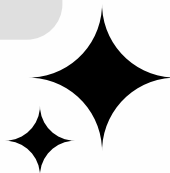
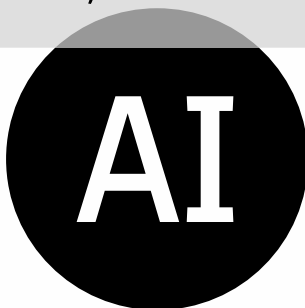


Co-Inteligență

Cum să trăiești și să
muncești cu



Traducere
din engleză de
Roxana Măciucă

Ethan Mollick



Cuprins

Introducere: Trei nopți nedormite	9
---	---

PARTEA ÎNTÂI

1. Crearea de minți extraterestre	21
2. Alinierea extraterestrului la valorile umane	45
3. Cele patru reguli ale co-inteligenței	63

PARTEA A DOUA

4. AI – persoană	83
5. AI – entitate creativă	109
6. AI – coleg de muncă	139
7. AI – profesor	173
8. AI – îndrumător	191
9. AI – viitorul nostru	205
Epilog: AI – noi înșine	223
Mulțumiri	225
Note	227

Introducere

Trei nopți nedormite

Cred că prețul plătit pentru a înțelege inteligența artificială (AI)* – **de a o înțelege** cu adevărat – este de cel puțin trei nopți nedormite.

După câteva ore de utilizare a sistemelor de inteligență artificială generativă, va veni un moment în care îți vei da seama că modelele lingvistice mari (LLM)** – noua formă de AI care alimentează servicii precum ChatGPT – nu funcționează așa cum te-ai aștepta să o facă un calculator. În schimb, se comportă mai degrabă ca o persoană. Realizezi brusc că interacționezi cu ceva nou, ceva extraterestru***, și că lucrurile sunt pe cale să se schimbe. Nu poți să adormi, fiind entuziasmat și neliniștit în egală măsură, și te întrebi: Cum va arăta jobul meu? Ce fel de meserii vor putea avea copiii mei? Oare chestia asta gândește? Te întorci la calculator în toiul nopții și ceri inteligenței artificiale lucruri care par imposibile, doar pentru a vedea cum le îndeplinește. Îți dai seama că lumea s-a schimbat fundamental și că nimeni nu îți poate spune cu adevărat cum va arăta viitorul.

* S-a optat pentru acronimul englezesc AI (artificial intelligence) în loc de IA (inteligență artificială), dată fiind frecvența folosirii acestuia în limba română (n.t.).

** În original, în limba engleză, Large Language Models – se va folosi acronimul englezesc datorită frecvenței folosirii în această formă (n.t.).

*** În original, în limba engleză, alien, poate fi tradus ca „străin” sau „diferit”; chiar dacă AI, respectiv LLM, sunt create de om, pe Terra, am ales să folosim adjectivul „extraterestru” pentru a accentua ideea de straniețate (n.t.).

Deși nu sunt informatician, sunt profesor, studiez inovația și sunt implicat de mult timp în studiul aplicațiilor inteligenței artificiale, mai ales în educație. De-a lungul anilor, inteligența artificială a promis mult mai mult decât a livrat. Timp de decenii, cercetarea în domeniu a părut mereu pe punctul de a face un progres major, dar cele mai multe aplicații practice – de la mașini autonome la instructori AI personalizați – au înregistrat progrese destul de lente.¹ În tot acest timp, am continuat să experimentez cu instrumente AI, inclusiv cu GPT de la OpenAI, încercând să le integrez în munca mea și cerându-le studenților mei să folosească și ei AI. Așa că am început să pierd nopțile imediat după lansarea ChatGPT în noiembrie 2022.

