza fiabilitatea sistemelor de inteligență artificială" (pct. 17).

Dispozițiile convenției „vizează să garanteze că activitățile duse în cadrul ciclului de viață al sistemelor de inteligență artificială sunt pe deplin compatibile cu drepturile omului, democrația și statul de drept". Pentru aceasta „fiecare Parte adoptă ori menține măsurile legislative, administrative ori altele apropriate pentru a da efect dispozițiilor Convenției; respectivele măsuri sunt gradate și diferențiate în funcție de gravitatea și de probabilitatea apariției de impacturi negative" asupra drepturilor umane, normelor democratice și valorile statului de drept. În vederea asigurării aplicării efective a dispozițiilor aferente se stabilește un mecanism de urmărire și se prevede o cooperare internațională adecvată (art. 1).

Apoi, este un tratat deschis ratificării pentru ansamblul statelor lumii, indiferent de localizarea lor geografică. Astfel, Convenția va fi deschisă semnăturii statelor membre ale Consiliului Europei, statelor nemembre care au participat la elaborare (precum Australia și Canada) și Uniunii Europene. După intrarea sa în vigoare Comitetul Miniștrilor CE va putea, după consultarea Părților și obținerea consimțământului lor unanim, să invite orice stat nemembru al Consiliului care nu a participat la elaborarea tratatului să adere la Convenția-cadru printr-o decizie luată cu majoritatea calificată de două treimi din voturile exprimate și unanimitatea reprezentanților Părților având dreptul de a face parte din Comitetul Miniștrilor (art. 30). Nu în ultimul rând, și ceea ce îi accentuează dimensiunea de globalitate, procesul de negociere a asociat multipli actori, cuprinzând membrii Consiliului Europei, state membre, de pe alte continente, Uniunea Europeană, actori ai industriei și ai societății civile. Suntem în prezența, așa cum îi arată și numele, unei *convenții-cadru* care valorifică experiențele de până acum în domeniu, stabilește jaloanele transnaționale bazate pe principii comune și deschide perspective de dezvoltare, completare și permanentă actualizare a sa prin elaborarea și adoptarea a noi reguli de aplicare sectoriale. E preconizată o Conferință a Părților, cu reuniuni periodice, menită să faciliteze utilizarea și aplicarea efectivă a convenției-cadru, să examineze posibilitățile de a o completa sau modifica, a formula recomandări particulare relative la interpretarea și aplicarea sa, a facilita cooperarea între state în domeniu (art. 23). *Marile principii ce vor trebui să guverneze activitatea statelor în materie sunt: demnitatea umană și autonomia individuală, transparență și control, obligația de a da seama și responsabilități, egalitate și nediscriminare, respectarea vieții private și protecția datelor cu caracter personal, prezervarea sănătății, fiabilitate și încredere, inovație sigură*.

d) Governing AI for Humanity, Actul digital mondial – adoptat de Adunarea Generală a ONU – 22/23.09.2024.

ONU este singura entitate mondială capabilă să reunească și să faciliteze colaborarea necesară, să-și asume responsabilitățile pentru a ajuta puterile publice, întreprinderile, experții și societatea civilă să coopereze veritabil, grație colectării de date, difuzării celor mai bune practici și asistență tehnică[50].

Un Pact digital mondial formulează o viziune comună pentru un viitor digital deschis, liber, sigur și centrat pe ființa umană, valorificând obiectivele și principiile expuse în Carta ONU, în Declarația Universală a Drepturilor Omului și în Programul 2030.

Conectivitatea numerică și întărirea capacităților. *Obiective*: combaterea prăpastiei numerice pentru conectarea întregii lumi, în special, a grupelor vulnerabile, a internetului într-o manieră veritabilă utilă și la un cost abordabil; dotarea tuturor în vederea deținerii de capacități numerice pentru a le da mijloacele de a participa deplin la economia numerică, de a se pregăti contra pericolelor și de a veghea la buna stare și dezvoltarea în planurile, fizic și mental.

Statele membre trebuie: a) să se angajeze să aplice politica și noile modele financiare de telecomunicații, să asigure o deservire la un comerț abordabil în zonele dificile de acces; și b) să se angajeze să întărească sau să instituie o educație publică privind cultura digitală și cunoștințe transdisciplinare și să încurajeze învățarea/perfecționarea în tot cursul vieții pentru toți lucrătorii.

Toate părțile interesate trebuie:

a) să convină ținte comune pentru o conectivitate universală și veritabilă, utilă și să se angajeze să urmeze și să aplice progresele realizate în raport cu aceste ținte;

b) să lărgească stabilimentele medicale și instituțiile publice privind activitățile de cartografie și de dezvoltare a conectivității, care sunt în curs pentru școli;

c) să se angajeze să coordoneze activitățile, subvențiile și incitațiile în favoarea formării tehnice și profesionale în digital și a echipamentelor de acces public, în special, pentru femei, fete și persoanele din zonele rurale; și

d) să fixeze o țintă constând într-un milion de ”campioni ai digitalului”, în serviciul obiectivelor de dezvoltare durabilă, la orizontul 2030.

[50] M. Voicu, *Pactul (digital) mondial – un viitor digital deschis, liber și sigur pentru umanitate (22-23.09.2024, ONU)*, www.juridice.ro

Organizațiile multilaterale trebuie: a) să fixeze o țintă revizuită de 100 miliarde de dolari pentru a se realiza obiectivele inițiativei Partener & Connect Digital Coalition, la Orizontul 2030 (Uniunea Internațională a Telecomunicațiilor); și b) să accelereze acțiunea vizând racordarea tuturor stabilimentelor școlare la internet, până în 2030, conform inițiativei ”Giga” a Uniunii internaționale a telecomunicațiilor și a Fondului națiunilor Unite pentru copilărie.

Cooperarea digitală in serviciul accelerării și realizării obiectivelor dezvoltării durabile.

Obiectivele: a) a se investi de o manieră țintă în infrastructurile și serviciile digitale comune și a acționa să progreseze cunoștințele mondiale și a acționa în comun a celor mai bune practici in materie de bunuri digitale comune, în scopul de cataliza realizarea obiectivelor dezvoltării durabile; b) a veghea ca datele amplificând realizarea obiectivelor dezvoltării durabile să fie reprezentative, compatibile și accesibile; c) să se elibereze frontierele și să se pună în comun datele, cunoștințele în materie de I.A. și infrastructurile pentru a facilita inovațiile, permițând atingerea țintelor obiectivelor de dezvoltare durabilă; și d) a ancora durabilitatea mondială în activitatea de concepție și aplicare a normelor de durabilitate digitală, armonizate la nivel mondial, ca balustrade, pentru protejarea planetei.

Statele membre trebuie: a) să elaboreze, cu diverse părți interesate, un CADRU enunțând principiile de concepție, care sunt fondate pe cele mai bune practici și un ansamblu de definiții pentru infrastructurile digitale comune, sigure, inclusive și durabile; b) să constituie și să țină la zi un referențial mondial de experiențe în materia infrastructurilor și serviciilor comune digitale și; c) să aloce un procent convenit, din totalul ajutorului internațional pentru dezvoltare și transformarea digitală, punând, în special, accent pe întărirea capacității administrațiilor publice.

Toate părțile interesate: a) să se angajeze să completeze recensământul lacunelor referitoare la datele relative la obiectivele de dezvoltare durabilă și să acopere 90% din datele privind obiectivele disponibile și accesibile publicate la orizontul 2030; b) să se angajeze să furnizeze instaurarea ecosistemelor de date deschise și accesibile care permit atenuarea efectelor catastrofelor și crizelor, mai rapid și de o manieră concretă, în special prin intermediul Fondului Națiunilor Unite pentru analiza riscurilor complexe și al Mecanismului de finanțare a observațiilor sistematice a Organizației Meteorologice Mondiale; c) să pună în practică inițiativele de cercetare colaborativă referitoare la datele și la aplicările exploatând I.A., care acționează în sensul obiectivelor dezvoltării durabile, în domeniile prioritare, precum agricultura, educația, energia, sănătatea și tranzițiile verzi; și d) să se angajeze, să creeze un referențial mondial în linie, reunind

datele de mediu fiabile și deschise pentru cercetătorii și decidenții politici, acompaniat de licențiați, de norme de calitate, de infrastructuri și de garanției pentru a susține transformarea digitală verde.

Organismele multilaterale și regionale: a) să pună în practică mecanismele de finanțare comună pentru a ajuta puterile publice în planificarea și conceperea infrastructurilor și serviciilor comune digitale; b) să lărgească codurile – obiect facultativ ale Comitetului de ajutor al dezvoltării Organizației de cooperare și dezvoltare economică în scopul de a urmări finanțarea datelor și transformarea digitală în ansamblul sectoarelor de dezvoltare și a activităților înfăptuite pentru realizarea obiectivelor dezvoltării durabil; și c) să adopte un plan comun în favoarea transformării digitale pentru a orienta toate aspectele dezvoltării digitale și pentru a tranșa o parte dintr-un non ghișeu axat pe digitalul propus de Fondul comun pentru obiectivele dezvoltării durabile, în scopul facilitării inițiativelor de transformare digitală ale țărilor, susținute prin coordonatele și coordonatorii rezidenți și de echipele țărilor Națiunilor Unite.

Apărarea drepturilor omului. *Obiective*: a) să se considere drepturile omului, drept fundamentul viitorului digital (deschis, sigur și securizat, în care demnitatea umană ocupă un loc central); b) să se acopere prăpastia digitală existentă, între femei și bărbați în privința a ceea ce spațiile de linie să fie discriminatorii și sigure pentru femei și să lărgească participarea femeilor în sectorul tehnologic și la elaborarea politicilor digitale; și c) să se aplice dreptul internațional relativ la muncă, independent de modalitățile de muncă și să protejeze lucrătorii contra supravegherii digitale, deciziilor algoritmice arbitrare și fărâmițării controlului pe care îl exercită asupra muncii lor.

Statele membre trebuie: a) să se angajeze să pună în practică un mecanism consultativ asupra drepturilor omului în domeniul digital, coordonat de Înaltul Comisariat al ONU pentru drepturile omului, care să furnizeze consiliere practică asupra chestiunilor relative la drepturile omului și la tehnologie și experții în materie de drepturile omului, punând în evidență bunele practici și reunind părțile interesate pentru ca ele să studieze răspunsurile eficace și coerente la problemele de ordin legislativ sau regulamentar.

Toate părțile interesate: a) să transpună angajamentele juridice existente în practică și în normele regionale, naționale și sectoriale și să ia măsurile necesare pentru a proteja femeile, copii, tinerii, persoanele în vârstă, pe cele cu handicap, populațiile autohtone și minoritățile etnice, religioase și lingvistice și să le acorde mijloacele pentru a obține deplin efectele din tehnologiile digitale și b) în cazul puterilor publice, angajatorii și lucrătorii, să se angajeze să respecte drepturile

relative la muncă, susținute prin Organizația Internațională a Muncii – OIM și să promoveze perspectivele de muncă, veritabile și echitabile, grație politicilor novatoare în materie de reglementare, de protecție socială și de investiții.

Un internet inclusiv, deschis, sigur și partajat. *Obiective*:

a) prezervarea naturii libere și partajate a internetului, ca un mediu comun mondial unic și b) întărirea guvernanței multipartite responsabile a internetului în scopul de a ajuta valorificarea potențialului rețelei serviciului pentru realizarea obiectivelor de dezvoltare durabilă și de a nu ignora nicio persoană.

Statele membre: c) să evite închiderile și să generalizeze a accesului la internet, să acționeze pentru a acoperi prăpastia digitală și să vegheze ca măsurile țintă să fie proporționale, nediscriminatorii, dacă sunt necesare pentru a atinge obiectivele legitime și conforme cu dreptul internațional al drepturilor omului, iar d) în cadrul mecanismului de cyberdiplomație al ONU să se abțină de la orice acțiune susceptibilă de a perturba, păgubi sau distruge infrastructurile critice care furnizează servicii transfrontaliere sau infrastructurile care asigură disponibilitatea generală și integritatea internetului.

Toate părțile interesate: e) să respecte neutralitatea rețelei, gestiunea nondiscriminatorie a traficului, normele tehnice interoperabilitatea infrastructurilor și a datelor, ca și neutralitatea platformelor și a aparatelor în scopul de a susține un internet deschis și interconectat.

Încredere și securitate digitală.

Obiectivele: a) întărirea cooperării între puterile publice, sector, experți și societatea civilă, în scopul stabilirii și aplicării normelor, orientărilor și principiilor relative la utilizarea responsabilă a tehnologiilor numerice; b) să elaboreze criteriile și normele de responsabilitate solide pentru platformele digitale și utilizării în scopul de a lupta contra dezinformării, discursurilor care propagă ura și conținutul în linie prejudiciabilă; c) întărirea capacităților și creșterea numărului de specialiști în cybersecuritate și punere în practică a virtuților de încredere și a sistemelor de certificare ale organelor de control regional și național eficace; d) integrarea chestiunilor de gen în politicile digitale și în concepția tehnologiilor și garantarea unei toleranțe zero în privința violenței fondate pe gen sau genul în scopul instituirii unei lumi mai egalitare și mai corectate pentru femei și fete.

Toate părțile interesate trebuie să se angajeze să elaboreze normele, orientările și codurile de conduită sectoriale, comune pentru lupta contra publicațiilor de conținut prejudiciabil pe

platformele digitale și să promoveze spațiile civice sigure, a) comisarii de securitate în linie, reprezentând diferite domenii de competență, trebuie să colaboreze pentru a susține interpretările comune și cele mai bune practici care respectă libertatea de expresie și accesul la informație, asigurând o protecție efectivă contra prejudiciilor; b) mediile sociale trebuie să se angajeze, să pună în practică mecanismele de reglementare, care garantează respectarea normelor convenite în ansamblul sectorului. Normele ținând de mecanisme vor putea să inspire proiectul de cod de conduită relativ la integritatea informațiilor asupra platformelor digitale și dezbaterile ținute cu ocazia Conferinței mondiale organizată de ONU pentru educație, știință și cultură-UNESCO, pe tema ”*Pentru un internet de încredere-către prejudiciile mondiale pentru reglementarea platformelor digitale privind informația ca bun comun*”; c) alianțele multipartite, precum Coaliția de acțiune asupra tehnologiilor și inovației privind serviciul de egalitate între femei și bărbați, trebuie să contribuie la elaborarea unei măsuri standard a violenței în linie în privința femeilor și fetelor și metodele care să permită cele mai bune măsuri, de a urmări și combate prejudiciile care denotă anumite tendințe, iar d) nevoile copiilor trebuie să fie o prioritate a politicilor și normelor de securitate, în special în ceea ce privește o concepție și un acces, adaptate la vârstă; platformele trebuie să difuzeze evaluările și datele relative la efectul asupra copiilor pe lângă organismele însărcinate cu reglementarea și cercetarea”.

Protecția datelor şi autonomizarea.

Obiective: a) să vegheze ca datele să fie gerate în interesul tuturor și într-o manieră de a nu dăuna persoanelor și comunităților; b) a da fiecăreia și fiecăruia capacități și cele utile necesare pentru a gera și controla datelor lor personale, inclusiv opțiunile și mijloacele referitoare la inscripția asupra platformelor digitale sau anularea inscripției și pentru a autoriza sau nu utilizarea datelor lor în scopurile de antrenament al algoritmilor și c) să elaboreze normele și cadrele multiniveluri și interoperabile, încadrând calitatea, măsura și utilizarea datelor, în deplin respect al drepturilor de proprietate intelectuală, în scopul de a permite un flux de date sigur și securizat și o economie mondială incluzivă.

Statele membre: a) să dateze textele care prevăd ca datele personale și viața privată să fie protejate, care să fie fondate, de exemplu, pe Convenția Uniunii africane asupra cybersecurității și protecției datelor cu caracter personal și pe Regulamentul UE privind protecția datelor. Aceste texte ar trebui să aibă următoarea redactare: a1) a da cetățenilor mijloacele de a face veritabil consimțământul lor, care să fie revocabil, astfel ca opțiunile lor să le permită să accepte sau nu ca

datele lor să fie utilizate; a2) să completeze măsurile de protecție juridică, punând în acțiune mediatorii și mecanismele de asigurarea independenței și ușor accesibile; și b) să adopte o declarație asupra drepturilor relative la datele care consacră transparența, în scopul de a garanta luarea deciziilor fondate pe date veritabile, interoperabilității și portabilității, asigurând protecțiile contra manipulării comportamentelor și discriminării.

Toate părțile interesate: c) să elaboreze definițiile comune și normele relative la interoperabilitate, acces la datele în funcție de tipul și calitatea și de măsura datelor și să asigure urmărirea și aplicarea; d) să întărească puterea și controlul exercitat prin fiecare asupra utilizării datelor lor personale, inclusiv a celor mai bune opțiuni, permițându-le să nu se asocieze la o astfel de activitate, de ameliorare a interoperativității, a portabilității datelor și opțiunilor de cifre și e) să examineze recomandarea Consiliului consultativ de înalt nivel pentru multilateralism eficace, referitoare la elaborarea multipartită a Pactului mondial pentru date care trebuie adoptat de Statele membre până în 2030.

Guvernanța dinamică a IA și a tehnologiilor emergente.

Obiective: a) a veghea la concepția și utilizarea IA și a altor tehnologii emergente, să fie transparente, fiabile, sigure și supuse unui control prin mijloace umane, asupra modului cum principiul responsabilității este aplicat; b) a plasa transparența, echitatea și aplicarea principiului responsabilității în inima guvernanței IA, ținând cont de faptul că incumbă puterilor publice să decelereze riscurile pe care sistemele de IA le pot comporta și remedia și că revine cercetătorilor și întreprinderilor care dezvoltă sistemele de IA de urmărire a acestor riscuri, de o manieră transparentă și de remediere; c) a combina orientările și normele internaționale, cadrele reglementărilor naționale și normele tehnice, într-un cadru facilitând guvernanța vie a IA, ca o difuzare activă a învățămintelor și a celor mai bune practici, care există la zi, la toate țările industriile și sectoarele, iar d) în cazul autorităților responsabile cu reglementarea, coordonarea politicilor relative la digital, la concurență, fiscalitate, protecția datelor, cât și cele privind drepturile la muncă, să urmărească dacă tehnologiile digitale emergente concordă cu valorile noastre umane.

Acțiuni. Statele membre:

a) să lanseze de urgență, în colaborare cu industria, un efort mondial de cercetare și de dezvoltare pentru a veghea la ceea ce sistemele de IA să fie sigure, echitabile, responsabile, transparente, interpretabile și demne de încredere și concordând cu valorile umane. Ele trebuie, de

asemenea, să vizeze impunerea unui procentaj minim de investiții în IA, alocat guvernanței IA și activităților de natură a garanta că sistemele de IA concordă cu valorile umane.

b) să creeze un organ consultativ de înalt nivel pentru IA în cadrul Pactului mondial pentru date, care să cuprindă experți din Statele membre și reprezentanții entităților componente ai ONU, sectorul de tehnologii, din mediile universitare și din societatea civilă, care să se reunească perioada pentru care a examina nivelurile mecanismelor de guvernanță a IA la nivel regional, național și sectorial;

c) să convină cu asociațiile industriale elaborarea de principii directoare sectoriale în scopul de a garanta că dezvoltatorii și utilizatorii dispun de orientări aplicabile și pertinente pentru concepția, aplicarea și audit util prin IA în context specific;

d) să se angajeze cu dezvoltatorii de tehnologii și de platforme digitale pentru a întări măsurile de transparență și de aplicare a principiului responsabilității, în special punând în practică echipele însărcinate cu chestiunile relative la drepturile omului și la etică și consilierii de control transdisciplinari și independenți;

e) să întărească capacitățile de reglementare interdisciplinare și multipartite în sectorul public, inclusiv capacitățile judiciare notate de Consiliul consultativ de nivel înalt pentru multilateralism eficace, în scopul de a garanta că reglementările și pieţele publice purtând asupra sistemelor IA și altor tehnologii emergente favorizând incluziunea, securitatea, siguranța și luarea în sarcină rapidă a riscurilor și

f) să interzică utilizarea aplicațiilor tehnologiilor ale căror efecte potențiale sau reale nu puteau fi justificate în privința dreptului internațional al drepturilor omului, inclusiv cele care nu satisfăceau criteriile de necesitate, distincții și de proporționalitate.

Bunurile digitale comune mondiale. Obiective:

a) a dezvolta și a încadra tehnologiile digitale pentru a favoriza dezvoltarea durabilă, a întări puterea de acțiune umană, a anticipa riscurile și prejudiciile și a le remedia eficace;

b) a veghea ca aceasta cooperare digitală să fie inclusivă și să permită ca toate părțile interesate să-și aducă contribuția la mandatele, funcțiunile și cunoștințele lor;

c) a conveni că fundamentele cooperării noastre sunt: Carta ONU, Programul de dezvoltare durabilă la orizont 2030 și textele relative la drepturile umane, universal recunoscute și la dreptul internațional umanitar; și

d) a permite schimburile regulate și susținute între State, regiuni, și sectoare de activitate, în toate problemele, în scopul de a favoriza achiziția de cunoștințe și cele mai bune practici, de a susține inovația și capacitățile în materie de guvernanță și de a veghea ca aceasta guvernanță digitală să nu înceteze a fi concordantă cu valorile noastre umane comune.

Acțiuni. Toate părțile interesate:

a) să se angajeze în a pune în comun datele de experiență în materia guvernanței și reglementării, a alinierii la principiile și cadrele internaționale asupra măsurilor naționale și a practicilor sectorului, a întări capacitățile reglementare și a aplica măsurile unei guvernanțe dinamice pentru a urmări evoluția rapidă a tehnologie și

b) să susțină avansarea principiilor, obiectivelor și acțiunilor enunțate în pactul digital mondial ca un mijloc al unui cadru de cooperare multipartită.

