

ÎN UMBRA TSUNAMI-ULUI AI-ULUI GENERATIV, DOAR ATÂT - BUNĂTATEA

Autor: Petru Dimitriu | 29 iulie 2025



Când eram mic și cu mintea fragedă, eram convins că virtutea umană ce merită cea mai mare admirație este inteligența. Virtutea de a fi savant în sensul pur, etimologic – cunoscător a cât mai multe, cât mai diverse, atât abstracte cât și concrete; virtutea de a ști cu adevărat, de a *înțelege* (de asemenea în sensul proto-etimologic¹), concretizată în a fi capabil să selectezi și să legi multiple cunoștințe între ele pentru a produce noi și noi informații. Polimatul, *homo universalis*, era imaginea omului ideal pe care aș fi invocat-o spontan, sub inerția curiozității înnăscute, a ținutelor trasate de școală și a eroului copilăriei mele, micul savant Dexter. Abia după câteva secunde de răgaz, rușinat de acest gând, aș fi adăugat sentinței mele despre omul ideal, careva calități morale: smerenia, blândețea, curajul...

Treptat, și în chip firesc, odată cu maturizarea, am realizat că trăsătura de căpătâi ce ne conferă umanitatea nu este inteligența, ci este dată de virtuțile morale, cărora cunoștințele, fie ele oricât de precise și corecte, le sunt inferioare în plan ontologic. E lesne de ajuns la această concluzie atunci când descoperim cum inteligența și răul pot coexista nestingherite și nu se anulează reciproc, ca de pildă atunci când întâlnim un om deopotrivă rău și inteligent (iar pentru aceasta, probabil cel mai la îndemână, ajunge să privim introspectiv) sau atunci când comparăm magnitudinea răului adus în lume pe calea ignoranței cu aceea a răului deliberat, pentru care cunoștințele reprezintă factorul agravant, nicidecum un har răscumpărător.

Deși se poate ajunge la această realizare pe calea intuiției și a observației psihologice, modernitatea ne oferă tuturor, de multe decenii deja, un studiu de caz elocvent – virusul informatic. Confruntându-ne, în calitate de victime sau de specialiști, cu virusurile informatice contemporane, putem să apreciem complexitatea lor și să afirmăm, poate forțând ușor termenul așa cum o făceau oamenii la apariția primelor calculatoare în anii 1950, că sunt *inteligente*: angrenajele informatice care le constituie iau decizii în mod autonom, efectuează computații, se multiplică, comunică, având ținta fixată explicit de către creatorii lor, o țintă ce nu poate fi numită decât malefică, căci nimeni care a pățit să-i fie criptate documentele personale și să fie extorcat pentru a le mai recupera (sau doar pentru o speranță deșartă) nu ar spune că experiența este produsul unui izvor de bunăcuviință. În jargonul specific, spunem despre un virus, fie el și informatic, că este un program *malitios* (în engleză, *malicious*), deși virusul informatic nu poartă

răspundere pentru „faptele” sale, însă valoarea sa etică este preluată de la intenția autorilor². Intelligent precum este, capabil să acumuleze și să întrebuințeze cunoștințe, virusul informatic este totuși, finalmente, o entitate non-umană, lipsită de suflet și de intenții propriu-zise.

Însă capacitatea unui virus informatic (ca și a oricărui program informatic) de stoca, colporta și amesteca cunoștințe este derizorie pe lângă aceea a sistemelor informatice care au luat cu asalt spațiul informatic în ultimii câțiva ani, cele cuprinse sub termenul-umbrelă, deja clișeic, de inteligență artificială (AI), mai precis inteligența artificială *generativă* (GenAI). Ceea ce nu mulți utilizatori ai sistemelor de AI precum ChatGPT, Midjourney sau Perplexity realizează – sau ceea ce adesea uită – este că, mai întâi de toate, aceste sisteme reprezintă o modalitate inedită de a *stoca* cunoștințe în cantități impresionante. Dacă o bază de date oarecare, precum cea care stochează paginile unui site web sau articolele de pe Wikipedia, reține, în general, înșirui de cuvinte, literă cu literă (iar la un nivel fundamental, înșirui de numere), arhitecturile informatice din spatele AI-ului generativ, modelele de limbaj, reușesc să stocchez, într-o manieră compactă, cantități mult mai mari de cunoștințe, împreună cu informații contextuale despre cum anume se înrudesce și se alătură toate aceste cunoștințe. Nu voi intra acum în alte explicații tehnice, dar rețin ideea principală: inteligența artificială generativă, în primul rând, *stochez* cunoștințe.

Constrâns de interogările pe care le primește de la utilizatori, un model de AI nu face decât să reproducă cunoștințele pe care le-a înmagazinat, însă nu oricum, ci „forțat” de contextul furnizat de interogare, „obligat” să reproducă aceste cunoștințe în așa fel încât să satisfacă informația contextuală furnizată de interogare. Contextul interogării, așadar, provoacă potrivirea cunoștințelor, generând răspunsuri ce produc aparența raționării.

