

ANDREI DRĂGULINESCU

Între libertate și captivitate digitală

**Cum tehnologia ne modelează viitorul,
dar ne și prinde în lanțurile sale**

*(Inteligența artificială, 5G și 6G
TikTok, Facebook, YouTube)*



Colecția *Dosarele Omega* - 10
Făgăraș, 2025

1. Inteligența artificială (IA) și ChatGPT: promisiuni și pericole ascunse / 4

1.1 IA: revoluția tăcută care schimbă lumea. Ce impact va avea ChatGPT asupra educației și viitorului nostru? / 4

1.2 Cum transformă ChatGPT viața cotidiană și știința: beneficii surprinzătoare și provocări viitoare / 10

1.3 Când ChatGPT generează fapte false: pericolele nevăzute ale IA / 14

1.4 Când IA îți spune că te iubește / 16

1.5 Pericolele și manipularea IA: Cum îți se poate afecta reputația în era chatbot-urilor / 20

1.6 Victima ChatGPT: Cum o acuzație falsă poate distruge o reputație / 24

1.7 Relația fatală cu un chatbot / 27

1.8 ChatGPT: Viitorul educației sau un risc pentru învățământ? / 31

1.9 Provocările ChatGPT în educație / 35

1.10 ChatGPT, integritatea academică și soluții viabile / 40

1.11 Viitorul educației: inteligența artificială în locul profesorilor? / 45

1.12 IA: promisiuni și pericole / 49

1.13 Elon Musk și alți experți de renume avertizează: viitorul periculos al IA / 53

1.14 Riscurile neașteptate ale IA / 59

1.15 Revoluția IA se scrie cu... biți cuantici / 62

1.16 Alte riscuri ascunse ale IA: amenințări și provocări / 69

1.17 Către o tiranie globală prin IA? / 73

1.18 De la Homo sapiens la Homo deus: Pericolele și promisiunile unui viitor transumanist / 76

2. YouTube, Social Media și școala online / 82

2.1 Exploatarea copiilor pe YouTube: fața ascunsă a „familiei perfecte” / 82

2.2 Cocaină electronică: Pericolele ascunse ale ecranelor pentru copii / 86

2.3 Dependența digitală: Cum tehnologia fură copilăria / 89

2.4 Rețelele sociale, democrația și cazul decapitării profesorului de istorie / 94

2.5 Algoritmii dezvăluie cele mai ascunse colțuri ale ființei noastre / 98

2.6 Soluții care contracarează expunerea excesivă la ecrane / 101

2.7 Cu sau fără telefon la școală? / 104

2.8 Măsuri de reducere a timpului petrecut online / 107

3. Smartphone, TikTok, Instagram...: dependență, pornografie, suicid / 112

3.1 „Melci care își poartă casa în buzunar” / 112

3.2 Obsedați de like-uri și răspunsuri / 116

3.3 „Pornografia ne-a corupt copiii. Iar noi permitem asta” / 119

3.4 „TikTok oferă minorilor droguri și sex” / 122

3.5 Impactul tragic al TikTok: adolescenți care ajung să-și pună capăt zilelor / 126

3.6 Un fenomen tot mai răspândit: sextingul / 129

3.7 Un post de dopamină / 133

3.8 Terapeuta care și-a șters conturile de pe *social media* / 136

4. Smartphone și social media: lanțurile se strâng / 141

4.1 „*Smartphonificarea* vieții” / 141

4.2 Autism virtual, singurătate și „spălarea creierului” / 144

4.3 De la „aventurieri pe Instagram” la conținuturi nepotrivite și deces / 148

4.4 Părinții sunt pe locul doi în creșterea copiilor. Cine este pe primul? / 151

4.5 Cu ce îi hrănesc pe copii rețelele de socializare / 154

4.6 Rețelele sociale schimbă definiția iubirii / 158

4.7 Nu le cumpărați copiilor smartphone! / 162

4.8 Reîntoarcerea în viața reală / 166

4.9 „Nivelul anxietății mele a scăzut la zero” / 169

4.10 Scrisoare deschisă către *smartphone*-ul meu / 174

5. Smartphone: dependență și libertate / 194