Aplicarea, efectele și examenul periodic al stadiului aplicării Pactului digital mondial. Pentru ca pactul să ”producă fructe”, trebuie ca activitățile să fie realizate la un înalt nivel, diversele părți interesate să fie însărcinate să urmărească Pactul la nivelurile național, regional și sectorial, ținând cont de contextele regionale și respectarea politicilor, mandatelor și competentelor internaționale.

Mecanismele de cooperare existente, în special:

– Forumul asupra guvernanței Internetului;

– Summit-ul mondial asupra societății de informație;

– Uniunea Internațională a Muncii/OIM;

– Înaltul comisariat al ONU pentru drepturile omului;

– CNUCED, UNESCO; și

– Programul ONU pentru dezvoltare.

se vor bucura de un rol major în susținerea punerii în practică a Pactului, furnizând cunoștințele asupra diferitelor chestiuni, sectoriale, de orientări și a unui experiențe practice pentru a facilita dialogul și a acționa în privința obiectivelor convenției.

Aplicarea multipartită.

Cadrul mondial trebuie să fie pilotat prin Statele membre, cu participarea actorilor din sectorul privat și a societății civile, care au un rol esențial în privința dezvoltării inovațiilor de care este nevoie în materie de guvernanță.

Pentru a evita crearea de structuri birocratice, foarte rigide și foarte lente, este esențial ca părțile interesate să pună în practică Pactul de o manieră dinamică.

Participarea sectorului privat trebuie să fie inclusivă și deschisă paletei marilor întreprinderi, celor mici și mijlocii, start-up-urilor, prin intermediul organelor reprezentative, iar participarea părților interesate va putea fi susținută printr-un Fond de afectațiune socială, care va servi, între altele, finanțarea unui program de burse în domeniul cooperării digitale și al participării societății civile, precum și al unui proiect relativ la crearea unei Adunări permanente a tinerilor în cadrul ONU.

Forumul de cooperare digitală.

Este esențial ca punerea în operă a Pactului să facă obiectul evaluărilor regulate, în principal, prin sesiuni anuale ale Forumului de cooperare digitală, cu misiunea de a susține angajamentul tripartit și de a asigura urmărirea aplicării Pactului. În acord cu orientările Consiliului consultativ de înalt nivel Forumul, la care au aderat Statele membre și părțile interesate din sectorul privat și din Societatea civilă trebuie să:

a) examineze stadiul aplicării principiilor și angajamentelor enunțate în Pactul digital mondial;

b) să faciliteze dialogul și colaborarea transparentă dintre diverse cadre multipartite privind digitalul și reducerea activităților numai de cele pertinente și adecvate;

c) să susțină punerea în comun a cunoștințelor și a informațiilor fondate pe datele probante referitoare la principalele tendințe în domeniul digital;

d) să pună în comun învățămintele și experiențele trase din promovarea formării transfrontaliere de guvernanță digitală;

e) să adopte și să promoveze soluțiile politice la problemele care apar, la zi, în domeniul digital și al lacunelor în materie de guvernanță; și

f) de a promova prioritățile în materie de participare, spre a facilita luarea deciziilor și a acțiunii părților interesate la nivel individual și colectiv.

Forumul va fi orientat către acțiune și se va concentra asupra evaluării progreselor și lacunelor în materie de guvernanță digitală și asupra comunicării informației pe acest subiect, ca și asupra transmiterii cunoștințelor și a schimburilor între părți, a tendințelor și a marilor probleme relative la tehnologiile emergente, mult mai rapid decât în prezent.

***e) SUA- Blueprint for an AI Bill of Rights*[51].**

Guvernul federal al SUA a adoptat inițial o abordare mai pasivă în materie de reglementare a IA în scopul de a favoriza inovația și a nu împiedica creșterea în domeniu. La apariția semnificativă a aplicațiilor noilor tehnologii ale IA cadrul normativ american era constituit în principal de legile și regulamentele deja existente și mai puțin pe o legislație globală și aplicabilă tuturor sistemelor de această natură. Au urmat apoi o serie de linii directoare neobligatorii, precum cele aferente Declarației drepturilor IA pentru conceperea și utilizarea responsabilă a noilor tehnologii, publicată de Casa Albă în octombrie 2022 și o serie de corpusuri de norme de autoreglare sectorială. Într-o perspectivă evolutivă, mai ales sub impactul evoluțiilor rapide în domeniul și apariția inteligenței artificiale generative, la 30 octombrie 2023, după ce în iulie același an 15 mari întreprinderi americane de profil au adoptat un angajament voluntar pentru a favoriza o dezvoltare sigură, securizată și demnă de încredere a IA, Președintele american a semnat un decret vizând a guverna dezvoltarea și utilizarea IA. Actul executiv urmărește în special să stabilească noi norme în materie de siguranță și securitate a inteligenței americane, a proteja viața privată a persoanelor, a face să promoveze echitatea și drepturile civile, a apăra consumatorii și lucrătorii, a promova inovarea și concurența și a face să progreseze leadershipul american în lume.

Cadrul reglementar astfel instituit are în vedere:

– să identifice clar conținutul generat de IA în scopul de a proteja populația contra fraudei și înșelătoriei bazate pe IA;

– a testa sistemele IA pentru a asigura că sunt sigure, securizate și demne de încredere;

– a partaja rezultatele testelor lor de securitate și a altor informații critice cu guvernul SUA.;

– a dezvolta investițiile în sectorul IA în scopul de a găsi și de a corija vulnerabilitățile programelor critice.

Pentru guvernul federal al SUA și agențiile sale se stabilesc următoarele exigențe generale:

– elaborarea unui memorandum privind securitatea națională care să orienteze acțiunile viitoare referitoare la inteligența artificială;

– întărirea și accelerarea cercetării și a dezvoltării instrumentelor pentru prezervarea vieții private;

– elaborarea de directive clare vizând a împiedica algoritmi de IA să fie utilizați spre a exacerba discriminarea;

[51] https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

– aplicarea celor mai bune practici în ansamblul sistemului de justiție penală și în special în materie de anchete și urmărire în caz de violare a drepturilor civile legate de noile tehnologii;

– stabilirea unui program de securitate pentru a primi semnalările de prejudicii ori de practici de îngrijire a sănătății periculoase implicând IA și stabilirea de măsuri pentru a le remedia;

– elaborarea de principii și de practici exemplare pentru a atenua prejudiciile și a maximiza avantajele IA pentru lucrători, descurajând deplasarea locurilor de muncă;

– concentrarea cercetării privind inteligența artificială de-a lungul Statelor Unite grație unui proiect pilot al National AI Research Resources.

Principii:

Sisteme sigure și eficiente. Oamenii trebuie protejați împotriva tehnologiilor care sunt nesigure sau ineficiente. Sistemele IA trebuie să fie testate riguros înainte de implementare și să respecte standarde specifice fiecărui domeniu pentru a preveni utilizările neprevăzute care ar putea cauza daune.

Protecție împotriva discriminării algoritmice. Sistemele IA nu trebuie să cauzeze discriminări nedrepte bazate pe factori precum rasă, sex, vârstă sau alte caracteristici protejate legal. Proiectarea trebuie să includă evaluări de echitate și să utilizeze date reprezentative.

Confidențialitate și protecția datelor. Utilizatorii trebuie să fie protejați împotriva utilizării abuzive a datelor. Sistemele IA trebuie să includă protecții implicite, iar colectarea datelor să fie limitată la strictul necesar.

Informare și explicare. Utilizatorii trebuie să fie informați despre funcționarea sistemelor IA și despre modul în care acestea iau decizii. Transparența este esențială pentru încrederea publicului.

Opțiuni umane și alternative. Oamenii trebuie să aibă acces la soluții alternative, inclusiv posibilitatea de a interacționa cu o persoană umană atunci când un sistem IA eșuează sau ia o decizie contestată.

Pe lângă aceste evoluții afirmate în plan federal și statele federale au încorporat treptat reguli privind IA și în legile mai vaste, vizând sectoare foarte specifice, precum cele ale locurilor de muncă, protecția vieții private și a consumatorilor. De remarcat faptul că, în contextul adoptării decretului din 30 octombrie 2023, Președintele S.U.A. preciza că, pentru a se realiza promisiunile inteligenței artificiale și a se evita riscurile aferente, această tehnologie trebuie guvernată și nu există o altă soluție decât încadrarea sa.

Capitolul IV.

Inteligența artificială și drepturile fundamentale (umane).

Cadrul general al drepturilor fundamentale care se aplică utilizării IA în UE constă în Carta drepturilor fundamentale a UE (Carta), precum și în Convenția europeană a drepturilor omului.

Mai multe alte instrumente ale Consiliului Europei și instrumente internaționale privind drepturile omului sunt relevante. Printre acestea se numără Declarația Universală a Drepturilor Omului din 1948 și principalele convenții ale ONU privind drepturile omului. În plus, legislația secundară sectorială a UE, în special *acquis*-ul comunitar în materie de protecție a datelor și legislația UE privind nediscriminarea, contribuie la garantarea drepturilor fundamentale în contextul IA. În cele din urmă, se aplică, de asemenea, legislația națională a statelor membre ale UE.

Atunci când introduc noi politici și adoptă noi acte legislative privind IA, legiuitorul UE și statele membre, acționând în cadrul domeniului de aplicare al dreptului UE, trebuie să se asigure că se ține seama de respectarea întregului spectru de drepturi fundamentale, astfel cum sunt consacrate în cartă și în tratatele UE. Politicile și legile relevante trebuie să fie însoțite de garanții specifice privind drepturile fundamentale. În acest sens, UE și statele sale membre trebuie să se bazeze pe dovezi solide privind impactul IA asupra drepturilor fundamentale, pentru a se asigura că orice restricții ale anumitor drepturi fundamentale respectă principiul necesității și principiul proporționalității.

Prin lege trebuie prevăzute garanții relevante pentru a proteja în mod eficace împotriva interferențelor arbitrare cu *drepturile fundamentale* și pentru a oferi securitate juridică atât dezvoltatorilor, cât și utilizatorilor de IA. Sistemele voluntare de respectare și garantare a drepturilor fundamentale în dezvoltarea și utilizarea IA pot contribui și mai mult la atenuarea încălcărilor drepturilor. În conformitate cu cerințele minime de claritate juridică – ca principiu de bază al statului de drept și ca o condiție prealabilă pentru garantarea drepturilor fundamentale –, legiuitorul trebuie să acorde atenția cuvenită atunci când definește domeniul de aplicare al unei astfel de legi privind IA.

Utilizarea sistemelor de IA implică o gamă largă de drepturi fundamentale, indiferent de domeniul de aplicare. Printre acestea se numără – dar și depășesc – viața privată, protecția datelor, nediscriminarea și accesul la justiție.

Carta drepturilor fundamentale a UE (Carta) a devenit obligatorie din punct de vedere juridic în decembrie 2009 și are aceeași valoare juridică ca tratatele UE. Acesta reunește într-un singur text drepturile civile, politice, economice și sociale.

Potrivit articolului 51 alineatul (1) din cartă, instituțiile, organele, oficiile și agențiile Uniunii trebuie să respecte toate drepturile consacrate în cartă. Statele membre ale UE trebuie să facă acest lucru atunci când pun în aplicare dreptul Uniunii. Acest lucru este valabil atât pentru IA, cât și pentru orice alt domeniu.

Tehnologiile în general, implică niveluri diferite de automatizare și complexitate. De asemenea, ele variază în ceea ce privește amploarea și impactul potențial asupra oamenilor. Utilizarea sistemelor de IA implică un spectru larg de drepturi fundamentale, indiferent de domeniul de aplicare. Acestea includ, fără limitare, protecția vieții private și a datelor, nediscriminarea și accesul la justiție. Atunci când se utilizează IA, trebuie luată în considerare o gamă mai largă de drepturi, în funcție de tehnologie și de domeniul de utilizare. Pe lângă drepturile privind viața privată și protecția datelor, egalitatea și nediscriminarea, precum și accesul la justiție, ar putea fi luate în considerare și alte drepturi. Printre acestea se numără, de exemplu, demnitatea umană, dreptul la securitate socială și asistență socială, dreptul la o bună administrare (în cea mai mare parte relevant pentru sectorul public) și protecția consumatorilor (deosebit de important pentru întreprinderi). În funcție de contextul utilizării IA, trebuie luat în considerare orice alt drept protejat în cartă.

Implementarea sistemelor de IA implică un spectru larg de drepturi fundamentale, indiferent de domeniul de aplicare. Potrivit articolului 51 alineatul (1) din cartă, statele membre ale UE trebuie să respecte toate drepturile consacrate în cartă atunci când pun în aplicare dreptul Uniunii. În conformitate cu standardele internaționale existente – în special Principiile directoare ale Organizației Națiunilor Unite privind afacerile și drepturile omului (UNGP) –, întreprinderile ar trebui să dispună de „un proces de diligență în materie de drepturi ale omului pentru a identifica, a preveni, a atenua și a explica modul în care abordează impactul acestora asupra drepturilor omului” (principiile 15 și 17).

Acest lucru este valabil indiferent de dimensiunea și de sectorul lor de activitate și cuprinde întreprinderile care lucrează cu IA. În îndeplinirea angajamentelor sale față de Principiile directoare ale ONU, UE a adoptat mai multe acte legislative care abordează instrumente sectoriale, în special în contextul obligațiilor legate de diligența necesară pentru drepturile omului. În prezent, au loc discuții cu privire la propunerea unei noi legislații secundare a UE.

Această legislație ar urma să impună întreprinderilor să efectueze diligența necesară în ceea ce privește impactul potențial al operațiunilor și lanțurilor lor de aprovizionare asupra drepturilor omului și a mediului[52].

În conformitate cu standardele internaționale consacrate în domeniul drepturilor omului – de exemplu, articolul 1 din Convenția europeană a drepturilor omului (CEDO) și articolul 51 din cartă –, statele sunt obligate să garanteze drepturile și libertățile persoanelor. Pentru a se conforma în mod eficace, statele trebuie, printre altele, să instituie mecanisme eficace de monitorizare și de asigurare a respectării legii.

În UE, există un set bine dezvoltat de organisme independente având mandatul de a proteja și de a promova drepturile fundamentale. Printre acestea se numără autoritățile pentru protecția datelor, organismele de promovare a egalității, instituțiile naționale pentru apărarea drepturilor omului și instituțiile Avocatului poporului.

Carta etică europeană cu privire la utilizarea inteligenței artificiale în sistemul judiciar și în legătură cu acesta.

În anul 2018, Comisia Europeană pentru Eficiența Justiției a Consiliului Europei a adoptat primul text european care stabilește principiile etice referitoare la utilizarea inteligenței artificiale în sistemele judiciare. Aceasta este „*Carta etică europeană cu privire la utilizarea inteligenței artificiale în sistemul judiciar și în legătură cu acesta*” („Carta”).

Carta este destinată părților interesate din domeniul public și privat, responsabile de proiectarea și de implementarea instrumentelor și serviciilor de inteligență artificială care implică prelucrarea hotărârilor judecătorești și a datelor (de învățare automată sau orice alte metode care derivă din știința datelor). Documentul se referă, de asemenea, la factorii de decizie publică care se ocupă de legislația sau reglementarea cadru, al dezvoltării, auditului sau utilizării unor astfel de instrumente și servicii.

[52] Agenția pentru Drepturi Fundamentale a Uniunii Europene, *Înțelegerea viitorului inteligența artificială și drepturile fundamentale*, 2021, ISBN 978-92-9461-230-4 doi:10.2811/087099 TK-04-20-657-RO-C

Utilizarea unor astfel de instrumente și de servicii în sistemele judiciare urmărește să îmbunătățească eficiența și calitatea justiției și ar trebui încurajată. Totuși, acest lucru trebuie să se desfășoare în mod responsabil, ținând seama de drepturile fundamentale ale persoanelor, astfel cum sunt prevăzute în Convenția Europeană a Drepturilor Omului (CEDO) și în Convenția privind protecția datelor cu caracter personal și în conformitate cu alte aspecte fundamentale, cum sunt principiile prezentate mai jos, care ar trebui să ghideze elaborarea politicilor de justiție publică în acest domeniu.

Prelucrarea deciziilor judiciare de către inteligența artificială, potrivit dezvoltatorilor acesteia, este de natură, în materie civilă, comercială și administrativă, să contribuie la îmbunătățirea predictibilității aplicării legii și coerența hotărârilor judecătorești, sub rezerva respectării principiilor expuse mai jos. În materie penală, utilizarea ar trebui însă luată în considerare cu cele mai mari rezerve pentru a preveni discriminarea bazată pe date sensibile, în conformitate cu garanțiile unui proces echitabil.

Fie că este concepută cu scopul de a asista în furnizarea de consultanță juridică, de a ajuta la redactare sau în procesul de luare a deciziilor ori de a consilia utilizatorul, este esențial ca prelucrarea să fie efectuată cu transparență, imparțialitate și echitate, certificată printr-o evaluare externă și independentă a unui expert.

Cele cinci principii adoptate în Cartă sunt:

1. *Principiul respectării drepturilor fundamentale*: proiectarea și implementarea instrumentelor de inteligență artificială și a serviciile aferente trebuie să fie compatibile cu drepturile fundamentale;

2. *Principiul nediscriminării*: prevenirea în mod specific a dezvoltării sau intensificării oricărei discriminări între indivizi sau grupuri de indivizi;

3. *Principiul calității și securității*: în ceea ce privește prelucrarea hotărârilor judecătorești și a datelor, se vor folosi surse certificate și date intangibile, cu modele elaborate într-o manieră multidisciplinară, într-un mediu tehnologic securizat;

4. *Principiul transparenței*, imparțialității și echității: metodele de prelucrare a datelor trebuie transformate în aspecte accesibile și ușor de înțeles, se vor autoriza audituri externe;

5. *Principiul „sub controlul utilizatorului*": excluderea abordărilor imperative și asigurarea că utilizatorii sunt actori informați și iau decizii informate.

Se recomandă totodată ca principiile Cartei să facă obiectul aplicării, monitorizării și evaluării periodice de către actorii publici și privați, în vederea asigurării continue a îmbunătățirii practicilor. În acest sens, este de dorit o revizuire regulată a implementării principiilor Cartei de către acești actori, explicând adecvat, unde este cazul motivele neimplementării sau implementării parțiale, revizuire ca va fi însoțită de un plan de acțiune pentru introducerea măsurilor necesare.

Autoritățile independente menționate în Cartă ar putea fi responsabile de evaluarea periodică a nivelului de aprobare a principiilor Cartei de către toți actorii și să propună îmbunătățiri pentru a o adapta la tehnologiile aferente și la evoluția acestor tehnologii.

Carta este menită să asigure trecerea spre digitalizarea tot mai accentuată a sistemului judiciar, în majoritatea domeniilor de drept.

Efectele implementării inteligenței artificiale în justiție asupra drepturilor fundamentale[53].

Conform Art. 6 alin. (1) din Convenția Europeană a Drepturilor Omului „Orice persoană are dreptul la judecarea cauzei sale în mod echitabil, în mod public și în termen rezonabil, de către o instanță independentă și imparțială, instituită de lege, care va hotărî fie asupra încălcării drepturilor și obligațiilor sale cu caracter civil, fie asupra temeiniciei oricărei acuzații în materie penală îndreptate împotriva sa.” În ce măsură sunt compatibile aceste cerințe cu o instanță formată din IA?

Accesul la o instanță.

Trebuie să avem în vedere existența unui echilibru între dreptul fundamental de acces la o instanță și cerințele formale procedurale. În acest sens, Curtea Europeană a Drepturilor Omului a stabilit că „instanțele trebuie, în aplicarea normelor de procedură, să evite un formalism excesiv care ar aduce atingere caracterului echitabil al procedurilor”[54], o interpretare prea restrictivă a normelor de procedura putând afecta însăși dreptul de acces la o instanță[55].

Așadar, simpla aplicare mecanică a unor norme procedurale nu satisface condițiile impuse de jurisprudența CEDO, ci este necesară o apreciere de la caz la caz a incidenței acestor norme. Pentru realizarea unei astfel de aprecieri factorul decizional uman este indispensabil. IA nu poate ține cont de multitudinea diferențelor aparent minore dintre cauze. Există astfel riscul de a aplica

[53] Ioana Ciutacu, *Efectele implementării inteligenței artificiale în justiție asupra drepturilor fundamentale și asupra drepturilor procedurale*, Revista Universul Juridic, 2022

[54] CEDO – Ghid privind art. 6 din Convenția europeană a drepturilor omului, Dreptul la un proces echitabil (aspectul civil), Actualizat la 30 aprilie 2018, p. 21, cu referire la cauza Hasan Tunç și alții împotriva Turciei.

[55] CEDO – Ghid privind art. 6 din Convenția europeană a drepturilor omului, Dreptul la un proces echitabil (aspectul civil), Actualizat la 30 aprilie 2018, p. 21, cu referire la cauza Hasan Tunç și alții împotriva Turciei.

normele de procedură fără luarea în considerare a situației concrete a părților sau într-o manieră prea formalistă.

Prin urmare, implementarea IA ca instanță ar crea dificultăți cu privire la respectarea dreptului de acces la o instanță.

Echitatea.

Art. 6 din Convenția Europeană a Drepturilor Omului face referire și la judecarea cauzei într-un mod echitabil, lucru greu de realizat în ipoteza unui judecător cu IA. Se ridică în primul rând întrebarea dacă este echitabil ca o mașină să decidă soarta unui om. Cu toate că IA ar fi mai obiectivă decât o persoană, fiind lipsită preconcepții, înclinații religioase, politice sau alte trăsături subiective ce caracterizează omul, IA este de asemenea lipsită de empatie și de etică.

