

Inteligența artificială: Zeu sau mașină?

Înțelegerea și demistificarea inteligenței artificiale: ce este, ce nu este, și ce ar putea fi. Plus impactul asupra drepturilor noastre (1)

Autor: Bogdan-Andrei Lungu

Introducere	1
Ce este, de fapt, inteligența artificială?	2
Ce este inteligent și artificial la inteligența artificială?	3
Dilema majoră a sistemelor de inteligență artificială de azi	6
Ce nu este inteligența artificială?	7
Inteligența artificială și drepturile omului	8
Viziuni alternative la adresa tehnologiei	12

Introducere

Termenul de inteligență artificială (IA), sau acronimul său "AI" (artificial intelligence) din limba engleză, a ajuns un loc comun pentru noi, astăzi. Îl auzim în presă, de la oameni îngrijorați de viitorul posibilei automatizări a muncii lor. Îl auzim de la pasionații de tehnologie, care spun că această tehnologie va schimba omenirea. Și îl mai auzim și de la elevi și studenți, spunând că ChatGPT le face viața mai ușoară (prea ușoară, se tem unii).

Cert este că termenul metaforic de inteligență artificială nu mai este restricționat și accesat doar de specialiști și entuziaști. Cei mai mulți dintre noi am auzit de el, într-o forma sau alta.

Doar că puțini dintre noi sunt conștienți de gradul în care suntem înconjurați de sisteme de inteligență artificială în tehnologiile pe care le folosim zilnic. Le găsim în asistenții noștri vocali (Alexa, Siri, Google Home), în tehnologiile Google (Google Search, Google Translate), în sectorul financiar ([personalizare de produse, știința datelor](#)), în educație ([scriere, comunicare și personalizarea conținutului](#)) și cercetare științifică (analiza de date și pentru modele statistice), medicină ([diagnosticare și analiză](#)), agricultură ([automatizare și eficientizarea producției agricole](#)), entertainment ([crearea de conținut, servicii de recomandare și identificarea trendurilor](#)), în servicii publice ([#Ion](#)), resurse umane ([economisire de timp și resurse](#)), în crearea feed-ului de pe social media ([filtrare conținut relevant pentru utilizatori](#)) și așa mai departe.

Dar ce este acest termen metaforic descris drept inteligență artificială? Ce este inteligent la aceste tehnologii și de ce este această inteligență artificială? Sau măcar putem vorbi astfel despre aceste tehnologii, care nu sunt atât de noi pe cât credem?

În acest articol vom explora ce este inteligența artificială, ce nu este, și ce ar putea fi. În unele articole viitoare o să punem pe tapet mai în detaliu problemele (noi sau vechi) pe care sistemele AI le ridică și riscurile pentru drepturile fundamentale. Facem această serie de articole (și materiale în forme mai simple) pentru a permite oamenilor să aibă un fundament pentru discuții, nu pentru a da noi lecții.

Ce este, de fapt, inteligența artificială?

A defini inteligența artificială este dificil. Termenul în sine este metaforic și nu descriptiv. Iar pe lângă asta mai este și foarte controversat între experți. Fiindcă IA nu este un lucru, ci multiple lucruri. Iar definiția pe care o dai schimbă perspectiva asupra acestei tehnologii - o poate mări, sau micșora, ne poate facilita o viziune amplă sau microscopică a inteligenței artificiale.

Pentru scopurile noastre, însă, am utilizat o definiție care încearcă să fie și precisă și clară pentru inteligența artificială.

Așadar, inteligența artificială este "un sistem bazat pe o mașină care poate, pentru un set dat de obiective definite de om, să facă predicții, recomandări sau decizii care influențează mediile reale sau virtuale".

Aceasta este definiția pe care o puteți găsi și în [legislația din SUA](#) privind reglementarea IA și, în linii mari, în definiția dată de [OECD](#). O definiție similară o să fie și în [Regulamentul IA](#), proaspăt agreat de Uniunea Europeană.