5.1 „Să șterg TikTok a fost aproape ca și cum aș renunța la alcool” / 194

5.2 „Cel mai puternic drog pe care îl avem” / 197

5.3 Rețelele de socializare - mai rele decât pornografia? / 201

5.4 Școala cu și fără telefoane mobile / 205

5.5 O schimbare radicală / 209

5.6 Povestea tinerilor care îmbrățișează viața fără smartphone / 212

5.7 Protecția copiilor pe rețelele sociale: o necesitate urgentă / 216

5.8 Cum să-ți recâștigi viața: Minimalismul digital și eliberarea de dependența de tehnologie / 220

5.9 Din experiențele minimaliștilor digitali / 223

6. 5G și 6G: binecuvântare sau blestem? / 228**6.1 „Viața mea a devenit un iad pe pământ” / 228****6.2 Efecte asupra sănătății / 231****6.3 Alte efecte, reacții critice și prețul plătit / 234****6.4 Plan de acțiune și alternativă mai sigură la 5G / 238****6.5 De la 5G la 6G și mai departe / 241*****Concluzii / 243******Bibliografie / 246***

1. Inteligența artificială (IA) și ChatGPT: promisiuni și pericole ascunse

1.1 IA: revoluția tăcută care schimbă lumea. Ce impact va avea ChatGPT asupra educației și viitorului nostru?¹

Inteligența artificială (IA) este o tehnologie care imită funcțiile creierului uman și deja transformă numeroase domenii ale vieții noastre, de la afaceri la educație. Cu fiecare zi, IA devine tot mai prezentă în activitățile noastre, iar unele dintre cele mai recente realizări, precum ChatGPT, au impresionat prin abilități uimitoare de a înțelege și genera limbaj natural. Dar cum va afecta aceasta viitorul educației? În cele de mai jos, vom explora atât beneficiile, cât și provocările pe care le aduce inteligența artificială în școli și în societatea noastră.

Ce este IA?

În cazul cel mai general, IA „se referă la capacitatea unui sistem informatic de a imita comportamentul creierului uman. Acest proces implică primirea de informații sub formă de date externe, învățarea prin antrenare și, pe baza acestei învățări, atingerea obiectivelor pentru care a fost proiectat.”²

IA se bazează pe rețele neuronale, sisteme care se aseamănă creierului uman în modul în care învață și procesează informațiile. Ele permit IA să învețe din experiență, așa cum ar face o persoană. Aceasta se numește învățare profundă (deep learning).³

Comaniile care dezvoltă IA au drept scop final să se ajungă la așa-numita inteligență artificială generală (IAG), adică un software capabil

¹ Capitolul a apărut mai întâi în formă ușor prescurtată, ca articol, în numărul pe februarie 2025 al revistei „Familia Ortodoxă”.

² Marta Montenegro-Rueda, José Fernández-Cerero, José María Fernández-Batanero, Eloy López-Meneses, *Impact of the Implementation of ChatGPT in Education: A Systematic Review*, Computers, vol. 12, nr. 8, iulie 2023.

³ Zoe Kleinman, Chris Vallance, *AI 'godfather' Geoffrey Hinton warns of dangers as he quits Google*, BBC News, 2 mai 2023.

KBIS | We know books

să învețe și să reacționeze la fel ca un om, acel „tip de IA capabilă să înțeleagă sau să învețe orice sarcină intelectuală pe care un om o poate realiza. Cu alte cuvinte, IAG este un tip de IA care poate îndeplini orice funcție cognitivă pe care o poate realiza un om, în loc să fie limitată la un set specific de sarcini.”⁴ IAG este deocamdată un concept ipotetic și constituie un obiectiv încă neatins în cadrul IA, dar spre care tind cercetările.