În al doilea rând, judecățile IA în aparență obiective, bazate doar pe situația de fapt și actele normative incidente, ar putea fi totuși lipsite de echitate. Aceasta se datorează faptului că echitatea este un concept nu poate fi cuprins doar în norme de drept și raționamente juridice, ci variază de la speță la speță. Echitatea este lăsată la aprecierea judecătorului care se bazează în luarea deciziei sale și pe anumite însușiri inerente omului – compasiune, empatie, integritate.

Astfel, chiar dacă IA este un judecător imparțial, aceasta nu înseamnă că poate judeca o cauză în mod echitabil.

Termenul rezonabil.

Respectarea unui termen rezonabil este esențială pentru desfășurarea procesului în conformitate cu exi-gențele impuse de Convenția Europeană a Drepturilor Omului. Întrucât oamenii sunt limitați din punct de vedere al capacității de muncă, timpul lor este de ase-menea limitat. Aceasta este principala cauză a supra-aglomerării instanțelor și implicit a duratelor lungi ale proceselor. În schimb, IA nu este limitată din punct de vedere al timpului de lucru, putând să funcționeze aproape neîntrerupt. Am putea avea în acest fel un judecător care nu obosește, care are ședințe în fiecare zi și redactează hotărâri fără pauze. O astfel de instanță ar reduce în mod semnificativ durata procesului, respectând cu siguranță standardul termenului rezonabil.

Independența instanței.

Independența este una dintre cele mai importante atribute ale unei instanțe. Acest aspect are în vedere independența individuală a oricărui judecător față de orice factor exterior. În aparență, IA este indepen-dentă față de oameni, aceasta fiind chiar una dintre caracteristicile de bază pentru ca o mașină să fie considerată IA. Cu toate acestea, trebuie să ținem cont de faptul că IA este creată

de oameni, evoluând ulterior în mod independent printr-un proces de învățare continuă. Programatorul IA, poate, teoretic, să intervină asupra modului în care acționează/gândește IA oricând.

Din moment ce unul sau mai mulți oameni au posibilitatea de a accesa și modifica codul ce stă la baza funcționării mașinii care este chiar judecătorul, în ce măsură mai există garanția independenței? Dacă pentru judecători independența este asigurată printr-un cadrul legal și prin valorile morale și etice, aceasta nu se poate aplica și în cazul roboților. În timp ce un judecător poate rezista influențelor externe, un calculator/robot nu poate să acționeze altfel decât a fost programat.

Prin urmare, standardul independenței IA este diferit de cel al unui judecător. Pe când „judecătorii sunt independenți și se supun numai legii", am putea spune că IA este independentă și se supune numai legii și programatorului.

Toți cetățenii au dreptul de a fi tratați în mod egal de către instituțiile statului. Interzicerea discriminării, consacrată de art. 14 din Convenția Europeană a Drepturilor Omului și de art. 16 din Constituția României, este cu atât mai importantă în cazul unui proces. Este de neconceput ca o persoană să fie tratată diferit pe simple considerente legate de naționalitate, rasă, sex, orientare sexuală ș.a.m.d.

Chiar dacă discriminarea ar părea un defect uman, și IA este capabilă de discriminare. Un exemplu în acest sens este sistemul de predictibilitate a gradului de recidivă folosit în Statele Unite ale Americii. Acest sistem, numit COMPAS, a fost conceput pentru a ajuta judecătorii să individualizeze pedepsele inculpaților, ținând cont de posibilitatea ca aceștia să recidiveze. S-a dovedit că acest sistem acorda un scor mai mare de recidivă indivizilor Afro-Americani decât celorlalți[56]. Chiar dacă a fost argumentat în literatura de specialitate faptul că această disproporție nu a fost bazată pe rasă, ci pe statisticile conform cărora persoanele de culoare recidivează într-o proporție mai mare decât alți indivizi, existența unui anumit grad de discriminare nu poate fi trecută cu vederea. Concluzionând, nu putem considera că un judecător cu IA care încalcă dreptul fundamental al egalității dintre cetățeni este o instanță imparțială[57].

[56] https://massivesci.com/articles/machine-learning-compas-racism-policing-fairness/
[57] Mirko Bagaric, Dan Hunter, Nigel Stobbs, *Erasing the Bias Against Using Artificial Intelligence to Predict Future Criminality: Algorithms are Color Blind and Never Tire,* 88 U. Cin. L. Rev. 1037 (2020), https://scholarship.law.uc.edu/uclr/vol88/iss4/3 p. 1067.

În ceea ce privește *dublul grad de jurisdicție*, un proces cu adevărat just nu poate fi imaginat într-o societate democratică fără acesta, mai ales în materie penală. Art. 2 alin (1) din Protocolul 7 al Convenției Europene a Drepturilor Omului prevede că „Orice persoană declarată vinovată de o infracțiune de către un tribunal are dreptul să ceară examinarea declarației de vinovăție sau a con-damnării de către o jurisdicție superioară".

În ipoteza unei instanțe cu IA, se mai justifică controlul jurisdicțional, având în vedere riscul redus al unei greșeli judiciare?. Necesitatea unui al doilea grad de jurisdicție s-ar putea impune atâta timp cât există posibilitatea ca primul verdict să fie greșit, oricât de redusă ar fi această posibilitate. În acest caz, cine va efectua controlul: IA sau o persoană?

Dacă am admite ca deciziile IA să fie cenzurate de o altă IA, se pune problema efectivității acestui control. În jurisprudența Curții Europene a Drepturilor Omului a fost consacrată ideea că pentru un mijloc de apel valabil, instanța ce efectuează controlul jurisdicțional trebuie să poată anula decizia atacată, aceasta ducând fie la pronunțarea unei noi decizii de către instanța de control, fie la o retrimitere către instanța de fond[58]. Această exigență nu poate fi respectată în ipoteza a două mașini cu IA, din moment de instanța de control judiciar raționează într-un mod foarte asemănător sau chiar identic cu instanța de fond. Neexistând diferențe majore între modul de gândire al celor două instanțe, aproape toate căile de atac ar fi respinse, iar dublul grad de jurisdicție ar deveni doar o formă fără fond.

O soluție ar fi ca deciziile IA să fie cenzurate de instanțe umane[59]. Însă, oamenii ar avea tendința de a modifica hotărârile emise de IA, chiar dacă există sau nu motive reale de cenzură, dată fiind neîncrederea oamenilor în algoritmul folosit. Aversiunea oamenilor față de deciziile IA ar duce la admiteri nefondate și excesive ale căilor de atac.

Prin urmare, controlul jurisdicțional al deciziilor emise de IA ar genera probleme, ducând fie la respingeri excesive ale căilor de atac, fie la admiteri și modificări nefondate. În oricare dintre cele două extreme, dreptul fundamental la justiție, ce include și un control jurisdicțional, ar fi afectat.

---

[58] CEDO, cauza Kingsley împotriva Regatului Unit – 35605/97, Hotărârea Marii Camere din 28 Mai 2002, pct. 32-34

[59] Berkeley J. Dietvorst, *Algorithms Versus Counter-Normative Reference Points. People reject (superior) algorithms because they compare them to counter-normative reference points,* The University of Chicago Booth School of Business, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2881503

Transparența decizională.

Pentru ca decizia unei instanțe să poată fi contestată, este necesar ca raționamentul din spatele deciziei să poată fi explicat. Acest lucru se transpune nu numai în motivarea din cuprinsul unei hotărâri judecătorești, dar și în operațiunile de logică juridică ce duc la o anumită concluzie.

Un profesionist al dreptului, sau chiar și un justițiabil, poate să înțeleagă raționamentul unui judecător, dar ar avea dificultăți în a îl înțelege pe cel al unui sistem de IA. Neînțelegerea argumentelor și a conexiunilor logice ce au fundamentat o decizie, creează o problemă din perspectiva posibilității de a ataca hotărârea judecătorească. Cu alte cuvinte, cum poți contesta ceva ce nu înțelegi?

Principiul explicabilității este esențial pentru o IA ce poate fi implementată în realitate. Prin urmare, este important ca mecanismele și procesele din spatele deciziilor mașinilor să poată fi explicate și înțelese, în special de către justițiabili.

Aşadar, principiile drepturilor (umane) se impun a fi incluse în colectarea și selectarea datelor, precum și în concepția, dezvoltarea, desfășurarea și utilizarea modelelor, instrumentelor şi serviciilor care rezultă. Relevantă este în acest sens rezoluția Consiliului pentru Drepturile Omului al ONU din 13 iulie 2023 care subliniază importanța "*garantării, promovării și protejării drepturilor omului de-a lungul ciclului de viață al sistemelor de inteligență artificială*". La rândul său, rezoluția Adunării Generale a ONU din 21 martie 2023 privind promovarea de sisteme ale inteligenței artificiale sigure, securizate și demne de încredere reclamă tuturor statelor membre și părților interesate "*să se abțină ori să înceteze să utilizeze sisteme de inteligență artificială care sunt imposibil de exploatat conform dreptului internațional al drepturilor umane ori care prezintă riscuri incluse pentru exercitarea drepturilor umane*".

În contextul interstatal, primul generator de juridicitate la nivel comun, internațional se poate constata că, în pofida multiplicării și diversificării inițiativelor interguvernamentale de gen (promovate în special în cadrul ONU, UNESCO, OCDE, Consiliului Europei și Uniunii Europene), principalele măsuri prevăzute de textele regulare ale inteligenței artificiale și ale utilizării sale reiau tradiționala dublă orientare, desigur cu unele nuanțări și particularizări de situație și de conjunctură. Se observă astfel că documentele adoptate atât de Adunarea Generală și Consiliul pentru Drepturile Omului ale ONU, UNESCO, cât și în Consiliul Europei, organizații de cooperare și coordonare interstatală politică insistă mai ales pe dimensiuni și aspecte așa-zis "fundamentaliste" ținând cu precădere de valorile de respectat și de cadrele de repere acţionale, în

timp ce OCDE şi UE, ca organizații de orientare și integrare economică, privilegiază (prioritar) chestiuni aferente securizării pieței și stimulării inovării[60].

Primul text mondial privind etica inteligenței artificiale, Recomandarea UNESCO din 7 noiembrie 2021 (*Artificial intelligence ethics recommendation*), "tinde a modifica în mod fundamental balanța puterilor ce există între oamenii de rând și guvernele și societățile ce dezvoltă instrumente de IA". El acordă întâietate a patru valori fundamentale, respectiv respectul, protecţia şi promovarea drepturilor umane şi demnităţii, diversitatea şi incluziunea, promovarea de societăţi paşnice juste şi interdependente, mediul şi ecosistemele care prosperă pe care statele membre se angajează să le respecte şi să le asigure respectarea la nivelul lor. O abordare fondată pe drepturile omului, întemeiată pe 10 principii: proporţionalitate şi inocuitate, siguranţă şi securitate, respectarea vieţii private şi protecţia datelor, guvernanţă şi colaborare multipărţi şi adaptative, responsabilitate şi redevabilitate, sensibilizare şi educaţie, transparenţă şi explicabilitate, supraveghere şi decizie umană, durabilitate, echitate şi nediscriminare. Documentul UNESCO listează, de asemenea, acţiuni pe care cele 193 de state membre aderente trebuie să le realizeze, în special stabilirea unui instrument legislativ spre a încadra şi monitoriza aplicaţiile IA spre "a asigura o securitate totală pentru datele personale şi sensibile", precum şi a educa masele asupra acestui subiect. Se enunţă valori şi principii comune care să ghideze instituirea infrastructurii juridice necesare pentru a asigura o dezvoltare sănătoasă şi echitabilă a IA[61].

În acest sens, al protejării drepturilor (umane) fundamentale, se propune o Cartă pentru a construi o etică a IA, cerându-se astfel o "veritabilă alianţă mondială" în materie. În această viziune, principiile guvernanţei în materie de IA ar trebui să se bazeze pe "Carta ONU şi Declaraţia universală a drepturilor omului din 1948", iar reflecţia cuvenită, pe măsură ce sistemele respective devin tot mai performante şi mai autonome, să se concentreze asupra modului de reglementare a conceperii şi utilizării lor, în scopul de a le fundamenta pe valori umane şi a preveni scăparea lor de sub controlul democratic, cu consecințe imprevizibile și ireparabile[62].

[60] Mircea Duțu, *Despre premisele, dezvoltările şi perspectivele unui drept al inteligenței artificiale*, Pandectele române 2/2024, p. 5 și urm.
[61] Idem
[62] Idem

Capitolul V.

IA și Dreptul penal. Riscuri pe care le poate prezenta IA.

I. Personalitatea juridică a IA

Având în vedere viteza de dezvoltare a tehnologiilor de inteligență artificială și nivelul posibilei dezvoltări viitoare, este esențial să se clarifice statutul juridic al acestor entități și să se discute diverse puncte de vedere. Potrivit lui Pagallo[63], cu privire la această problemă, există trei dezbateri fundamentale. Prima este dacă roboții au personalitate juridică și drepturi constituționale. A doua este dacă au dreptul de a contracta și de a fi incluse în dreptul comercial și civil. A treia discuție este dacă ele conduc la noi tipuri de responsabilități (umane).

Dicționarul Oxford[64] definește persoana juridică ca „*O persoană fizică, companie sau altă entitate care are drepturi legale și este supusă unor obligații*". Conform acestei definiții, fiecare entitate cu drepturi și obligații legale are personalitate juridică și este considerată din punct de vedere juridic, o persoană. În plus, așa cum se precizează în § 6 din Declarația Universală a Drepturilor Omului și §16 din Pactul Internațional cu privire la Drepturile Civile și Politice, „*Orice om are dreptul de a i se recunoaște pretutindeni personalitatea juridică*". În timp ce fiecare persoană care are drepturi responsabilități care decurg din personalitatea juridică și este considerată subiect juridic datorită faptului că este persoană fizică, în doctrină mai este cunoscut un termen, cel de „agent juridic" ca alternativă la termenii de persoană juridică[65]. Mai mult, este discutată și problema dacă „*personalitatea este dată datorită faptului că este supusă unor drepturi și responsabilități, sau că drepturile și responsabilitățile sunt recunoscute datorită personalității juridice*". O problemă complicată se ridică și în legătură cu inteligența artificială.

Pentru a rezuma, personalitatea juridică este un statut care indică recunoașterea unei persoane juridice și capacitatea acesteia de a avea anumite drepturi și obligații. Astfel, entităților cu unele aptitudini esențiale li se conferă personalitate juridică în urma acestor capacități. Ca

---

[63] Ugo Pagallo, Vital, *Sophia, and Co. – The Quest for the Legal Personhood of Robots*, Information 9, no. 9 (September 2018), p. 3, https://doi.org/10.3390/info9090

[64] Legal Person, Meaning of Legal Person by Lexico, Lexico Dictionaries – English, https://www.lexico.com/definition/legal_person

[65] Chopra/White, *Artificial Agents – Personhood in Law and Philosophy*; Steffen Wettig/Eberhard Zehendner, *The Electronic Agent: A Legal Personality under German Law?*, 2003; Pagallo, *Killers, Fridges, and Slaves*, November 2011, AI & SOCIETY 26(4):347-354, DOI: 10.1007/s00146-010-0316-0; Karolin Kemper, *Rechtspersönlichkeit für Künstliche Intelligenz*? URL: cognitio-zeitschrift.ch/2018-1/Kemper, DOI: https://doi.org/10.5281/zenodo.1928171 ISSN: 2624-8417, p. 6.

rezultat al deținerii personalității juridice, aceste entități își pot exercita unele drepturi și își pot asuma obligații. În ceea ce privește ființele umane, datorită prezumției de a avea abilități particulare, ele sunt considerate persoane fizice (în sens juridic).

În ceea ce privește IA, pe plan filosofic, s-a spus că o "*persoană morală*" este așa dacă poate diferenția binele de rău și că a fi o "*persoană morală*" este o precondiție a dobândirii statutului de personalitate juridică. Cu toate acestea, întrebarea despre statutul juridic al IA ridică unele chestiuni interesante și complexe legate de responsabilitate, proprietate intelectuală, drepturile de autor și alte aspecte legale. De exemplu, *Responsabilitatea*: cine este responsabil pentru acțiunile sau deciziile luate de un sistem de IA? Este vorba de creatorii săi, utilizatorii săi sau poate chiar IA însăși? *Proprietatea intelectuală*: pot fi protejate juridic algoritmii sau modelele de inteligență artificială ca proprietate intelectuală? Cine deține drepturile de autor sau brevetele pentru lucrările produse de IA?; *Confidențialitatea și datele personale*: Cum trebuie să fie gestionate datele personale colectate sau procesate de sistemele de IA conform legilor privind confidențialitatea și protecția datelor?; *Drepturile de muncă și economice*: pot fi recunoscute drepturile de muncă ale sistemelor de IA sau pot exista cerințe legale pentru compensarea utilizării lor?

Discuția despre personalitatea juridică ar trebui să înceapă de la John Chipman[66], care, în cartea sa despre natura și izvoarele Dreptului, arată că persoana, în limbaj comun, se referă la ființa umană. În sens juridic, persoana este subiect de drept cu drepturi și obligații. Potrivit lui Solum[67], întrebarea dacă o entitate ar trebui să fie considerată persoană juridică se poate reduce la alte întrebări referitoare la: dacă entitatea poate și ar trebui să facă obiectul unui set de drepturi și îndatoriri legale. Ansamblul special de drepturi și îndatoriri care însoțește personalitatea juridică variază în funcție de natura entității. Atât corporațiile, cât și persoanele fizice au personalitate juridică, dar cu seturi diferite de drepturi și obligații legale. Cu toate acestea, personalitatea juridică este de obicei însoțită de dreptul de a deține proprietate și capacitatea de a acționa în judecată și de a fi dat în judecată. În istoria omenirii[68], lucrurile neînsufletite au deținut drepturi legale în diverse timpuri. Templele din Roma și clădirile bisericii din Evul Mediu erau considerate drept

[66] John Chipman Gray, *The Nature and Sources of the Law*, Ed. Peter Smith, Gloucester, 1972, (MacMillan 1921), p. 27 și urm.
[67] Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, North Carolina Law Review vol. 70 nr. 4, p. 1239 și urm.
[68] John Chipman Gray, *op. cit.*, p. 28

subiect al drepturilor legale. În dreptul maritim, nava în sine devenea subiectul unei proceduri reale și putea fi găsită „vinovată”[69]; autorul menționat a relatat un caz (indian) din secolul al XX-lea în care un avocat a fost numit de către o instanță de apel pentru a reprezenta un idol al familiei într-o dispută cu privire la cine ar trebui să aibă custodia acestuia. John Chipman Gray este categoric în a afirma că persoana juridică este o ficțiune, atâta timp cât ea nu are nici inteligență și nici voință. Dar tocmai aceste atribute ale personalității juridice sunt în dispută în legătură cu inteligența artificială.

În prezent, apreciem că Testul Turing[70] nu mai este suficient în evaluarea Inteligenței Artificiale. Oamenii de știință au propus alte metode de evaluare cum ar fi: recunoașterea vorbirii, jocuri de strategie, pentru a determina ”autonomia” inteligenței artificiale dar și modele de evoluție a acesteia.

Pentru a-i fi acordată personalitate juridică Inteligenței Artificiale sunt necesare însă mai multe elemente. Simpla abilitate de comunicare (Siri sau Alexa de ex.) nu este suficientă. Sunt sisteme mult mai avansate decât Siri sau Alexa care nu au abilități de comunicare dar care nu trebuie ignorate. Pe de altă parte, dacă IA nu are abilități de comunicare, nu ne putem da seama cât este de inteligentă. Alături de abilitatea de a comunica, inteligența este o altă condiție esențială atunci când discutăm de o eventuală personalitate juridică a IA. În planul dreptului penal însă, inteligența omului sau lipsa acesteia nu influențează capacitatea sa de a răspunde pentru o faptă atâta timp cât fapta comisă se încadrează în cerințele legale. Cu alte cuvinte, nivelul de inteligență al unui om nu afectează statutul său juridic. Astfel, nu distingem persoanele după niveluri de inteligență, cum ar fi o persoană cu nivelul IQ 70 sau nivelul IQ 150 în cauze penale (în ipoteza în care nu vorbim de vreo cauză de iresponsabilitate sau altă circumstanță personală). Prin urmare, *doar* inteligența unei entități artificiale (bazate pe algoritmi) nu poate fi acceptată ca punct de reper în acordarea sau nu de statut juridic.

Liberul arbitru (capacitatea de a decide independent de orice influență exterioară) este o altă componentă esențială care trebuie avută în vedere în acordarea sau nu de statut juridic

[69] Christopher Stone, *Should Trees Have Standing? -Toward Legal Rights for Natural Objects*, 45 S. CAL. L. REv. p. 450, 453-57 (1972)

[70]Testul Turing este un experiment creat de un om de știință britanic numit Alan Turing în anul 1950 și constă într-un experiment în care un evaluator uman interacționează cu un sistem de inteligență artificială și cu un interlocutor uman printr-o interfață de text. Turing a vrut să afle dacă o mașină sau un computer poate gândi la fel ca oamenii. Ca exemplu, este dat un copil de 10 ani și un robot. Amândoi răspund la întrebări, dar nu știi care este copilul și care este robotul. Dacă nu poți să îți dai seama care dintre ei este mașina, înseamnă că robotul a trecut testul Turing și este considerat inteligent ca un om.

entităților IA. Pe plan juridic, actele încheiate de o persoană care nu are libertate de voință sunt nule. Unele idei exprimate de doctrină, Mulligan[71] de ex. sugerează că, dacă se poate face o elaborare a criteriilor și o entitate AI are liber arbitru din cauza acestor criterii de elaborare, aceste entități au personalitate juridică și sunt răspunzătoare din punct de vedere juridic. Potrivit Solum[72], o persoană trebuie să aibă liberul arbitru pentru a avea personalitate juridică. Pe de altă parte, Sullins[73] consideră că nu ar fi plauzibil să se aplice aceste concepte la IA, deoarece oamenii nu pot explica concepte precum liberul arbitru și intenționalitatea nici în prezent. Sullins a declarat că entitățile AI care pot interpreta mediul în mod independent și pot fi adaptative ar trebui considerate agenți. În prezent nu putem spune cu certitudine dacă entitățile IA au conștiință de sine. De fapt întrebarea care se pune este cum putem dovedi că avem liber arbitru? Potrivit lui Gless și Weigend, se poate presupune că ”roboții” decid cu efectul liberului arbitru dacă entitățile AI își pot recunoaște judecățile în timpul procesului de decizie, sunt conștienți de comportamentul lor că are importanță socială și afectează viața oamenilor[74]. Ca urmare, dacă putem determina existența liberului arbitru, atunci acesta ar trebui să fie un criteriu al personalității juridice. După această concretizare, următorul pas va fi ce drepturi putem acorda entităților de inteligență artificială cu liber arbitru. În plus, controlul comportamentului lor poate juca un rol semnificativ în corelarea capacității liberului arbitru[75]. În acest sens, raportul Parlamentului European a inclus și cuvintele: „*Dacă nu există un control din partea unui om, acesta (IA) devine actorul însuși*”[76]. Problema existenței liberului arbitru în cazul entităților IA începe să fie tot mai discutată în contextul în care noile evoluții ale IA ne arată că deși sunt programate de ingineri și funcționează pe bază de algoritmi, acestea pot învăța și pot lua decizii în mod autonom.