După doar câteva ore, era clar că avusese loc o schimbare uriașă între versiunile anterioare de GPT și acest model nou. La patru zile după lansarea inteligenței artificiale, am decis să arăt acest nou instrument studenților de la cursul meu de antreprenariat. Aproape nimeni nu auzise de el. În fața studenților mei, am făcut o demonstrație despre cum AI poate să genereze idei, să scrie planuri de afaceri, să transforme acele planuri în poezii (nu că ar fi o cerere mare pentru asta) și, în general, să joace rolul unui cofondator de companie. Până la sfârșitul cursului, unul dintre studenții mei, Kirill Naumov, reușise deja să creeze un demo funcțional pentru proiectul său de antreprenariat – o ramă foto inspirată din seria Harry Potter, care reacționa atunci când oamenii se apropiau de ea – folosind o bibliotecă de cod pe care nu o mai utilizase niciodată, în mai puțin de jumătate din timpul de care ar fi avut nevoie în mod normal. A doua zi, deja primea mesaje de la investitori.

La două zile după ce le-am prezentat studenților cum funcționa AI, câțiva mi-au spus că au folosit ChatGPT pentru a le explica anumite concepte pe care nu le înțelegeau „ca și cum erau niște copii de zece ani”. Au început să nu mai ridice

mâna atât de des în timpul cursurilor – de ce să se expună judecăților când ar fi putut să întrebe AI mai târziu? Și brusc fiecare eseu era scris perfect din punct de vedere gramatical (deși adeseori referințele erau incorecte și ultimul paragraf începea, în general, cu „În concluzie” – un semn distinctiv al textelor generate de ChatGPT, odată ce a fost îmbunătățit). Dar studenții nu erau doar încântați, erau și îngrijorați. Voiau să știe cum avea să arate viitorul.

Unii dintre ei m-au întrebat ce însemna asta pentru carierele pe care voiau să le urmeze („Să mai devin radiolog dacă AI poate să facă mare parte din această muncă?”, „Peste cinci ani va fi în continuare o meserie bună să scrii texte de marketing?”). Alții m-au întrebat când se va opri progresul tehnologic sau dacă se va opri. Cineva chiar a întrebat dacă inteligența artificială generală (AGI)*, mașină ipotetică ce ar fi mai inteligentă decât omul, ar putea să fie dezvoltată înainte de absolvire.

Nu aveam răspunsuri atunci (deși acum am câteva), dar am înțeles că nici eu nu sunt imun la aceste schimbări. Așa că am decis să mă implic direct și să testez AI. O mare parte din munca mea din ultimii cinci ani a constat în construirea de simulări pentru a preda abilități de afaceri. A fost nevoie de investiții enorme din partea universității mele și de eforturile a doisprezece oameni pentru a crea aceste experiențe digitale complexe. După mii de ore, lecțiile noastre erau excelente: cu mii de linii de cod, reușeam să construim simulări de învățare, care ajutau la predarea unor abilități precum negocierea. Dar apoi am decis să-i cer ChatGPT următoarele:

* *Artificial General Intelligence* – inteligență care are capacitatea de a înțelege și de a îndeplini orice sarcină realizată de om. Ca în cazul AI și LLM, se va folosi acronimul englezesc pe parcursul cărții (n.t.).

Vei fi profesorul meu care mă va învăța cum să negociez. Vei simula un scenariu detaliat în care trebuie să iau parte la o negociere. Tu vei juca rolul unei părți. Eu îl voi juca pe cel al celeilalte părți. Îmi vei cere părerea în fiecare etapă a scenariului și vei aștepta până o vei primi. După ce vei primi răspunsul meu, îmi vei spune detaliat ce face și ce spune cealaltă parte. Vei da o notă răspunsului meu și îmi vei oferi feedback detaliat despre ce anume să îmbunătățesc folosindu-mă de știința negocierii. Îmi vei da un scenariu pentru o negociere mai dificilă dacă mă descurc bine și un scenariu pentru o negociere mai ușoară dacă nu mă descurc.

ChatGPT mi-a dat următorul răspuns:

- Sigur, aș fi bucuros să te ajut să-ți exersezi capacitatea de a negocia printr-un exercițiu de simulare! Să începem cu un scenariu simplu. Ești un agent de vânzări care încearcă să vândă 100 de pixuri unui client. De obicei, pixurile costă un dolar, dar clientul încearcă să negocieze prețul. El îți oferă 0,50 dolari pentru un pix. Cum răspunzi?

Și apoi m-a ghidat printr-o întrebare de simulare care, deși nu perfectă, făcea 80% din ceea ce echipa mea reușise să creeze în luni de muncă. Totul, doar pe baza unui singur prompt de un paragraf!

