

INTELIGENȚA ARTIFICIALĂ

Manfred Spitzer (n. 1958) este unul dintre cei mai cunoscuți specialiști în neuroștiințe din Germania, o personalitate foarte implicată în dezbaterile despre tehnologie și educație. A studiat medicina, psihologia și filozofia la Universitatea din Freiburg, iar în prezent este profesor de psihiatrie la Universitatea din Ulm, unde conduce din 1997 Clinica psihiatrică universitară și din 2004 Centrul pentru neuroștiințe și învățare. A fost în două rânduri profesor invitat la Universitatea Harvard. Este gazda emisiunii TV săptămânale *Geist und Gehirn* (*Minte și creier*), în care prezintă unui public larg rezultatele cercetărilor sale în domeniul neuroștiințelor.

A publicat numeroase cărți, dintre care *Lernen* (*Învățarea*), *Vorsicht Bildschirm!* (*Atenție, ecran!*) și *Digitale Demenz* (*Demența digitală*, trad. Humanitas, 2020, reed. 2024) au devenit bestsellere.

MANFRED SPITZER

**INTELIGENȚA
ARTIFICIALĂ**

**CUM NE AJUTĂ ȘI CUM NE AMENINȚĂ
O CREAȚIE SUPERIOARĂ OMULUI**

Traducere din germană
de Alexandru Bejinariu

HUMANITAS

Redactor: Adina Săucan
Coperta: Ioana Nedelcu
Tehnoredactor: Manuela Măxineanu
Corector: Marilena Vasile
DTP: Iuliana Constantinescu, Veronica Dinu

Tipărit la Livco Design

Manfred Spitzer
Künstliche Intelligenz
Copyright © 2023 by Droemer Verlag.
An imprint of Verlagsgruppe
Droemer Knaur GmbH & Co. KG, Munich

© HUMANITAS, 2025, pentru prezenta versiune în limba română

Descrierea CIP a Bibliotecii Naționale a României
Spitzer, Manfred
Inteligența artificială: cum ne ajută și cum ne amenință
o creație superioară omului / Manfred Spitzer;
trad. din germană de Alexandru Bejinariu. –
București: Humanitas, 2025
Conține bibliografie
ISBN 978-973-50-8934-4
I. Bejinariu, Alexandru (trad.)
004.8

EDITURA HUMANITAS
Piața Presei Libere 1, 013701 București, România
tel. 021/408 83 50, fax 021/408 83 51
www.humanitas.ro

Comenzi online: www.libhumanitas.ro
Comenzi prin e-mail: vanzari@libhumanitas.ro
Comenzi telefonice: 0723 684 194

Pentru Manfred Neumann

Cuprins

Cuvânt înainte	9
1. ChatGPT: duhul din lampă	15
2. Firul evenimentelor	38
3. Primii informaticieni construiesc primele calculatoare	61
4. Computerul-creier	74
5. Științele naturii: intuiția artificială și expertii umani.	103
6. Inteligența artificială: o realitate cotidiană	124
7. Științele umaniste: de la scrierea cuneiformă la hermeneutică. ...	144
8. Inteligența artificială în medicină	165
9. Fascinație și frică	186
10. Consecințe sociale: prejudecăți, manipulare, pierderea adevărului și a temelor pentru acasă	204
11. Dimensiunea militară a inteligenței artificiale	220
12. Inteligența artificială: salvare și amenințare	239
<i>Note</i>	269
<i>Referințe</i>	281
<i>Credite fotografice</i>	315

