LIBRIS | We know books

INTELIGENȚA ARTIFICIALĂ – CUM? DE CÂND? DE CE?

Coordonatori:
Petru Dimitriu
Andrei Marin

PRESA UNIVERSITARĂ CLUJEANĂ
2024

LIBRIS | We know books

Cuprins

Cuvânt înainte 7

Între alchimie și matematică.
O scurtă istorie a inteligenței artificiale 13
Petru Dimitriu

De la creierul uman la rețele neuronale –
bazele inteligenței artificiale 67
Dragoș Manea

Despotul fără corp: transformări politico-sociale
și antropologice sub impactul inteligenței artificiale 99
Andrei Marin

Eu, tu și chatbotul.
Sau despre noua inginerie sufletească 151
Victoria Maria Deliu

Deschiderile inteligenței artificiale
către științele exacte 183
Andrei Marin

Despre efectele subtile ale tehnologiei:
între determinism și neutralitate tehnologică 227
Teodora Nichita

Cum (nu) gândește ChatGPT 251
Petru Dimitriu

Despre autori 271

LIBRIS | We know books

Între alchimie și matematică. O scurtă istorie a inteligenței artificiale

Petru Dimitriu

Introducere

Cariera mea de programator – amator sau profesionist – care anul acesta își împlinește majoratul, s-a întrevăzut de timpuriu. Încă de la mijlocul anilor '90, eroul animat al copilăriei mele, copilul-geniu Dexter, era sprijinit în aventurile sale din vastul laborator secret de nelipsitul său *Computer*. Realizând o predicție a viitorului apropiat de o remarcabilă acuratețe, desenul animat regizat de animatorul Genndy Tartakovsky îl portretiza pe asistentul virtual al lui Dexter drept o entitate amorfă, răspândită în întreg laboratorul – și care, probabil, *era însuși laboratorul* – care îi furniza micului savant zăvorât în mijlocul pustiei tehnologice tot felul de date, informații sintetizate, rapoarte în timp real și, ocazional, chiar și sfaturi, toate acestea redate într-un limbaj natural impecabil sonorizat de o reconfortantă voce feminină.

Așadar, deloc surprinzător, primul meu contact cu informatica, petrecut în clasa a cincea, nu a putut decât să-mi stârnească fascinația debordantă pentru această știință aplicată care se constituie, de altfel, într-un veritabil meșteșug modern. Țintind, ca prim reper, participarea la olimpiadele de informatică – care, în România, au fost mereu considerate competiții prestigioase –, m-am angajat cu entuziasm în asceza mentală

LIBRIS We know books

a studiului aplicat al algoritmilor. Totuși, caracterul abstract al algoritmilor pe care simțeam că îi exersam ca într-un soi de culturism cognitiv fără un obiect clar m-a determinat, după câțiva ani, să-mi deplasez interesul înspre realizarea unor proiecte concrete și să trec competițiile de informatică axate pe algoritmică într-un plan secundar.

Urmărind să-mi diversific abilitățile de programator, mi-am propus la începutul anului 2012 să deprind pe cont propriu o paradigmă de programare utilizată pe scară largă, mai ales în context profesional, denumită *programarea orientată pe obiecte*. Implementarea soluțiilor informatice în această paradigmă presupune că programatorul exprimă funcționarea soluției software descriind, în limbajul codului, concepte sub forma unor abstractizări informatice denumite *obiecte* ce sunt *instanțe* ale unor *clase*, definite de un set de atribute și interacțiuni esențiale. Această foarte populară filozofie a programării prezintă avantajul de a facilita redactarea unui cod sursă modular, mai lizibil și mai organizat.

În primul program pe care l-am scris folosind programarea orientată pe obiecte am ales să modelez, pentru început, părțile de vorbire ale limbii române. Am descris o clasă denumită *Substantiv*, care avea drept atribute genul gramatical, numărul, cazul, articolul (hotărât sau nehotărât) și lista tuturor declinărilor. Pe lângă acestea, am implementat și un subalgoritm dedicat (în limbaj tehnic, o *metodă*) care putea fi folosit pentru a determina declinarea unui substantiv, în funcție de atributele unei anumite instanțe concrete a clasei *Substantiv*. Similar, am conceput clasa *Adjectiv*,

LIBRIS We know books

care avea, printre altele, și o metodă dedicată acordării în gen și număr cu o instanță de *Substantiv*; apoi clasa *Pronume*, clasa *Verb*, clasa *Adverb*, clase pentru părțile de propoziție și așa mai departe. Din aproape în aproape, am realizat un program capabil să reprezinte fraze întregi în structuri de date informatice dedicate și să le exprime textual, acordând și articulând cuvintele, respectând fidel regulile gramaticale ale limbii române.

