

Monede-putere, monede-vanitate, monede-respect, monede-sex, monede-trib etc



Într-o prezentare recentă (din luna iunie, anul acesta)ⁱ, Ray Kurzweil, unul dintre cei mai cunoscuți viitorologi și experți din lume în calculatoare, își menține predicția din 1999ⁱⁱ ca se va ajunge la nivelul de inteligența artificială generală (AGI) în 2029, adică peste 5 ani.

În esență, inteligența artificială generală ar putea înțelege și învața orice sarcină umană. În diverse forme robotice, ar putea să facă mult mai bine, mai repede și mai ieftin orice fac oamenii în prezent (și mai mult decât atât), permițând abundența totală.

El anunța cu ocazia respectivă că numeroși alți specialiști consideră deja că este probabil să se ajungă la acel nivel în jurul anului 2035, chiar dacă, la momentul inițial al predicției sale, majoritatea credeau că este vorba de ceva posibil de atins abia în jurul anului 2100.

Cu alte cuvinte, ceilalți experți din 1999 se înșelasera cu mult, avansând între timp cu 65 de ani termenul anticipat de ei atunci, și, chiar și așa, Ray Kurzweil consideră în iunie 2024 că e probabil că el să aiba până la urmă dreptate „cu totul” pe subiect, adică și 2035 să fie pesimist, progresul rămas de făcut fiind realizabil până în 2029.ⁱⁱⁱ

Amintea în prezentarea respectivă și modul său de estimare – practic presupusese în 1999 că ritmul anual de creștere a puterii de calcul, versus cost, din deceniile anterioare se va păstra cel puțin la fel și în viitor. Aceasta însemna că în 2029 ar trebui să se ajungă la calcule de 1 trilion de biți pe secundă la un cost foarte redus, nivel de complexitate și eficiență la care inteligența artificială generală s-ar atinge.

În prezent, și alți experți recunoscuți, precum Ben Goertzel, îi împărtășesc optimismul,^{iv} apreciind că am fi în grafic în această privință și că evoluțiile recente, gen ChatGPT, ar confirma relevanța predicției sale.

În orice caz, ca să se atingă inteligența artificială generală în 2029 sau în 2035 – nu ar schimba fundamental gradul de reușită în analiză și în anticiparea viitorului al predicției de mai sus și nici implicațiile așteptate de la rezultatul în sine.

Iar, în privința implicațiilor AGI, Ray Kurzweil **anticipează efectiv un fel de rai pe pamânt pentru omenire cam imediat după acel moment**. De la bunuri și servicii la discreție și moduri de a acoperi alte nevoi/dorințe fără a fi necesar să se mai muncească - la posibilitatea „înghețării” vârstei și chiar a reversării ei în pasul doi^v.

Adică, de la roboți care să funcționeze ca furnizori/prestatori de bunuri, servicii și emoții (produs mâncare, plimbat câinele, oferit prietenie și chiar tandrețe) - la tinerețe fără bătrânețe/nemurire.

De altfel, Ray Kurzweil ia peste 80 de diverse suplimente zilnic^{vi} pentru a prinde de la cei 76 de ani ai sai din prezent momentul reușitei respective și a putea beneficia el însuși de avantajele acesteia.

În urma cu aproximativ o săptămâna, China a anunțat,^{vii} însa, ca va mari gradual (pe o durata de 15 ani) vârsta de pensionare cu 3 ani, pe fondul declinului demografic în curs și al estimarilor cu privire la necesarul de fonduri, chiar daca pensiile sale publice sunt reduse față de media UE.

Predicțiile lui Ray Kurzweill mi s-au parut în contradicție și cu faptul ca Japonia are și ea în vedere creșterea vârstei de pensionare. Consiliul Fiscal japonez a propus deja în 2023 aducerea acesteia la 68 de ani (chiar daca și pensiile publice japoneze sunt relativ reduse față de media UE).^{viii}

Atât Japonia, cât și China, au o speranța de viața ridicată și în creștere și o structura demografică în deteriorare rapidă (mai ales Japonia), dar sunt țări foarte avansate tehnologic, inclusiv în inteligența artificială, și cu cetățeni, în medie, bine educați și cu o bună etică a muncii.

Or, dacă astfel de țări se considera nevoite să crească gradual vârsta de pensionare (nimeni nu face așa ceva ușor politic), este clar că estimarea acestora cu privire la momentul obținerii inteligenței artificiale generale și a beneficiilor sale este mult mai puțin optimista decât cea a lui Ray Kurzweill.

Cred că cele de mai sus merita privite din două perspective – una legată direct de contradicția de poziționari de mai sus și una legată indirect de aceasta, dar, așa zice, chiar mai importantă.