Cheia de boltă a inteligenței artificiale generative, și mai cu seamă a modelelor de limbaj ce stau în spatele aplicațiilor precum ChatGPT, o reprezintă, așadar, cunoștințele. Această situație nu este întrutotul nouă. Încă de la inventarea lor, calculatoarele au fost întrebuințate ca depozitare de cunoștințe. După ce timp de ani de zile, aplicațiile informatice au fost din ce în ce mai potente în ceea ce privește *eficiența* stocării informațiilor, proliferarea modelelor de limbaj a reprezentat un salt în *eficacitatea* stocării cunoștințelor, întrucât se poate argumenta că acestea reușesc mai bine ca oricând să surprindă semantica datelor pe care le ingerează, chiar dacă aceasta presupune costuri semnificative sub alte aspecte, mai cu seamă pecuniare.

Pe măsură ce modelele de limbaj evoluează iar halucinațiile și costurile de operare ale acestora se diminuează, devine din ce în ce mai evident că o cantitate de cunoștințe imposibil de reținut de către o ființă umană pe parcursul unei vieți obișnuite poate fi consumată și reprodușă de un model de limbaj mai elocvent decât ar putea o persoană.

Putem afirma, aşadar, că asistăm la un eveniment fără precedent în istoria umanităţii, acela al deplasării, cel puţin parţiale, a autorităţii epistemice dinspre experţii umani înspre modelele de inteligenţă artificială generativă.

Nu este fantasmagoric să ne imaginăm că peste un deceniu vom umbla cu modelul de AI la purtător. Miniaturizat, susţinut de un hardware mai puternic, probabil nici măcar dependent de o conexiune la reţea, modelul de limbaj personal ne va ajuta să punem cap la cap cunoştinţe care altminteri ne-ar lua semnificativ mai mult timp să le procesăm cu propriul creier. Considerentele practice îi vor obliga pe cei activi să procedeze astfel, cel puţin ocazional, dintr-un motiv banal: eficienţa; aceia dintre noi care vor învăţa să mânuiască cu dibăcie instrumentele de AI se vor bucura de avantaje competitive zdrobitoare în viaţa de zi cu zi.

Până atunci însă, să ne amintim că destoinicele modele de AI nu sunt decât descendente ale rudimentarelor aplicaţii şi virusuri, biete entităţi non-umane mănate, în mod fundamental, de date (în engleză: *data-driven*). Şi odată ce actuala întrupare a inteligenţei artificiale ne demonstrează că nu puterea de a reţine cunoştinţe abstracte şi nici măcar capacitatea de a le combina nu reprezintă apanajul omului, căci efortul acestora este, mai nou, delegat din ce în ce mai eficient, se ridică întrebarea: care sunt totuşi virtuţile care, în calitate de oameni, ne mai rămân de cultivat? Răspunsul e limpede: virtuţile morale.

Reuşitele tehnologiei care ne degrevează tot mai mult de balastul mundan ne oferă tot mai mult timp şi tot mai multe oportunităţi să facem, pur şi simplu, *ce vrem*. De aici apare problema. *Ce* vom alege să facem cu toate posibilităţile care ni se deschid la orizont? Înzestraţi cu tot mai multe instrumente tot mai potente, *încotro* ne vom îndrepta intenţiile, despovărate de greutatea efortului? Inteligenţa artificială va fi un factor potenţiator al tuturor trăsăturilor omeneşti – bune sau rele, fie că le recunoaştem în acest fel sau nu. Accesul facil şi universal la cunoaştere vor aplatiza şi apropia condiţiile erudiţiei şi ale ignoranţei, aşa încât principalul discriminant între indivizi va fi, mai limpede ca niciodată, orizontul moral. Amoralitatea, vecină cu relativismul moral, nu mai este o scuză, pentru că ne apropie de condiţia ingrată, inferioară, a propriei noastre creaţii.

Acum este timpul să căutăm mai acerb şi să regăsim reperele morale, să redescoperim eroii de la care merită să ne revendicăm, să spicuim măcar din marile texte şi lucrări ale omenirii, în căutarea înţelepciunii care ne va împiedica să decădem la starea unor simple modele de limbaj întrupate (şi încă unele de foarte slabă performanţă). Acum avem încă un argument în sprijinul acestui demers şi o demonstraţie sub ochii noştri a tezei că datoria existenţială a omului, în sensul asumării şi desăvârşirii umanităţii sale, este să dobândească ceea ce maşinile nu vor putea dobândi niciodată – iar aceasta, într-un singur cuvânt, este bunătatea.

NOTE

1. Românescul *a înțelege* se trage din latinescul *intelligere*, compus din *inter* și *lego* (a selecta). ↑
2. Unii ar putea argumenta că există și virusuri informatice „bune”, un exemplu putând fi considerat faimosul virus *Stuxnet*, despre care se presupune că a fost conceput de actori statali pentru a sabota programul nuclear al Iranului. ↑