Încântat de rezultat, am plusat și am realizat și un algoritm care analiza sintactic și morfologic o frază scrisă în limba română și o transpunea în structura care îi permitea programului să o reproducă, am adăugat o bază de date pentru stocarea cuvintelor împreună cu câteva „cunoștințe" rudimentare despre noțiunile pe care le reprezintă, iar în final am implementat câteva reguli de extragere și actualizare a acestor cunoștințe ca răspuns la frazele introduse de agentul uman. Am obținut un program capabil să analizeze sintactic și morfologic fraze scrise în limba română și să efectueze silogisme simple. Pe baza cunoștințelor acumulate, programul putea răspunde la întrebări de tipul „Sunt câinii animale?" efectuând deducții logice simple și putea înregistra automat noi cunoștințe atunci când recepționa de la utilizator afirmații exprimate categoric. De pildă, propoziția „Lupii sunt verzi." determina stocarea în baza de date a proprietății „verde" atașată noțiunii „lup".

În câteva luni de muncă, obținusem un (rudimentar și prototipic) *robot conversațional* – în engleză, un *chatbot*. În buna tradiție a numirii chatboților cu prenume omenești, am pus programului meu numele *Cătălin* (un

LIBRIS | We know books

prenume românesc și o trimitere la Luceafărul lui Eminescu, cel „nemuritor și rece") și am participat cu el la câteva concursuri de creație software, unde am reușit să impresionez, în ciuda faptului că programul meu era deosebit de limitat și instabil.

Deși am avut cu consecvență grijă să nu omit să prezint întotdeauna realizarea mea drept ceea ce credeam că este, respectiv un simplu robot conversațional care efectuează silogisme, m-am simțit de câteva ori surprins să aud din partea membrilor juriilor că ceea ce realizasem eu era, de fapt, „un fel de inteligență artificială". Pe atunci, nu consideram câtuși de puțin că modestul program *Cătălin* merita o astfel de descriere, fiindcă știam prea bine că era nu doar extrem de sărac în abilități și cunoștințe, ci și îngrădit la nivel structural de o funcționare eminamente deterministă. Cu toate acestea, o privire în istoria de mai puțin de un secol a subdomeniului din informatică intitulat „inteligența artificială" relevă că o astfel de descriere nu este neapărat eronată, ci poate doar anacronică.

Spectaculoasele reușite ale ultimilor (doar!) trei ani au încetățenit accepțiunea din acest moment a publicului larg față de inteligența artificială, care o asociază, de obicei, cu soluțiile generatoare de conținut multimedia ultrarealist contrafăcut (imagini, clipuri video/audio *fake*), dar mai ales cu oracolul virtual ChatGPT. Spre exasperarea experților IT care se încăpățânează să nu renunțe la viziunea clasică, mai largă asupra domeniului, pare că această accepțiune implicită câștigă tot mai mult teren, inclusiv în tagma programatorilor, care au găsit în arhitectura informatică ce stă la baza acestor

LIBRIS We know books

realizări, denumită *rețea neuronală*, un instrument cu aplicabilitate universală, pe care o adaptează pentru a rezolva pragmatic o cvasiinfinitate de probleme.

De fapt, însăși ideea de inteligență artificială a cunoscut, de-a lungul scurtei sale istorii, sensuri multiple, unele devenind rapid desuete, iar altele coexistând în paralel și uneori în conflict. Cu o evoluție sinuoasă, scrutată de filozofi și intens speculată academic și economic, pare că astăzi, inteligența artificială traversează o nouă epocă de aur în care nu mai reprezintă doar un domeniu de cercetare, ci a devenit și un furnizor redutabil de instrumente de cercetare și lucru pentru savanți, artiști și oameni de toate profesiile. Ultimul deceniu a adus o succesiune amețitoare de noi reușite, iar astăzi, experții se întrec în predicții despre următoarele redute ce vor fi cucerite de noua tehnologie.

În timp ce giganții IT concurează pentru a îndeplini aceste predicții, oamenii obișnuiți interacționează deja de ani buni cu programe și site-uri îmbogățite cu sisteme de inteligență artificială care le facilitează sarcini computerizate ce au devenit deja banale – culegerea de informații de pe internet, traducerile automatizate în timp real, recunoașterea imaginilor și a fragmentelor muzicale. Se ridică deja întrebări legitime: cât timp va mai accelera evoluția tehnologică la care asistăm și care va fi impactul socioeconomic al dezvoltării inteligenței artificiale în următorii ani?