Definiția de mai sus este cea la care facem trimitere când folosim IA în acest articol. Acest termen ambiguu are, totuși, o istorie. Inteligența artificială a fost botezată astfel de un mare informatician pe nume [John McCarthy, în vara anului 1956](#) (SUA), când s-a ținut prima conferință academică despre o astfel de nouă tehnologie la Dartmouth College. McCarthy voia ca acești noi roboți să fie la fel de inteligenți ca omul - să-i oglindească inteligența. Și poate și să o depășească. Entuziasmul era și atunci, ca și azi, la maxim în legătură cu aceste tehnologii.

Originile termenului sunt importante, fiindcă eroarea fundamentală - anume că omul este o mașinărie sau mașinăria poate fi om - a rămas, și încă ne bântuie și [distorsionează discursul](#) despre inteligența artificială (și umană) și astăzi.

Ce este inteligent și artificial la inteligența artificială?

Nu există, în realitate, nimic inteligent și nimic artificial la ce numim noi sisteme de inteligență artificială.

În primul rând, astfel de tehnologii au o realitate fizică, materială. Sistemele IA sunt antrenate, corectate, verificate și reverificate [de oameni reali](#), în mare parte din țări în curs de dezvoltare, care lucrează în condiții crâncene de muncă, [pentru un salariu extrem de mic](#). De altfel, la fel facem și noi, gratuit, când selectăm în [testele CAPTCHA](#) unde sunt semafoarele sau bicicletele din imagine ca să "dovedim că suntem oameni". Mai mult, aceste

sisteme există undeva, pe un hard disk, consumând [foarte multă energie și poluând](#).

În al doilea rând, **inteligența despre care se vorbește este de cele mai multe ori [definită mecanic](#)**, în termeni generali, drept rezolvarea de probleme într-un mod eficient și cu resurse scăzute. Dar și un ciocan face în esență același lucru - ne ajută să batem cuie, eficient, cu resurse scăzute (forță + ciocan + cui). Totuși noi nu numim ciocanul inteligent. Iar această definiție nu se apropie de inteligența umană pe care sistemele IA ar trebui să le oglindească.

Întrucât mașinile devin din ce în ce mai capabile, sarcinile considerate a necesita „inteligență” sunt deseori eliminate din definiția IA, un fenomen cunoscut sub numele de [efectul IA](#). O observație în teorema lui Tesler spune că „IA este ceea ce nu a fost încă făcut”. De exemplu, [recunoașterea optică a caracterelor](#) este adesea exclusă din domeniul IA, după ce a devenit o tehnologie de rutină. Cu cât ne obișnuim mai mult cu unele tehnologii, cu atât "inteligenta" lor - odată fascinantă - devine invizibilă.

În al treilea rând, **sistemele IA nu sunt artificiale**, pentru că folosesc o tonă de resurse: energie electrică, investiții financiare care sprijină cercetarea și antrenarea lor, informații folosite pentru antrenare și oameni care le corectează în mod manual, dar și echipe de profesioniști care le tot îmbunătățesc și monitorizează. Termenul "inteligent" este doar o metaforă proastă pentru un fenomen pe care nu-l înțelegem complet - sau poate mai bine zis “o bună etichetă de marketing”.

De [aceea unii experți](#) preferă termenul de *machine learning* (adică mașinărie care învață), un subdomeniu al cercetării în domeniul sistemelor IA, care funcționează cu ajutorul *rețelelor neurale* (largi rețele statistice, suprapuse, matematice). *Machine learning* înseamnă, în termeni generali, o mașinărie care generalizează un răspuns posibil pe baza unor date de antrenare, printr-un proces inductiv. Rețelele neurale sunt și ele un subdomeniu al *machine learning*, care imită [în structură creierul uman](#). Subdomeniul tehnologiilor *machine learning* este unul extraordinar de dificil. Acesta este o aplicare a inteligenței artificiale, care se bazează pe baze de date gigantice pentru a produce un rezultat, cu ajutorul unui alt proces numit *deep learning* (învățare profundă). Astfel de tehnologii complicate au ajuns în produsele digitale pe care le folosim zilnic. Le găsiți în sistemele de [recomandare de filme și seriale de la Netflix](#), de exemplu.