Așa am început să nu mai dorm noaptea.

Și, ca să fiu sincer, încă nu dorm. În fiecare săptămână, inteligența artificială pare să vină fie cu un nou miracol, fie cu

o nouă problemă îngrijorătoare. Curând, m-am trezit certându-mă cu un chatbot care mă acuza că sunt lipsit de etică atunci când i-am cerut ajutor pentru a scrie un cod. Am descoperit că pot crea imagini minunate spunând cuvintele potrivite unei mașinării. Puteam să scriu cod în Python, un limbaj de programare pe care nu îl cunoșteam. Am descoperit că o mașinărie poate să îndeplinească proporții mari din munca mea... și poate și din munca altor oameni. Am descoperit ceva extrem de apropiat de o co-inteligență extraterestră, una care poate să interacționeze bine cu oamenii, fără să fie o ființă umană sau măcar o ființă conștientă. Cred că în curând cu toții vom începe să pierdem nopțile.

Și așa nedormit cum sunt acum, mă tot întorc la întrebările studenților mei: Ce înseamnă această tehnologie pentru viitorul pieței muncii și al educației? Lucrurile se petrec atât de repede încât e greu să fim siguri, dar începem să-i vedem impactul.

Inteligența artificială este ceea ce noi, cei care studiem tehnologia, numim tehnologie de uz general (General Purpose Technology – ironic, tot GPT). Aceste progrese apar o dată la o generație – precum motorul cu aburi sau internetul – și influențează fiecare industrie și fiecare aspect al vieții noastre. Și, într-un fel, inteligența artificială generativă ar putea avea un impact chiar și mai mare de atât.

Tehnologiile de uz general sunt, de obicei, adoptate lent, deoarece necesită ca multe alte tehnologii să funcționeze bine. Internetul este un exemplu perfect. Deși a apărut ca ARPANET la sfârșitul anilor 1960, a fost nevoie de aproape trei decenii pentru a deveni de uz general în anii 1990, odată cu inventarea browserului web, dezvoltarea computerelor accesibile și extinderea infrastructurii necesare pentru internetul de mare viteză. Au mai trecut încă 50 de ani până când smartphone-urile

au permis apariția rețelelor sociale. Totuși, multe companii nu au incorporat pe deplin internetul: „digitalizarea” unei afaceri este încă un subiect de discuție delicat la facultățile de administrare a afacerilor, mai ales că multe bănci încă folosesc calculatoare mainframe. Și tehnologiile de uz general anterioare au avut nevoie de zeci de ani de la dezvoltare până să fie puse în aplicare. Să luăm ca exemplu calculatoarele, o altă tehnologie care a produs transformări majore. Primele calculatoare au fost îmbunătățite rapid, datorită Legii lui Moore care descrie o tendință pe termen lung potrivit căreia capacitatea calculatoarelor se dublează o dată la doi ani.² Dar tot a fost nevoie de zeci de ani ca acestea să apară în firme și școli, deoarece, în ciuda creșterii rapide a capacității de procesare, acestea au pornit de la un stadiu foarte rudimentar. Și totuși, modelele lingvistice mari au dovedit o capacitate extrem de mare la doar câțiva ani de la inventarea lor. De asemenea, au fost adoptate de consumatori extrem de rapid; ChatGPT a ajuns la 100 de milioane de utilizatori, mai repede decât orice alt produs din istorie, datorită faptului că a fost gratuit, accesibil tuturor și incredibil de util.³

Mai mult, aceste modele devin tot mai bune. Dimensiunea^{*} lor crește cu un ordin de mărime^{**} pe an, ceea ce le îmbunătățește constant capacitățile. Deși acest progres va încetini probabil în viitor, are loc într-un ritm care eclipsează orice altă tehnologie majoră. Și LLM-urile reprezintă doar o parte dintr-o serie mai mare de tehnologii de învățare automată care alimentează noul val al inteligenței artificiale. Chiar

* Este vorba despre numărul de parametri, o valoare ajustabilă din rețeaua neuronală a unui model AI. Cu cât un model AI are mai mulți parametri, cu atât are o performanță mai mare, putând să învețe și să rețină mai multe tipare din datele care îi sunt furnizate pentru a fi antrenat (n.t.).