1. Prima perspectivă este aceea pur tehnică – personal, chiar cred că inteligența artificială va ajunge la unele reușite cu totul deosebite foarte curând, dar, pe ansamblu, pare mult mai prudent de mers pe abordarea Chinei și Japoniei pe acest subiect, decât pe cea a lui Ray Kurzweill.

Dacă am merge pe cea din urmă, ar decurge logic că majoritatea oamenilor nu ar mai prea trebui să muncească sau să mai învețe ceva cam de pe acum, pe ideea că doar producerea de mâncare și a ceva haine, utilități minimale și grija medicală ar mai fi necesare în mod continuu pe termen scurt. În rest, am putea trăi relativ decent câțiva ani din rezerve și aștepta momentul „lămpii lui Aladin”.

Statele ar și putea tipări bani între timp și mai abitir, inteligența artificială generală urmând să rezolve problema resurselor de orice fel, inclusiv financiare, în 2029 sau 2035.

De altfel, numeroase politici publice din România după 1989 (duse pe noi „culmi” în ultimii ani), par să fi fost elaborate de mari cititori și fani ai lui Ray Kurzweill.

Dacă însă China și Japonia au „dreptate”, cei care ar merge pe varianta cealaltă ar risca să ajungă foarte rapid în săracie și chiar să moară de foame la propriu în următoarea decadă.

De altfel, unii oameni de știință mult mai sceptici cu privire la momentul atingerii nivelului inteligenței artificiale generale atrag atenția că acesta (și posibilitatea fructificării sale efective) nu ar depinde doar de nivelul de putere de calcul în sine.

De exemplu, necesarul estimat de consum de energie și de volum de date ar depăși cu mult ceea ce pare realizabil în următorii 10 ani (unele modele consumă deja anual echivalentul consumului de energie al unor țări mai mici doar pentru antrenare^{ix}).

Chiar dacă pe partea de volume de date, există deja modalități de a genera artificial date suplimentare, cele două constrângeri, împreună, ar putea împinge semnificativ în timp momentul atingerii AGI.

2. A doua perspectivă vizează mai pe ansamblu dimensiunile modului de organizare a societăților actuale din lume și ale naturii umane.

Sa presupunem ca în 2029 sau 2035 chiar s-ar atinge nivelul inteligenței artificiale generale și s-ar putea face roboți de un tip apropiat de „lampa lui Aladin”.

Daca ne uitam la **organizarea standard a societăților și formelor de putere în lume** și ne gândim și la istoria speciei, inclusiv recenta, **în primul rând**, pare improbabil ca tehnologia respectiva sa fie facuta automat accesibila întregii omeniri sau macar unui procent mare.

În al doilea rând, chiar daca tehnologia respectiva ar fi facuta disponibila tuturor, **daca privim nivelul pur individual** - ar fi mult mai bine pe ansamblu decât anterior, dar tot ar ramâne probabil destule aspecte deconectate de ce se înțelege îndeobște prin rai (pe pamânt) având în vedere configurarea actuala a valorilor umane.

Sa le luam pe rând.

Pe dimensiunea economica privata a **nivelului de organizare societala**, practic, Ray Kurzweil pare sa se așteapte cumva ca unele dintre cele mai bogate și puternice companii și unii dintre cei mai bogați și puternici oameni din lume sa puna la dispoziția tuturor roboți care ar anula cam tot avantajul bogației și puterii respective (de vreme ce ar aduce abundența generala).

Adica, sa renunțe de buna voie la influențarea economicului și a politicului (și parțial și a altor paliere) tocmai la momentul la care ar atinge un nivel de putere/intervenție cvasideniurgic.

Desigur, simpla mulțumire dezinteresata de a face un mare bine semenilor și chiar vanitatea pot ajuta, dar cred ca este riscant de presupus ca majoritatea respectivilor o sa considere ca merita.

Nu trebuie ignorat nici ce s-ar întâmpla pe **plan academic, artistic, cultural mai larg, tehnic etc (planul mai oricarei meserii sau activități umane, de fapt)**. Practic, mai toți cei care sunt peste medie în vreo privința sau se pricepe bine la ceva o sa aiba în diverse grade o dilema pentru ca susținerea adoptarii și accesibilității generale a tehnologiei respective ar veni cu prețul pierderii rapide a relevanței propriilor competențe/talente (cel mai muncitor și inteligent profesionist ar fi surclasat ușor și de robotul celui mai mare leneș sau prostovan de pe planeta).