Problema IA reprezintă un subiect cu impact global, care a început să capteze atenția încă din anii 1950. În ultimii ani, interesul pentru utilizarea IA în diferite domenii și cu precădere în domeniul educațional a crescut semnificativ. Cu toate acestea, există și motive de îngrijorare.⁵

Alan Turing este considerat părintele IA. Studiul său, „Computing Machines and Intelligence”, publicat în 1950, începe cu întrebarea „Pot mașinile să gândească?”, în cadrul căreia a încercat să definească termenii „mașină” și „gândire” și a propus așa-numitul test Turing, ca metodă de determinare a inteligenței unei mașini. Astfel, Turing a adus în discuție un subiect inovator pentru vremea sa, referindu-se la mașinile capabile să simuleze comportamente umane și să realizeze acțiuni inteligente, precum jocul de șah. Conceptul de „inteligentă artificială” a fost introdus pentru prima dată de John McCarthy în 1956, atunci când a organizat prima conferință academică pe această temă. De-a lungul anilor, cercetările în domeniul inteligenței artificiale au trecut prin perioade de creștere și declin. În 1966 a fost introdusă Eliza, unul dintre primii roboți asistenți virtuali (chatbot), care realiza simulări ale unor conversații cu un psihoterapeut. În anul 1997, computerul Deep Blue reușea să îl învingă pe campionul mondial la șah Gari Kasparov. În 2011, era lansat Siri, asistentul virtual integrat pentru prima oară într-un telefon iPhone 4s, capabil să utilizeze comenzi vocale și o interfață de limbaj natural pentru a răspunde întrebărilor, a face recomandări și a îndeplini acțiuni prin accesarea serviciilor online, adaptat la preferințele și comportamentele individuale ale utilizatorilor, oferind rezultate personalizate. În același an, supercomputerul Watson câștiga în competiția cu doi rivali umani în cadrul show-ului Jeopardy. În 2014, asistentul virtual Alexa, utilizat inițial de Amazon, începea să fie folosit pe scară largă. În 2016, chatbot-ul Tay lansat de Microsoft a fost oprit la

⁴ Gareth Corfield, Matthew Field, *Meet ChatGPT, the scarily intelligent robot who can do your job better than you*, The Telegraph, 6 decembrie 2022.

⁵ Drd. Irena Apolzan (Arădăvoaicei), dr. Marilena-Jeana Cîmpineanu, *Benefits and Challenges of Using Artificial Intelligence in Education*, Euro-Atlantic Resilience Journal, vol. 2, nr. 3, 2024, p. 49.

doar 16 ore după lansare, în urma unor controverse legate de remarci rasiste lansate de robot în conversații cu persoane de pe Twitter. În 2017, AlphaGo, creat de Google, reușea să îl învingă pe campionul mondial la jocul Go. În 2018, Uniunea Europeană stabilea o serie de recomandări privitoare la etica utilizării IA. În ultimele decenii, algoritmi de învățare profundă au permis calculatoarelor să proceseze și să interpreteze date complexe, ducând la progrese semnificative în domeniile viziunii computerizate și procesării limbajului natural. În 2022, potențialul uriaș al IA de a schimba modul în care interacționăm și comunicăm cu mașinile a fost evidențiat prin lansarea ChatGPT (dezvoltat de laboratorul de cercetare Open AI din San Francisco), un model lingvistic de dimensiuni mari (LLM: large-scale language model), adică un model de învățare automată antrenat pe volume mari de date text pentru a înțelege și genera limbaj natural.⁶ Odată cu lansarea ChatGPT, care a devenit aplicația cu rata de creștere cea mai rapidă de până acum din lume, a atras atenția publicului așa-numita Inteligență Artificială Generativă (GenAI). Aceste aplicații GenAI au capacitatea de a imita abilitățile umane pentru a genera rezultate precum imagini, texte, muzică, videoclipuri și coduri software.⁷

Ce este ChatGPT?