Termenii de mai sus (*machine learning*, rețele neurale, *deep learning*) arată că aceste tehnologii doresc să imite procesul prin care oamenii învață, gândesc și ajung la anumite rezultate. Doar că acești termeni sunt mai mult metaforici decât descriptivi, moșteniți de-a lungul timpului, din partea unor ciberneticieni și informaticieni foarte entuziaști din secolul trecut, care s-au văzut pe sine în propriile creații.

De exemplu, [Marvin Minsky \(1927-2016\)](#), pionier în domeniul IA, un fel de erou al domeniului, a fost întrebat dacă mașinăriile pot gândi. [Răspunsul lui](#): ”Bineînțeles că mașinile pot gândi; noi putem gândi și suntem ”mașini de carne””. Metafora omului ca "mașină de carne" arată, încă o dată, că pentru a ridica aceste tehnologii IA în slăvi omul ajunge să fie dezumanizat drept o halcă de carne.

Cât de “zeu” este ChatGPT?

Ultimul trend exploziv din domeniul IA a fost cel al [inteligenței artificiale generative](#) (ChatGPT, Bard de la Google, Midjourney, etc). IA generativă se referă la modele de inteligență artificială concepute pentru a genera conținut nou sub formă de text scris, audio, imagini sau videoclipuri la cererea unei persoane. Aplicațiile și cazurile de utilizare sunt foarte largi, de aici și generalitatea termenului. De fapt, termenul de *model statistic computațional larg* descrie mult mai precis fenomenul. Este destul de general încât să cuprindă mai tot ce trece drept sistem IA generativ. Și destul de precis încât să facă trimitere la tehnologii specifice.

Modul în care funcționează astfel de tehnologii trebuie considerat mai puțin demne de titlul de “zeu” și descrise mai precis drept “mașină”. Pe baza datelor de antrenare (training data) și a întrebării/prompt-ului utilizatorului (tot input-ul dat de om), mașinăriile produc un model al realității (output). De exemplu când cerem ChatGPT-ului, - care este un model lingvistic extins (large language model) - , să ne dea răspunsul la o simplă întrebare, acesta asta face - produce un răspuns (output) în funcție de datele de antrenare și întrebare. Are [date de antrenare vaste](#), cum ar fi cărți, articole de pe internet, comentariile utilizatorilor, website-uri, printre altele.

Pe baza a ce găsește, ne dă un răspuns care s-ar putea să ne placă sau nu. E un model probabilistic - adică se bazează pe logica probabilității matematice, simplu spus. Încearcă să ghicească răspunsul corect pe baza a ce știe deja (datele de antrenare). De aia câteodată "[halucinează](#)" spunând tot felul de bazaconii - un model bazat pe șansă o va da în bară, în unele situații.

Din exemplul dat mai sus, putem observa cum nu este nimic inteligent sau artificial la ChatGPT. În schimb, avem o mașinărie foarte bună la sintetizarea informațiilor folosite pentru antrenare (în marea majoritatea preluate de pe internet și din cărți), lucru inițial antrenat inclusiv manual (și automatizat în timp) de către cercetătorii de la compania mamă OpenAI. E un fel de motor de căutare mult mai avansat și mai rapid și care, pe baza a ce-i ceri, îți poate da diverse răspunsuri personalizate.