Nu în ultimul rând, același lucru ar fi relevant pentru **dimensiunea puterii statale**. Este foarte discutabil daca aceasta ar mai subzista cu adevarat în cazul în care inteligența artificiala generala ar fi larg disponibila populației pentru ca aceasta ar trebui sa poata acoperi orice serviciu public și alte nevoi mult mai bine decât statele actuale.

Acesta este motivul pentru care chiar și aspecte mult mai restrânse par puțin probabil sa fie permise politic, cel puțin în urmatorii 25-30 de ani.

Este, de exemplu, puțin probabil în urmatoarele decenii în multe țari ca promovările și evaluările judecătorilor și procurorilor din domeniul penal sa fie lasate doar analizei inteligenței artificiale chiar daca aceasta ar demonstra un nivel mare de corectitudine și competența în materie. Sau ca în destule țari ar fi lasați doar roboții sa elaboreze programa școlara și sa educe copiii în spiritul cunoașterii și gândirii critice față de guverne și politicieni.

Desigur, de la un punct încolo, nu va depinde doar de state (depinde și ce state etc), dar lucrurile se anunța semnificativ mai complicate decât sunt sugerate de experții în IA menționați.

Ideea nu este deloc aceea ca accesul general la AGI ar fi imposibil, ci ca nu pare probabil fara pregatire și efort proactiv contrar din timp.

Cei care simt frustrare la gândul ca vreun politician mediocru sau hoț ar putea întârzia sau împiedica accesul la inteligența artificiala generala trebuie sa aiba în vedere ca și daca roboții respectivi ar lua „problema” în propriile „mâini”, nu putem deloc anticipa doar scenarii roz.

Nu degeaba, într-o cunoscuta poveste science fiction, primul robot cu inteligența artificiala inventat primește sarcina sa faca oamenii fericiți și îi prinde pe toți și îi „lobotimizeaza”^{xx}, ajungând la concluzia ca e singurul mod în care poate sa-i faca

fericiți.

Aici ajungem și la nivelul pur individual uman. În aceasta privință, predicția menționată pare să aibă în vedere doar marginal diversitatea nevoilor și dorințelor/emoțiilor umane (conform lui Maslow etc)^{xi}.

Or, de exemplu, faptul că marea majoritate a oamenilor din țările OECD nu mai suferă de foame nu pare să fi făcut societățile respective mai calme și mulțumite, segmente importante devenind arțagoase și frustrate concentrându-se mai mult pe nevoia de a fi ascultate și respectate, bagate în seama etc – aspecte din ce în ce mai dificile în vacarmul general.

De asemenea, faptul că numărul celor care au sute de milioane sau miliarde de euro, dolari rimibi sau yeni în lume a crescut foarte mult în ultimii 50 de ani nu i-a făcut pe mulți dintre aceștia să nu se mai implice în viața economică sau în influențarea puterii.

Un alt aspect de avut în vedere ar fi acela că un robot de tip „lampa lui Aladin” ar putea aduce beneficii enorme utilizatorilor săi, inclusiv pe partea de alinare a singurătății, îngrijire, posibil chiar și tandrețe etc, dar ar crește probabil mult numărul celor care nu ar mai putea construi „natural” relații **umane** suficiente/satisfacătoare.

Nu în ultimul rând, după cum menționăm și la punctul anterior, legat de modul de organizare valoric și de putere al societății, destui oameni nu ar fi probabil mulțumiți nici de „nivelarea” implicată și ar putea fi tentați să-și folosească roboții pentru diverse avantaje de statut într-o lume în care „lampa lui Aladin” le-ar face teoretic foarte dificile.

În acest sens, deși tindem să ne concentrăm atenția zilnică mai mult pe scandaluri și pe piața economică și să privim critic „goana după profit” **monetar, fiecare din nevoile și dorințele umane reprezintă, de fapt, o “piață” separată în sine a carei (ne)satisfacere poate stimula un anumit tip de comportament.**

Or, într-o lume a inteligenței artificiale generale, satisfacerea unora dintre planuri ar rămâne incompatibilă cu satisfacerea altora.

De la monede-monedă, operăm ca omenire mult mai des decât ne dăm seama de obicei și pe piețele cu monede-putere, monede-vanitate, monede-respect (macar aparent), monede-sex, monede-trib, monede-reciprocitate etc.

Dacă de la „nevoi” în sens (psihologic) ne uităm și la emoții și dorințe, plaja tuturor „piețelor” alternative devine foarte largă.

De altfel, **experiența lumii ultimilor 20 de ani sugerează că atunci când un tip de nevoie este aproape satisfăcut în societate, alte tipuri de nevoi se exprimă, de fapt, mai acut**, ca și cum am avea, în medie, un fel de volum total care se redirecționează, pe principiul vaselor comunicante.

