CHRISTINA LEUKER ȘI WOUTER VAN DEN BOS: DE CE TREBUIE SĂ PROTEJĂM IGNORANȚA DE AVÂNTUL INTELIGENȚEI ARTIFICIALE

Autor: Iulian-Alfred Husac | 18 ianuarie 2022

Un articol de Christina Leuker și Wouter Van Den Bos pentru Nautilus

Articol original: *We Need to Save Ignorance From AI*

După căderea Zidului Berlinului, cetățenilor Germaniei de Est li s-a dat dreptul de a-și citi dosarele create de serviciile secrete germane (STASI). Se estimează că până astăzi doar 10 procente dintre ei și-au exercitat acest drept.

În 2007, James Watson, unul dintre savanții care au descoperit structura ADN, a refuzat să i se ofere vreo informație despre gena sa APOE, a cărei alelă este cunoscută pentru declanșarea bolii Alzheimer.

Majoritatea oamenilor răspund persoanelor care fac sondaje de opinie că, dacă ar putea alege să-și cunoască ziua morții sau data unor evenimente viitoare fericite, ar prefera să nu o facă.

Fiecare dintre situațiile de mai sus relevă o ignoranță voită. Deși Socrate a spus că o viață lipsită de întrebări este o viață irosită, iar Hobbes a susținut primatul curiozității ca pasiune a oamenilor, multe dintre cele mai vechi întâmplări descriu pericolele ce derivă din a ști prea mult. De la Adam și Eva și pomul cunoașterii, la Prometeu și furtul secretului focului, ni se arată că deciziile din viața reală trebuie să mențină echilibrul dintre a alege să știi sau să rămâi ignorant.

Ce-ar fi dacă ar apărea o tehnologie care să încline balanța într-un mod imprevizibil, complicând modul și momentul în care decidem să rămânem ignoranți? Această

tehnologie e deja aici și se numește inteligența artificială.

Inteligența artificială sau AI poate găsi tipare și face inferențe pornind de la o bază relativ mică de informații. Spre exemplu, câteva aprecieri pe Facebook sunt destule pentru ca tehnologia AI să-ți poată identifica personalitatea, rasa și sexul. Despre un alt algoritm se susține că poate distinge între bărbații heterosexuali și homosexuali cu o precizie de 81 de procente, iar cât privește orientarea sexuală în cazul femeilor, cu o precizie de 71 de procente, pornind de la o simplă fotografie.[1] Un alt algoritm, denumit COMPAS, poate prezice recidivismul unei persoane pornind de la informații precum arestări juvenile, antecedente penale în familie, educație, gradul de izolare socială și activitățile din timpul liber, cu o precizie de 65 de procente.[2]

În toate aceste cazuri, concluzia poate fi foarte diferită de natura informațiilor culese și folosite de algoritm (chiar dacă validitatea unor rezultate poate fi contestată), făcând ca ceea ce știm să fie dificil de controlat. Legile în vigoare menite să ne protejeze ignoranța sunt rare sau chiar inexistente. Nu există o protecție a „dreptului de a nu ști”. Această situație creează mediul perfect în care, la fel ca în motto-ul vechi al Facebook, suntem predispuși să „ne mișcăm repede și să revoluționăm”. Dar când vine vorba despre viața noastră privată, este faptul de „a revoluționa” chiar ceea ce vrem să facem?

Guvernele și parlamentele știau de zeci de ani că cutia Pandorei nu trebuie deschisă. Numeroase legi de protecție a ignoranței individului au fost dezbătute încă din anii ‘90. Convenția Europeană asupra Drepturilor Omului și Biomedicinei din 1997 statuează, spre exemplu, că „Orice persoană are dreptul să cunoască orice informaţie culeasă cu privire la sănătatea sa. Cu toate acestea, dorinţa unei persoane de a nu fi informată trebuie respectată.” La fel, Declarația Asociației Mondiale a Medicinei asupra Drepturilor Pacientului din 1995 statua că „pacientul are dreptul de a nu fi informat [medical] la cererea sa expresă, cu excepția cazului în care informarea se face cu scopul protejării vieții persoanei.”