Cuvânt înainte

Am început să mă ocup cu rețelele neuronale acum mai bine de 30 de ani. Pe atunci, acestea erau modele computaționale care funcționau pe baza a câtorva zeci de neuroni interconectați. În ciuda dimensiunii lor reduse față de creierele reale, dotate cu mult mai multe celule nervoase (musculița de oțet: 250.000; șoarecele: 71 de milioane; șobolanul: 200 de milioane; macacul: 6,3 miliarde; omul: 86 de miliarde), chiar și aceste modele computaționale ne-au ajutat să înțelegem lucruri uimitoare. Dintr-odată a devenit clar cum funcționează învățarea, cum pot neuronii să simbolizeze ceva (problema reprezentării obiectelor în interiorul creierului), să stocheze și să proceseze informații (v. capitolul 4) și cum se face că ei, asemenea creierelor, se degradează cu demnitate, spre deosebire de un computer care „cade” brusc sau „iar se blochează”. „*Graceful degradation*” se numește proprietatea conform căreia, chiar și în ciuda pierderii a 50% până la 70% dintr-un creier sau dintr-o rețea neuronală, funcționarea ei se conservă uluitor de bine. Am sintetizat mai apoi numeroasele noi descoperiri din cercetarea rețelelor neuronale, promovată mai ales în SUA, în prima mea carte de popularizare *Geist im Netz* (*Mintea din rețea*), menită să familiarizeze publicul larg cu o tematică în mare măsură încă necunoscută la acea vreme în spațiul european. În ciuda realizărilor impresionante publicate deja pe atunci, nici nu aș fi visat că toate acestea vor defini într-o bună zi modelul de afaceri al celor mai bogate companii din lume (Alphabet, Amazon, Apple, Meta și Microsoft). În prezent, rețelele neuronale pe care aceste companii sunt capabile să le simuleze sunt uriașe și vor depăși în câțiva ani numărul de neuroni

din creierul uman. Cu ajutorul lor se evaluează comportamentul fiecărui individ dintr-o mare parte a omenirii, în vederea, printre altele, a eficientizării publicității personalizate.

Pentru mine, rețelele neuronale au fost un instrument pentru înțelegerea creierului, iar ele joacă în continuare acest rol pentru neuroștiințe: dacă astăzi putem identifica percepțiile și gândurile unei persoane (fără să o întrebăm) pe baza activității sale cerebrale măsurate cu un scanner (RMN), acest lucru este cu puțință numai datorită utilizării rețelelor neuronale (Gaziv *et al.*, 2022; Tang *et al.*, 2023). În trecut, ca și în prezent, acest fapt și-a dovedit relevanța cu totul deosebită pentru psihiatrie. Psihiatri precum Ralph Hoffman de la Yale, Jonathan Cohen și David Servan-Schreiber din Pittsburgh, precum și mulți alții au făcut pentru prima dată legătura între rețelele neuronale și fenomene psihopatologice precum halucinațiile (Hoffman, 1987) sau tulburările formale de gândire (Servan-Schreiber, Cohen, 1990). Cu toate acestea, propriile mele lucrări despre delir (Spitzer, 1995) și tulburările formale de gândire (1997) nu m-au făcut să întrevăd la momentul respectiv că, 25 de ani mai târziu, *Computational Psychiatry* (psihiatria computațională) avea să devină o veritabilă ramură de cercetare în domeniul psihiatriei (Macpherson *et al.*, 2021; Friston, 2023).

Elementul deosebit în scrierea acestei cărți a constat în aceea că obiectul cercetării (și, așadar, conținutul cărții) prindea contur în timp ce scriam. Multe dintre exemple, de la utilizarea IA în războiul din Ucraina și până la prognozele meteo sau cele seismice, sunt de anul acesta*. Așa s-a mai întâmplat doar în cazul unei singure alte cărți, și anume la *Pandemie*, pe care am scris-o în toilul primului val de Covid-19 din primăvara anului 2020. Pe atunci, totul se schimba de la o zi la alta, atât în sfera profesională, cât și în cea privată. Cum în cazul celor cu afecțiuni grave sistarea oricărei activități nu era nicidecum o variantă posibilă, am fost nevoiți în spital să reinventăm psihiatria pentru a mai putea să o practicăm – în rutina clinică de zi cu zi, ea presupune o serie de activități de grup, deoarece aproape toate tulburările psihice se

* Este vorba despre anul 2023 (n. tr.).

asociază cu probleme legate de faptul de a fi împreună. Fărăma de viață privată rămasă după foarte multe ore petrecute în clinică era, de asemenea, în condițiile izolării cu totul altfel – așa cum știe oricine a trăit acele vremuri.

Din fericire, IA nu este o pandemie, dar, în primul rând, excep-tând războiul din Ucraina și criza climatică, despre nimic altceva nu se discută la fel de des. În al doilea rând, multe studii despre IA tratează informații sau posibilități cu totul noi sau chiar „progrese” din cadrul celor mai diferite domenii. Este astfel limpede, în al treilea rând, că IA ne va schimba tuturor viața în cel mai mare grad.

