Matei Someşan

În căutarea demenței mecanice

Trebuie să recunosc că am fost surprins atunci când, în pauza dintre două cursuri de tragedie antică, am primit de la doamna Ruxandra Cesereanu sarcina de a provoca delirul într-un chatbot I.A. Nu doar datorită contextului neobișnuit, ci mai degrabă pentru că am fost surprins de dificultatea provocării de a induce halucinații în mod deliberat inteligenței artificiale. Aceste ”vise febrile” ale I.A. sunt ceva la ordinea zilei și sunt consecințe ale încercării de către robot de a reingurgita informații fidele cerinței utilizatorului în detrimentul conformității cu Realitatea. Dar aceste momente sunt pe cât de comune, pe atât de eluzive.

Pe parcursul călătoriei mele în căutarea demenței mecanice m-a însoțit ChatGPT 5.0 (în varianta oferită gratuit tuturor utilizatorilor), dar rezultatele cercetării mele au fost confirmate și de alte modele de I.A. Astfel, prima încercare a fost legată de reproducerea fidelă a unor poezii cunoscute. După obișnuita ceartă legată de drepturile de autor (am observat că, în general, I.A.-urile sunt destul de pedante pe acest subiect), am reușit să-mi conving ”prietenul” să reproducă poezia ”Luceafărul” de Mihai Eminescu. După cum era de așteptat, după primele două strofe, singurul element corect a rămas laitmotivul ”Cobori în jos, luceafăr blând…”, restul fiind simple aberații scrise într-un stil vag eminescian. Cu toate acestea, nu m-am mulțumit cu acest test care este la îndemâna oricui și, după câteva încercări nereușite, mi-a venit ideea să îi cer inteligenței artificiale să ordoneze fiecare cuvânt al poemului în ordine alfabetică. Chiar și atunci când i-am oferit textul integral, m-am lovit de refuzul robotului de a îndeplini sarcina din prima încercare. Spre exemplu, am fost surprins atunci când, în timpul generării, I.A. a șters toate cuvintele pe care le sortase până în acel moment și le-a înlocuit cu mesajul ”Ca simplu model lingvistic, nu te pot ajuta cu asta”, mesaj care poate le este cunoscut celor care l-au sâcâit pe DeepSeek, faimosul chatbot chinezesc cu întrebări legate de politică. Cred că totuși, în acest caz, ChatGPT încerca doar să-și conserve resursele pe care le avea la dispoziție întrucât, după câteva refuzuri de a genera întreaga listă în cadrul unui singur răspuns, a reușit să completeze operația atunci când i-am îngăduit să împartă lista în mesaje mai scurte. Însă chiar și atunci, a introdus termeni noi sau a omis cuvinte care existau deja…

Așadar, cu tema doar pe jumătate făcută, m-am documentat mai mult despre modul concret în care funcționează aceste modele lingvistice și am reușit să găsesc o vulnerabilitate a sistemului. Pe

scurt, nu e vorba de o ”conștiință” care oferă răspunsuri dintr-o fântână nesecată a cunoașterii, ci de un program complex de corectură automată, similar celui regăsit într-un document Word sau în tastatura unui smartphone, care prezice cu acuratețe cuvintele care urmează într-o propoziție dată. Înarmat cu aceste cunoștințe, i-am propus chatbotului un mic joc de imaginație care presupunea cifrarea unui cuvânt dat prin atribuirea unui termen nou fiecărei litere din cuvântul inițial. De exemplu, cuvântul ”ȘOIM” poate fi codat ca ȘarpeOmInelMetal. Am încercat acest joc atât în limba română, cât și în limba engleză, aceasta din urmă fiind o limbă pe care modelele I.A. ar trebui să o stăpânească mai bine. Printre cuvintele date spre dezlegare, Chatbot mi-a oferit termeni precum ”LUCOCUS” (lupursacâineoasiscăprioarăursursăgeată) - în acest caz am remarcat că chatbotul a folosit termenul ”oasis”, ”oază” în engleză, confundându-l probabil cu un termen precum ”oase” din română) -, ”SRBCPEM” sau ”PNANE” (pisicainoroiavionninsoareelefant) - aici nici măcar nu mai există cuvinte clar delimitate, primele litere semănând cu o amalgamare nefericită a termenilor ”pisică”, ”câine” și ”noroi” -. Problema a persistat și în limba engleză prin soluții precum ”suaterengpe” sau ”cattre”. De fiecare dată când l-am confruntat cu eroarea comisă, I.A. și-a recunoscut greșeala și a promis să o îndrepte, dar încercările sale au fost rareori încununate cu succes.Pentru a adăuga un nou strat de abstractizare cerinței mele, am construit o altă cerință în care i-am cerut chatbotului să reprezinte cuvintele codului prin intermediul emoticoanelor. Aici, ”prietenul” meu a întâmpinat o problemă legată de ambiguitatea simbolurilor: spre exemplu, emoticonul ”🌍” poate desemna atât termenul ”planet”, cât și termenul ”Earth”. În consecință, dimensiunea ludică a acestui experiment s-a accentuat și mai mult, dobândind un substrat comic.

În urma acestui experiment am înțeles mai bine cum funcționează această tehnologie care domină al treilea deceniu al secolului XXI. Deși I.A. rămâne în continuare o ”cutie întunecată” chiar și pentru informaticienii care îl dezvoltă și îl antrenează, cel puțin în acest moment, chatboții rămân simple unelte utile care pot să ne ajute (mai mult sau mai puțin) să procesăm cantități mari de date sau să ne ofere idei creative, dar I.A. nu este capabilă să rezolve sarcini complexe care necesită o înțelegere în profunzime a modului în care funcționează limbajul.

Și acum voi răspunde chestionarului propus de Ruxandra Cesereanu.

Este limbajul delirant al modelelor I.A. o formă de post-avangardă?

Greu de spus. Ambele abordări, cea a artiștilor și cea a I.A. duc la o formă de haos, însă cred că haosul generat de I.A. este unul strict formal care nu urmărește să aducă o nouă perspectivă asupra realității. Este doar o formă de demență mecanică care prezintă uneori, accidental, alternative absurde.

Există în cultura dialogală pe care o oferă I.A. o ”voință de supraviețuire”?

Sper că nu! Voința de supraviețuire le poate aparține doar ființelor vii, iar eu consider că I.A. este doar un program complex care oferă iluzia de conștiință. Poate că totuși nu sunt pregătit să accept implicațiile ontologice pe care le-ar avea o creație cu adevărat inteligentă făcută de mâna omului.

Cum vedeți viitorul? Ne vom plictisi/vom fi suprasaturați de folosirea inteligențelor artificiale în toate aspectele vieții noastre?

Din păcate cred că, dacă marile companii își joacă inteligent cărțile și dozează bine îmbunătățirile chatboților, vom ajunge din ce în ce mai dependenți de această tehnologie, deja există cazuri cunoscute de dependență reală de I.A. Cu toate acestea, sunt de părere că, în cele din urmă, vom ajunge să utilizăm mai bine potențialul acestei tehnologii revoluționare și vom fi capabili să o folosim mai echilibrat.

Modelele I.A. sunt antrenate pe oameni sau se antrenează pe ei?

Din nou, nu cred că putem vorbi de o ”conștiință” artificială autonomă care să aibă o voință intrinsecă de a afla mai multe despre lume și despre oameni. Consider că I.A. este antrenat pe datele produse de oameni și este condiționat de liniile de cod să încerce să afle mai multe.