Într-adevăr, ChatGPT, aplicația lansată pe 30 noiembrie 2022 (versiunea 3.5), înregistra deja în ianuarie 2023 un număr de 100 de milioane de utilizatori activi. Prin comparație, TikTok, considerat anterior un adevărat fenomen digital, avusese de nouă luni pentru a ajunge la același nivel. Un studiu din ianuarie 2023, la care au participat 1.000 de elevi de liceu din Statele Unite, a constatat că 89% dintre aceștia au folosit ChatGPT pentru a-i ajuta la teme pentru acasă. Tehnologia, cu toate imperfecțiunile sale, a fost adoptată într-un timp incredibil de rapid.⁸ În martie 2023 a fost lansată versiunea ChatGPT-4, ce prezintă îmbunătățiri semnificative în ceea ce privește raționamentele efectuate și concizia frazelor, cu prețul unui timp ceva

⁶ Drd. Irena Apolzan (Arădăvoaicei), dr. Marilena-Jeana Cîmpineanu, *art. cit.*, pp. 51-53.

⁷ Drd. Irena Apolzan (Arădăvoaicei), dr. Marilena-Jeana Cîmpineanu, *art. cit.*, p. 57.

⁸ Chris Moran, *ChatGPT is making up fake Guardian articles. Here's how we're responding*, The Guardian, 6 aprilie 2023.

mai lung de oferire a răspunsului în comparație cu ChatGPT-3.5.⁹ Ambele versiuni (3.5 și 4), în comparație cu cea de bază de la care au pornit (ChatGPT-3), au obținut o îmbunătățire și ajustare a conversațiilor în scopul de a reduce numărul răspunsurilor false, toxice sau inutile. Abordarea folosită pentru ajustarea ChatGPT se numește învățare prin întărire cu feedback uman (RLHF – reinforcement learning with human feed-back).¹⁰

ChatGPT este o tehnologie de IA creată pentru a imita conversația și limbajul uman, folosind o vastă cantitate de informații pentru a răspunde la întrebări și a rezolva probleme.¹¹ ChatGPT este conceput pentru a îndeplini sarcini complexe, precum redactarea de articole, poezii, eseuri și chiar coduri de programare originale, dar și pentru a traduce, rezuma sau extinde texte.¹²

Este aproape imposibil de crezut că Chat GPT a reușit, într-un interval atât de scurt, să revoluționeze lumea așa cum o știm. Acum, oricine are acces la această platformă poate introduce ce întrebare dorește și primește instantaneu răspunsuri generate pe baza cererii sale. Cu doar câteva cuvinte, poți obține informații și referințe în câteva secunde, pe care le poți vizualiza imediat pe ecran. Proiecte de afaceri sau chiar manuale întregi pot fi generate rapid, iar cu câteva ajustări, pot fi finalizate într-un timp mult mai scurt decât ar fi fost necesar doar pentru a face partea de documentare.¹³

Deși tehnologiile care stau la baza IA există de ceva vreme, abia la sfârșitul anului 2022, odată cu lansarea ChatGPT de către OpenAI, s-a simțit cu adevărat că începem să ne apropiem de momentul apariției unei tehnologii care să imite inteligența. Într-un raport din 2023, președintele Microsoft Canada, Chris Barry, a declarat că „era inteligenței artificiale a sosit, aducând cu ea o schimbare majoră ce poate afecta fiecare aspect al vieților noastre” și că „nu este doar un avans tehnologic, ci o schimbare a societății”. Aceasta a fost una dintre reacțiile mai echilibrate. În schimb, mulți artiști și scriitori sunt

⁹ Tareq Rasul et al., *The role of ChatGPT in higher education: Benefits, challenges, and future research directions*, Journal of Applied Learning & Teaching, vol. 6, nr. 1, mai 2023, pp. 2-3.

