

# O SCURTĂ INCURSIUNE ÎN VIITORUL INTELIGENȚEI ARTIFICIALE

Autor: Andrei Marin | 22 ianuarie 2022



Avansul inteligenței artificiale pe parcursul ultimilor ani ne pune în fața unor întrebări semnificative despre capacitatea omenirii de a controla, pe termen lung, dezvoltarea acestui domeniu, precum și despre conservarea autonomiei umanului în fața progresului tehnologic. Făcând abstracție de scenariul în care se renunță în totalitate la dezvoltarea inteligenței artificiale, progresul natural - mai lent sau mai rapid - va afirma superioritatea computerelor asupra oamenilor în diferite domenii.

O privire aruncată asupra jocului de șah ne arată că, de la victoria istorică a lui *Deep Blue* contra lui Garry Kasparov, șansele jucătorilor umani în fața computerelor s-au diminuat progresiv. Capacitatea de calcul a inteligenței artificiale este exploatată de o serie de algoritmi din ce în ce mai performanți, care au înclinat definitiv balanța în favoarea computerelor. În prezent, inteligența artificială pătrunde și în alte jocuri, precum Go sau Poker. Avantajul său pornește de la o deosebire fundamentală între om și computer. Inteligența umană este constrânsă de parametrii fizici ai creierului, care, spre deosebire de hardware-ul inteligenței artificiale, nu pot fi ameliorați cu ușurință.

Pe termen lung, transformările produse de implementarea sistemelor inteligente vor fi din ce în ce mai însemnate. Deja industria beneficiază din plin de inteligența artificială, capabilă să eficientizeze liniile de producție, iar apoi să prelucreze imense baze de date în scopuri publicitare. Suntem, din acest punct de vedere, într-o plină revoluție industrială, greu de comparat cu alte procese istorice, din cel puțin două considerente. În primul rând, nu există un consens asupra definiției unei revoluții industriale, deci nu putem spune câte asemenea revoluții au avut loc. În al doilea rând, procesul căruia îi suntem martori nu dă semne de încetinire, astfel că nu poate fi echivalat unor procese istorice încheiate de zeci sau sute de ani. Avansul tehnologic a modelat masiv obiceiurile noastre de consum, fapt repercutat asupra industriei, care se schimbă rapid sub acțiunea noilor tehnologii (o privire asupra evoluției automobilelor în ultimii ani este mai mult decât edificatoare).

Discuția despre pierderea locurilor de muncă ca urmare a creșterii ponderii inteligenței

artificiale este una cât se poate de realistă. Numărul de sarcini lucrative pe care le poate întreprinde inteligența artificială este în creștere, pătrunderea acesteia în diferite domenii fiind deja exemplificată anterior în cazul jocurilor. S-au încercat diferite ierarhizări ale locurilor de muncă în baza criteriului siguranței față de avansul computerelor, însă indiferent de acuratețea acestor predicții, probabil că – într-un viitor nu foarte ușor de precizat – economia va trece printr-o schimbare de paradigmă în ceea ce privește relațiile de muncă. Un prim semn al acestei schimbări îl constituie discuțiile despre venitul minim garantat, o sumă de bani pe care să o primească fiecare adult, independent de activitatea sa profesională. Dacă în prezent asemenea venituri ar fi acoperite, în final, de munca altor oameni, există posibilitatea ca în viitor sistemele inteligente să preia acest rol. Momentan, programele experimentale introduse în diferite state vizează impactul psihologic al acestor venituri, căutând să vadă cum își va organiza viața o persoană care ar putea să supraviețuiască fără a munci. Indiferent de rezultatele pozitive sau negative ale unor asemenea experimente, este posibil să ajungem în situația de a introduce un asemenea venit pentru a preveni tulburări sociale de o gravitate nemaîntâlnită.

