

**ERIK J. LARSON**

**MITUL  
INTELIGENȚEI  
ARTIFICIALE**

**De ce computerele  
nu pot gândi la fel ca noi**

Traducere de Ovidiu-Gheorghe Ruța.

POLIROM  
2022

# Cuprins

Introducere .....	9
-------------------	---

## Partea întâi

### Lumea simplificată

Capitolul 1. Eroarea legată de inteligență.....	17
Capitolul 2. Turing la Bletchley.....	26
Capitolul 3. Eroarea legată de superinteligență.....	39
Capitolul 4. Singularitatea, trecut și prezent.....	50
Capitolul 5. Înțelegerea limbajelor naturale .....	56
Capitolul 6. Inteligența artificială drept kitsch tehnologic .....	66
Capitolul 7. Simplificări și mistere.....	74

## Partea a doua

### Problema inferenței

Capitolul 8. Nu calcula, analizează.....	93
Capitolul 9. Puzzle-ul lui Peirce (și nedumerirea lui Peirce).....	98
Capitolul 10. Probleme ale deducției și ale inducției .....	108
Capitolul 11. Învățarea automată și big data.....	133
Capitolul 12. Inferența abductivă. Originile inferenței ca presupunere .....	156
Capitolul 13. Inferența și limbajele (partea I) .....	187
Capitolul 14. Inferența și limbajele (partea a II-a) .....	200

## Partea a treia

### Viitorul mitului

Capitolul 15. Mituri și eroi .....	231
Capitolul 16. Mitologia inteligenței artificiale invadează neuroștiințele .....	239
Capitolul 17. Teoriile neocorticale despre inteligența umană .....	257
Capitolul 18. Sfârșitul științei? .....	263
<i>Note</i> .....	275
<i>Mulțumiri</i> .....	293
<i>Index</i> .....	295

## Puzzle-ul inferenței

Naratorul lui Edgar Alan Poe își găsea cu greu cuvintele pentru a descrie ceva despre care Peirce avea să scrie mai apoi volume întregi – inferența abductivă. Însă inferența abductivă e un tip de inferență. Ce e inferența? În primul rând, un substantiv. Verbul e „a infera”, care denotă o acțiune. Etimologic, „a infera” înseamnă „a face să se întâmple, a produce”, de la cuvintele latinești *in*, „în”, și *ferre*, „a aduce”. După *Oxford English Dictionary*, inferența e ceva ce facem din punct de vedere cognitiv, cu mintea noastră: „a-ți forma o opinie sau a decide că ceva e adevărat pe baza informațiilor disponibile”.

Din păcate, același dicționar ne spune că „a deduce” e sinonim cu „a infera”, ceea ce nu ne ajută deloc (deoarece deducția e doar un tip de inferență).

Ne oferă și câteva exemple, care subliniază caracterul general al termenului în uzul comun:

*A infera ceva (din altceva):* O mare parte a sensului trebuie inferată din context. Cititorii trebuie să infereze mobilul ucigașului.

*A infera că:* E rezonabil să inferăm că guvernul știa de aceste înțelegeri.

Inferența e producerea unui nou gând, ceea ce, în logică, înseamnă a trage o concluzie și, mai general, implică a folosi ceea ce știm deja și ceea ce vedem sau observăm pentru a ne aduce la zi credințele anterioare. Ca să folosim exemplul din dicționar, am putea infera mobilul ucigașului pe baza a ceea ce știm deja și a ceea ce am citit în ziare (cum a făcut Dupin).

Inferența e și un fel de salt, considerat rațional, ca atunci când inferăm că „guvernul știa de aceste înțelegeri” – din nou, pornind de la ceea ce știam anterior (de pildă, informații publice sau aflate de la alții) și de la informațiile aflate citind (sau observând) știrile noi.

Inferența e un act cognitiv de bază pentru mințile inteligente. Dacă un agent cognitiv (o persoană sau un sistem de inteligență artificială) nu e inteligent, va face inferențe proaste. Dar orice sistem care face inferențe trebuie să posede o inteligență elementară, fiindcă însuși actul de a folosi ce e cunoscut și ce e observat pentru a ne aduce la zi credințele e intrinsec legat de ceea ce numim „intelență”.