Dilema majoră a sistemelor de inteligență artificială de azi

Cea mai mare dilemă provocată de sistemele de inteligență artificială din ziua noastră este faptul că cele mai multe dintre ele sunt un fel de *cutii negre* ale căror procese de funcționare și de producție de rezultate nu sunt înțelese, în totalitate, nici de creatorii lor, chit că sunt extraordinar de eficiente și productive. De aici dilema: eficiență pe de o parte, inexplicabil pe de altă parte. O cutie neagră este un concept care se referă la o tehnologie pe care n-o înțelegem complet și a cărui conținut rămâne un mister pentru noi. În acest stadiu se află multe sisteme de inteligență artificială din ziua de azi, în special cele bazate pe "rețele neurale".

Nici creatorii și nici utilizatorii nu înțeleg pe deplin ce fac aceste tehnologii, cum ajung la rezultatele la care ajung, de ce privilegiază un rezultat și nu un altul, și cum își construiesc modelele. Asta se întâmplă fiindcă rețelele neurale de mai sus sunt extraordinar de complexe și neinteligibile unui om (inclusiv expertului sau echipei de experți care le-a făcut).

Iar atunci când nu sunt înțelese nici măcar de creatori, cine ar fi responsabil pentru consecințele nefaste ale deciziilor unor astfel de mașinării? Asta aduce o sumă mare de probleme, despre care vom discuta mai jos în articol.

Ce nu este inteligența artificială? (sau despre *inteligența artificială ca risc existențial*)

În lumea științifico-fantasticului, inteligența artificială este văzută ca un risc la adresa umanității. De exemplu, în filmul de la Marvel, *The Age of Ultron* (2015), un super-robot este inventat de unul din cei mai inteligenți oameni din lume, Tony Stark. Ultron, robotul, învață tot ce este de învățat, având acces la toate bazele de date ale lumii. Prima lui concluzie? Umanitatea trebuie eliminată.

În zilele noastre, datorită progresului tehnologic rapid din domeniul inteligenței artificiale, o astfel de frică, care era rezervată filmelor SF, a ajuns să fie o realitate pentru multe persoane. Progresul tehnologic din domeniul inteligenței artificiale pare de nestăvilit iar unele persoane se tem, pe bună dreptate și într-un mod complet legitim, că aceste sisteme IA s-ar putea să scape de sub control, similar cu cazul altor tehnologii precum rachetele nucleare căzând în mâinile cui nu trebuie. Printre țările în care un număr considerabil de oameni consideră că IA s-ar putea să scape de sub control în [viitor se numără](#): India, SUA, UK, China, Germania, Brazilia, Spania, Japonia, și Canada.

Institutul privind studiul riscurilor [existențiale din UK](#) studiază care sunt cele mai mari provocări la adresa vieții omului pe pământ. E vorba, pe lângă o catastrofă nucleară sau climatică de *inteligența artificială ca [risc existențial la adresa umanității](#)*. Motivele enumerate de cercetători pentru care ar trebui să fim atenți la dezvoltarea IA sunt siguranța oamenilor și securitatea. Aceștia argumentează că, dacă un sistem IA superintelligent (capabil de propriile decizii fără input uman), folosit în scopuri militare, [ajunge în mâinile cui nu trebuie](#), societatea va avea de suferit.

Riscurile identificate de către cercetători pot fi reale. Asta nu înseamnă că panica exagerată este justificată, în prezent. În viitor, dacă astfel de tehnologii vor avansa tot mai mult, o astfel de reacție s-ar putea să fie complet justificată.

Totuși, astăzi, inteligența artificială nu este superintelligentă, conștientă și super puternică. [Unii cercetători](#) se tem de inteligența artificială generală

(*Artificial General Intelligence, AGI*) care ar trebui să fie mai inteligentă decât omul - sau toată omenirea - și să dorească să ne șteargă de pe fața pământului. Le este frică de un fel de Ultron, doar că acest personaj rămâne, pentru moment, izolat în lumea SF-ului.

În lumea tehnologiei și printre specialiști, există o altă mare frică legată de problema alinierii valorilor sistemelor IA cu cele ale oamenilor ([the alignment problem](#)). Pe foarte scurt, oamenilor din industrie le este frică că inteligența artificială ar deveni atât de puternică încât nu și-ar mai alinia valorile înscrise în cod cu cele ale creatorilor lor.