Există câteva joburi care în mod sigur vor supraviețui avansului inteligenței artificiale. Multe dintre acestea se bazează pe atașamentul nostru nu neapărat rațional față de anumite produse sau servicii. Chiar dacă inteligența artificială va realiza anumite sarcini mai bine (după niște criterii cuantificabile), nu este obligatoriu necesar ca aceasta să înlocuiască oamenii în slujbele corespunzătoare acelor sarcini. Deși un profesor simulat de inteligența artificială ar putea preda neîntrerupt, cu costuri mai mici, obiectivitate totală și multe alte aparente avantaje, există un mare dubiu că părinții vor prefera să-și înscrie copiii la o școală ce apelează la o asemenea soluție. Fizicianul Michio Kaku folosește adesea sintagma „principiul omului cavernelor” pentru a desemna asemenea comportamente aparent iraționale fundamentate pe instincte și pe tradiții. Dezvoltarea societății umane arată numeroase exemple de manifestare a acestui principiu, mai ales când tensiunile dintre opțiunile avute la dispoziție se măresc. Într-o formulare mai strictă, oamenii, obișnuiți să lucreze cu semeni de-ai lor și să se încreadă în aceștia, acceptă mai greu forme alternative de interacțiune. În această categorie intră orice de la întâlnirile online în locul celor fizice, la contactul cu inteligența artificială<sup>1</sup>.

O altă problemă semnificativă se referă la autonomia inteligenței artificiale în întreprinderea diferitelor sale sarcini, mai precis la capacitatea sau incapacitatea noastră tehnică de a constrânge o super-inteligență (o inteligență artificială superioară celei umane) la activitatea ce i-a fost trasată. Să ne imaginăm, de pildă, o inteligență artificială capabilă să preia controlul asupra unui alt computer. În acest sens, este suficient ca aceasta să fie capabilă să trimită un e-mail cu un mesaj înșelător scris coerent, însoțit de un link a cărui accesare să compromită computerul pe care a fost accesat. Mecanismul acesta este deja folosit de către infractorii cibernetici umani, deci nu se află cu nimic dincolo de posibilitățile prezentului. Numai că, în slujba inteligenței

artificiale, efectul ar putea fi diferit. Hackerii umani pretind adeseori bani pentru recuperarea accesului la propriul computer, însă o inteligență artificială ar putea fi mai degrabă preocupată de sporirea capacităților sale computaționale, ceea ce i-ar conferi reziliență în fața unor tentative umane de închidere forțată<sup>2</sup>. O asemenea inteligență scăpată de sub control ar putea deruta industriei întregi. Nu este de neimaginat o super-inteligență capabilă să-și creeze o identitate umană falsă în mediul virtual, în baza căreia să obțină bani pe care să-i investească cu un succes mai mare decât ar avea un investitor tradițional.

Societatea umană trebuie să-și pună asemenea întrebări la modul cel mai serios, deoarece nu este deloc limpede că primul nostru contact cu o super-inteligență va fi unul pașnic. Experiențele anterioare ale interacțiunii dintre două societăți umane cu niveluri semnificativ diferite de dezvoltare sunt deosebit de elocvente în acest sens. Frontiera umanului este permeabilă de către inteligența artificială, iar discuția pe care o purtăm nu este deloc futuristă. Cultura este deja afectată de avansul inteligenței artificiale, în sensul în care forme noi de divertisment se datorează acesteia. De zeci de ani există jocuri video în care rivalul nostru este o inteligență artificială. Totodată, prin algoritmi de predicție ai marilor servicii de difuzare de muzică sau conținut video, ne este influențat comportamentul de consum de conținut cultural. Și aici, inteligența artificială își face simțită prezența în noi zone. Anul trecut, inteligența artificială a extrapolat finalul Simfoniei a X-a a lui Ludwig van Beethoven<sup>3</sup>. Printr-o analiză a compozițiilor anterioare, algoritmul a învățat tehnicile compozitorului, pe care le-a aplicat pentru a simula stilul acestuia. De aici și până la compoziții proprii ale inteligenței artificiale mai este un pas - prelucrarea unei cantități mult mai mari de muzică deja scrisă. Ne putem însă aștepta să ascultăm compozițiile unor computere în viitor. Fizicianul Max Tegmark își deschide cartea dedicată inteligenței artificiale<sup>4</sup> printr-un pasaj literar, în care imaginează o super-inteligență artificială capabilă să creeze filme fără actori, strict pe baza simulărilor figurilor și a vocilor umane. Interesant la o asemenea creație ar fi posibilitatea de a produce același conținut video în toate limbile cunoscute de către computer, cu mișcări realiste ale buzelor pentru fiecare dintre acestea.