Dacă un sistem de inteligență artificială nu face deloc inferențe, nu merită să i se atribuie statutul de inteligență artificială. (Deși am putea spune că până și un sistem care etichetează imagini cu pisici face inferența că ceea ce „vede” e o pisică – așa că ștacheta poate fi lăsată destul de jos.)

E imposibil să înțelegi o glumă, să descoperi un nou vaccin, să rezolvi un caz de omucidere, ca Dupin, sau chiar să ții pasul cu diverse evenimente și informații din lume fără să fii capabil să faci inferențe. Sigur, știm multe lucruri, dar numai inferențele ne poartă spre noi cunoștințe (sau credințe). Știm că soarele va răsări mâine, deci nu trebuie să inferăm că e așa. La fel, nu ne obosim să inferăm că mâna ne e în continuare atașată de antebraț. Acestea sunt cunoștințe pe care le avem deja, convingeri deja formate. Dar ceea ce știm se schimbă și e adus la zi permanent. Dacă afară se întunecă în mod misterios prea devreme, am putea infera că are loc o eclipsă de soare sau poate că o puternică furtună de praf acoperă soarele la vest sau poate că are loc un holocaust nuclear. Depinde. Ce știm în momentul de față? Ce răspuns pare cel mai logic, dat fiind ceea ce vedem?

Într-un sens general, facem întotdeauna inferențe – e ca o condiție pentru a fi treaz. Se poate să intru în bucătărie, să găsesc o sticlă de Pepsi pe jumătate goală și să inferez că sora mea a lăsat-o acolo, fiindcă ea bea Pepsi și tocmai ne vizitează. Pe de altă parte, niște muncitori schimbă blaturile și am remarcat că mai devreme unul dintre ei bea Pepsi. De fapt, și eu am băut Pepsi mai devreme și am lăsat sticla, băută pe jumătate, pe verandă, așa că poate soția mea a adus-o înăuntru. Până la urmă, facem conjecturi pentru a obține o explicație care are noimă, pe baza a ceea ce știm și a contextului în care ne aflăm. Aceasta e inferența „în timp real”, fiindcă tragem concluzii în timp ce pătrundem în încăpere. În lumea reală, circumstanțele se schimbă permanent, deci inferențele în timp real sunt des folosite. La urma urmei, gândim în timp. Un program informatic care poate rezolva o problemă după zece miliarde de ani nu e absolut deloc inteligent, cum nu e inteligent nici unul care, în timp real, intră într-un zid.

Natura provizorie a multor inferențe înseamnă că inferențele inițiale pot fi greșite, mai ales dacă sunt făcute în pripă. Dacă ajung cu întârziere la birou, șeful ar putea infera că nu iau lucrurile în serios, când de fapt circulația a fost îngreunată din cauza unui accident. Cu alte cuvinte, șeful trage o concluzie pe baza unei

impresii preexistente sau a unei prejudecăți în ce mă privește. În conversațiile de zi cu zi, oamenii folosesc cuvântul „inferență” în acest sens, referindu-se la o concluzie neîntemeiată trasă în pripă: „Așa e Suzy, face tot felul de inferențe aberante despre tine după lucrurile pe care le-ai spus aseară”. Și e adevărat că, tehnic vorbind, Suzy face inferențe, dar sensul aici e că ele sunt părtinitoare, că Suzy se grăbește să facă presupuneri nedrepte (poate fiindcă e prost dispusă sau nu te place).

Într-un sens mai restrâns, inferența a intrat în vocabularul matematic cu multă vreme în urmă și, mai recent, a apărut în discuțiile despre computație și inteligența artificială. În acest context, „inferența în timp real” se poate referi la un robot care străbate un mediu dinamic, cum ar fi o stradă aglomerată. „Inferența probabilistică” trage concluzii pe baza datelor statistice, cu aplicații evidente în abordările bazate pe date ale inteligenței artificiale.