Aceasta **nu înseamnă că rezultatul global al inteligenței artificiale generale nu ar fi în mod normal mult mai bun decât situația actuală**, pentru că nevoile de bază și unele plăceri/dorințe standard sunt mai importante pe ansamblu decât alte tipuri de nevoi și dorințe (unele „piețe” fiind mult „mai” benigne/”mai puțin” maligne decât altele).

Numeroase sondaje și arată că viața de „pensionar” oferă multe satisfacții în medie, mai ales dacă e una cu o sănătate bună.

Dar nu ar fi „raiul pe pământ”, potențialul de frustrare și resentiment rămânând ridicat.

De exemplu, ar fi mai greu să te simți respectat cu adevărat (de alți oameni) dacă nu te poți diferenția prin mai nimic (decât, eventual, prin efort pur benevol, dar într-o lume în care ceilalți nu ar mai cam avea nevoie de rezultatul efortului tău).

Asta dacă nu se lucrează chiar asupra stimulării valorilor și așteptărilor echilibrate la nivel individual (și nu în moduri distopice).

În concluzie, evoluțiile tehnologice actuale oferă foarte multe oportunități și vin și cu ceva riscuri mari.

Să sperăm că vom avea curând cât mai multe avantaje pentru ansamblul populației și cât mai puține dezavantaje – undeva

între Kurzweil și China/Japonia.

Este necesara, însa, atât multa prudența în privința rapidității cu care s-ar ajunge la AGI, cât și o implicare bine organizata din partea societății, dar și a țarilor mai mici, în urmarirea și controlul dezvoltarilor tehnologice pentru creșterea șanselor unei distribuii cât mai largi a beneficiilor acestora în pasul doi.

Scenariile și exemplele de mai sus arata, cred, destul de clar, ca va conta enorm ce fel de oameni ajung mai întâi la nivelul respectiv și în ce context (între „rai” și „iad” poate fi o distanța mica).

Între timp, ar trebui sa acordam mult mai multa atenție celorlalte „piețe” pe care activeaza omenirea – a „piețelor” monedelor-putere, monedelor-vanitate, monedelor-respect, monedelor-sex, monedelor-trib etc.

„Piața monedelor-monedes” nu trebuie deloc ignorata, dar este, de fapt, printre cele mai eficient așezate și mai transparente dintre ele.

O mult mai mare atenție concentrata pe „piețele alternative” respective este un aspect important nu doar pentru progresul real al societății, ci și pentru tipul de „parere” pe care ar putea sa și-l faca un robot despre noi ca specie din datele disponibile public (foarte puține din „piețele alternative” respective fiind cu totul benigne).

i. <https://www.youtube.com/watch?v=uEztHu4NHrs> The Last 6 Decades of AI — and What Comes Next.

ii. Predicția a fost facuta în 1999. A fost dezvoltata în cartea sa "The Singularity is Near" (Viking USA, 2005).

iii. Unii experți deja susțin ca și 2027 ar putea fi posibil. Pentru stadiul actual, este util de citit raportul menționat la : <https://www.weforum.org/agenda/2024/04/stanford-university-ai-index-report/>

iv. <https://www.livescience.com/technology/artificial-intelligence/ai-agi-singularity-in-2027-artificial-super-intelligence-sooner-than-we-think-ben-goertzel>

v. În plus de prezentarea video, a se vedea, de exemplu : <https://www.economist.com/by-invitation/2024/06/17/ray-kurzweil-on-how-ai-will-transform-the-physical-world>

vi. <https://www.wired.com/story/big-interview-ray-kurzweil/>

vii. Pentru unele detalii și comentarii, a se vedea, de exemplu :

<https://www.nytimes.com/2024/09/13/world/asia/china-retirement-age.html#:~:text=This%20is%20the%20first%20time%20China%20has%20raised>

viii. <https://thediplomat.com/2023/01/surviving-old-age-is-getting-harder-in-japan/>

ix. Desigur, în aceasta privință poate să treacă prin gând și scenariul distopic din Matrix în care oamenii erau folosiți de inteligența artificială pe post de baterii pentru a-și acoperi consumul de energie.

x. De cele mai multe ori, termenul lobotomie se aplica operației chirurgicale care constă în secționarea, în encefal, a unei părți din fibrele nervoase care leagă lobul prefrontal (sediul ideatiei - formarea și înlanțuirea ideilor) cu restul creierului.

xi. Este cunoscut că piramida lui Maslow este contestată în diverse privințe, dar cred că reflectă suficient de bine liniile mari ale nevoilor umane și ce dorește să sublinieze articolul.