Inteligența artificială, așa cum se întrevide astăzi, nu poate face așa ceva. Iar inteligența artificială generală nu există. Este, la momentul actual, doar un vis tehnologic al unei superinteligențe care ne va depăși din toate punctele de vedere. Dacă va exista sau nu, este alt subiect. Momentan, cu facilitățile tehnologice pe care le avem, și cu cunoștințele din domeniu, o astfel de inteligență artificială rămâne doar un vis.

Și ne petrecem prea mult timp discutând despre ceva cu un risc aproape insignifiant, în loc să studiem mai degrabă problemele reale ale AI, unele din ele deja care sunt deja documentate.

Inteligența artificială și drepturile omului

Ca în cazul oricărei noi tehnologii, sau a unei tehnologii dominante precum inteligența artificială care ne înconjoară la aproape fiecare pas în viața noastră tehnologizată, drepturile omului pot să fie afectate. **În cazul de față, s-a dovedit deja ca sistemele de IA ajung să discrimineze sau să perpetueze prejudecățile programatorilor sau cele din datele de antrenare.**

De exemplu, în 2015, una din cele mai mari companii de e-commerce din lume, [Amazon, a testat un nou sistem de angajare automatizat](#), care funcționa cu ajutorul inteligenței artificiale (machine learning). Sistemul avea o problemă ușoară de rezolvat, trebuia să aleagă cel mai bun candidat pentru un post la Amazon. Sistemul scana CV-urile aplicanților, în căutarea celui mai potrivit candidat pentru post, conform unor cerințe interne stabilite de Amazon. Din 100 de candidați, sistemul trebuia să vină înapoi cu

cei mai buni 5 oameni. Problema cu acest sistem a fost identificată imediat odată ce oamenii din HR de la Amazon au observat o lipsă evidentă la participanții selectați: lipsa femeilor.

Sistemul de inteligență artificială operat de Amazon pentru angajare a discriminat puternic în [defavoarea femeilor](#). Aproape toți posibili noi angajați au fost bărbați. Când au explorat de ce s-a întâmplat asta, răspunsul a fost unul cât se poate de previzibil: sistemul IA a învățat care candidați merită să fie aleși pe baza tiparelor anterioare de aplicare pentru astfel de posturi de muncă. [A aflat că bărbații sunt cei care aplică](#) cel mai des și ajung să fie angajați în domeniul ingineriei software. Așa că “a învățat” că bărbații sunt pur și simplu angajați mai potriviți prin prisma faptului că sunt bărbați și pentru că au fost acceptat în trecut.

Acest sistem a perpetuat o problemă a discriminării pe baze de sex și gen. A demonstrat că sistemele IA perpetuează ceea ce știu din datele lor de antrenare. Amazon a încercat să soluționeze o problemă nu tehnică, ci cu implicații de natură socială și politică prin tehnologie de tip IA. Nu a reușit să facă asta.

Pe lângă acest caz particular, se poate identifica o întreagă **taxonomie**, care nu este exhaustivă, cu problemele ridicate la adresa drepturilor noastre fundamentale de noile sisteme de inteligență artificială folosite astăzi:

- **Limitarea libertății de exprimare:** sisteme de inteligență artificială care filtrează și, eventual, cenzurează conținutul definit în mod netransparent ca fiind ilegal sau contrar termenilor de utilizare sau nepotrivit în anumite servicii online, de obicei rețele sociale. Exemplu: 41 de guverne au blocat în 2023 prin sisteme IA conținut politic, social sau religios, care ar fi fost în limitele libertății de exprimare conform normelor internaționale. ([vezi Raport Freedom of the Net 2023](#)).
- **Stereotipizare:** perpetuarea unor prejucții și stereotipuri despre anumite categorii de persoane și catalogarea oamenilor în mod automat pe baza unui deficit de date și/sau cunoștințe despre individ pentru livrarea de reclame personalizate. Exemplu: aplicația text-imagie Stable Diffusion [reprezintă bărbații de culoare ca fiind criminali, în timp ce bărbații albi sunt directori executivi](#) perpetuând, astfel, stereotipuri rasiale și de clasă.