Cândva, cercetătorilor din domeniul inteligenței artificiale le-a pus mari probleme o condiție a inferenței, folosirea inteligentă a ceea ce știm deja – problema „cunoașterii”. Sistemele care nu știu nimic nu prea pot face inferențe. Așa încât cercetătorii din perioada de început au încercat să introducă cunoștințe în sistemele de inteligență artificială, pentru a le ajuta să înțeleagă datele primite de la senzori sau sub formă de text. Dar s-a descoperit (pe calea anevoioasă, prin eșecuri repetate) că sistemele de inteligență artificială cu arhive mari ce conțineau fapte concrete și reguli tot trebuiau să utilizeze cunoștințele în context pentru a trage concluzii relevante. Această „folosire” a cunoștințelor e ceea ce face ca inferența să fie atât de dificilă. Care părțică de cunoaștere este relevantă din carul cu fân al memoriei mele, aplicată la lumea din jur aflată într-o schimbare dinamică?

Capacitatea de a stabili care cunoștințe sunt relevante nu e o competență ce ține de puterea de calcul. Poe insistă că pe tărâmul „analiticului” nu se ajunge la intuițiile umane pe bază de formule; ele sunt „chestiuni aflate dincolo de limitele regulii ca atare” sau ale calculelor. Ba chiar Dupin pare să găsească autorul omorurilor – urangutanul – folosindu-se de o conjectură norocoasă, pe care mai apoi o confirmă întâlnindu-l pe stăpânul animalului pierdut. Deci: făcea doar conjecturi? Într-un sens important, da. Dar asta nu înseamnă că tot acel proces nu e o inferență. Dimpotrivă, înseamnă că e una de un tip important.

## Încă ceva despre Turing

În influentul său articol din 1950 „Computing Machinery and Intelligence”, Turing respinge întrebările privitoare la mașini care chiar ar gândi, ironizându-și propriul titlu și spunând că „gândirea” e iremediabil neștiințifică și subiectivă. A vorbi despre computere care gândesc e ca și cum ai vorbi despre submarine care înoată. Când spunem „înoată”, deja le antropomorfizăm. Delfinii înoată, submarinele nu. Turing credea că la fel se întâmplă și când folosim cuvântul „a gândi”. Dacă un computer ar juca șah, cine ar putea spune dacă acesta gândește sau doar calculează?

Pe Turing îl interesa o minte complet programabilă. Prin urmare, a renunțat la deosebirea pe care o făcea inițial între intuiție și ingeniozitate, aducând intuiția – orice ar fi ea – în sfera computației. În felul acesta, a făcut problema inteligenței artificiale complet testabilă. Teza era radicală chiar și după standardele lui anterioare, dar nu îl putem blama că a adoptat-o, fiindcă ea a creat condițiile necesare pentru ca ulterior, în același deceniu, cercetătorii din domeniul inteligenței artificiale să-și înceapă munca fără ca preocupările filosofice să îi împiedice să avanseze.

Din păcate, nu s-a discutat niciodată serios cum anume ar putea inferența computațională să fie – sau să devină – similară inferenței umane. Domeniul nu a pornit de la o teorie a inferenței, care ar fi oferit un plan pentru dezvoltarea ulterioară (sau o dovadă că acest lucru e imposibil). Faptul că cercetătorilor din domeniul inteligenței artificiale le lipsește o teorie a inferenței e ca și cum inginerii din domeniul nuclear ar începe lucrul la bomba nucleară fără să se fi elucidat mai întâi detaliile reacțiilor de fisiune. E clar că nu e suficient să se cunoască ecuația lui Einstein. Tot așa cum nu e suficient ca oamenii pasionați de inteligența artificială să cunoască doar teoria computațională – întrucât cercetătorii din domeniu trebuie să afle tocmai cum poate computația să fie convertită în gama și tipurile adecvate de inferențe folosite de mintea umană. Întrebarea trebuia pusă direct. Ignorând-o sau ocolind-o, domeniul a făcut ca speranțele false, piste care se termină în niște fundături și pierderea de timp să fie inevitabile.