Cunoașteți pe cineva care vrea să știe dacă inteligența artificială poate face cafea? Nu? Nici eu. Însă unii oameni o consideră o între-bare importantă, de exemplu Steve Wozniak, cofondatorul Apple. El a susținut că i-ar atribui unei IA „inteligență generală” doar atunci când un robot controlat de IA ar putea să intre într-o casă străină, să găsească bucătăria și să facă acolo o ceașcă de cafea (va-săzică ar trece *testul cafelei*). De la începutul anilor '50, când mate-maticianul și informaticianul britanic Alan M. Turing (1912–1954) s-a întrebat dacă și cum s-ar putea distinge între un computer și un om, comunicând cu ei doar prin intermediul unui teleimprimator, în virtutea acestui scop, s-au tot propus asemenea noi „teste” – nu-mite *teste Turing*. În cazul *testului IKEA* (numit și *Flat Pack Furniture Test*), IA trebuie să asambleze o piesă de mobilier ale cărei părți se află, laolaltă cu instrucțiunile de montaj, într-un pachet sigilat. Trecerea probei de admitere la o universitate americană sau a tes-tului de angajare într-o firmă au fost propuse, de asemenea, drept criteriu pentru inteligența reală (deosebită de inteligența mimată) a mașinilor. Și, cu toate că doar în ultima vreme 442 de autori au discutat în total 204 de asemenea posibile teste (Srivastava *et al.*, 2023), nu cred că ne ajută la înțelegerea inteligenței artificiale doar să ne întrebăm ce ar trebuie să facă o IA pentru a părea umană.

Mulți sunteți probabil deja familiarizați cu procesul invers – un test prin care omul trebuie să convingă computerul că nu este și el tot un computer –, și l-ați întâlnit ori de câte ori ați avut de rezolvat un CAPTCHA pentru a naviga spre următoarea pagină

pe internet. „CAPTCHA“ este un acronim pentru *Completely Automated Public Turing test to tell Computers and Humans Apart* („test Turing public complet automatizat pentru a distinge între computere și oameni“). Este vorba despre sarcini (prezentate în formă grafică) ce pot fi rezolvate cu ușurință de către oameni, dar nu și de computere. Menirea lor este să asigure securitatea și să împiedice alte computere sau roboți să se dea în mod abuziv drept oameni și să completeze formulare pe internet. Ne spune aceasta ceva despre felul în care suntem diferiți de IA? – Dacă vă gândiți că „da“, atunci ar trebui să știți că între timp CAPTCHA-urile au ajuns să fie rezolvate cu succes și de IA (George *et al.*, 2017).

Același lucru este valabil, din punctul meu de vedere, și pentru întrebarea dacă IA are conștiință. Pe 20 iunie 2023, specialistul în neuroștiințe Christof Koch a pierdut un pariu făcut în urmă cu 25 de ani cu filozoful Chalmers (Finkel, 2023; Cogitate Consortium *et al.*, 2023; Melloni *et al.*, 2023; Wnuk, 2023). La acea vreme, el pariașe că întrebarea „Ce este conștiința?“ va primi un răspuns definitiv de la neuroștiințe în următorul sfert de veac. Dar s-a înșelat, iar lada de vin roșu i-a revenit filozofului. Cât timp este evident că nu știm ce este conștiința, orice discuție legată de faptul că cineva sau ceva are sau nu conștiință este sortită eșecului. Situația e asemănătoare și în cazul altor afirmații făcute recent: IA nu are emoții; oamenii dispun, spre deosebire de IA, de intuiții. Cea dintâi afirmație depinde foarte mult de ce anume se înțelege prin „emoție“, iar cea de-a doua este pur și simplu falsă.

În fine, această problemă se ivește și sub denumirea de inteligență artificială *generală* (Artificial General Intelligence, AGI). Astfel, compania DeepMind a publicat o IA cu numele de *Gato*, care poate să joace jocuri Atari, să producă titluri pentru imagini, să stea de vorbă cu oameni, să clădească cu ajutorul unui braț robotic cărămizi și să îndeplinească alte 596 de sarcini (Sparks, 2022). Nimeni altul decât cercetătorul-șef al Facebook în domeniul IA, Yann LeCun, s-a pronunțat în legătură cu acest meșter priceput la toate (*eierlegende Wollmilchsau*), spunând că nu este foarte impresionat de faptul că se pun în aceeași oală mai multe abilități și apoi se afirmă că în oală există deodată inteligență generală.