Un alt criteriu discutat este cel al intenției (sau raționament), specific sistemului de drept continental și nu anglo-saxon[77]. Unele expresii, cum ar fi „intenționat”, „cu scopul” și „acțiune cu

---

[71] Christina Mulligan, *Revenge Against Robots*, South Carolina Law Review 69 (2018): 14, https://doi.org/10.2139/ssrn.3016048; ”if an elaboration can be made and an AI entity has free will because of these elaboration criteria, these entities have legal personhood and are legally liable”

[72] L. Solum, *op. cit*, p. 1240

[73] John P. Sullins, *When Is a Robot a Moral Agent*?, în: Machine Ethics, Michael Anderson/Susan Leigh Anderson (eds.) (Cambridge: Cambridge University Press, 2011), p. 25–27, https://doi.org/10.1017/CBO9780511978036.013

[74] Gleß/Weigend, *Intelligente Agenten und das Strafrecht*, Intelligent Agents and Criminal Law, January 29, 2015, DOI 10.515/zstw-2014-0024, p. 572

[75] Schirmer, *Artificial Intelligence and Legal Personality*, p. 4;

[76] Nathalie Nevejans, *European Civil Law Rules in Robotics* (European Parliament, Legal and Parliamentary Affairs, December 2016), 15, http://www.europarl. europa.eu/committees/fr/supporting-analyses-search.html.

[77] Pagallo, Vital, *Sophia, and Co. – The Quest for the Legal Personhood of Robots*, p. 5–6

un scop", sunt, de asemenea, utilizate în contextul legal al intenției. *Mens rea* este formată din combinația dintre intenție și conștientizarea acțiunii ilicite. În cazul oamenilor, intenția înseamnă că a luat hotărârea în cunoștință de cauză, dirijându-și în mod conștient voința. La fel, nu discutăm despre persoanele incapabile (sau care au o altă circumstanță personală) care le împiedică să răspundă penal. Legea le acordă acestora alt statut. Entitățile IA autonome, în luarea deciziilor independente de algoritm, nu au intenția definită ca parte componentă a *mens rea*[78]. Existența liberului arbitru, în acest context, ar fi o condiție a intenției (sau a acțiunii). Este importantă așadar ordinea în care tratăm aceste criterii; liberul arbitru și intenția. Se conturează tot mai mult ideea că ar trebui să ne concentrăm mai mult pe liberul arbitru decât pe intenție, ca unul dintre criteriile care trebuie îndeplinite dacă vrem să acordăm entităților IA personalitate juridică.

În ceea ce privește capacitatea unei entități inteligente de a acționa în mod intenționat, există argumentul extrem de popular al "Camerei Chineze" (Chinese Room), un experiment de gândire aparținând lui Searle, din anul 1980[79]. În acest studiu academic, Searle a susținut că înțelegerea și gândirea sunt mai complicate decât urmărirea unui set de algoritmi, contrar părerii că trecerea testului Turing de inteligență artificială este suficientă și potrivită pentru ca acea entitate să fie o „mașină care gândește". Acest experiment de gândire s-a derulat astfel: Să presupunem că un vorbitor nativ de engleză, fără cunoștințe de chineză, este închis într-o cutie (o bază de date) plină cu simboluri chinezești, împreună cu un manual de instrucțiuni (program) pentru a manipula simbolurile. În acest experiment, oamenii din afara sălii trimit întrebări cu caractere chinezești (input). Persoana din cameră poate combina și forma simboluri chinezești care sunt răspunsuri corecte la întrebări, urmând instrucțiunile din program. Deși acest program permite persoanei să treacă testul Turing pentru a înțelege chineza, ea nu rostește un singur cuvânt în chineză. Ca rezultat al acestui experiment de gândire, argumentează Searle, deși trec testul Turing, sistemele de inteligență artificială ajung la această concluzie folosind doar 1-urile și 0-urile din algoritm (caracterele chinezești în experimentul gândirii), fără a avea efectiv conștiința și conștientizarea răspunsurilor. Alte criterii avute în vedere sunt empatia, un corp biologic dar și mortalitatea. Apreciem că aceste ultime criterii ar restrânge nejustificat această discuție. Ar însemna că *ab initio* nicio entitate IA nu ar putea vreodată să aibă drepturi și obligații din moment ce trebuie investită atributele unei entități biologice (om), lucru imposibil din moment ce IA este un produs tehnologic

[78] Gleß/Weigend, *Intelligente Agenten und das Strafrecht*, p. 572.
[79] John Searle, *Minds, Brains, and Programs*, Behavioral and Brain Sciences 3, no. 3 (September 1980), p. 417–24.

al omenirii și nu o altă entitate biologică. Chiar definiția artificialului menționată la începutul acestui studiu arată: ”*făcut de oameni, adesea ca o copie a ceva natural*”.

Capacitatea juridică a persoanei fizice.

Noțiunea de „persoană fizică” și „ființă umană” sunt de obicei tratate ca sinonime. Jurisconsultul Visa A.J. Kurki afirmă că, în sistemele juridice occidentale, persoanele fizice sunt ființe umane care s-au născut, sunt în viață și sunt conștiente. Mai mult, pentru a fi o persoană juridică, este necesară abilitatea de a juca un rol social, fie în prezent, fie în viitor. Aceasta poate fi definită ca abilitatea de a comunica și de a exprima idei într-un mod care este înțeles de majoritatea lumii.

Capacitatea juridică a persoanei juridice.

Majoritatea jurisdicțiilor din lume permit anumitor entități, altele decât persoanele fizice, să dobândească drepturi și obligații, cum ar fi corporațiile. Aceste entități care nu sunt umane, dar cărora li se acordă capacitate juridică, sunt denumite persoane juridice. O persoană juridică, asemenea persoanelor fizice, poate deține proprietăți, încheia contracte, să dea în judecată și să fie dată în judecată. Totuși, nu au dreptul de a vota sau de a încheia căsătorii, deoarece acestea sunt considerate drepturi inerente doar *ființelor umane*.

Așa cum se întâmplă cu toate domeniile nereglementate, reglementarea prezintă provocări deosebite pentru legiuitori, deoarece implementarea unor astfel de măsuri de reglementare poate duce la efecte imprevizibile care ar putea împiedica dezvoltarea viitoare.

Teoreticienii proprietății intelectuale se gândesc de cel puțin treizeci de ani dacă un sistem informatic poate fi considerat autor în raport cu drepturile de autor. Mai mult, dispozitivele de tip „asistent virtual”, cum ar fi Siri de la Apple, devin din ce în ce mai autonome și sunt capabile să producă conținut mai departe de controlul direct al omului, ceea ce ridică întrebarea dacă asistenții virtuali AI ar trebui să beneficieze de libertatea de exprimare. O altă controversă se referă la răspunderea penală a subiecților AI. Dacă un dispozitiv autonom AI, care nu este controlat de om, comite o infracțiune, cine ar trebui să fie tras la răspundere? Hallevy consideră că dispozitivele AI ar trebui să aibă răspundere penală, dacă pot înțelege că acțiunile pe care le întreprind sunt contrare legii.

Acum mai bine de douăzeci de ani, Lawrence Solum a abordat problema dacă sistemele de IA ar trebui să primească drepturi legale, ajungând la concluzia că, atâta timp cât științele cognitive confirmă că procesele care produc comportamentele IA sunt similare cu cele ale minții umane,

există un motiv foarte bun pentru a trata IA-urile ca pe oameni și, prin urmare, a le acorda drepturi. Mai mult, persoana juridică include deja non-oameni, cum ar fi ”părțile juridice”, așadar se poate concluziona că statutul uman nu este o condiție necesară pentru a obține personalitatea juridică. Personalitatea juridică, prin crearea unor noi clasificări, poate fi folosită pentru a acorda potențial drepturi non-omului. În acest fel, IA-urile ar putea beneficia de drepturile și responsabilitățile unei persoane juridice, fără a utiliza drepturile persoanelor fizice, care sunt în prezent considerate a fi exclusiv inerente naturii umane.

Capacitatea penală, este necesară? Unii autori afirmă că este nevoie de o capacitate penală pentru a ține ritmul cu noile tehnologii care apar[80]. Acești autori susțin că noțiunile clasice de responsabilitate, răspundere pentru produse și altele asemenea nu mai au capacitatea de a asigura înfăptuirea justiției și de a proteja interesele legitime relativ la această tehnologie. Totodată, acești autori arată că unghiul din care este văzută inteligența artificială începe să se schimbe, de la simple unelte în mâinile unui utilizator, la ceva care depășește acest concept. În viziunea acestor autori, schimbarea percepției are loc întrucât tehnologia exhibă o inteligență și o autonomie din ce în ce mai impresionantă, precum și abilități sociale, capacitate de a percepe și de a arăta empatie.

Alți autori susțin că, în realitate, ceea ce determină schimbarea nu este evoluția tehnologiei, ci factorii economici[81]. În viziunea lor, schimbările majore în drept apar atunci când tehnologia creează niște riscuri economice crescute. În continuare, acești autori explică felul în care a evoluat inteligența artificială de-a lungul decadelor și concluzionează că, în prezent, sume importante de bani sunt investite în acest sector.

„O specie este definită de ceea ce este și de locul în care acționează”[82], iar un subiect de drept în raporturile juridice care se nasc și se desfășoară în lumea finită și bidimensională este „prin excelență, individul”, considerându-se, pe alocuri la nivel axiomatic, că „omul este singurul în măsură să participe la un raport juridic”. Această axiomă juridică nu este altceva decât expresia unui primordial și indubitabil dat aprioric și anume, prezența omului în lume, ca punct de plecare al obligațiilor derivate din conduita agentului uman. Însă dacă evoluția își urmează cursul neîntrerupt în care homo sapiens l-a înlocuit pe homo neanderthalensis, înseamnă că putem presupune că și homo sapiens va fi înlocuit de o nouă specie, *homo digitalis*, „iar activitatea umană

[80] Laukyte, Migle, (2019), *AI as a Legal Person*. 209-213 DOI: 10.1145/3322640.3326701.
[81] Curtis E.A. Karnow, *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Pub (2018), p. XIX și urm.
[82]

din lumea reală va fi înlocuită cu activitatea digitală din lumea artificială și din cea virtuală. Acesta este viitorul nostru: Inteligența Artificială (IA)".

Termenul de persoană electronică a fost utilizat mai recent în proiectul de raport privind regulile de drept civil în domeniul roboticii, al Comisiei pentru Afaceri Juridice a Parlamentului European și se consideră în literatura juridică recentă că „este doar o chestiune de timp până când va fi acceptat conceptul de personalitate electronică și va fi atribuit sistemelor AI cu consecințele juridice inerente acestui fapt".

II. Noțiunea de responsabilitate a Inteligenței Artificiale.

În ceea ce privește o eventuală răspundere penală, alte probleme se ridică. În cazul *deciziilor autonome*: dacă un sistem de IA ia o decizie care duce la un rezultat negativ sau ilegal, cine este responsabil? Este creatorul sau operatorul sistemului, sau sistemul însuși? *Erori și biase ale algoritmilor*: Sistemele de IA pot fi susceptibile la erori sau biase care pot duce la rezultate nedrepte sau discriminatorii. Cine este responsabil pentru aceste erori sau biase? Este vorba de programatorii care au dezvoltat algoritmii, utilizatorii care i-au implementat sau chiar de sistemul de IA în sine? *Neglijența în dezvoltarea sau utilizarea* IA: Dacă o organizație dezvoltă sau utilizează un sistem de IA fără să ia măsuri adecvate pentru a preveni efectele negative sau ilegale, poate fi considerată responsabilă penal pentru consecințele acestor acțiuni? Ce se întâmplă în cazul în care, să spunem că am rezolvat aceste probleme de reglementare juridică, este condamnată într-un caz penal IA. Ce fel de pedeapsă îi aplicăm? Măsuri educative? Închisoare?

În literatura juridică că răspunderea juridică survine numai atunci când sunt întrunite anumite condiții: conduita ilicită; existența vinovăției; prejudiciul, rezultatul socialmente dăunător sau urmarea; legătura dintre conduita ilicită și rezultatul produs; inexistența unor împrejurări care exclud răspunderea juridică. Răspunderea juridică nu poate fi desprinsă de premisa responsabilității, ambele constituind fenomene legate comportamentul social, de cunoașterea reciprocă și alteritate, de normele sociale și juridice, de realizarea şi aplicarea dreptului, dar și a resorturilor legate de societate, autoritate și de interacțiunile cu personalitatea umană, în accepțiunea sa de persoană responsabilă și persoană-în-drept.

Faptul juridic ilicit reprezintă o varietate a faptelor juridice în general. Faptele juridice sunt acele evenimente și acțiuni omenești care, potrivit legii, dau naștere, modifică sau sting un raport juridic. Din această precizare rezultă că faptele juridice se exprimă sub două forme: evenimentele

și acțiunile omenești. Evenimentele sunt acele împrejurări care nu depind de voința omului (nașterea, moartea, calamitățile naturale), iar acțiunile omenești sunt cele care apar ca rezultat al exprimării voinței umane, de exemplu încheierea unui contract sau săvârșirea unei infracțiuni, ele pot fi deci licite și ilicite[83]. Sensul cel mai larg al noțiunii de conduită, care privește până în prezent numai ființa umană și numai prin derivație conduita unui lucru (sau bun juridic) este acela de voință și conștiință obiectivate prin dreptul pozitiv, iar nota definitorie a conduitei este obiectivarea conștiinței în acte și fapte deliberate[84]. Așadar, ansamblul faptelor concrete ale individului uman, aflate sub controlul voinței și rațiunii acestuia desemnează conduita umană, iar calificarea lor ca ilicite se produce dinspre dreptul pozitiv numai atunci când realizarea lor are loc prin încălcarea dispozițiilor normelor juridice în vigoare.

Într-o formulă sintetică, trăsăturile de maximă generalitate ale conduitei ilicite receptate de teoria generală a dreptului sunt următoarele: a) – antisocialitatea, adică trăsătura generală a tuturor categoriilor de fapte ilicite, deoarece orice acțiune ilicită prejudiciază o anumită relație socială, b) – antijuridicitatea, ca expresie a contradicției dintre conduita individului și o normă de drept prin care se interzic sau se impun anumite comportamente; c) – imoralitatea, ca expresie a încălcărilor concomitente în multe situații prin faptele săvârșite atât a ordinii de drept, cât și a ordinii morale.

Acestor trăsături ale conduitei ilicite, acceptate aproape unanim în literatura juridică română, dar și străină, li se mai adaugă, uneori – nefiind acceptată de toți autorii - și trăsătura de imputabilitate, ca în concepția lui T. Pop, care considera că: „întotdeauna o activitate efectivă poate fi caracterizată numai atunci ca o violare a dreptului, când și în măsura în care ea este imputabilă”[85].

Pentru a impune răspunderea juridică (penală) în cazul sistemelor AI acestea ar trebui să întrunească atât cerințele elementului obiectiv, (*actus reus*), cât mai ales cerințele elementului subiectiv (*mens rea*)[86]. Efectele deciziilor sau acțiunilor întreprinse de sistemele AI sunt adesea rezultatul unor interacțiuni nenumărate între mulți actori incluzând programatori, dezvoltatori, utilizatori, adică ceea ce doctrina denumește ,,problema mâinilor multiple”, (dar și a „lucrurilor

[83] Boboș Gheorghe, Vlădica Rațiu, George, *Răspunderea,responsabilitatea și constrângerea în domeniul dreptului*, Editura ARGONAUT, Cluj-Napoca, 1996, p. 39
[84] Mihai, Gheorghe, *Fundamentele dreptului. Teoria răspunderii juridice*, Volumul V, Editura C.H. Beck, București, 2006, p. 174
[85] Avrigeanu, Tudor, *Pericol social, vinovăție personală și imputare penală*, Editura Wolters Kluwer, București, 2010, p. 125
[86] L. Stănilă, *Inteligența artificială, dreptul penal și sistemul de justiție penală. Amintiri despre viitor*, UJ, 2020, p. 118

multiple”[87], ceea ce sporește nu numai ambiguitatea imputării faptei, ci mai ales a cerinței legăturii indisolubile dintre autor și faptă, precum și a libertății de acțiune și decizie a acestuia.

Cu alte cuvinte, trebuie rezolvată în continuare ecuația aristotelică a condițiilor referitoare la controlul acțiunii și condițiile epistemice ale acesteia, respectiv capacitatea de cunoaștere a sensului și consecințelor acțiunii, în sensul în care acțiunea trebuie să aparțină agentului, iar acesta trebuie să înțeleagă semnificațiile legale și morale ale acesteia.

Vinovăția sau culpabilitatea poate fi definită ca ansamblul proceselor psihice care fundamentează corelația dintre fapta ilicită și autor (în concepția psihologică) sau ca un reproș adresat acestuia din partea societății de a nu-și fi adaptat conduita cerințelor ordinii juridice, un concept normativ care exprimă contrarietatea dintre voința subiectului și norma de drept (în concepția normativă)[88].

Culpabilitatea sau vinovăția reprezintă expresia sintetică a aspectului subiectiv al faptei pentru că implică un act de conștiință, o atitudine a conștiinței în raport cu urmările faptei și un act de voință sub impulsul căruia este realizată fapta. Într-o altă opinie, vinovăția este un atribut al atitudinii unei ființe umane responsabile, liberă în spirit și în act, față de vorba, fapta sau gândul cărora le imprimă o interpretare subiectuală valorizatoare incompatibilă cu valorile universale.

În doctrina common law, vinovăția este asimilată culpabilității reclamă o „verificare din partea societății” din moment ce autorul faptei a comis în mod voluntar o faptă cu o consecință periculoasă. În teoria normativă, vinovăția este văzută ca „legătura internă între autor – ca destinatar și legitimitatea normei”, în virtutea căreia s-ar manifesta o anumită „componentă emoțională a decepției pentru violarea normei”. Din această perspectivă, specifică doctrinei juridice germane, vinovăția este pusă în relație cu indicatorul de „eficacitate motivatorie a normelor”, astfel încât incidența acestuia ar fi dependent de o prealabilă evaluare a conduitei persoanei în raport cu gradul de atașament al persoanei față de valorile juridice protejate de norme.

În sistemul penal englez, vinovăția este desemnată prin expresia *mens rea*, sintagmă ce ar cuprinde o sumă de elemente psihice prezente în momentul comiterii actului și care constau în funcțiile sau stările de prevedere, cunoaștere, acceptare, neglijență, ușurință. În acest sens, englezii obișnuiesc să afirme: „*There is no crime without guilty mind*”, (nu există infracțiune fără conștiință

[87] Coeckelbergh Mark, *Artificial Intelligence, Responsibility Attribution and a Relational Justification of Explainability,* Science and Engineering Ethics (2020) 26:2051–2068, available online: https://doi.org/10.1007/s11948-019-00146-8

[88] L. Stănilă, *Răspunderea penală a persoanei fizice*, UJ, București, 2012, p. 53

vinovată) ceea ce confirmă avertismentul trasat de importantul teoretician al dreptului penal, R. A. Duff, cu privire la imposibilitatea trasării unei distincții clare între *actus reus* și *mens rea.* Aspectul cel mai dificil al problemei constă în încercarea de a identifica și de a izola un actus distinct de orice element mental până la o simplă mișcare fizică a trupului. Din această etapă vom fi tentați să introducem un minim element mental ca voința sau dorința, ceea ce ne va conduce la un nou element compozit și anume, „o acțiune ca mișcare a trupului intenționată sau dorită", ceea ce nu ar rezolva problema în sensul dihotomiei urmărite.

Pentru ca o entitate inteligentă (AI) să poată fi trasă la răspundere penală aceasta ar trebui să îndeplinească cinci cerințe cumulative: capacitatea de comunicare – cu cât este mai facilă comunicarea cu o entitate cu atât mai inteligentă este aceasta; existența conștiinței – se presupune că o entitate inteligentă este conștientă de sine și este capabilă de introspecție; autodeterminarea și 4) independența – se presupune că o entitate inteligentă își poate stabili scopuri și obiective și poate întreprinde acțiuni în mod independent pentru a-și atinge obiectivele și a-și realiza scopurile; creativitatea – se presupune că o entitate inteligentă posedă un anumit grad de creativitate, nu neapărat în sensul de a crea un lucru nou ci mai degrabă să fie capabilă să întreprindă acțiuni alternative atunci când acțiunea inițială eșuează[89].