Mai putem face și alte exerciții de imaginație benefice înțelegerii noastre asupra posibilităților computerelor. Dacă am menționat capacitatea pe care ar putea să o aibă inteligența artificială de a crea de la zero chipuri și voci pentru filme sau seriale, este limpede că aceasta ar putea să simuleze oameni reali pe baza unor informații despre aceștia<sup>5</sup>. O simplă analiză a tuturor celor cunoscute despre noi (date biografice, filmări, înregistrări audio etc.) ar putea permite unei mașini să ne simuleze prezența inclusiv după moarte, prin același mecanism prin care l-a simulat pe Beethoven în actul creației. Într-un asemenea moment, realitatea fizică ar face posibile experiențe pe care astăzi le asociem exclusiv cu dimensiunea spirituală, ceea ce ar avea consecințe greu de imaginat. Dezvoltarea, în paralel, a interfețelor om-mașină și a implanturilor bionice ne

conduce cu gândul la posibilitatea unor hibrizi de tip *cyborg*, ceea ce ar reprezenta o altă transgresiune a computerelor în zona umanului. Dispozitivele artificiale de redare a vederii sunt un domeniu în care se lucrează de zeci de ani. La fel sunt și protezele robotizate, pentru care se caută o reproducere cât mai bună a funcțiilor membrilor naturale.

Suntem în fața a două fenomene diferite: simbioza om-mașină și simularea umanului. Primul dintre acestea dă toate semnele că va conserva pentru un timp îndelungat dimensiunea umană. Pătrunderea mașinilor se face de la exterior către interior, de la membre și organe de simț, către organe interne. Creierul, principal responsabil pentru individualitatea noastră, se distinge de celelalte organe prin complexitatea sa, astfel că nu este vizat în viitorul previzibil de înlocuire, ci doar de o mai bună racordare la computer<sup>6</sup>. Chipurile implantate pe creier sunt în mod cert o extensie a capacităților umanului, nicidecum o substituție a acestora. Domeniul acesta este unul de pionierat, deoarece în prezent capacitatea sistemelor de citire a gândurilor este destul de limitată. Nu se poate pune încă problema transferului întregului ansamblu de gânduri către o mașină. Tehnica actuală a reușit să decodifice gândurile legate de scrierea unor litere, permițând unor persoane lipsite de mobilitate să scrie fără a-și folosi mâinile<sup>7</sup>. Nu este greu să ne imaginăm ce cale lungă mai este de străbătut până la o prezervare computerizată a gândurilor noastre după moarte. În același timp, tentativele computerelor de a simula comportamentul uman au avut numai succese parțiale. Testul Turing<sup>8</sup>, care își păstrează autoritatea în domeniu, pare încă distant pentru inteligența artificială. În esență, a trece acest test înseamnă, pentru un computer, a da răspunsuri indiscernabile de ale unui om la o serie de întrebări.

Totuși, și aici problema pare a se pune în termeni de timp, nu de posibilitate. Să ne gândim la capacitatea de a dezvolta inteligențe artificiale, pe care în acest moment o are doar omul. Acesta a creat deja algoritmi de *machine learning*, folosiți în cele mai diverse condiții. De exemplu, un aspirator inteligent îi poate folosi pentru a evita să lovească obiecte din camera prin care se deplasează. Asemenea algoritmi ar putea permite unei inteligențe artificiale apropiate de capacitatea umană să se dezvolte prin forțe proprii. Cu alte cuvinte, dacă inteligența artificială va atinge inteligența umană, atunci o va și depăși. Consecințele sunt, bineînțeles, greu de anticipat și vor acoperi întreaga plajă de activitate.

Putem însă să ne menținem în cazul particular al consecințelor legale ale unei asemenea situații. Omul este capabil să eludeze legea sau să o citească în cheia care îi este favorabilă, iar atunci ne putem aștepta la un acțiuni similare și din partea inteligenței artificiale. Putem deja observa, *in nuce*, asemenea comportamente într-un exemplu simplu - funcțiile de autocorectare. Dincolo de utilizarea lor în scrierea unor texte, acestea joacă un rol și în programare, acolo unde un cod scris de un utilizator uman poate rula uneori chiar și dacă prezintă greșeli minore. Compilatorul a trecut de stadiul

În care citește și execută codul (dacă îl înțelege, adică dacă este corect scris). Acum îl citește și îndreaptă greșelile cu soluția cea mai probabilă să constituie adevărata dorință a programatorului. Ne putem gândi numai ce ar însemna în viitor un asemenea gest de „rezistență“ a unei inteligențe artificiale față de ceea ce i-ar apărea drept o negliobie umană. Proporțiile sunt cu totul diferite, dar esența este aceeași: o greșeală umană – reală sau imaginată – este corectată de computere. În prezent, corectura se referă la modificarea unei instrucțiuni, astfel ca programul să ruleze în loc să returneze o eroare. În viitor, inteligența artificială ar putea lua în calcul și alte considerente atunci când stabilește ce anume este o greșeală. Printre acestea s-ar putea număra, de exemplu, consecințele posibile ale comenzii ce i-a fost dată.