¹⁰ Tim Fütterer et al., *ChatGPT in education: global reactions to AI innovations*, Scientific Reports, vol. 13, nr. 1, septembrie 2023, p. 2.

¹¹ Gareth Corfield, Matthew Field, *art. cit.*

¹² Maria Radeva, *GUEST POST: The Benefits and Risks of ChatGPT for Education*, The Learning Scientists, februarie 2024, <https://www.learningscientists.org/blog/2024/2/15-1>.

¹³ *The Impact of Chat GPT on Education: The Good and the Bad*, Digital Learning Institute Blog, <https://www.digitallearninginstitute.com/blog/the-impact-of-chat-gpt-on-education>.

îngrijorați că vor deveni depășiți, guvernele încearcă să reglementeze rapid această tehnologie, iar oamenii de știință sunt într-un continuu dezacord pe această temă. Marile companii sunt foarte dornice să profite de pe urma acestui entuziasm. Mari firme globale, cum ar fi Microsoft, Meta și Alphabet, susțin activ dezvoltarea IA. Pe lângă miliardele investite de giganții din tehnologie, finanțările pentru start-up-urile IA au ajuns aproape la 50 de miliarde de dolari în 2023. La un eveniment de la Universitatea Stanford din aprilie 2023, CEO-ul OpenAI, Sam Altman, a spus că nu-i pasă dacă firma cheltuiește 50 de miliarde de dolari pe an pe IA. Viziunea sa include crearea unui „super-asistent” care ar fi un „coleg extrem de competent, care știe totul despre viața mea – fiecare e-mail, fiecare conversație avută – dar nu îl percepi ca pe o extensie”.¹⁴

În ultimii ani, progresele în prelucrarea limbajului natural (NLP) au condus la realizări remarcabile în aplicații practice precum recunoașterea vorbirii și entităților, rezumarea, traducerea limbajului și generarea de texte. Cercetătorii au folosit, de asemenea, modele de rețele neuronale recurente (RNN) în diverse domenii, datorită structurii lor care le permite să memoreze dependențele din texte, adică relațiile dintre cuvinte și alte entități din cadrul unei fraze sau al unui pasaj mai lung. Totuși, modelele RNN au limite, mai ales când trebuie să proceseze dependențe pe distanțe foarte lungi în texte scrise în limbaj natural. De aceea, a fost introdusă arhitectura Transformer pentru a rezolva această problemă. Aceasta face ca modelul să fie foarte eficient în înțelegerea relațiilor dintre cuvinte într-o propoziție, indiferent de poziția lor. Modelele GPT (prescurtare de la Generative pretrained transformer), bazate pe arhitectura Transformer, au avut un succes considerabil în sarcinile NLP. Modelele GPT sunt antrenate pe cantități masive de date textuale (aproape 175 de miliarde de parametri de antrenare și 570 GB de text pentru GPT-3, care a stat la baza versiunii GPT-3.5 lansate în noiembrie 2022), având capacitatea de a genera texte asemănătoare celor scrise de oameni și de a efectua alte sarcini legate de limbaj cu un grad înalt de acuratețe. Spre deosebire de alte modele lingvistice bazate pe IA, ChatGPT generează și prezintă conținut complet nou într-o conversație în timp real cu utilizatorul. În plus, ChatGPT poate menține constant un stil de dialog care implică utilizatorul într-un mod mai natural, oferind răspunsuri relevante pentru fiecare întrebare, ceea ce îl

¹⁴ Navneet Alang, *No god in the machine: the pitfalls of AI worship*, The Guardian, 8 august 2024.