- **Dezumanizare:** reducerea omului la date de antrenare pentru hrănirea sistemelor de inteligență artificială; transformarea unei individualității umane în date, reprezentate numeric (data points), și confundarea unui om cu datele pe care le lasă pe internet.
- **Discriminare:** privilegierea unor anumite categorii de persoane pe baza datelor de antrenare (vezi de ex. cazul Amazon de mai sus); ignorarea completă a altor categorii de oameni; discriminarea activă (dezinformare, propagandă) împotriva unor categorii de oameni
- **Dezinformare, propagandă:** Sisteme de IA folosite pentru a [dezinforma în mod activ](#) Exemplu: tehnologii de tip ChatGPT, Mid Journey, Stable Diffusion folosite pentru a genera informații și imagini false. Propagandă generată în mod automatizat. Exemplu: conflictul Israel - Palestina, [o parte din conținutul](#) care circula pe social media despre război a fost generat automat de roboți.
- **“Gargară automatizată”:** Gargara este echivalentul românesc a termenului bullshit(ing) [teoretizat de filosoful Harry Frankfurt](#) - înseamnă discurs gol care ne umple capul, dar care nu înseamnă nimic. Astfel de discurs domină deja rețelele de socializare, și mai mult de încurcă decât să ne informeze.
- **Încălcarea vieții private prin monitorizarea în masă cu sisteme IA:** Monitorizarea, până de curând, necesita muncă umană manuală. De exemplu, astăzi, cu ajutorul sistemelor IA de recunoaștere facială, cuplat cu un sistem de CCTV larg, [monitorizarea se poate face la cu totul altă scară](#). Guvernele pot fi seduse de ușurința cu care se poate monitoriza un cetățean sau o populație de cetățeni prin intermediul tehnologiilor care-i înconjoară, [sugerează Bruce Schneier](#), expert în securitate cibernetică.
- **Predicție comportamentală netransparentă:** sistemele IA care iau decizii automatizate despre comportamentul indivizilor în mod netransparent - și câteodată în mod neinteligibil pentru inginerii care le-au făcut - care periclitizează autonomia indivizilor în cauză. Exemplu: [predicția comportamentală](#) a unui angajat la locul de muncă, în sănătate, educație, etc.

- **Reducerea autonomiei decizionale:** automatizarea proceselor decizionale implica reducerea autonomiei decizionale ale indivizilor implicați - mașină ajunge să fie cea care decide, nu omul. Exemplu: a pune ChatGPT (sau alte tehnologii) sau alt chatbot [să aleagă pentru tine](#).
- **Transferul responsabilității pe mașină:** responsabilitatea (implicit și consecințele) pentru o decizie algoritmică luată de un sistem de inteligență artificială ajunge să fie pusă pe sistemul IA și nu pe un individ, o echipă de indivizi sau o companie care a construit/înțeles/folosit sistemul. De exemplu, [sistemul COMPAS folosit în instanțele judecătorești din SUA](#) pentru predicția recidivei printre persoanele acuzate de crimă, care au greșit de nenumărate ori predispoziția la recidivă a unui individ.

Așa cum am mai precizat mai sus, în opinia noastră, atât exemplele de mai sus dar și altele sunt o consecință a uneia din marile dileme actuale a sistemelor de inteligență artificială - că sunt **cutii negre**, a căror proces (și rezultat) decizional nu-l putem înțelege și descifra. Caracterul de tip **cutie neagră** a sistemelor de inteligență artificială poate duce la multe atingeri la adresa drepturilor noastre fundamentale, iar cel mai mare pericol actual este faptul că multe nu pot fi înțelese, replicate sau raportate și că nu am avea pe cine să tragem la răspundere pentru consecințele nefaste ale unor astfel de sisteme automatizate sau să găsim soluțiile eficiente pentru repararea lor.