Totuși, în literatura de specialitate cea mai recentă se arată că deși tehnologia modelează acțiunea umană într-un mod care depășește un rol doar instrumental și că tehnologiile AI avansate pot crea aparența de a fi agenți responsabili, acestea nu îndeplinesc criteriile tradiționale pentru a fi considerate deplin responsabile din perspectiva dreptului și prin urmare ele nu pot fi trase la răspundere juridică în nume propriu.

Marea provocare a dreptului în stabilirea răspunderii juridice pentru o faptă juridică ilicită constă în detectarea și stabilirea vinovăției unui sistem AI, ca legătură internă dintre acesta, ca autor al unei fapte prohibite de lege și legitimitatea normei juridice, ca rezultat al unei aprecieri proprii sau ca un comportament simulat ori asociat cu comportamente arhivate în memoria sa (a se vedea conceptul testului Turing, Imitation Game). Prin urmare, în prezent, singura soluție este aceea de a atribui răspunderea juridică oamenilor pentru faptele tehnologiei AI.

Prejudiciul sau urmarea socialmente periculoasă. Conform „Rezoluției Parlamentului European din 20 octombrie 2020 conținând recomandări adresate Comisiei privind regimul de

[89] L. Stănilă, Inteligența artificială, dreptul penal și sistemul de justiție penală. Amintiri despre viitor, UJ, 2020, p. 112

răspundere civilă pentru Inteligența Artificială" dispozitivele sau procesele fizice sau virtuale care sunt dirijate de sistemele AI pot fi, din punct de vedere tehnic, cauza directă sau indirectă a prejudiciului sau a daunei, dar, cu toate acestea, sunt aproape întotdeauna rezultatul acțiunii cuiva care a construit, implementat sau perturbat sistemele (https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_RO.html).

În consecință, prejudiciul civil nu comportă dificultăți de configurare, definire sau de reparare deoarece răspunderea civilă este mai flexibilă și deține instrumente de reparație a pagubei produse practic inepuizabile, indiferent cine sau ce este autorul faptei prejudiciabile.

Cu toate acestea, adecvarea regulilor de răspundere civilă delictuală poate fi problematică în materia pagubelor create de sistemele dotate cu inteligență artificială, având în vedere modelul antropocentric al răspunderii delictuale pentru repararea faptelor prejudiciabile (delictelor) format încă din dreptul roman și care se întemeiază pe o cauză unică a producerii pagubei[90].

În dreptul penal, urmarea sau rezultatul dăunător produs se concretizează în lezarea unui interes, public sau privat. Fiind o ramură de drept public, rezultă că el ocrotește interesul public prin formula „principalele valori ale societății". Fie că este vorba despre un interes de ordin public sau despre unul privat, eficiența garantării sale legale rezidă în dubla ipostază a expresiei sale sociale, de generalitate a conviețuirii și de individualitate în conviețuire ori aceasta presupune receptarea și acceptarea alterității și o perspectivă intersubiectivă a relațiilor dintre actorii sociali.

Problema aristotelică a non-ignoranței primește noi semnificații ale cunoașterii, ce presupun conștientizarea acțiunii, conștientizarea semnificațiilor morale și chiar a alternativelor acelei acțiuni deoarece interesul, indiferent de titularul lui, are o triplă ipostază: cognitiv, afectiv și pragmatic. Acestea sunt valori-instrumente aflate în raport cu orizontul valorilor-scop, respectiv Adevărul, Binele, Frumosul, Dreptatea, Utilul, într-o multitudine de întrepătrunderi și potențări reciproce. *În acest context, se pune problema dacă valorile sociale protejate de dreptul pozitiv sunt opozabile sistemelor AI, fie și autonome, iar dacă ținta precisă a prejudiciului sunt drepturile subiective se ridică întrebarea dacă sistemele AI pot fi titulare de asemenea drepturi și obligații corelative.* Cu alte cuvinte, chestiunea la care trebuie soluționată filosofic, dar și pragmatic juridic, este aceea dacă sistemele AI pot fi considerate actori sociali (ca să evităm termenul „ființe").

[90] Liability for Artificial Intelligence and Other Emerging Digital Technologies – Report from the Expert Group on Liability and New Technologies – New Technologies Formation, PDF ISBN 978-92-76-12959-2 doi:10.2838/573689 Catalogue number: DS-03-19-853-EN-N, https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en , p. 19).

În ceea ce privește legătura de cauzalitate dintre acțiunea sistemelor AI și rezultat dificultatea constă în aceea că nu sunt doar mai multe „mâini" (umane) ci există de asemenea ceea ce am putea numi multe lucruri, adică multe tehnologii diferite, fiind implicate diverse programe, adică lucruri care sunt relevante deoarece contribuie în mod cauzal la acțiunea sistemului AI în ansamblu.

Se estimează în literatura de specialitate că problema cauzalității în materia faptelor realizate de sistemele AI este și mai complexă datorită dependenței acestora de date și input-uri externe, ceea ce va relativiza și mai mult legătura directă dintre faptă și rezultat în sensul existenței unei singure cauze sau a interacțiunii unor cauze multiple actuale sau potențiale[91] În vederea stabilirii legăturii de cauzalitate dintre fapta în care a fost implicată Inteligența Artificială și consecința socialmente negativă este necesară clarificarea tuturor interacțiunilor și a rolurilor avute atât de oameni, cât și a celor dintre oameni și sistemele AI sau numai dintre acestea pentru că în materia răspunderii tehnologiei sau a celei pentru tehnologie, aceasta este o problemă a ceea ce se vede, dar a și sub-entităților tehnologice care nu se văd.

În cazul sistemelor AI ar fi necesară o programare anterioară în sensul implementării sistemului normelor juridice, altfel sistemul ar acționa ca „agent inocent". Dacă, însă, sistemul AI ar acționa „pe baza propriilor experiențe, cunoștințe și decizii", ar fi posibilă angajarea răspunderii sale penale.

*La 4 iulie 1981[92], a fost raportată prima omucidere provocată de un robot. Un inginer, Kenji Udara efectua niște lucrări de întreținere la un robot la uzina Kawasaki Heavy Industries. Victima a omis să oprească complet robotul. Când a intrat într-o zonă restrânsă a liniei de producție, robotul a detectat-o ca un obstacol în linia de producție și a aruncat-o pe o mașină adiacentă, folosind brațul său hidraulic puternic, victima decedând instantaneu. Acest incident a fost primul de acest fel care ne-a arătat o formă de interacțiune a Inteligenței Artificiale cu omul. În alt caz, oamenii de știință au dezvoltat cu succes computere care pot învinge pe cel mai mare dintre Marii Maeștri de șah. Putem considera această inteligență echivalentă cu cea a oamenilor? Gary Kasparov a fost bătut de Deep Blue[93] care a demonstrat că joacă șah mai bine ca Marii

[91] *Liability for Artificial Intelligence and Othe Emerging Digital Technologies*-Report from the Expert Group on Liability and New Technologies, 2019, p. 22

[92] Paul S. Edwards, *Killer robot: Japanese worker first victim of technological revolution*, Deseret News Dec. 8, 1981, at A1.

[93] Deep Blue also makes mistakes: V. Anand 'More Questions than Answers', http://www.research.ibm.com/deepblue/home/may11/story_3.html (14 June 2017).

Maeștri. Cu toate acestea, Deep Blue a funcționat pe baza algoritmilor programați de către cercetătorii care s-au ocupat de acest proiect. Dar se ridică întrebarea, legat de acest exemplu, dacă cercetătorii care au programat Deep Blue nu au nicio șansă să-l bată pe G. Kasparov la șah, atunci cum a putut un computer (chiar și programat de ei)? Și-a dezvoltat inteligența pe baza algoritmilor? Se poate deduce din acest exemplu că Deep Blue gândește ca un om?

În legătură cu problema unei eventuale responsabilități penale a IA, unii autori[94] clasifică IA în Inteligență Artificială Generală (bazată pe programe de computer, egală sau superioară inteligenței umane) și alt tip de Inteligență Non Biologică (bazată pe programe de computer sau non terestră). Dar, acest tip de inteligență nu se limitează la noțiunea de robot, din exemplele de mai sus.

III. Răspunderea pentru fapta ilicită comisă de o mașină/robot/ dirijat de inteligența artificială; mașinile autonome: responsabilitate.

Prin noțiunea de mașini autonome se înțelege acele autovehicule care sunt capabile să analizeze mediul extern și să se deplaseze dintr-un punct în altul fără intervenția umană sau cu o intervenție minimă[95].

Inteligența artificială din cadrul mașinilor autonome funcționează utilizând viziunea computerizată și recunoașterea imagistică pentru a putea înțelege mediul înconjurător şi deep learning pentru a putea conduce mașina astfel încât să respecte regulile de circulație.

Mașinile autonome nu au aceleași limitări pe care la au simțurile umane, componenta de machine vision putând fi astfel concepută încât să vadă inclusiv prin ziduri. În acest sens, maşinile se bazează pe mai multe camere, senzori, sonar, GPS şi alte asemenea tehnologii. Printre ele se remarcă în mod special LiDAR, care este un acronim de la light detection and ranging. Această tehnologie permite maşinii să emită mai multe impulsuri laser şi, în funcţie de timpul necesar pentru ca lumina să se întoarcă şi de lungimile de undă ale luminii, să creeze imagini 3D ale mediului înconjurător. Odată creată această imagine a mediului înconjurător, inteligenţa artificială analizează această informaţie şi ia decizii pe baza ei[96].

---

[94] Tyler L. Jaynes, *Legal personhood for artificial intelligence: citizenship as the exception to the rule*, AI & SOCIETY (2020) 35, p. 343–354, https://doi.org/10.1007/s00146-019-00897-9

[95] C. Ching-Yao, *Advancements, prospects, and impacts of automated driving systems*, în International Journal of Transportation Science and Technology, vol. 6, nr. 3/2017, p. 208-216.

[96] C. Ching-Yao, *Advancements, prospects, and impacts of automated driving systems*, International Journal of Transportation Science and Technology, vol. 6, nr. 3/2017, p. 208-216.

Deep learning, este un mecanism ce permite inteligenței artificiale să analizeze mediul înconjurător, să extragă anumite informații și să creeze date predictive (spre exemplu, să estimeze unde se va afla o mașină peste câteva secunde), urmând să ia decizii în mod independent, pe baza acestor informații.

Convenția de la Viena din 1968 asupra circulației rutiere, ratificate și de România și publicată în Buletinul Oficial nr. 86 din 20 octombrie 1980[97] conține unul dintre principiile fundamentale pe care se întemeiază această convenție și anume: *șoferul este operatorul mașinii și trebuie să se afle întotdeauna în controlul autovehiculului și al sursei de pericol pe care o creează*.

Tot în aceeași situație este și actualul Cod rutier15, care prevede la art. 1 că circulația pe drumurile publice a vehiculelor, pietonilor şi a celorlalte categorii de participanți la trafic, drepturile, obligațiile şi răspunderile care revin persoanelor fizice şi juridice, precum şi atribuțiile unor autorități ale administrației publice, instituții şi organizații sunt supuse dispozițiilor prevăzute în prezenta ordonanță de urgență. Tot astfel, noțiunea de conducător, potrivit art. 6 pct. 12 din actul menționat, este definită drept persoana care conduce pe drum un grup de persoane, un vehicul sau animale de tracțiune, animale izolate sau în turmă, de povară ori de călărie. Este evident că legiuitorul nu a avut niciodată în vedere mașinile autonome atunci când a elaborat respectivele acte legislative (legislația europeană începe să permită mașinilor autonome să circule pe drumurile publice. Sistemele care sunt în curs de modificare a legislației sau care şi-au modificat în mod expres legislația pentru a permite circulaţia lor pe drumurile publice sunt: Olanda, Marea Britanie, Franţa, Germania și altele.

Așadar, având în vedere actuala legislație, trebuie concluzionat că simpla punere în circulaţie a unui autovehicul autonom este ilicită, pentru simplul motiv că se creează un risc pentru valoarea socială ocrotită care nu este acceptat de societate şi autorizat de lege ca atare (risc nepermis). Având în vedere că autovehiculele autonome nu sunt autorizate de către lege, rezultă că nu se poate vorbi despre un risc permis. De altfel, acest fapt este şi normal, întrucât societatea nu acceptă încă aceşti roboţi ca pe nişte elemente inerente stilului de viaţă. Chiar şi în măsura în care nu se acceptă teoria imputării obiective a rezultatului, răspunderea penală poate fi înlăturată pentru lipsa vinovăţiei, în astfel de cazuri[98].

---

[97] http://www.monitoruljuridic.ro/act/conventie-din-8-noiembrie-1968-asupra-circulatiei-rutiere-emitent-consiliul-de-stat-publicat-n-buletinul-oficial-nr-31023.html

[98] Georgian Marcel Husti, *Vinovăţia şi alte elemente privind răspunderea penală în cazul maşinilor autonome,* Caiete de drept penal nr. 3/2029, p. 73 și urm.

Sistemul de drept penal românesc consacră exclusiv răspunderea penală subiectivă, bazată pe vinovăţie. Pentru a exista o infracţiune, trebuie să se comită o faptă tipică, comisă cu vinovăţia cerută de lege, antijuridică şi imputabilă persoanei vinovate. Dintre aceste elemente, cele mai problematice pentru a stabili răspunderea în cazul infracţiunilor comise de maşinile autonome sunt legătura de cauzalitate şi vinovăţia. O primă problemă a actualului sistem referitor la vinovăţie este că, în prezent, majoritatea legislaţiilor şi a autorilor de drept consideră că nicio persoană nu ar trebui să creeze o sursă de pericol pe care să nu o poată controla. Or, dacă maşinile autonome îşi ating ţelul de a fi mai puţin periculoase decât cele conduse de oameni, atunci ar fi chiar dezirabil să se creeze aceste surse de pericol pe care să nu le poată controla oamenii, ci doar inteligenţa artificială. Alternativ, se poate considera că sursa de pericol este controlată prin faptul că producătorii, atunci când pun în circulaţie maşinile autonome, le testează şi ulterior le supraveghează.

Cu privire la identificarea unei culpe, apar diferenţe pentru următoarele motive: dacă pentru un eveniment rutier clasic se analizează dacă şi în ce fel a fost neglijent conducătorul auto, pentru un eveniment rutier care implică un vehicul autonom, pe lângă faptul că este necesar a se analiza raportul de cauzalitate direct, referitor la producerea evenimentului, trebuie verificat şi raportul de cauzalitate dintre comportamentul autovehiculului şi fabricaţia sau programarea sa. În cazul unui vehicul autonom, pe lângă aceste elemente ale stării de fapt, mai trebuie dovedită şi legătura de cauzalitate cu o persoană umană sau o persoană juridică, întrucât doar acestea pot răspunde penal.

Faptele intenţionate care se comit în legătură cu maşinile autonome nu pun prea mari probleme dreptului penal. În aceste cazuri, persoana se foloseşte de autovehicul ca de un simplu instrument, o unealtă necesară pentru atingerea ţelului. În mod evident, în aceste cazuri va răspunde persoana care a avut conduita infracţională. Astfel, dacă o persoană sabotează o maşină autonomă pentru a omorî un şofer, această persoană va răspunde pentru o infracţiune de omor. În schimb, nu va putea răspunde pentru un omor acea persoană care a îndemnat o alta să se deplaseze folosind un vehicul autonom tocmai în speranţa că va avea loc o defecţiune şi va deceda (fiind în ipoteza unui risc inexistent)[99].

Probleme mult mai mari apar, în schimb, în cazul infracţiunilor din culpă. Pentru a exista vinovăţie, este necesar ca autorul faptei să fi avut libertatea de a acţiona diferit faţă de modul în

[99] Idem

care a acţionat în mod efectiv. Prin urmare, pentru a reproşa o culpă persoanei chemate să răspundă pentru infracţiunile comise de maşinile autonome, este necesar a se stabili că ea avea obligaţia şi libertatea de a acţiona într-un alt mod. Juridic, aceste trăsături se regăsesc în faptul că (i) trebuie să existe o obligaţie de diligenţă care să fi fost încălcată de către cel chemat să răspundă şi (ii) că rezultatul era previzibil şi evitabil din punct de vedere subiectiv.

Referitor la primul element din structura culpei, anume cel al obligaţiei de diligenţă, se remarcă faptul că traficul rutier este, în sine, o sursă de pericol care este reglementată în vederea reducerii riscurilor pe care le presupune această activitate. Aceste norme pornesc de la premisa că şoferul trebuie să se afle întotdeauna în controlul autovehiculului. În consecinţă, prin ipoteză, simpla punere în circulaţie a maşinilor autonome, în actualul cadru legislativ, reprezintă încălcarea unei obligaţii de diligenţă.

Al doilea element pune probleme în mod special în ceea ce priveşte etalonul la care trebuie să aibă loc raportarea în ceea ce priveşte previzibilitatea şi evitabilitatea urmării. Având în vedere că şoferul autovehiculului este înlocuit de un sistem informatic, la ce standarde trebuie ţinut acesta? În cazul în care se produce un accident pe care un şofer mediu nu îl putea evita, dar un şofer foarte bun l-ar fi putut evita, fapta va constitui infracţiune? Ceea ce un om nu poate să prevadă ar putea să fie prevăzut de inteligenţa artificială. De altfel, aceasta este şi raţiunea pentru care au fost dezvoltate maşinile autonome. Pornind de la premisa că peste 90% dintre accidentele auto au drept cauză culpa umană, se încearcă eliminarea elementului uman din ecuaţie. Tehnologia înlocuieşte şoferul. În aceste condiţii, trebuie modificate şi standardele la care este ţinută conduita „şoferului". Maşinile autonome sunt dotate cu tehnologie superioară simţurilor umane, care le permite să fie mai atente şi să observe mai bine mediul înconjurător. Este elocvent, în acest sens, accidentul Uber din 2016. Un pieton a traversat strada neregulamentar, pe timp de noapte. Maşina autonomă nu a reuşit să îl observe şi l-a accidentat mortal[100]. Rapoartele preliminare au ajuns la concluzia că nimeni nu ar fi putut evita accidentul. Între timp însă au apărut numeroase articole în care diverşi specialişti afirmă că accidentul ar fi putut fi evitat de maşina autonomă. În timpul accidentului nu au funcţionat anumite sisteme de siguranţă şi frânare care ar fi trebuit să funcţioneze. Producătorul senzorului LiDAR, despre care am vorbit supra, susţine că această tehnologie ar fi trebuit să identifice pietonul care trecea neregulamentar, arătând că senzorul funcţiona corespunzător. Mai

[100] Idem, https://www.forbes.com/sites/meriameberboucha/2018/05/28/uber-self-driving-car-crash-whatreally-happened/#5a4dfca04dc4

mult, se discută despre dotarea, pe viitor, a maşinilor autonome, cu tehnologie capabilă de imagistică termică. Această tehnologie ar trebui să fie mai mult decât capabilă de a detecta din timp orice fiinţă, inclusiv animalele, în timp suficient pentru a evita accidentele pe timp de noapte. Aşadar, pentru tehnologie, rezultatul era previzibil şi evitabil. Tehnologia are potenţialul de a fi mai capabilă decât o fiinţă umană. În consecinţă, dacă pentru un om este imposibil de evitat un astfel de accident în acele condiţii de noapte, pentru inteligenţa artificială acest lucru era posibil.

În măsura în care maşina autonomă ar fi trebuit într-adevăr ca, în mod normal, să detecteze respectivul pieton şi să activeze sistemul de frânare, programatorul va răspunde în măsura în care putea să prevadă şi să redacteze codul astfel încât să ofere maşinii suficiente repere în funcţie de care să ia decizia corectă (de a frâna). Dacă un programator aflat în poziţia agentului, având experienţa şi capacitatea sa de a prevedea posibileleerori, ar fi putut să prevadă un asemenea accident, atunci acesta ar fi răspuns penal, iar în caz contrar – nu. În esenţă, sistemul de apreciere a previzibilităţii şi evitabilităţii respectivului eveniment rămâne la fel, doar că acesta nu se aplică sistemului informatic (maşinii), ci persoanei din spatele acesteia. Pe măsură ce tehnologia va avansa şi va fi din ce în ce mai greu de făcut un asemenea examen, apreciem că etalonul poate fi utilizat prin raportarea la modul în care ar fi trebuit să se comporte maşina autonomă dacă ar fi funcţionat corect sistemele sale. Desigur, acest etalon îşi dovedeşte limitele în măsura în care eroarea provine tocmai dintr-o eroare de design, caz în care se va aplica acest etalon persoanei responsabile cu designul maşinii.

Totuşi, pentru a analiza poziţia subiectivă a autorului, este necesar ca raportarea să se facă la momentul specific la care are loc acţiunea sau inacţiunea. Din acest punct de vedere, este discutabil în ce măsură programatorul sau operatorul ar putea să prevadă conduita maşinii. Odată setate regulile, folosind machine learning, spre exemplu, inteligenţa artificială se dezvoltă singură, pe baza tiparelor pe care învaţă să le recunoască şi să le reproducă, luând decizii pe baza acestor informaţii. Odată ce programatorul a stabilit mai multe reguli şi condiţii, inteligenţa artificială se dezvoltă în mod independent. În esenţă, folosind machine learning, inteligenţa artificială se dezvoltă fără intervenţie externă (fără a fi programată). O maşină autonomă care are deja o experienţă de câţiva ani este semnificativ mai dezvoltată decât una recent pusă în mişcare. În aceste condiţii, se mai poate reţine vinovăţia programatorului-operator, având în vedere că maşina a evoluat fără intervenţia sa externă? Este important de reţinut că inteligenţa artificială se poate dezvolta doar atât cât îi permit bazele de date la care are acces. Bazele de date la care are acces

sunt furnizate şi organizate de către operatorul uman. Pentru aceste motive, apreciem că şi în astfel de cazuri există în continuare elementul previzibilităţii şi al evitabilităţii, existând, pe cale de consecinţă, şi infracţiune. Desigur, analiza cu privire la posibilitatea de prevedere şi evitare rămâne, în continuare, să fie făcută de la caz la caz. Cu toate acestea, simplul fapt că inteligenţa artificială a evoluat de una singură nu este apt să înlăture vinovăţia celui chemat să răspundă, întrucât acesta, alegând bazele de date, are posibilitatea să prevadă comportamentul maşinii. În schimb, în măsura în care inteligenţa artificială se va dezvolta suficient de mult cât să poată intra în clasificările theory of mind sau chiar self-aware, prezentate supra, devine din ce în ce mai greu de acceptat faptul că, la nivel subiectiv, rezultatul era previzibil şi, mai ales, evitabil.