Astfel, dacă aducem în discuție problema drepturilor, ar trebui să privim relația dintre oamenii și inteligențele artificiale drept bilaterală. Max Tegmark, pe care l-am invocat și mai devreme, discută mai multe scenarii – foarte diferite unul de altul – pentru conviețuirea oamenilor cu o super-inteligență artificială. Ideea pe care trebuie să o reținem este că se impune să acordăm atenție reglementării inteligenței artificiale, fiindcă aceasta poate ajunge să fie o influență dominantă în societate. Aceasta este deja folosită în publicitatea electorală și – din păcate – în industria știrilor false, în slujba diferitelor tabere politice. O super-inteligență ar putea să-și formeze însă propria sa tabără politică, poate chiar propria sa ideologie.

Asemenea situații nedorite pot fi prevenite printr-o serie de reglementări, echivalente cu legile. Asemenea reglementări se pot referi la conexiunea dintre super-inteligența artificială și internet, de exemplu. A ține departe o entitate digitală de spațiile eminentamente umane poate fi o precauție necesară. Similar, ne putem pune problema închiderii forțate a sistemelor inteligente, moment în care discuția despre drepturi devine esențială. Primul dintre drepturile omului este dreptul la viață, care s-ar traduce pentru computere în dreptul la alimentare cu electricitate, reparații etc. O super-inteligență care se vede în fața unor întreruperi nedorite ale activității ar putea lua decizii imprevizibile pentru a se menține funcțională. Exemplul dat mai devreme cu preluarea controlului asupra unui număr mare de computere individuale este banal prin comparație cu ce ar putea născoci o entitate cu o inteligență superioară celei umane.

Concluzia acestor explorări este că trebuie să admitem că inteligența artificială se va dezvolta și că acest lucru va afecta din ce în ce mai semnificativ activitatea umană. De aceea, sunt necesare niște discuții globale despre circumstanțele în care aceste procese se vor produce. Iar în orice proces decizional trebuie avută în vedere menținerea unui echilibru care să prevină amânarea exagerată a adoptării unor măsuri necesare, dar și luarea unor decizii pentru un viitor prea îndepărtat și incert. Politica va juca, desigur, un rol în această dezbatere, însă acesta va fi mai degrabă unul de reglaj fin, în sensul găsirii soluției preferabile dintr-o mulțime de soluții acceptabile, care să mențină autonomia umanului și mecanismele sale de control asupra computerelor.

## NOTE

1. Să ne gândim că tehnologia activității online, folosită mai intens după declanșarea pandemiei în 2020, există de peste zece ani, însă atenția dedicată dezvoltării unor programe dedicate acestor scopuri a fost redusă. ↑
2. Este tentant să credem că vom putea tăia alimentarea cu electricitate unei inteligențe „rebele“. Este însă probabil ca aceasta să reușească prin diferite mecanisme (inclusiv cel imaginat de noi) să lucreze descentralizat, ocupând rețele întregi. ↑
3. Beethoven’s last symphony finished by AI | Music | DW | 09.10.2021; ↑
4. Max Tegmark, *Viata 3.0*, Ed. Humanitas, 2019; ↑
5. Fenomenul are deja loc – la o scară mică – prin așa-numitele conținuturi *deepfake*. Un articol cuprinzător este *What are deepfakes – and how can you spot them?* | Internet | The Guardian. De asemenea, un articol despre tentativele din domeniul prezenței digitale a persoanelor decedate este *Microsoft patent would reincarnate dead relatives as chatbots* – The Washington Post. ↑
6. O discuție interesantă găsim în Michio Kaku, *Viitorul minții umane*, Ed. Trei, 2016; ↑
7. *New Brain Implant Turns Visualized Letters into Text* – Scientific American; ↑
8. *Turing test* | Definition & Facts | Britannica. ↑

*Imagine: Unsplash*