IBRIS | We know books
face un model unic în comparație cu alte modele lingvistice mari (LLM-uri).¹⁵

Ceea ce numim acum IA se concentrează în mare parte pe LLM-uri. Aceste modele sunt antrenate pe seturi uriașe de date – de exemplu, ChatGPT a „absorbit” aproape întregul internet public – și sunt programate să identifice tipare în ele. Unități de semnificație, cum ar fi cuvinte, părți de cuvinte și caractere, devin „tokenuri” și sunt asociate cu valori numerice. Modelele învață cum se raportează aceste tokenuri între ele și, în timp, ajung să înțeleagă ceva asemănător cu contextul: unde ar putea apărea un cuvânt, în ce ordine și așa mai departe. De decenii, testul standard pentru a evalua dacă tehnologia se apropie de inteligență a fost testul Turing, creat de matematicianul britanic Alan Turing. Testul presupune ca un interogator uman să pună întrebări unor subiecți invizibili – un computer și o persoană – prin mesaje text, pentru a stabili care este mașina. Dacă o proporție suficientă de interogatori sunt înșelați, mașina poate fi considerată inteligentă. ChatGPT poate deja să înșele unele persoane în anumite situații. Nu știm însă exact cum funcționează sistemele de inteligență artificială. Profesorul Leif Weatherby de la Universitatea New York sugerează că modelele procesează atât de multe combinații de date încât este imposibil pentru o singură persoană să le înțeleagă pe deplin. Remarca Navneet Alang, autor de articole despre dimensiunea socială și filosofică a tehnologiei: „Computerele ar putea ajunge la o formă de gândire, dar ele nu visează, nu doresc sau nu au sentimente, iar acest aspect este mult mai important decât lasă să se înțeleagă promotorii IA. Când folosim inteligența pentru a găsi soluții la crize economice sau pentru a combate rasismul, o facem dintr-un simț al moralității și dintr-un sentiment de obligație față de cei din jur – din convingerea că avem responsabilitatea de a îmbunătăți lucrurile într-un mod semnificativ moral.” Și oferea drept exemplu „Deep Thought”, computerul din „Ghidul autostopistului galactic” de Douglas Adams. Când a fost întrebat „care este răspunsul la întrebarea supremă a vieții, universului și a tuturor celorlalte”, a dat celebrul răspuns „42”. Comenta Alang: „Absurdul acestui răspuns subliniază un adevăr simplu dar ușor uitat: viața și sensul ei nu pot fi reduse la o propoziție simplă sau la o listă de nume, la fel cum gândirea și simțirea umană nu pot fi limitate la o reprezentare binară, doar cu cifre de 0 și 1. Dacă întrebi inteligența artificială despre sensul vieții, nu

¹⁵ Md. Mostafizer Rahman, Yutaka Watanobe, *ChatGPT for Education and Research: Opportunities, Threats, and Strategies*, Applied Sciences, vol. 13, nr. 9, mai 2023, pp. 1-2.

răspunsul este greșit, ci întrebarea.”¹⁶

În cele ce urmează, vom explora atât beneficiile, cât și riscurile și pericolele pe care le poate ascunde ChatGPT. De asemenea, vom analiza avantajele și dezavantajele utilizării acestui chatbot în școli, vom discuta dacă inteligența artificială în școli va putea ajunge să înlocuiască profesorii, precum și posibilele avantaje, dezavantaje și riscuri pe care le-ar putea aduce inteligența artificială în viitor.

1.2 Cum transformă ChatGPT viața cotidiană și știința: beneficii surprinzătoare și provocări viitoare

IA devine tot mai prezentă în viața cotidiană și în profesiile diverse, iar ChatGPT este deja un instrument utilizat în multe domenii, de la sănătate și educație, până la cercetarea științifică și gestionarea problemelor personale. În continuare, vom explora câteva dintre cele mai notabile beneficii ale ChatGPT, evidențiate prin exemple concrete și studii recente, fără a pierde însă din vedere nici provocările și riscurile asociate.