Prin urmare, vinovăţia s-ar putea întemeia pe faptul că programatorul-operator, cu ocazia acestor controale, ar fi putut să prevadă faptul că există anumite erori de programare care ar putea conduce la un accident. Cu atât mai mult se va putea reproşa producătorului faptul că a eliberat maşina pe piaţă fără a o testa suficient sau că nu a evaluat corect feedbackul primit de la utilizatori, că nu a informat publicul cu privire la anumite informaţii esenţiale ş.a.m.d. Proprietarul maşinii poate răspunde dacă nu întreţine maşina în mod corespunzător. Beneficiarul poate răspunde dacă manevrează incorect autovehiculul (atât timp cât neglijenţa nu se produce ca urmare a lipsei de informare).

* Potrivit lui Bernard Marr, o faţetă a inteligenţei umane este capacitatea de a învăţa din experienţă[101]. Învăţarea automată, pe de altă parte, este o aplicaţie a AI care imită învăţarea şi permite software-urilor să înveţe din practică. Nu reprezintă o noutate faptul că, prin intermediul AI, se pot identifica potenţiali criminali[102]. De pildă: „risk assessment tools", o clasă de instrumente algoritmice, numite instrumente de evaluare a riscurilor (RAI), sunt concepute pentru a prezice viitorul risc al unui inculpat de a comite fapte antisociale[103]. Aceste previziuni ajută organele judiciare, spre exemplu, dacă inculpatul trebuie încarcerat sau nu. De altfel, putem menţiona chatbot-ul Sweetie, un program AI, a cărui menire este să combată pedofilia online (acte cu conţinut sexual explicit care implică copii, realizate prin intermediul webcam-ului) prin

[101] Bernard Marr, What Is the Difference Between Deep Learning, Machine Learning and AI?, Forbes (2016), disponibil la https://www.forbes.com/sites/bernardmarr/2016/12/08/what-is-the-difference-between-deep-learning-machine-learning-and-ai/?sh=3781af0726cf

[102] A Recommendation System for Statutory Interpretation in Cybercrime" at the University of Pittsburgh, NIJ award number 2016-R2-CX-0010, disponibil la https://nij.ojp.gov/funding/awards/2016-r2-cx-0010

[103] Alex Chohlas-Wood, Understanding risk assessment instruments in criminal justice, în The Brookings Institution (2020), disponibil la https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

identificarea suspecţilor, infractorilor şi a victimelor. Acest chatbot a fost creat de organizaţia Terre des Hommes[104] din Olanda, în 2003. De la prima utilizare, Sweetie a condus la condamnarea mai multor cetăţeni englezi, danezi, olandezi şi belgieni pentru pedofilie online[105].

Sistemele AI ar putea permite actorilor umani să comită infracţiuni „invizibile". De exemplu, majoritatea oamenilor nu sunt capabili să imite vocile altor persoane în mod realist sau să creeze manual fişiere audio care să se asemene cu înregistrări ale vorbirii umane. Cu toate acestea, recent s-au înregistrat progrese semnificative în dezvoltarea sistemelor AI de sinteză a vorbirii care învaţă să imite vocile indivizilor. Nu există niciun motiv evident pentru care rezultatele acestor sisteme nu ar putea deveni imposibil de distins faţă de înregistrările autentice, în absenţa măsurilor de protecţie special concepute.

[104] www.terredeshommes.nl
[105] Terre des Hommes, *First conviction for child abuse in Belgium thanks to Sweetie*, 2015, disponibil la https://www.terredeshommes.nl/en/news/first-conviction-child-abuse-belgium-thanks-sweetie

Capitolul VI.

Impactul tehnologiei (inclusiv a Inteligenței Artificiale) asupra rețelelor de socializare; hărțuirea pe rețelele de socializare; Cyberbullying-ul

Utilizarea tehnologiilor informației și comunicațiilor a revoluționat modul în care oamenii comunică și formează relații între ei. De la copii mici până la persoane în vârstă, oamenii nu sunt niciodată departe de tehnologie. Cu toate acestea, forma pe care o ia tehnologia și scopurile pentru care este utilizată variază în funcție de vârsta indivizilor[106]. De exemplu, în timp ce copiii de școală elementară folosesc adesea tablete pentru a viziona videoclipuri sau pentru a juca jocuri online, adolescenții sunt atrași mai mult de smartphone-uri, de pe care pot accesa rețelele sociale pentru a stabili și a menține relații cu colegii lor[107].

Adulții folosesc din ce în ce mai mult tehnologia (computer, tabletă, telefon smart) ca vehicul pentru stabilirea unor relații personale (după cum demonstrează proliferarea site-urilor de întâlniri online), dar și în scopuri mai practice, cum ar fi obținerea de indicații, urmărirea de știri sau divertisment, ori îndeplinirea sarcinilor la locul de muncă, iar tendința generală este în creștere.

Într-o societate democratică unde fiecărui cetățean îi este garantată respectarea drepturilor și libertăților fundamentale, este important să remarcăm că unul dintre aceste drepturi pe care absolut orice persoană este liberă să-l manifeste este dreptul la libera exprimare a opiniei, care este invocat ori de câte ori sunt expuse puncte de vedere antagonice, ori de câte ori cineva încearcă să impună un punct de vedere sau să nege dreptul la existență a unei alte opinii, ori de câte ori aceea idee exprimată lezează demnitatea altuia, dar nu fiecare conștientizează că acest drept trebuie raportat și la ceilalți care la fel, au dreptul să-și exprime liber opinia într-o polemică, și aceste opinii trebuie respectate și acceptate, chiar dacă nu corespund cu cea a emițătorului inițial[108]. Se pune problema de a ști ce se întâmplă când în urma expunerii unei opinii, anumiți indivizi încep a insulta sau a scrie mesaje private care au un conținut agresiv, ofensator, violent sau amenințător, și cum se identifică astfel de acțiuni. Mai mult, ce facem atunci când persoana din spatele

[106] Robin Kowalski, Susan P. Limber, Annie McCord, *A developmental approach to cyberbullying: Prevalence and protective factors.* Aggression and Violent Behavior, 2017, p.3 și urm., doi: 10.1016/j.avb.2018.02.009

[107] Lenhart, A. (2015), *Teens, social media & technology overview 2015* (Internet American Life Project). Pew Research Center; http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/

[108] Vlada Usatâi, Cristina Lazariuc, *Cyberbullying: dominanții anonimi*, Technical-Scientific Conference of Undergraduate, Chisinau, 23-25 March 2021, Vol. II, p. 647

monitorului se ascunde sub anonimat care nu ne permite să aflăm absolut nici un fel de informație veridică despre acesta, ceea ce ne face vulnerabili și neprotejați?

Acest comportament în mediul on-line se numește cyberbullying (hărțuire cibernetică), ce reprezintă, în sine, un sistem acțiuni care utilizează tehnologiile informației și comunicației pentru a susține un comportament deliberat, repetat și ostil al unui individ sau grup de indivizi, care este destinat să dăuneze unei persoane anumite sau unui grup de persoane[109].

Istoricul Howard Segal[110] sugerează că toate evoluțiile tehnologice sunt ”binecuvântări amestecate” (mixed blessings), prezentând societății atât beneficii extraordinare, cât și neașteptate poveri[111]. Similar cu formele tradiționale de agresiune, agresiunea cibernetică este adesea deliberată și neobosită, dar poate fi și mai deranjantă din cauza naturii anonime a atacului. Potențialii agresori cibernetici își pot ascunde identitatea folosind nume de ecran (screen name) și adrese de protocol de internet bine ascunse, lăsând victima vulnerabilă și neliniștită.

Deși există foarte multe studii referitoare la cyberbullying, elaborate din diferite perspective (din perspectiva agresorului, a victimei, a unui grup țintă – minori, spre exemplu) niciunul nu explică noțiunea de agresiune cibernetică (cyberbullying). În anul 2009[112] a fost lansată o teorie care presupune că pentru a avea loc un comportament deviant trebuie să existe o convergență în timp și spațiu a trei elemente minime: un agresor probabil, o țintă adecvată și absența unui apărător capabil[113]. Cyberbullyingul (agresiunea cibernetică) este semnificativ legată de delincvență și interiorizarea unor forme de devianță precum auto-vătămarea intenționată și ideea suicidară.

Pentru a înțelege ce reprezintă cyberbullingul, în esență, este necesar mai întâi a explica ce reprezintă însuși bullying-ul, astfel va fi mai ușor de determinat care este rădăcina problemei, cyberbullying-ul nefiind altceva decât o formă a bullying-ului, desigur, cu trăsături specifice. Fenomenul de „bullying” reprezintă orice lezare fizică, verbală și socială, având efecte devastatoare atât asupra psihicului unui copil sau adolescent, cât și asupra unei persoane adulte. Acesta îmbracă forma agresiunilor directe sau indirecte, excluderi, farse cu caracter denigrator,

[109] Idem, p. 648
[110] Nagle, M. (2008), *Casualties of bullying*, UMaine Today, Vol. 8 No. 1, pp. 2-9.
[111] Dianne L. Hoff and Sidney N. Mitchell, *Cyberbullying: causes, effects, and remedies*, Journal of Educational Administration Vol. 47 No. 5, 2009 pp. 652-665;
[112] Mesch, G.S. (2009), *Parental Mediation, Online Activiteis, and Cyberbullying*, CyberPsychology & Behavior 12 (4): 387-393
[113] Felson, M., and Clarke R.V. (1998), *Opportunity Makes the Thief: Practical theory for crime prevention*, Police Research Series Paper 98, London: Home Office

umilitor sau intimidant. Pentru ca un abuz fizic sau emoțional să fie considerat ca aparținând fenomenului de bullying, acesta trebuie să prezinte câteva caracteristici, și anume: să fie intenționat, să fie consecvent sau altfel spus să aibă caracter repetitiv și să există o inegalitate clară între forțe, agresorul considerându-se superior victimei, adică având capacitatea de a o domina. Bullying-ul este prezent practic în toate sferele sociale[114]. O caracterizare a agresiunii ”tradiționale” (de tip bullying) o găsim și la Olweus[115]: ”*agresiunea tradițională este definită ca fiind un act agresiv care provoacă rău sau suferință, care se repetă de obicei în timp și care apare în rândul indivizilor a cărei relație se caracterizează printr-un dezechilibru de putere*”. Deseori termenul de bullying (engl. to bully – a teroriza), se confundă cu termenul de mobbing (engl. mob - „mulţime dezorganizată, angajată în violență fără reguli) care se referă tot la hărțuirea psihologică. Leymann consideră însă, că în ciuda faptului că există multe elemente comune, semnificativă rămâne diferenţa între cei doi termeni: bullying-ul este specific mai ales mediului şcolar, iar mobbing-ul – mediului organizaţional. În vreme ce mobbing-ul vizează victime ce provin din rândurile persoanelor calificate peste medie, în cadrul unui proces de bullying agresorul testează terenul pentru a vedea dacă între angajaţii noi există, aşa-numitele „ţinte uşoare. Mobbing-ul presupune o agresiune mai puţin exprimată fizic, spre deosebire de bullying care recurge mai degrabă la agresiunea de tip fizic. Din acest motiv, în cazurile de bullying, agresorul are şanse mai mari să fie pedepsit[116].

Comportamentul agresiv este așadar o acțiune negativă atunci când cineva provoacă intenționat sau încearcă să provoace, rănire sau disconfort asupra altei persoane. În această definiție, fenomenul agresiunii este astfel caracterizat prin trei criterii specifice: (a) Este un comportament agresiv sau „a face rău” intenționat (b) care se desfășoară de obicei cu o anumită repetitivitate (c) relație interpersonală caracterizată printr-un dezechilibru de putere, favorizând

[114] Idem, p. 648

[115]Dan Olweus, *School bullying: Development and some important challenges. Annual Review of Clinical Psychology,* 2013, nr. 9, p. 1-14. doi:10.1146/annurev-clinpsy-050212-185516; Olweus explică modul în care a apărut acest fenomen: Un puternic interes social pentru fenomenul a ceea ce s-a numit ulterior Bullying-ul a început mai întâi în Suedia la sfârșitul anilor 1960 și la începutul anilor 1970, sub denumirea de „mobbing”. După două publicări a cercetării sale pe această temă, una în 1973 în Suedia apoi în SUA în anul 1978 (cu denumirea: *Aggression in the Schools: Bullies and Whipping Boys)* autorul a decis schimbarea denumirii de ”mobbing” cu cea de ”bullying”, în anul 2013.

[116] Heinz Leymann, *Mobbing and Psychological Terror at Workplaces,* Violence and Victims nr. 5 (1990), p. 119-126.

făptuitorul (făptuitorii). Tot potrivit lui Olweus[117] sunt șapte niveluri diferite în cadrul scării de intimidare: persoanele care doresc să intimideze și să inițieze acțiunea, adepții lor, susținătorii sau agresorii pasivi, susținătorii pasivi sau eventualii agresori, privitorii dezactivați, posibilii apărători și apărătorii cărora nu le place acțiunea de intimidare și îi ajută pe cei victimizați.

În acest context, al violenței/agresiunii tradiționale, se ridică problema dacă criteriile menționate mai sus, în definirea acestui tip de comportament, pot fi folosite și în definirea cyberbullying-ului. Plecând de la această definiție a agresiunii ”tradiționale”, Peter Smith[118] și colegii săi au definit agresiunea cibernetică ca fiind „*un act agresiv, intenționat, efectuat de un grup sau de o persoană,* ***utilizând forme electronice de contact****, în mod repetat și în timp împotriva unei victime care nu se poate apăra cu ușurință pe ea însăși”.*

Dan Olweus face și el o caracterizare a cyberbullying-ului ca fiind acea ”*hărțuire cu forme electronice de contact sau comunicare, cum ar fi telefoanele mobile / celulare și Internet-ul*”[119]. O altă definiție este dată de profesorul canadian Bill Besley, potrivit căruia cyberbullying-ul este ”*utilizarea tehnologiilor informaționale și de comunicare precum e-mailul, telefonul mobil și mesaje text pager, mesagerie instantanee (Instant Message), site-uri web personale defăimătoare și site-uri web de sondaje personale online defăimătoare, pentru a sprijini deliberat, repetat un comportament ostil al unui individ sau grup, care este destinat să dăuneze altora*”[120]. Alți termeni folosiți sunt: intimidarea/hărțuirea electronică, e-intimidarea/hărțuirea sau intimidare/hărțuire tehnologică (bullying tehnologic). Se pare că există două medii principale care sunt implicate în agresiunea cibernetică, telefonul mobil/celular și computerul. În timp ce Internetul a fost descris ca fiind transformarea societății prin furnizarea de comunicare persoană-persoană, telefonul mobil a transformat grupul de prieteni/colegi într-un grup conectat la rețea. Conceptualizarea hărțuirii cibernetice este agravată de faptul că hărțuirea cibernetică poate avea multe forme diferite și poate apărea în multe locuri diferite. Drept exemple de forme de manifestare a hărțuirii cibernetice pot servi formele de comunicare care urmăresc să intimideze, să controleze, să manipuleze, să discrediteze sau să umilească destinatarul. În ceea ce privește *dominația anonimilor*, aceștia sunt

[117] Olweus, D. (2001), *Peer harassment: a critical analysis and some important issues*. (pp. 3-20). New York: Guilford Publications. Retrieved from http://www.olweus.org/public/bullying.page
[118] Smith, P. K., *The nature of cyberbullying and what we can do about it,* Journal of Research in Special Education Needs, 2015, nr. 15, p. 176-184. doi: 10.1111/1471-3802.12114
[119] Olweus, D., *Cyberbullying: A critical overview.* In B. J. Bushman (Ed.), Aggression and Violence: A Social Psychological Perspective, 2017, pp. 225- 240, New York: Routledge.
[120] Bill Belsey*, Always on? Always Aware!' Cyberbullying* at 17 January 2007, para 8

persoanele care se consideră îndreptățite să realizeze acțiunile enumerate mai sus, protejându-și în același timp propria persoană de un efect asemănător[121].

Practica hărțuirii cibernetice nu se limitează la copii și adolescenți, comportamentul este identificat prin aceeași definiție, atunci când este practicat de adulți, distincția în grupele de vârstă se referă uneori la abuz ca cyberstalking sau cyberharment atunci când este comis de adulți față de adulți. Tacticile obișnuite folosite de cyberstalker sunt efectuate în forumuri publice, rețele sociale sau site-uri de informații online și sunt menite să amenințe câștigurile, angajarea, reputația sau siguranța victimei. Mulți cyberstalkers încearcă să distrugă reputația victimei lor și să întoarcă alte persoane împotriva acesteia.

Criteriul dezechilibrului de putere poate fi evaluat „în termeni de diferențe în cunoștințele tehnologice dintre făptuitor și victimă, anonimat relativ, statut social, număr de prieteni sau poziția grupului marginalizat”[122]. Criteriul repetării ar trebui să fie înțeles într-un mod oarecum diferit, cu accent pe cât de mulți indivizi pot fi contactați cu un mesaj negativ, mai degrabă decât pe comportamentul cibernetic al făptuitorului, care constă adesea într-*un singur act*.

Hărțuirea cibernetică diferă de hărțuirea tradițională într-o varietate de moduri[123]: a) depinde de un anumit grad de expertiză tehnologică; b) este în primul rând un mod indirect mai degrabă de a hărțui victima, decât față în față și, prin urmare, făptuitorul poate fi anonim; c) în mod similar, făptuitorul nu vede de obicei reacția victimei, cel puțin pe termen scurt; d) varietatea rolurilor de spectatori în hărțuirea cibernetică este mai complexă decât în majoritatea hărțuirilor tradiționale (spectatorul poate fi cu făptuitorul atunci când este trimis sau postat un act; cu victima; sau cu niciunul, când se vizitează doar anumite pagini de internet cu un anumit mesaj, vizibil multor persoane); e) se consideră că un motiv al agresiunii tradiționale este statutul câștigat prin arătarea puterii (abuzive și fizice) asupra altora, în fața martorilor, însă agresorului cibernetic îi va lipsi adesea acest lucru; f) amploarea publicului potențial este crescută, deoarece hărțuirea cibernetică poate ajunge la un public deosebit de mare într-un grup egal (peers), comparativ cu grupurile mici care sunt publicul obișnuit în agresiunea tradițională; g) este dificil să scapi de

[121] Vlada Usatâi, Cristina Lazariuc, *op. cit*., p. 649
[122] Smith, P. K., del Barrio, C., & Tokunaga, R., *Definitions of bullying and cyberbullying: How useful are the terms*? In S. Bauman, J. Walker, & D. Cross (Eds.), *Principles of cyberbullying research: Definition, measures, and methods,* 2013, pp. 26–40, Philadelphia, PA: Routledge.
[123] Smith, P. K. (2012), *Cyberbullying and cyber aggression*, In S. R. Jimerson, A. B. Nickerson, M. J. Mayer, & M. J. Furlong (Eds.), Handbook of school violence and school safety: International research and practice (pp. 93–103). New York, NY: Routledge

hărțuirea cibernetică (nu există „nici un refugiu sigur"), deoarece victimei i se pot trimite mesaje pe mobil sau computer sau poate accesa comentarii urâte ale site-ului, oriunde s-ar afla.

Hărțuirea cibernetică *directă* se referă la acte agresive limitate doar la făptuitor și victimă. Pe de altă parte, hărțuirea cibernetică *indirectă* are loc pe mai multe *platforme media* și are potențialul de a implica un public mult mai mare decât doar victima și făptuitorul. Tot în acest context, cercetători ca Giumetti, McKibben, Hatfield, Schroeder, and Kowalski[124] au definit cyber agresiunea și "cyber incivility" ca fiind "*comportamente nepoliticoase/grosolane care apar prin intermediul tehnologiei informației și comunicațiilor, cum ar fi e-mailul sau mesajele text*". S-a propus și o definiţie a reţelelor sociale (SNSs) ca „servicii bazate pe Internet care permit indivizilor să (1) îşi construiască un profil public sau semi-public într-un sistem de sine stătător, (2) îşi articuleze o listă de alţi utilizatori cu care au o anumită relaţie şi (3) să vadă şi să traverseze lista lor de contacte şi a altora în interiorul sistemului"[125]. Unele dintre trăsăturile SNS pot facilita comportamentul agresiv online, cum ar fi profilul public al indivizilor (ca „regulă" a acestor site-uri, deşi în momentul creării profilului sau ulterior utilizatorul şi-l poate restricționa), vizibilitatea reţelei de prieteni a fiecăruia (chiar dacă profilul e restricționat, ca în cazul Facebook), posibilitatea de a posta comentarii la profil sau la poze sau, mai recent, aplicațiile de chat/mesagerie instant integrate profilului individual[126]. Controlabilitatea mediului online a fost folosită drept explicaţie pentru victimizarea în relaţiile/ interacţiunile online. Adolescenţii care comunică mai mult pe Internet percep într-o mai mare măsură acest tip de comunicare ca fiind mai controlabilă, caracterizată de o mai mare reciprocitate, mai extinsă şi mai profundă decât comunicarea faţă în faţă[127]. Prin urmare, asocierea dintre competenţele sociale şi cyberbullying poate furniza anumite explicaţii pentru fenomenul agresivităţii online prin identificarea patternurilor (tipare) de vulnerabilitate socială. Formele indirecte de intimidare/hărțuire cibernetică pot avea loc și fără

[124] Giumetti, G. W., McKibben, E. S., Hatfield, A. L., Schroeder, A. N., & Kowalski, R. M., *Cyber incivility @ work: The new age of interpersonal deviance. Cyberpsychology,* Behavior, and Social Networking, 2012, nr. 15, pp.148-154. doi:10.1089/cyber.2011.0336

[125] Ellison, N. B. (2007), *Social network sites: Definition, history, and scholarship,* Journal of Computer-Mediated Communication, 13(1), article 11. Available from http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[126] Monica Barbovschi, *A fi social şi nesocial online. Explorarea comportamentului agresiv online al copiilor și adolescenţilor (cyber-bullying*), Acest articol este rezultatul proiectului de cercetare Riscuri și efecte ale utilizării Internetului în rândul copiilor și adolescenților. Perspectiva evoluției spre societatea cunoașterii (knowledge society) finanțat de Ministerul Educației, grant CNCSIS tip A (nr. 1494/2007); echipa de cercetare coordonată de Prof. Dr. Maria Roth, Universitatea Babeș-Bolyai, Cluj-Napoca, p. 5

[127] Peter, J., Valkenburg, P.M. (2006), *Research Note: Individual Differences in Perceptions of Internet Communication,* European Journal of Communication 21(2). London: Sage Publications; Monica Barbovschi, *op. cit.*, p. 5

implicarea directă a victimei. Aceste forme de intimidare cibernetică sunt identificate ca fiind „behind my back bullying”[128]. Răspândirea în mediul online de bârfe și zvonuri despre o persoană sau postări publice de comunicări potențial jenant sau imagini ale victimelor în spațiul cibernetic sunt exemple de forme de intimidare cibernetică indirectă.