Încotro? Cum reglementăm inteligența artificială?

În Uniunea Europeană reglementarea inteligenței artificiale a trecut de la simplă discuție la realitate concretă prin noul [regulament privind sistemele de inteligență artificială](#). Acest regulament clasifică sistemele IA în trei categorii mari și late (risc inacceptabil, risc moderat și risc scăzut). Este primul regulament de genul acesta care încearcă să reglementeze tehnologii IA.

Ce a înțeles bine regulamentul privind sistemele IA de la UE este că nu există determinism tehnologic. Adică tehnologia nu ne dictează prezentul și nici viitorul. Totul depinde de oamenii care creează sau implementează o anumită tehnologie și pentru ce scopuri o folosesc.

Mai mult, fiecare tehnologie aduce cu sine o perspectivă asupra lumii, o ideologie, un mod de a vedea. Inteligența artificială încurajează o perspectivă, care poate fi (momentan) negativă în unele contexte, una **unde oamenii pot fi reduși la date de antrenare, dezumanizati de dragul de a avea sisteme automatizate tot mai eficiente**. Dar asta nu este un dat de la natură. Poate fi schimbat, aceste unelte pot fi folosite și pentru sporirea drepturilor fundamentale ale omului, scopuri inerent pozitive, și nu doar în moduri mercantile, pentru interese private, sau în moduri negative.

Viziuni alternative la adresa tehnologiei

Viziunile despre tehnologie de obicei se pot împărți pe un spectru cu 2 extreme:

- cele ale teho-optimiștilor (utopie), care spun că tehnologia ne va face viața mai bună în foarte multe aspecte.
- cele ale teho-pesimiștilor (distopie), cei de sus cu riscurile existențiale, care spun că tehnologia pe care noi am creat-o ne va aduce sfârșitul (sau pur și simplu ne-ar face viața dificilă).

Avem nevoie, în schimb, de un teho-realism, un punct de mijloc critic, mai potrivit decât viziunile utopice și distopice. Tehno-realismul susținut aici înseamnă adoptarea atât a perspectivelor optimiste cât și conștientizarea eșecurilor tehnologiilor de tip IA.

Tehno-realismul este pragmatic, practic, pune accent pe reglementarea democratică a unor astfel de tehnologii (cum ar fi regulamentul UE privind inteligența artificială) și susține protecția și sporirea drepturilor noastre fundamentale.

În viziunea noastră teho-realistă, inteligența artificială poate fi folosită pentru sporirea drepturilor noastre fundamentale, cum ar fi accesul la educație, prin facilitarea rapidă, acasă la toată lumea, de cunoștințe noi și livrate în mod personalizat și adaptat nevoilor fiecăruia (cu grijă la detalii factuale).

Putem înscrie, dacă dorim, în mod colectiv, un set de valori profund în favoarea drepturilor fundamentale în tehnologiile noastre: transparență decizională, acces la datele de antrenare, încurajarea autonomiei utilizatorilor, și altele. Asta ar însemna, în practică, o restructurare majoră a modului în care ne construim tehnologia, cum o dezvoltăm și cum o folosim.

Acest proces necesită deliberare democratică intenționată și de lungă durată. În joc este viitorul societății noastre, a democrației și a sănătății internetului. Iar când drepturile noastre fundamentale au de suferit de pe urma unei tehnologii, toate trei mai devreme menționate au de suferit la rândul lor.

Lecturi suplimentare

Kate Crawford "[The Atlas of AI](#)"

Răzvan Rughiniș "[Societatea Digitală](#)"

Mark Coeckelbergh "[AI Ethics](#)"

Ultima actualizare: Decembrie 2023