Înainte de a fi tentați să răspundem impulsiv acestor scenarii apocaliptice, o privire în trecutul recent ne încredințează că merită să privim lucrurile cu mai

LIBRIS We know
multă prudență: fervoarea din jurul celei mai recente renașteri a IA este, cu siguranță, cea mai intensă de până acum, dar nu este nicidecum prima la care asistă comunitatea științifică. Pentru a înțelege mai bine contextul, mijloacele și țintele dezvoltării mașinilor inteligente, în cele ce urmează voi realiza o trecere în revistă a istoriei inteligenței artificiale și a conceptelor care au contribuit la multiplele sale definiri și redefiniri.

Gândește sau ne păcălește?

„Pot mașinile să gândească?"– cu această întrebare, care deschide faimoasa lucrare intitulată *Computer Machinery and Intelligence*[1], publicată în anul 1950, matematicianul englez Alan Turing a pornit dezbaterea științifico-filozofică despre capacitățile cognitive ale calculatoarelor. La acel moment, primul calculator complet electronic de uz general, ENIAC, fusese realizat de câțiva ani, timp în care ziarele din toată lumea anunțaseră deja – adesea bombastic și hipersimplificat – apariția noii invenții, descriind-o în termeni antropomorfici, îndeosebi drept „creierul electronic"[2]. Presa românească de atunci, cu o prudență greu de imaginat astăzi, a relatat și ea despre acest „așa-zis creer electric"[3] (sic), precizând că mașina de calcul „întrece creerul omenesc în iuțeală, logică și memorie"[4] (sic), însă, foarte important, fără a omite „excepția înclinării omului către gândirea creatoare"[5]. Pe un ton cel puțin la fel de sceptic, Ziarul Științelor și al Călătoriilor sugera, aproape un an mai târziu, că ENIAC „nu raționează"[6], pentru ca și mai târziu să acorde spațiu unui amplu editorial în care inventatorul britanic A. M.

LIBRIS We know books

Low afirma răspicat că „nu există nici cea mai mică posibilitate de a construi vreodată"[7] mașini electronice „care să gândească pentru noi"[8]. Impactul primelor anunțuri despre capacitățile noilor calculatoare a rămas totuși sesizabil și la decenii distanță, iar opinia populară a întreținut în continuare ideea că „ele pot gândi ca o ființă umană"[9].

În contextul acestor rumori, alimentate, fără îndoială, și de imaginația publicului, lucrarea lui Turing propune în premieră un „test de inteligență" destinat calculatoarelor. Demersul lui Turing nu este însă unul de factură raționalistă. Deviind de la întrebarea originală, care este împovărată de dificultățile de a defini ce este o mașină și ce înseamnă a gândi, Turing propune o problemă alternativă „înrudită îndeaproape" și „exprimată în termeni relativ clari". *Testul Turing* afirmă că dacă un actor uman, după ce a conversat cu un alt actor uman, respectiv cu o mașină, printr-o interfață identică (de pildă, un ecran și o tastatură), nu poate afirma cu certitudine care dintre cei doi actori este omul și care este mașina, atunci putem afirma că mașina manifestă un comportament inteligent.

Deși naiv și ușor de fentat, după cum vom vedea în continuare, Testul Turing realizează o serie de presupoziții importante, printre care se numără aceea că umanitatea unei entități (fie ea și o mașină) este dată nu de înfățișarea umană, înzestrată cu trăsături biologice, precum trupul sau vocea, ci de capacitățile intelectuale pe care le manifestă; că posibilitatea unei mașini de a gândi „cu adevărat" este o problemă secundară; și că ceea ce este mai abordabil în evaluarea

LIBRIS | We know books

inteligenței este *ceea ce* mașina realizează, nu *cum* realizează – mașina putând fi, din perspectiva evaluatorului, o *cutie neagră*. Alan Turing furnizează, astfel, o primă viziune modernă despre esența inteligenței: inteligența presupune *a acționa* omenește[10].

După publicarea articolului, considerat primul din istorie care a tratat chestiunea inteligenței artificiale, chiar dacă fără a o numi ca atare, nu a trecut mult până când savanții au remarcat că Testul Turing presupune o țintă prea ușoară, întrucât lipsa unor criterii legate de funcționarea internă a mașinii face îndeplinirea condiției testului realizabilă facil prin subterfugii în spatele cărora în mod evident nu se află un mecanism de efectuare a raționamentelor.