Deși hărțuirea cibernetică și cyber incivility sunt corelate între ele și ambele sunt asociate cu rezultate negative, aceste fenomene sunt conceptual distincte una de alta, în special în ceea ce privește dezechilibrul de putere și intenția de a face rău, care este considerat a fi mult mai ambiguă în cazul cyber incivility decât în cazul agresiunii cibernetice[129]. Agresiunea cibernetică se referă la acțiunile dăunătoare intenționate care sunt comise cu ajutorul sau prin tehnologie dar care nu includ un dezechilibru de putere sau repetare[130]. Unii cercetători[131] au susținut că dificultățile în conceptualizarea și măsurarea adecvată a dezechilibrului de putere și/sau repetarea în cazul hărțuirii cibernetice trebuie să se concentreze mai degrabă asupra agresiunii cibernetice decât asupra cyberbullying. Din cauza importanței dezechilibrului de putere dintre victimă și făptuitor păstrăm utilizarea termenului de cyberbullying (sau intimidare/hărțuire cibernetică). Așa cum a remarcat Olweus[132] „*argumentele conceptuale și constatările empirice raportate susțin în mod clar utilitatea unei distincții pe linia cercetării* ***perpetuării*** *agresiunii/victimizare pe de o parte, și o linie generală de cercetare, a agresiunii/victimizării, în general, pe de altă parte*”. Cyberbullying-ul (hărțuirea cibernetică) are în comun cu hărțuirea tradițională trei caracteristici principale: este un act de agresiune; apare la indivizi între care există un dezechilibru de putere; iar comportamentul se repetă adesea. O caracteristică specifică este aceea că agresiunea sau hărțuirea are loc prin intermediul unui sistem informatic sau de comunicare la distanță (telefon mobil, tabletă). Faptul că o persoană este mai pricepută din punct de vedere tehnologic decât alta poate crea un dezechilibru de putere. Mai mult, anonimatul inerent în multe situații de cyberbullying poate crea un sentiment de neputință din partea victimei.

---

[128] Walrave, M., Demoulin, M., Heirman, W. and A. van der Perre, (2009) *Cyberpesten: Pesten in Bits & Bytes* [Cyberbullying: bullying in bits and bytes], Observatorium van de Rechten op het Internet, p. 27

[129] Idem

[130] Bauman, S., Baldasare, A., *Cyber aggression among college students: Demographic differences, predictors of distress, and the role of the university*, Journal of College Student Development, 2015, nr. 56, pp. 317-330.

[131] Grigg, D. W., *Cyber-aggression: Definition and concept of cyberbullying*, Australian Journal of Guidance Counselling, 2010, nr. 20, pp. 143-156

[132] Olweus, D., *School bullying: Development and some important challenges,* Annual Review of Clinical Psychology, 2013, nr. 9, p. 1-14. doi:10.1146/annurev-clinpsy-050212-185516

În ciuda similitudinilor dintre agresiunea cibernetică (cyberbullying) și cea tradițională, este important să observăm trăsăturile distinctive în cadrul celor două comportamente: *Anonimatul* este una dintre modalitățile cheie prin care hărțuirea cibernetică și hărțuirea tradițională diferă între ele. Internetul și comunicarea la distanță facilitează posibilitatea de a fi anonim. Agresorii, aflați la adăpostul anonimatului, se simt mai liberi să agreseseze, sub diverse forme, victimele. La urma urmei, riscurile de a fi prinși și pedepsiți pentru intimidarea cibernetică a unei alte persoane, de exemplu prin utilizarea unei identități fictive, par să fie scăzute. Mai mult, riscul de a fi atacat fizic de către victimă nu există pentru agresorii care își păstrează identitatea secretă. În plus, caracteristicile spațiului cibernetic determină ceea ce Walrave și colab. au denumit „the cockpit effect"[133]. Datorită „absenței unor indicii relaționale, precum și a apropierii fizice de o altă persoană" în spațiul cibernetic, infractorii au impresia că acțiunile lor au un impact redus asupra victimelor. Din această cauză, agresorii par nemiloși și par să aibă o lipsă de empatie față de victimele lor[134]. Mai mult, spațiul cibernetic oferă un public infinit pentru agresori. De exemplu, odată ce un agresor încarcă în spațiul virtual (internet) un site web pentru a-și bate joc de o victimă, oricine are o conexiune la internet îl poate vedea. Prin urmare, spațiul cibernetic oferă o oportunitate relativ ușoară pentru infractori de a-și răni intenționat victimele. Nu în ultimul rând, spațiul cibernetic (cyberspace) permite ca agresiunea să aibă loc indiferent de timp și spațiu. Deși oamenii nu sunt niciodată la fel de anonimi online așa cum se cred în viața offline, dificultățile în identificarea unui făptuitor online poate face victimele să se simtă neputincioase, în raport cu făptuitorul (sau agresorul lor). Acest anonimat lărgește în mod semnificativ grupul de potențiali autori ai hărțuirii cibernetice, în comparație cu hărțuirea tradițională. De exemplu, agresorii nu sunt nevoiți să se îngrijoreze dacă statura lor fizică este mai mică decât cea victimei. Anonimatul are și un alt efect advers. În cazul agresiunii față în față (tradiționale), agresorii pot observa impactul comportamentului lor asupra victimei. Pentru unii agresori, recunoașterea faptului că și-au rănit victima este suficient pentru a descuraja comportamentul de agresiune, pe viitor. Odată cu apariția cyberbullying-ului, nu există o modalitate directă pentru ca făptuitorul să știe și să perceapă efectul comportamentului său asupra victimei. Astfel, șansele de empatie și remușcări sunt semnificativ reduse.

[133] Walrave, M., Demoulin, M., Heirman, W. and A. van der Perre, *op. cit*., p. 17
[134] Idem

În plus, în timp ce cele mai multe agresiuni tradiționale apar la școală, în timpul zilei școlare, ori la serviciu, în timpul orelor de program, hărțuirea cibernetică, prin natura sa, poate apărea în orice moment al zilei sau al nopții[135]. De exemplu, mii de oameni pot vedea postări jignitoare online, în timp ce doar câțiva pot vedea un incident de intimidare la școală. Și consecințele punitive legate de hărțuirea tradițională și hărțuirea cibernetică diferă, de asemenea. Persoanele care hărțuiesc cibernetic nu pot vedea efectele imediate ale agresiunii lor asupra victimei. Orice răspuns care poate fi oferit de victimă poate fi întârziat în funcție de momentul în care victima devine conștientă de hărțuirea cibernetică (de exemplu, verifică mesajul text, vizualizează site-ul web). Această problemă de sincronizare sugerează că poate există motive diferite în spatele celor două tipuri de comportament de agresiune. Adică, motivele pentru săvârșirea hărțuirii cibernetice pot fi mai intrapersonale, în timp ce cele pentru hărțuirea tradițională pot fi mai interpersonale[136]. Există însă o situație în care agresorii pot vedea reacția victimei în momentul în care aceasta citește mesajele denigratoare sau vede pozele cu ea. Ne referim la hacking, atunci când un agresor, cu o foarte bună pregătire tehnică, sparge parolele de pe calculatorul victimei, în acest mod căpătând acces la camera web a calculatorului (desktop) sau laptop. Astfel, agresorul de tip stalker vede în timp real reacțiile victimei.

Dar, cum devine o persoană agresor cibernetic? La baza acestui tip de comportament stau factori de ordin intern dar și factori exteriori. Acești factori afectează starea internă actuală a individului, activând poate gânduri ostile, afectare negativă și excitare sporită. Starea internă actuală este, de asemenea, legată de procesele de evaluare și decizie. În esență, teoria alegerii raționale[137] afirmă că comportamentul deviant este rezultatul unei evaluări a costurilor și beneficiilor în care beneficiile sunt mai mari decât costurile. Oamenii iau decizii raționale în jurul problemelor de risc, efort și recompensă. În cazul hărțuirii cibernetice, riscurile par a fi reduse: infractorii pot opera anonim, există o absență a indicilor relaționali și a proximității fizice și pare a exista o lipsă de supraveghere în spațiul cibernetic. Prin urmare, deoarece majoritatea tinerilor au acces la internet, efortul necesar pentru a intimida pe cineva este mic. De exemplu, individul

[135] Bradshaw, C.P., Sawyer, A.L., & O'Brennan, L.M., *Bullying and peer victimization at school: Perceptual differences between students and school staff*, School Psychology Review, 2007, nr. 36(3), p. 361-382.
[136] Robin M. Kowalski, Gary W. Giumetti, Amber N. Schroeder, Micah R. Lattanner, *Bullying in the Digital Age: A Critical Review and Meta-Analysis of Cyberbullying Research Among Youth*, Psychological Bulletin, 2014, Vol. 140, No. 4, p. 1073–1137
[137] Clarke, R.V. and Felson, M. (2008), *Routine Activity and Rational Choice: Advances in Criminological Theory* (Volume 5), New Jersey: Transaction Publishers

poate aprecia o situație în care un răspuns agresiv este adecvat și decide să se angajeze într-o acțiune impulsivă, trimițând rapid un mesaj text urât către o altă persoană. La fel, dacă individul apreciază că situația nu cere o acțiune imediată, impulsivă, el sau ea poate decide că o acțiune mai atentă ulterioară este adecvată și poate decide astfel să creeze o pagină web spre exemplu, cu impact direct negativ asupra altei persoane (persoana respectivă să fie ridiculizată în diverse moduri). Acest comportament de cyberbullying poate alimenta o personalitate și așa agresivă a făptuitorului, care, sub umbrela anonimatului, poate determina sau provoca și alte persoane să aibă acel tip de comportament fie cu privire la aceeași victimă fie cu privire la alte victime. Agresorii adolescenți, atunci când se maturizează nu sunt atât de dezvoltați social și emoțional în comparație cu colegii lor. Sunt obișnuiți să-i degradeze pe alții pentru propria lor satisfacție și au nevoie de ajutor de la profesioniști pentru a-și schimba interacțiunile negative în altele mai pozitive[138]. Ciberspațiul oferă, de asemenea, agresorului un „public infinit". Astfel, fără mari eforturi sau riscuri, agresorul cibernetic este capabil să-și rănească în mod intenționat și în mod repetat victima, confirmând dezechilibrul de putere. Așadar, agresiunea cibernetică este o alegere rațională. La fel stau lucrurile și în ceea ce privește victima acestui tip de comportament. La baza statutului de victimă pot sta mulți factori care pot predispune o persoană să devină victima acestui tip de agresiune.

În cadrul cercetărilor empirice[139], s-a constatat că ratele de prevalență ale hărțuirii cibernetice sunt foarte *variabile,* în mare parte, datorate modului în care este definită hărțuirea cibernetică, parametrul de timp utilizat pentru a determina când hărțuirea cibernetică a avut loc (de exemplu, ultimele două luni; șase luni; ani), criteriul conservator versus liberal folosit pentru a determina dacă a avut loc hărțuirea cibernetică (de exemplu, cel puțin o dată; de două până la trei ori pe lună sau mai mult), precum și caracteristicile demografice ale eșantionului investigat (de ex. vârstă, sex, rasă).

Cu ce este diferită și/sau poate mai periculoasă hărțuirea de tip cyberbullying? Un răspuns ar fi că temerile, consecințele negative sunt resimțite de victimă pe un termen mai lung și chiar și atunci după ce făptuitorul a fost prins, victima unei asemenea agresiuni/hărțuiri adâncindu-se și

[138] Berger, K. (2007), *Update on school bullying at school: Science forgotten*? Development Review, nr. 27, pp. 90-126

[139] Brochado, S., Soares, S., & Fraga, S., *A scoping review on studies of cyberbullying prevalence among adolescents,* Trauma,Violence & Abuse, 2016, nr.18, p. 523-531.

mai mult în depresie, cu un respect de sine aflat în scădere. Cercetările confirmă că atât victimele agresorilor, cât și agresorii sunt afectați emoțional de actul de intimidare cibernetică[140].

Un adevăr care devine din ce în ce mai ubicuu[141] este faptul că tinerii dobândesc un anumit tip de putere (empowerment) prin noile tehnologii şi noile abilități tehnologice-sociale asociate, iar relația acestora cu comportamentul agresiv necesită investigaţii amănunţite.

Forme de intimidare/hărțuire/agresiune în mediul on line. Adolescenții (și nu numai) folosesc aceste tehnologii pentru a-și intimida colegii, prietenii sau alte persoane, în moduri neintenționate de proiectanții acestor dispozitive. Metodele sunt multiple și includ trimiterea de mesaje text de hărțuire, trimiterea de e-mail-uri fie de intimidare fie către mai multe persoane, în cazul în care e-maiul-ul era confidențial etc. Desigur, pe lângă metodele ”tradiționale” de cyberbullying (fotografierea de ex. a unei persoane și postarea pozei pe internet, pentru a stârni reacții negative sau a ridiculiza respectiva persoană) trebuie să menționăm o nouă formă de hărțuire sau agresiune în mediul on line, cea făcută pe rețelele de socializare, pe site-urile de dating sau în chat-room. Toate metodele de cyberbullying sunt în egală măsură nocive și pot avea consecințe dezastruoase asupra persoanei hărțuite (niveluri crescute de depresie, anxietate și simptome psihosomatice), putând duce inclusiv la suicid.

Au fost identificate următoarele moduri în care comportamentul de cyberbullying se poate manifesta[142]: (a) *flaming* (agresiune verbală) – implică trimiterea de mesaje vulgare, furioase sau ofesatoare unei persoane sau unui grup online; b) *harassment (*hărţuire) – implică trimiterea repetată de mesaje ofensatoare unei persoane; (c) *denigration* (denigrare) –trimiterea sau postarea de afirmaţii răutăcioase, neadevărate sau dăunătoare despre o persoană; (d) *cyberstalking* (urmărire online) – comportament de hărţuire ce include ameninţări sau este extrem de intimidant; (e) *masquerading* (mascaradă) – a pretinde că eşti altcineva şi a posta materiale care denigrează

[140] Ericson, N., U.S. *Department of Justice, Office of Juvenile Justice and Delinquency Program.* (2001). *Addressing the problem of juvenile bullying (FS-200127)* https://www.ncjrs.gov/pdffiles1/ojjdp/fs200127.pdf, p. 1-2

[141] Monica Barbovschi, *A fi social şi nesocial online. Explorarea comportamentului agresiv online al copiilor și adolescenţilor (cyber-bullying*), Acest articol este rezultatul proiectului de cercetare Riscuri și efecte ale utilizării Internetului în rândul copiilor și adolescenților. Perspectiva evoluției spre societatea cunoașterii (knowledge society) finanțat de Ministerul Educației, grant CNCSIS tip A (nr. 1494/2007); echipa de cercetare coordonată de Prof. Dr. Maria Roth, Universitatea Babeș-Bolyai, Cluj-Napoca, p. 5

[142] Willard, N. (2005), *Cyberbullying and Cyberthreats*, OSDFS National Conference. Available from http://csriu.org/cyberbullying.pdf; Monica Barbovschi, *A fi social şi nesocial online. Explorarea comportamentului agresiv online al copiilor și adolescenţilor (cyber-bullying*), Acest articol este rezultatul proiectului de cercetare Riscuri și efecte ale utilizării Internetului în rândul copiilor și adolescenților. Perspectiva evoluției spre societatea cunoașterii (knowledge society) finanțat de Ministerul Educației, grant CNCSIS tip A (nr. 1494/2007); echipa de cercetare coordonată de Prof. Dr. Maria Roth, Universitatea Babeș-Bolyai, Cluj-Napoca, p. 2

persoana respectivă sau o plasează într-o situaţie într-un potenţial pericol; (f) *outing şi trickery* (păcălire şi dezvăluire) – implică folosirea unor trucuri pentru a solicita informaţii compromiţătoare despre o persoană şi publicarea respectivelor informaţii; şi (g) *exclusion* (excludere) – descrie acţiuni menite a exclude în mod intenţionat o persoană dintr-un grup online (a cere colegilor să blocheze/excludă pe cineva din listele de Yahoo Messenger); h) *sexting* (adică distribuirea de imagini nud ale altei persoane fără acordul acelei persoane), j) *trolling* (întărâtarea) – reprezintă insultarea violentă a unei persoane, în mediul online, atât de tare încât să provoace victima să răspundă. De obicei, aceste atacuri instigă mânia victimei, făcând-o să se piardă cu firea și să răspundă într-un mod negativ; k) *comments* (comentarii) – presupune postarea de răspunsuri negative și denigrante la adresa unor fotografii, clipuri video sau mesaje, postate de o anumită persoană în mediul online; l) *fraping* (sabotarea) – presupune modificarea neautorizată a informațiilor de pe pagina personală a unei persoane, într-un context offline, atunci când victima își lasă telefonul/computerul deblocat sau în condițiile în care nu își restricționează accesul la dispozitive. Sabotarea include și comunicarea din numele victimei cu persoane cunoscute/necunoscute, pentru a-i denigra reputația; m) *fake profiles* (profiluri false) – sunt profilurile create de către agresorii cibernetici ce își însușesc identitatea altor persoane, cu scopul de a facilita comunicarea cu victimele lor. Cele mai multe forme arată că agresorii cibernetici preferă ”*agresiunea publică*”. Aceștia nu s-ar mulțumi să ducă convorbiri unilaterale cu o singură persoană din cel puțin două motive: 1) pentru că riscă să dea peste un agresor asemeni lui și să nu facă față unei asemenea provocări, și 2) într-o discuție privată, victima poate oricând să-l blocheze, ceea ce s-ar solda cu un eșec total al intențiilor pe care acesta/aceasta le-a avut inițial. În plus, e mult mai ușor de agresat pe cineva atunci când te susțin un număr de „n-zeci” de persoane.

Deși comportamentele de intimidare/hărțuire cibernetică și comportamentele tradiționale de intimidare/hărțuire sunt în mare măsură similare, din cel puțin două aspecte variabilele de intimidare/hărțuire cibernetică par să se comporte diferit față de variabilele tradiționale de intimidare. Corelația dintre victimizarea cibernetică și hărțuirea cibernetică este de obicei mai strânsă decât corelația dintre variabilele corespunzătoare pentru diferite forme traditionale de agresiune/hărțuire[143]. Cyberbullying-ul are o varietate de atribute care pot accentua impactul comportamentului de intimidare: de exemplu, potențialul de a include un public mai larg; anonimat; natura mai durabilă a cuvântului scris; și capacitatea de a atinge ținta în orice moment

[143] Idem

și în orice loc, inclusiv mediile sigure considerate anterior, cum ar fi casa persoanei hărțuite. În plus, cyberbullies pot fi mai îndrăzneți, deoarece ei nu își pot vedea victimele și cred că anonimatul le protejează de detectare.

Așadar, atâta timp cât există internet și persoane care nu sunt de acord cu părerile sau modul nostru de a fi, suntem permanent expuși riscului de a fi victima cyber-bullying-ului. Internetul a transformat modul în care interacționăm cu semenii noștri și a deschis căi de comunicare agresivă și de hărțuire în mediul virtual, ceea ce a generat preocupări legate de siguranța acestui mediu. Garantarea unui mediu virtual sigur necesită o înțelegere aprofundată a riscurilor cu care se confruntă oamenii în regim online, precum și a soluțiilor posibile pentru reducerea acestora.

O linie de cercetare interesantă explorează modalitățile de a integra fenomenul cyberbullying-ului cu noile tehnologii big data.

Învățarea Automată și Cyberbullying-ul. Conform proiectului *Pew Internet and American Life,* 80% dintre adolescenții care utilizează internetul sunt utilizatori ai rețelelor sociale. Numărul mare de adolescenți online a condus la o creștere a riscurilor pentru experiențe negative, nesănătoase și periculoase în mediul online. Literatura de specialitate descrie diverse intervenții pentru prevenirea cyberbullying-ului. Totuși, activarea programelor de prevenție secundară rămâne o problemă, deoarece majoritatea victimelor și martorilor nu raportează episoadele de cyberbullying către adulți.