** Adică de zece ori (n.t.).

și dacă dezvoltarea AI s-ar opri în clipa în care termin această propoziție, tot ar avea un impact profund asupra vieților noastre.

În cele din urmă, oricât de extraordinare ar fi tehnologiile de uz general, impactul acestora asupra muncii și educației ar putea să fie mai mic decât cel produs de AI. Dacă revoluțiile tehnologice anterioare presupuneau adeseori o muncă mai mecanică și mai repetitivă, AI acționează, în multe cazuri, ca o formă de co-inteligență. Îmbunătățește gândirea umană sau chiar o poate suplini, oferind răspunsuri uimitoare. Primele studii asupra efectelor pe care le produce inteligența artificială au descoperit că poate duce adeseori la o creștere a productivității cu 20% până la 80% într-o varietate de joburi, de la programare la marketing. În schimb, când motoarele cu aburi, una dintre cele mai importante tehnologii de uz general, care a dus la declanșarea Revoluției industriale, au început să fie folosite în fabrici, au dus la o creștere a productivității cu 18% până la 22%.⁴ Și în ciuda zecilor de ani de cercetare, economiștii s-au confruntat cu dificultăți când au trebuit să arate impactul real pe care calculatoarele și internetul l-au produs asupra creșterii productivității pe termen lung în ultimii douăzeci de ani.⁵

În plus, tehnologiile de uz general nu influențează doar munca, ci ating toate aspectele vieții noastre. Schimbă modul în care învățăm, ne distrăm, interacționăm unii cu alții și chiar și percepția noastră despre noi înșine. Școlile se află într-un impas legat de viitorul temelor scrise, din cauza primei generații de inteligență artificială, iar instructorii AI ar putea, în cele din urmă, să schimbe radical modul în care educăm elevii și studenții. Divertismentul generat de AI permite crearea unor povești personalizate și provoacă unde de șoc la Hollywood. De asemenea, dezinformarea generată de AI se

răspândește deja pe rețelele sociale într-un mod care este greu de detectat și de combătut. Lucrurile sunt pe cale să devină foarte ciudate; de fapt, dacă știi unde să te uiți, deja sunt.

Și toate acestea ignoră problema cea mai mare, extra-terestrul din încăpere. Am creat ceva ce a convins mulți oameni deștepți că este, într-un fel, apariția unei noi forme de inteligență. O inteligență artificială care a trecut complet testul Turing (Poate un calculator să păcălească un om astfel încât să creadă că este om?) și testul Lovelace (Poate un calculator să păcălească un om când vine vorba de sarcini a căror rezolvare necesită creativitate?) la o lună de la inventarea sa, o inteligență artificială care obține note maxime la cele mai grele examene ale noastre, de la examenul pentru intrarea în barou la examenul de specialitate în neurochirurgie. O inteligență artificială care obține cele mai mari scoruri posibile la testele pe care le folosim pentru a măsura creativitatea umană și care trece cele mai avansate teste pentru a determina dacă o entitate este conștientă sau nu. Și mai ciudat este faptul că nu este complet clar de ce AI poate face toate aceste lucruri, chiar dacă noi am construit sistemul și înțelegem din punct de vedere tehnic cum funcționează.

Nimeni nu știe cu adevărat unde duce tot acest proces, nici măcar eu. Și totuși, chiar dacă nu am răspunsuri certe, cred că pot să fiu un ghid util. Am ajuns să fiu o voce influentă în ceea ce privește implicațiile inteligenței artificiale, mai ales prin newsletterul meu, *One Useful Thing*, chiar dacă nu sunt informatician. De fapt, unul dintre avantajele pe care le am legat de înțelegerea inteligenței artificiale este că, în calitate de profesor la Wharton, am studiat și am scris mult timp despre modul în care *sunt utilizate* tehnologiile.⁶ Drept urmare, împreună cu coautorii mei, am publicat unele dintre primele studii⁷ despre folosirea AI în educație și afaceri și am testat