La mijlocul anilor 1960, informaticianul germano-american Joseph Weizenbaum a realizat un program capabil de a simula conversații (în termeni contemporani, un *chatbot*) în limba engleză ca parte dintr-un studiu privitor la comunicarea între om și mașină prin intermediul limbajului natural – o noutate absolută la acea vreme în ceea ce privește interacțiunea om-mașină. Funcționarea programului, denumit *ELIZA*[11], era surprinzător de simplă: fiecare mesaj introdus de utilizator era comparat cu mai multe șabloane ordonate după o prioritate prestabilită într-un set de reguli, până când acesta găsea o potrivire. În funcție de șablonul determinat, programul răspundea cu un mesaj prestabilit: fie unul static precum „I see"[12], fie unul dinamic, care relua cuvinte sau secvențe din interogările precedente ale utilizatorului – de pildă, la o interogare precum „I am often overly anxious"[13],

LIBRIS We know books

răspunsul ar fi putut fi „How long have you been often overly anxious?"[14]. Cel mai cunoscut set de reguli de conversație, denumit *DOCTOR*, simula o conversație cu un psihoterapeut care invită pacientul la autoreflecție, cel mai adesea reformulându-i afirmațiile sub forma unor întrebări clarificatoare.

În urma experimentelor, fără ca aceasta să fi fost intenția inițială a studiului, profesorul Weizenbaum a constatat cu surprindere că unii dintre utilizatorii care interacționau cu programul ELIZA au început să manifeste un atașament emoțional față de acesta, aparent trecând cu vederea că dincolo de ecran nu se află un actor uman și cu atât mai puțin unul rațional. Mai mult, într-o antologică anecdotă, secretara sa personală i-ar fi cerut profesorului să părăsească încăperea pentru a putea purta o discuție „adevărată" cu robotul.

Experimentul ELIZA a demonstrat fără echivoc trecerea Testului Turing prin tertipuri de inginerie socială și a fost consemnat drept un moment marcant în istoria inteligenței artificiale, dar și un exemplu reprezentativ de abordare superficială față de problema realizării unor mașini inteligente. Totodată, el a furnizat un răsunător studiu de caz pentru evidențierea tendinței oamenilor de a atribui calculatoarelor cu interfețe complexe, cel puțin la nivelul discursului, dacă nu și la nivel cognitiv, trăsături și acțiuni tipic omenești, un fenomen ce a primit numele *Efectul ELIZA*. Astăzi este atât de comun să ne raportăm în acest fel la calculatoarele din jurul nostru, încât un astfel de gest trece practic neobservat – de câte ori nu ne surprindem

LIBRIS We know books

spunând despre un program afectat de erori că „nu *vrea* să meargă" sau că „face fițe"; ori despre un program că „nu înțelege" o comandă; ori despre un sistem de operare că „nu știe" să deschidă un fișier; ori, după cum vom vedea mai departe, despre un program că „învață"?

Distincția dintre imitarea comportamentului inteligent și gândirea în adevăratul sens al cuvântului în contextul inteligenței artificiale a fost discutată, mai târziu, și de filozoful american John Searle, care, într-o lucrare[15] din 1980, a denumit în premieră cele două abordări, *inteligența artificială slabă*, respectiv *inteligența artificială tare* (identificată mai târziu și drept *inteligența artificială generală*). Tot atunci, Searle a avansat și un argument în favoarea tezei că un computer care procesează instrucțiuni nu poate dobândi o inteligență veritabilă, indiferent cât de bine reușește să o imite. Acest argument, denumit *Camera chinezească*, propune următorul experiment de gândire: să presupunem că un agent uman fără a avea studii de limbă chineză, aflat într-o cameră izolată primește o interogare scrisă în limba chineză și, după un timp, urmând un set de instrucțiuni la care are acces (ne putem imagina, unul destul de amplu!), produce un răspuns tot în limba chineză. În acest caz, persoana din cameră *cunoaște* limba chineză sau doar *mimează* cunoașterea ei? Mai mult, poate un calculator digital să producă o minte „adevărată" sau este acesta limitat structural la a produce cel mult o *proiecție* a minții? Deși dezbaterea filozofică în jurul chestiunii continuă, orizontul vag definit al dezvoltării IA a primit un nume: inteligența artificială generală.