În acest context, algoritmii de învățare automată (ML) ar putea fi utilizați pentru a detecta precoce și automat episoadele de cyberbullying, contribuind astfel la activarea în timp util a protocoalelor de intervenție. În ultimul deceniu, cercetătorii au testat o varietate de tehnici ML, cum ar fi analiza informată de sentimentul victimelor, analiza textuală și semantică, precum și analiza caracteristicilor utilizatorilor. Aceste metode au permis detectarea unor diverse rezultate ale cyberbullying-ului, inclusiv: Clasificări binare (de exemplu, implicat sau neimplicat); Identificarea rolurilor în dinamica cyberbullying-ului; Severitatea consecințelor.

În ciuda oportunităților oferite de acești algoritmi în domeniul prevenirii și intervenției, științele sociale și instituțiile educaționale par să neglijeze contribuția potențială a acestor tehnici.

Cele mai comune caracteristici utilizate în modelele predictive de detectare a cyberbullying-ului sunt caracteristicile bazate pe conținut. *O abordare tipică bazată pe conținut constă într-o analiză a valenței cuvintelor scrise în postările de pe rețelele sociale.* Cuvintele sunt

extrase din rețelele sociale și utilizate ca variabile de intrare (sau predictori) pentru a prezice cyberbullying-ul. De exemplu, cuvintele injurioase reprezintă un exemplu prototipic de cuvinte care prezic în mod semnificativ cyberbullying-ul.

Al-Garadi[144], Singh și Kaur[145] au arătat că metoda bag-of-words (analiză textuală în care un text este segmentat într-un „sac" de cuvinte) este cea mai frecvent utilizată pentru a analiza valența textelor în detectarea cyberbullying-ului. În această abordare, textele sunt transformate într-un vector al numărului de cuvinte, ceea ce facilitează efectuarea operațiilor matematice.

Aceste abordări bazate pe limbaj natural au permis, de asemenea, extragerea caracteristicilor comportamentale din conversațiile online. Aceste caracteristici includ atât tiparele comportamentale observate ale utilizatorilor (de exemplu, numărul de întrebări), cât și intențiile lor latente (de exemplu, umilirea, grooming-ul etc.). Tiparele comportamentale pot fi deduse prin calcularea unor metrici specifice, precum:

Frecvența termenilor (Term Frequency - TF): indică frecvența cu care un cuvânt apare într-un document, raportată la numărul total de cuvinte din text.

Frecvența inversă a documentului (Inverse Document Frequency - IDF): calculează importanța unui cuvânt într-un grup de texte. Mai exact, dacă „A" reprezintă numărul total de documente, iar „B" este numărul de documente în care apare un cuvânt „W", IDF este logaritmul raportului dintre „A" și „B".

Utilizarea acestor tehnici computaționale au evidențiat temele postărilor asociate cu cyberbullying-ul.

Conform lui Can și Atalas[146] , cyberbullying-ul are loc în principal atunci când textele de pe rețelele sociale includ teme specifice de conversație, cum ar fi moartea, aparența, religia și conținutul sexual. Acest aspect poate fi amplificat de efectul de dezinhibiție oferit de mediul online, care îi determină pe utilizatori să exprime mai multă violență decât ar face-o în interacțiunile față în față. Majoritatea autorilor s-a concentrat pe detectarea cyberbullying-ului din date textuale, mai degrabă decât din imagini. Această disproporție ar putea fi datorată unei limitări

[144] M.A. Al-Garadi, M.R. Hussain, N. Khan, G. Murtaza, H.F. Nweke, G. Ali Mujtaba, H. Chiroma, H. Ali Khattaki, and A. Gani, *Predicting cyberbullying on social media in the big data era using machine learning algorithms: review of literature and open challenges*, IEEE Access 7, 2019, pp 70701–70718.

[145] A. Singh, and M. Kaur, *Content-based cybercrime detection: A concise review.* International Journal of Innovative Technology and Exploring Engineering 8, no. 8, 2019, pp. 1193–1207. https://doi.org/ 10.1007/s11227-019-03113-z

[146] U. Can, and B. Alatas, *A new direction in social network analysis: Online social network analysis problems and applications*, Physica A: Statistical Mechanics and Its Applications 535, 2019, 1-38. https://doi.org/10.1016/j.physa.2019.122372

a datelor textuale comparativ cu imagini sau videoclipuri în ceea ce privește disponibilitatea datelor stocate. Această problemă devine și mai relevantă având în vedere dovezile privind existența a două tipare distincte de cybervictimizare (scrisă și vizuală). Prin urmare, trebuie acordată o atenție deosebită modului în care aceste două forme distincte de cybervictimizare sunt transmise diferit pe site-urile de social media.

De exemplu, atacurile de cyberbullying prin imagini și videoclipuri ar putea fi cel mai frecvent întâlnite pe platformele TikTok și Instagram, deoarece aceste site-uri sunt în principal vizuale. Pe de altă parte, cyberbullying-ul text-based ar putea fi mai comun pe Twitter și Whatsapp, având în vedere că majoritatea conținutului de pe aceste platforme este scris. Având în vedere că utilizatorii adolescenți ai internetului percep cyberbullying-ul vizual ca fiind mai periculos decât cel textual, aplicațiile de învățare automată par să nu fi întâmpinat încă urgența de a investiga cybervictimizarea prin imagini și videoclipuri.

Unii autori au luat în considerare caracteristicile bazate pe profil, cum ar fi informațiile din profilul utilizatorului (de exemplu, vârsta și genul) și informațiile de pe rețelele sociale (de exemplu, numărul de like-uri și de urmăritori). În ceea ce privește primele, dovezile din studiile psihosociale subliniază necesitatea includerii informațiilor personale, precum genul. Studiile anterioare au demonstrat că utilizatorii de sex masculin și feminin acționează diferit atunci când sunt implicați în dinamica cyberbullying-ului. De exemplu, utilizatorii de sex feminin tind să adopte stiluri de comunicare agresive (de exemplu, excluderea altor utilizatori dintr-un grup sau conspirațiile împotriva lor), în timp ce bărbații tind să folosească cuvinte mai amenințătoare. Pe de altă parte, așa cum au subliniat Can și Atalas (2019), informațiile de pe rețelele sociale au oferit indicii interesante pentru o mai bună înțelegere a comportamentelor utilizatorilor online asociate cu cyberbullying-ul.

Combinarea cyberbullying-ului cu tehnologiile big data implică utilizarea unor abordări bazate pe date pentru a aborda și a diminua acest fenomen. În acest sens, unele sarcini relevante, organizate clar în subtasks, care pot fi incluse într-o metodologie generală, sunt următoarele:

(i) Curățarea și Anotarea Seturilor de Date:

(i.a) Colectarea datelor relevante de pe platforme precum YouTube, rețele sociale sau forumuri de discuții unde apare fenomenul de cyberbullying.

(i.b) Anotarea datelor cu ajutorul participanților umani pentru a eticheta cazurile de cyberbullying.

(i.c) Abordarea provocărilor precum prejudecățile annotatorilor, barierele lingvistice și subiectivitatea în procesul de etichetare.

(ii) Modele de Învățare Automată:

(ii.a) Utilizarea algoritmilor de învățare automată pentru a identifica și filtra comentariile neambigue din setul de date adnotat.

(ii.b) Aplicarea metodelor de filtrare prin consens pentru a clasifica comentariile ca fiind neambigue, pe baza acordului dintre annotatori umani și etichetele generate de algoritmi.

(ii.c) Crearea de seturi de date noi, conținând comentarii neambigue, pentru instruirea modelelor.

(ii.d) Utilizarea rețelelor neuronale artificiale pentru a îmbunătăți performanța în identificarea cazurilor de cyberbullying.

(iii) Tehnici de Învățare Profundă:

(iii.a) Explorarea abordărilor bazate pe învățarea profundă, precum rețelele neuronale recurente (RNN) și rețelele neuronale convoluționale (CNN), pentru detectarea cyberbullying-ului.

(iii.b) Aplicarea acestor tehnici pe platformele de social media.

(iv) Implicații Practice:

Luarea în considerare a implicațiilor practice ale abordărilor bazate pe învățarea automată pentru familii, clinicieni și eforturile de prevenire/intervenție împotriva cyberbullying-ului.

Aspecte de cercetare viitoare:

Modele Hibride de Învățare Profundă: Investigarea modelelor hibride care combină diferite arhitecturi de învățare profundă pentru o detectare îmbunătățită a cyberbullying-ului. Explorarea tehnicilor care pot gestiona cyberbullying-ul în diverse modalități, precum vorbire, videoclipuri sau deepfakes.

Abordări Multimodale: Extinderea cercetării dincolo de text și imagini pentru a include alte forme de comunicare (de exemplu, audio, video). Dezvoltarea modelelor care pot detecta cyberbullying-ul pe diferite platforme și tipuri de media.

Augmentarea și Preprocesarea Datelor: Îmbunătățirea tehnicilor de preprocesare a datelor pentru a gestiona datele zgomotoase și nestructurate. Explorarea metodelor avansate de preprocesare pentru creșterea performanței modelelor.

Atacuri Adversariale și Robustețe: Investigarea atacurilor adversariale asupra modelelor de detectare a cyberbullying-ului. Dezvoltarea unor modele robuste, capabile să reziste manipulării adversariale.

Considerații Etice: Abordarea preocupărilor etice legate de confidențialitate, părtinire și echitate în detectarea cyberbullying-ului. Asigurarea că sistemele de detectare nu dăunează involuntar utilizatorilor nevinovați.

BIBLIOGRAFIE

Waseem, Akram, and R. Kumar. (2017) "*A Study on Positive and Negative Effects of Social Media on Society."* International Journal of Computer Sciences and Engineering, 5(10): 351–354.

Hugo, Rosa, N.S. Pereira, R.D. Ribeiro, P. da Costa Ferreira, J.P. Carvalho, S. Oliveira, L. Coheur, A.M. Veiga Simão, and I.M. Trancoso. (2019) "*Automatic Cyberbullying Detection: A Systematic Review*." Computers in Human Behavior, 93: 333–345.

Nandhini, Sri B., and J.I. Sheeba. (2015) *"Online Social Network Bullying Detection Using Intelligence Techniques.*" Procedia Computer Science, 45: 485–492.

Kumari, Kirti, and J.P. Singh. (2020) *"AI ML NIT Patna@ TRAC-2: Deep Learning Approach for Multi-Lingual Aggression Identification*." In: Proceedings of the 2nd Workshop on Trolling, Aggression and Cyberbullying, pp. 113–119.

Hochreiter, Sepp, and J. Schmidhuber. (1997) "*Long Short-Term Memory*." Neural Computation, 9(8): 1735–1780.

Graves, Alex. (2012) "*Long Short-Term Memory*." In: Supervised Sequence Labelling with Recurrent Neural Networks, pp. 37–45.

Elsayed, Mahmoud S., N.-A. Le-Khac, S. Dev, and A.D. Jurcut. (2020) "*Network Anomaly Detection Using LSTM Based Autoencoder*." In: Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks, pp. 37–45.

Church, Kennet W. (2017) "Word2Vec." *Natural Language Engineering*, 23(1): 155–162.

Devlin, Jacob, M.-W. Chang, K. Lee, and K. Toutanova. (2018) "*BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding."* arXiv preprint, arXiv:1810.04805.

Lagler, Klemens, M. Schindelegger, J. Böhm, H. Krásná, and T. Nilsson. (2013) "*GPT2: Empirical Slant Delay Model for Radio Space Geodetic Techniques*." Geophysical Research Letters, 40(6): 1069–1073.

Mandic, Danilo P., and J.A. Chambers. (2001) "*Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*."Wiley.

Simon, Hyellamada, B.Y. Baha, and E.J. Garba. (2022) *"Trends in Machine Learning on Automatic Detection of Hate Speech on Social Media Platforms: A Systematic Review.*" FUW Trends in Science & Technology Journal, 7(1): 001–016.

Castaño-Pulgarín, S. Andrés, N. Suárez-Betancur, L.M. Tilano Vega, and H.M. Herrera López. (2021) "*Internet, Social Media and Online Hate Speech. Systematic Review.*" Aggression and Violent Behavior, 58: 101608.

Ghosh Roy, Sayar, U. Narayan, T. Raha, Z. Abid, and V. Varma. (2021) "*Leveraging Multilingual Transformers for Hate Speech Detection*." arXiv preprint, arXiv:2101.03207.

Sadiq, Saima, A. Mehmood, S. Ullah, M. Ahmad, G. Sang Choi, and B.-W. On. (2021) "*Aggression Detection Through Deep Neural Model on Twitter*." Future Generation Computer Systems, 11: 120–129.

Kumar, Akshi, and N. Sachdeva. (2022) "*A Bi-GRU with Attention and Capsnet Hybrid Model for Cyberbullying Detection on Social Media.*" World Wide Web, 25(4): 1537–1550.

Alam, K. Saeed, S. Bhowmik, and P.R. Kumar, Ritesh, A. Kr. Ojha, S. Malmasi, and M. Zampieri. (2018) "*Benchmarking Aggression Identification in Social Media*." In: Proceedings of the 1st Workshop on Trolling, Aggression and Cyberbullying, pp. 1–11.

Herremans, Dorien, and C.-H. Chuan. (2017) "*Modeling Musical Context with Word2vec.*" arXiv Preprint, arXiv:1706.09088.

Mut Altın, Lütfiye Seda, A. Bravo, and H. Saggion. (2020) "*LaSTUS/TALN at TRAC-2020 Trolling, Aggression and Cyberbullying*." In: Proceedings of the 2nd Workshop on Trolling, Aggression and Cyberbullying, pp. 83–86.

S.F. Chan, A.M. La Greca, and J.L. Peugh, "*Cyber victimization, cyber aggression, and adolescent alcohol use: Short-term prospective and reciprocal associations*." Journal of adolescence 74, 2019, pp. 13-23. https://doi.org/10.1016/j.adolescence.2019.05.003.

R. Del Rey, R. Ortega-Ruiz, and J.A. Casas, "*Asegúrate: An intervention program against cyberbullying based on teachers' commitment and on design of its instructional materials."* International journal of environmental research and public health 16, no. 3, 2019, pp. 434. https://doi.org/10.3390/ijerph16030434.

Ch. Derave, B. Frydman, N. Genicot (dir.), *L'intelligence artificielle face a l'Etat de droit*, Editions Bruylant, Bruxelles, 2024, p. 15.

J. Gerards, R. Xenidis, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law*, Raport special al reţelei europene de experți în egalitate de gen şi nediscriminare (Comisia Europeană), 2021

J. Senechal, *L'AI Act dans sa version finale - provisoire -, une hydre a trois tetes*, Dalloz Actualite, 11 mars 2024.

Ugo Pagallo, Vital, Sophia, and Co. – *The Quest for the Legal Personhood of Robots*, Information 9, no. 9 (September 2018), p. 3, https://doi.org/10.3390/info9090

Legal Person, Meaning of Legal Person by Lexico, Lexico Dictionaries – English, https://www.lexico.com/definition/legal_person

Chopra/White, Artificial Agents – *Personhood in Law and Philosophy*;

Steffen Wettig/Eberhard Zehendner, *The Electronic Agent: A Legal Personality under German Law?*, 2003;

Pagallo, *Killers, Fridges, and Slaves*, November 2011, AI & SOCIETY 26(4):347-354, DOI: 10.1007/s00146-010-0316-0;

Karolin Kemper, *Rechtspersönlichkeit für Künstliche Intelligenz*? URL: cognitio-zeitschrift.ch/2018-1/Kemper, DOI: https://doi.org/10.5281/zenodo.1928171 ISSN: 2624-8417

John Chipman Gray, *The Nature and Sources of the Law*, Ed. Peter Smith, Gloucester, 1972, (MacMillan 1921), p. 27 și urm.

Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, North Carolina Law Review vol. 70 nr. 4

J. Gobin, *L'individu, fin de parcours? Le piege de l'intelligence artificielle*, Collection Le Debat, Editions Gallimard, Paris, 2024

Teubner, G. *Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law*, Journal of Law and Society, vol. 33 (4), 2006, p 508.

Calo, R. *Artificial Intelligence Policy: A Primer and Roadmap*, vol. 51 U.C.D. L. Rev. 399, 2017

Edwina L. Rissland, *Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning.* Yale Law Journal vol. 99, no. 8, 1990

Griffin, A. *Saudi Arabia grants citizenship to a robot for the first time ever*. The Independent, 2017. https://www.independent.co.uk/life-style/gadgets-and-tech/news/saudi-arabia-robot-sophiacitizenship-android-riyadh-citizen-passport-future-a8021601.html

Grimmelmann, J. *There's No Such Thing as a Computer-Authored Work – And It's a Good Thing, Too*, 39 Columbia Journal of Law & the Arts, 2016

Kasper, A (2014). *The Fragmented Securitization of Cyber Threats*, In: Kerikmäe (Ed.) Regulating eTechnologies in the European Union – Normative Realities and Trends, Springer, p 158

Kerikmäe, T.; Hoffmann, T.; Chochia, A. (2018). *Legal Technology for Law Firms*: *Determining Roadmaps for Innovation*. Croatian International Relations Review, 24 (81), 91−112

Kurki, V. *Revisiting legal personhood: Paper for Spanish-Finnish Seminar* in Legal Theory, 2016

Larson, D. *Artificial Intelligence: Robots, Avatars, and the Demise of the Human Mediator*, 25 Ohio St. J. on Disp. Resol. 105, 2010

Massaro, M., Norton, H. *Siri-Ously? Free Speech Rights and Artificial Intelligence*, 110 Nw. U. L. Rev. 1169, 2016

Naucius, M. *Should Fully Autonomous Artificial Intelligence Systems Be Granted Legal Capacity,* 17 Teises Apzvalga L. Rev. 113, 2018

Pan, Y. *Heading towards Artificial Intelligence* 2.0. Engineering 2, 2016

Russel, S., Norvig, P. *Artificial Intelligence: A Modern Approach*. Prentice Hall 1995

Semmler, S., Rose, Z. *Artificial Intelligence: Application Today and Implications Tomorrow,* 16 Duke L. & Tech. Rev. 85 2017-2018

Sepp, Paula-Mai; Vedeshin, Anton; Dutt, Pawn (2016). *Intellectual property protection of 3D printing using secured streaming*. In: Kerikmäe, T.; Rull, A. (Ed.). The Future of Law and eTechnologies (81−109). Cham: Springer

Sobel, B. *Artificial Intelligence's Fair Use Crisis*, 41 Colum. J.L. & Arts 45, 2017

Teubner, G. *Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law,* Journal of Law and Society, vol. 33 (4), 2006

Vladeck, D. *Machines without principals: liability rules and artificial intelligence*, Washington law review vol. 89, 2014, pp 122-123

B.C. Trandafirescu, *Raportul juridic în mediul digital – o abordare din perspectiva teoriei generale a dreptului* (21.12.2021);

G. Manu, *Impactul digitalizării asupra acțiunii administrative și dreptului administrative*;

R. Matefi, *Impactul Inteligenței Artificiale asupra drepturilor la egalitate și nediscriminare în lumina Propunerii 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

F. Kort, *Predicting Suprême Court Decisions Mathematically: a Quantitative Analysis of the "Rightto Counse*l" Cases, American Political Science Review 1957, n° 5.;

J. A. Segal, *Predicting Supreme Court Cases Probabilistically: the Search and Seizure* Cases (1962-1981), American Political Science Review 1984, p. 891 s.;

S. S. Nagel, *Applying correlation analysis to case prediction*, Texas Law Review 1963, p. 1006 s., citați de B. Barraud, *Un algorithme capable de prédire les décisions des juges: vers une robotisation de la justice?, Les cahiers de la justice* 2017

R.C. Lawlor, *What computers can do: analysis and prediction of judicial decisions*, American Bar Association Journal, 1963, 49, citat de B. Dondero, *Justice prédictive: la fin de l'aléa judiciaire?*, Recueil Dalloz 2017

A. Aft, J. Blackman, C. M. Carpenter, *FantasySCOTUS: Crowdsourcing a Prediction Market for the Supreme Court*, Northwestern Journal of Technology & Intellectual Property 2012, n° 10

J. Blackman, M. J. Bommarito, D. M. Katz, *Predicting the Behavior of the Supreme Court of the United States: A General Approach,* SSRN Electronic Journal, 21.07.2014.

S. Chassagnard-Pinet, *Les usages des algorithmes en droit: prédire ou dire le droit*?, Dalloz IP/IT 2017

A. Garapon, J. *Lassegue, Justice digitale, accepteriez-vous d'être jugés par des algorithmes?*, PUF, 2018.

A. Pembellot, *Justice prédictive, solution ou simple reproduction du passé*?, pe https://www.cyberjustice.ca, 2019.

F. Pasquale, Black Box Society. *Les algorithmes secrets qui contrôlent l'économie et l'information,* Fyp Edition, citat de A. Pembellot pe https://www.cyberjustice.ca/en/2019/07/18/justice-predictive-solution-ou-simple-reproduction-du-passe/ .

B. Barraud, *La croisée des savoirs – Un algorithme capable de prédire les décisions des juges: vers une robotisation de la justice*?, Les cahiers de la justice 2017

C. O'Neil, *Weapons of Math Destruction*, Crown, 2016 – https://www.theguardian.com/books/2016/oct/27/cathy-oneil-weapons-of-math-destruction-algorithms-big-data.

Rapp. Villani, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, Doc. fr., mars 2018, disponibil pe https://www.enseignementsup-recherche.gouv.fr/cid128577/rapport-de-cedric-villani-donner-un-sens-a-l-intelligence-artificielle-ia.html.

J. Lassègue, *L'Intelligence artificielle, technologie de la vision numérique du mond*, Les cahiers de la justice, 2019

J. Atif, J.P. Burgess, I. Ryl, *Geopolitique de l'IA. Les relations internationales a l'ere de la mise en donnees du monde*, Editions Le Cavalier Bleu, Paris, 2022.

Anu Bradford, *Digital Empires - The Global Battle to Regulate Technology*, Oxford University Press, 2023, pp. 68-91.