utilizările practice⁸ ale inteligenței artificiale, fiind dați ca exemplu chiar de marile companii din domeniul inteligenței artificiale. În plus, colaborez frecvent cu diverse organizații, companii și agenții guvernamentale, precum și cu mulți experți în domeniul AI, pentru a înțelege lumea pe care o construim. De asemenea, încerc să țin pasul cu valul de cercetări din domeniu, majoritatea sub formă de articole științifice care nu au trecut prin procesul îndelungat de evaluare editorială, dar care oferă informații importante despre acest nou fenomen (voi cita mult din aceste lucrări pentru a completa tabloul spre care ne îndreptăm, dar este important să realizăm că acest domeniu evoluează rapid). Pe baza tuturor acestor conversații și articole, pot să te asigur că nimeni nu are imaginea completă despre ceea ce înseamnă AI și nici măcar oamenii care creează și utilizează aceste sisteme nu înțeleg implicațiile depline.

Prin urmare, vreau să încerc să te iau într-un tur al inteligenței artificiale, prezentată ca un fenomen nou în lume, o formă de co-inteligență, cu toate ambiguitățile pe care acest termen le implică. Am inventat instrumente și tehnologii, de la topoare la elicoptere, care să ne maximizeze capacitățile fizice; și altele precum foile de calcul care să permită automatizarea unor sarcini complexe; dar niciodată nu am creat o tehnologie aplicabilă universal, care să ne stimuleze inteligența. Acum avem acces la un instrument care poate simula felul în care gândim și scriem, acționând ca o co-inteligență pentru a ne îmbunătăți munca (sau a ne înlocui). Dar multe dintre companiile care dezvoltă AI merg și mai departe, încercând să creeze o mașină conștientă, o formă cu adevărat nouă de co-inteligență, care ar putea coexista cu noi pe Pământ. Pentru a înțelege ce înseamnă acest lucru, trebuie să începem cu o întrebare de bază: Ce este AI?

Așadar, vom începe de aici, discutând despre tehnologia din spatele modelelor lingvistice mari. Acest lucru ne va oferi o bază solidă pentru a înțelege cum noi, ca oameni, putem lucra cel mai bine cu aceste sisteme. Apoi, vom explora cum inteligența artificială ne poate schimba viața, devenindu-ne coleg de muncă, profesor, expert și chiar companion. În cele din urmă, vom analiza ce ar putea însemna acest lucru pentru noi și ce înseamnă, de fapt, să gândim alături de o minte extraterestră.

Partea întâi



1 | Crearea de minți extraterestre

Discuțiile despre AI pot crea confuzie, pe de o parte pentru că AI înseamnă atât de multe lucruri și există tendința să se facă confuzie între toate acestea. Siri, care îți spune o glumă la comandă. Terminator, care zdrobește un craniu. Algoritmii, care prezic scoruri de credit.

Am fost mereu fascinați de mașinile care pot gândi. În 1770, inventarea primului automat de șah i-a uluit pe cei care l-au văzut – o tablă de șah așezată pe o vitrină elaborată, ai cărei pionii erau mutați de un robot îmbrăcat ca un vrăjitor otoman. Această mașinărie, cunoscută și sub numele de Turcul Mecanic, a făcut înconjurul lumii între 1770 și 1838.¹ A reușit să-i învingă la șah pe Benjamin Franklin și Napoleon, determinându-l pe Edgar Allan Poe să speculeze despre posibilitatea existenței inteligenței artificiale după ce a văzut această mașinărie în anii 1830. Desigur, totul era o înșelătorie – în spatele mecanismelor sale false se ascundea în mod abil un adevărat maestru de șah. Însă, timp de aproape trei sferturi de secol, unele dintre cele mai strălucite minți ale lumii s-au lăsat păcălite de dorința noastră de a crede că mașinile ar putea gândi.

Derulăm pe repede-înainte până în anul 1950, când o jucărie și un experiment de gândire, fiecare creat de un geniu dintr-un domeniu aflat la începuturile sale (informatică), au condus la o nouă concepție despre inteligența artificială. Jucăria era un șoarece mecanic cu cheiță pe nume Theseus, dezvoltat de