ChatGPT și gestionarea ADHD-ului

Vanessa Badham, o scriitoare australiană, recunoștea că „inteligența artificială (IA) poate fi atât dăunătoare, cât și extrem de utilă”, iar în cazul ei s-a dovedit benefică mai ales după ce a fost diagnosticată cu ADHD și a început să încredințeze prioritizarea sarcinilor „robotului” ChatGPT. A constatat că suferă de „time blindness”, afecțiune specifică ADHD-ului, ceea ce îi afectează capacitatea de a estima corect timpul, ducând la aglomerarea programului, întârzieri frecvente și dificultăți în luarea deciziilor. Astfel, a luat decizia să delege către ChatGPT toate sarcinile legate de prioritizare. Acesta calculează timpul de călătorie, sugerează rute și îi oferă informațiile necesare înainte de întâlniri, ajustându-și planurile în funcție de noile date. În plus, îi amintește dieta preferată, analizează mesele din restaurante, îi optimizează programul de exerciții, creează liste de lucruri de luat pentru weekend și o sfătuiește ce să poarte în funcție de vreme. „Una dintre cele mai grele torturi provocate de ADHD o constituie rușinea de a cere ajutor altora pentru sarcini simple, care par copleșitoare. Robotul nu mă judecă

¹⁶ Navneet Alang, *art. cit.*

niciodată și rezumă scrisorile de la bancă cu aceeași ușurință cu care îmi amintește că am programul prea încărcat”¹⁷, scria Badham. Astfel, IA, în special prin ChatGPT, i-a oferit Vanessei Badham un sprijin eficient în gestionarea simptomelor ADHD, ajutând-o să-și prioritizeze sarcinile și să-și îmbunătățească viața de zi cu zi.

Îmbunătățirea rezolvării problemelor științifice și matematice

Un alt beneficiu al ChatGPT îl constituie ajutorul pe care îl poate oferi în cazul unor probleme științifice sau matematice de complexitate ridicată. Este adevărat că interfețele de comunicare de tip chatbot, precum ChatGPT de la OpenAI și Gemini de la Google, întâmpină uneori dificultăți cu probleme simple de matematică, iar codul generat de acestea poate fi adesea defectuos și incomplet. În anumite cazuri, ele pot chiar inventa informații. Însă în septembrie 2024 OpenAI a lansat o versiune nouă a ChatGPT, denumită ChatGPT-4 Turbo, care ar putea rezolva aceste deficiențe. Compania a anunțat că noul chatbot, susținut de o tehnologie de IA numită OpenAI o1, este capabil să „raționeze” în rezolvarea problemelor de matematică, programare și știință. „Modelele anterioare, precum ChatGPT, răspundeau imediat ce le puneai o întrebare”, a declarat Jakub Pachocki, unul dintre principalii cercetători de la OpenAI. „Acest model poate să aibă nevoie de mai mult timp. Poate gândi problema – în limba engleză – și încearcă să o descompună și să găsească diverse unghiuri de abordare pentru a oferi cel mai bun răspuns.” În cadrul unei demonstrații, chatbot-ul a răspuns corect la o întrebare de chimie la nivel de doctorat și a diagnosticat o boală pe baza unui raport detaliat despre simptomele și istoricul unui pacient. Compania OpenAI a declarat că noua tehnologie ar putea ajuta fizicienii să creeze formule matematice complexe și ar putea sprijini cercetătorii în domeniul sănătății în experimentele lor. Deoarece internetul este plin de informații false, tehnologia a învățat să repete aceste neadevăruri, uneori chiar inventându-le. Dr. Pachocki și colaboratorii săi au încercat să reducă aceste erori, construind noul sistem al OpenAI folosind învățarea prin întărire (sau învățarea prin recompensă – reinforcement learning). În cadrul acestui proces, care poate dura săptămâni sau luni, sistemul învață prin încercări și erori extensive. De exemplu, rezolvând diverse probleme matematice, poate învăța ce metode duc la răspunsul

¹⁷ Van Badham, *Fearing AI, I was reluctant to use ChatGPT. But friends, it changed my life*, The Guardian, 11 septembrie 2024.