Cu această carte, încerc să îi întâmpin pe cititori tocmai acolo unde ei se află. Toți ne confruntăm aproape zilnic cu comunicate de presă despre succesele, „progresele” sau „revoluțiile” din cercetarea în domeniul inteligenței artificiale. Ce este IA, ce este în stare să facă ea, cum funcționează și unde sunt problemele – toate acestea vor fi discutate în cartea de față. Spre deosebire de un manual despre inteligența artificială care ar începe cu multă matematică, eu am ales un alt mod de prezentare. Utilizând exemple din domenii cât se poate de diferite, voi arăta care sunt proprietățile specifice ale IA, clarificându-se totodată și modul ei de funcționare. În acest fel, IA poate fi înțeleasă cu totul intuitiv – la fel cum se poate înțelege intuitiv, adică fără apel la teoria evoluției sau la teorema lui Pitagora, ce este o broască, un copac sau un triunghi dreptunghic.

Le mulțumesc lui Georg Grön, Thomas Kammer, Dennis Kätzel, Doreen Scheiwe și Friedrich Uehlein, care au citit fiecare mai multe capitole ale cărții, oferindu-mi sugestii valoroase. Jürgen Bolz a redactat textul și l-a făcut mai plăcut de citit. Margit Ketterle de la Editura Droemer a însoțit volumul încă de la început. Mi-am asumat eu traducerea citatelor din limba engleză, fără să mai menționez de fiecare dată acest lucru. Multe dintre ilustrații sunt de asemenea făcute de mine, iar indicele l-am făcut de mână. Toate greșelile îmi aparțin, căci a fi autor înseamnă să-ți asumi responsabilitatea pentru text. Cartea îi este dedicată lui Manfred Neumann, care m-a introdus în domeniul rețelelor neuronale și, în decursul a nenumărate conversații, m-a făcut să văd mai clar multe lucruri.

Ulm, iulie 2023
Manfred Spitzer

ChatGPT: duhul din lampă

Pe 30 noiembrie 2022, compania americană OpenAI a oferit acces liber publicului la o inteligență artificială (IA) cu care se poate conversa. De atunci, întreaga lume vorbește despre *intelența artificială*, prescurtată cu majuscule ca „IA” deoarece este vizat un *terminus technicus*. La doar cinci zile după activarea sa, ChatGPT avea un milion de utilizatori, iar în ianuarie 2023, așadar după mai puțin de două luni, a ajuns la 100 de milioane! Făcând o comparație, Instagram, rețeaua de socializare pentru vizionarea și distribuirea imaginilor (sau clipurilor), care apăruse cu zece ani mai devreme și care aparține de Facebook, a avut nevoie de aproximativ doi ani și jumătate pentru a ajunge la 100 de milioane de utilizatori. Chiar și TikTok, aplicația de smartphone (app) lansată în anul 2018 pentru vizionarea și distribuirea videoclipurilor scurte, a avut nevoie de nouă luni până să atingă cota de 100 de milioane de utilizatori (Anon, 2023b).

ChatGPT nu este un program de computer, ci o rețea neuronală (v. capitolul 4) pentru procesarea (*processing*) limbajului natural (*natural language*). Se mai vorbește și despre modele pentru *Natural Language Processing (NLP)*¹, iar până astăzi au fost dezvoltate numeroase sisteme de acest fel. Ele diferă în funcție de mărimea rețelei – ChatGPT are 175 de miliarde de sinapse* – și de

* Mai exact, parametri. Cifrele sunt valabile pentru GPT-3/GPT-3.5. La momentul august 2025, se estimează că datele de antrenare pentru GPT-4 sunt de circa 1,7-1,8 trilioane de parametri, iar pentru GPT-4o, de circa 200 miliarde de parametri (sursa: <https://aiexpjourney.substack.com/p/the-number-of-parameters-of-gpt-4o>) (n. ed. rom.).