Fiindcă sunt multe de analizat. Să ne gândim, de exemplu, la numeroasele inferențe din istoria științei. Savanții construiesc ipoteze, iar apoi le supun testelor. Dar ipotezele nu sunt create mecanic; e de notorietate că uneori apar brusc în mintea savanților (de obicei după ce aceștia își stăpânesc perfect domeniul). La fel ca Turing odinioară, cei care analizează cum s-au făcut descoperirile științifice tind să lase aceste salturi intelectuale în afara procedurilor practicii științifice, așa că actul central al inteligenței „călătorește fără bilet”, nefiind analizat el însuși. Dar asemenea ipoteze sunt acte autentice ale minții, esențiale pentru orice știință și adesea imposibil de explicat pe baza datelor, a probelor sau a oricăror lucruri evidente sau programabile<sup>7</sup>.

Când Copernic a afirmat că Pământul se rotește în jurul Soarelui, și nu invers, a ignorat o mulțime de dovezi și date acumulate de-a lungul secolelor de astronomi care au lucrat cu mai vechiul model al lui Ptolemeu. Copernic a redesenat totul cu Soarele în centru și a creat un model heliocentric utilizabil. Lucru important, modelul copernician inițial a fost de fapt mai puțin predictiv, deși era corect. La început, acesta a fost doar o schemă care, dacă era dusă până la capăt, putea furniza soluții elegante în locul celor tot mai complicate, cum ar fi mișcarea retrogradă a planetelor, care grevau modelul lui Ptolemeu. Copernic a putut să respingă modelul geocentric și să infereze o structură radical nouă pentru sistemul solar doar ignorând mai întâi toate datele sau reconceptualizându-le. (Observați că asta ridică o întrebare: cum l-ar fi ajutat „big data”? Toate datele se potriveau cu modelul greșit.)

Saltul copernician care a lansat revoluția științifică ar putea fi mai bine definit drept o coniectură inspirată. La fel s-ar putea spune și despre faptul că Kepler a ales o elipsă pentru a descrie mișcarea planetelor, fiindcă un număr extrem de mare (practic infinit) de forme geometrice pot fi adaptate orbitelor planetare (poate cu excepția celor transcendente, cum sunt sinusoidale). Elipsa nu era mai simplă decât toate celelalte – aceasta nu era o explicație de tip „briciul lui Ockham”. Coniectura lui Kepler a fost pur și simplu o explicație care lui i „s-a părut corectă”.

Faptul că coniecturile duc la descoperiri nu se potrivește cu prezentările mecanice ale științei. Dimpotrivă, le contrazice. Dar munca de detectiv, descoperirile științifice, inovațiile și simțul comun țin toate de funcționarea minții umane. Toate sunt inferențe pe care



cercetătorii care vor să ajungă la inteligența generală artificială trebuie să le ia cumva în calcul.

După cum vedeți, modelarea cognitivă – construirea unui computer care să gândească, să facă inferențe – generează nedumeriri. Cercetătorii din domeniul inteligenței artificiale (cel puțin deocamdată) ar trebui să își îndrepte cu precădere atenția spre inferență în contextul ei cotidian. De ce? Pentru că majoritatea inferențelor pe care le facem sunt aparent banale, precum diversele salturi și presupuneri pe care le facem în timpul conversațiilor obișnuite. Din păcate pentru acești cercetători, nici inferențele banale nu sunt ușor de transpus în programe. Testul Turing, de exemplu, e dificil în esență pentru că înțelegerea limbajelor naturale necesită multe inferențe bazate pe simțul comun, care nu sunt nici certe din punct de vedere logic, nici (adesea) foarte probabile. Cu alte cuvinte, necesită numeroase abducții.

De obicei, nici măcar nu remarcăm astfel de inferențe, ceea ce e bine: dacă le-am remarca, am tinde să ne blocăm în bucle solipsiste, urmărindu-ne propriile gânduri. Asta ne aduce înapoi la Peirce. Mai exact, ne aduce la structura tripartită a inferențelor pe care se sprijină inteligența: deducția, inducția și abducția.