Redactarea unor legi care să apere „dreptul la ignoranță” de inteligența artificială pune alte probleme. În timp ce informațiile medicale sunt strict reglementate, informațiile folosite de AI se află în mâinile sectorului tehnologic, recunoscut pentru reglementarea slabă și orientarea spre profit. De asemenea, informațiile colectate de AI sunt diverse, astfel că legile viitoare trebuie înzestrate cu o definiție mai largă a „dreptului la ignoranță”. În acest sens, cercetarea în domeniul psihologiei ignoranței voluntare ar putea fi de ajutor. Din păcate, subiectul a fost mult timp desconsiderat, mai ales ca domeniu de cercetare științifică riguroasă, poate și din cauza presupunerii implicite că este irațional să vrei să nu fii informat.

De curând, psihologul Ralph Hertwig și marele jurist Christoph Engel au publicat o taxonomie extinsă de argumente în favoarea ignoranței voluntare. Ei au identificat două seturi de argumente care au o relevanță specială în raport cu nevoia de ignoranță în fața

inteligenței artificiale. Primul set de argumente se învârte în jurul imparțialității și corectitudinii. Mai precis, cunoașterea poate uneori să corupă judecata, iar de aceea alegem adesea să rămânem în mod voit ignoranți. Un exemplu la îndemână este reprezentat de evaluarea *peer review* a lucrărilor academice care este de cele mai multe ori anonimă. La fel, în majoritatea țărilor, societăților de asigurări nu li se permite să știe toate detaliile despre sănătatea clienților lor înainte de a încheia contracte cu aceștia; acestea au dreptul să cunoască doar factorii generali de risc. Astfel de observații sunt relevante mai ales în cazul inteligenței artificiale, deoarece există riscul ca, în urma prelucrării informațiile rezultate, să devină prejudiciabile pentru indivizi.

Al doilea set de argumente face referire la stabilitatea emoțională și evitarea regretelor. Ignoranța voluntară, în accepțiunea lui Hertwig și Engel, poate ajuta oamenii să mențină „convingeri intime" și să evite „disconfortul mental, frica și disonanța cognitivă".[3] Răspândirea ignoranței voluntare este mare. Aproximativ 90 de procente dintre germanii chestionați vor să evite sentimente negative ce ar putea rezulta din „cunoașterea de dinainte a unor evenimente triste, precum moartea sau divorțul", iar începând cu 40 până la 70 de procente dintre aceștia nu vor să știe dinainte despre evenimente fericite, pentru a se putea bucura de „surprizele și suspansul" care ar rezulta, spre exemplu, din cunoașterea sexului copilului la momentul nașterii.[4]

Aceste argumente ne pot ajuta să înțelegem de ce trebuie să protejăm ignoranța de avântul inteligenței artificiale. Algoritmul *gaydar*, spre exemplu, pare să aibă aproape zero beneficii și este foarte costisitor când vine vorba de imparțialitate și corectitudine. Citând publicația *The Economist*, „în părțile lumii unde a fi gay este inacceptabil din punct de vedere social sau ilegal, un astfel de algoritm reprezintă o amenințare serioasă la adresa siguranței persoanei". În aceeași măsură, așa-zisele beneficii ale unui program care identifică etnia, precum cel aflat în prezent sub dezvoltarea celor de la NtechLab, sunt nesemnificative în raport cu vicierea adusă imparțialității și corectitudinii. Utilizarea sistemului COMPAS de prezicere a recidivismului are o acuratețe mai mare decât ar avea o ființă umană, dar, precum au scris Dressel și Farid, „nu este precizia pe care ne-o dorim atunci când este în joc viitorul unui suspect". Algoritmele care prezic speranța de viață a individului, ca cele dezvoltate de Aspire Health, nu contribuie în mod necesar la stabilitatea emoțională.

Astfel de exemple ilustrează utilitatea identificării unor motive de natură individuală în favoarea ignoranței și demonstrează cât de complexe pot fi întrebările despre cunoaștere și ignoranță, în special când vine vorba de inteligența artificială. Nu se poate răspunde clar când ignoranța colectivă este benefică sau adecvată din punct de vedere etic. O abordare potrivită ar fi luarea fiecărui caz individual în parte și supunerea lui la o analiză de tip risc-beneficiu. De preferat ar fi ca o astfel de analiză să fie publică, date fiind complexitatea dezbaterii și amploarea consecințelor, și ar trebui să includă toate părțile interesate, opiniile unor experți, precum și luarea în considerare a tuturor

rezultatelor posibile, chiar și a celor mai pesimiste scenarii.

O astfel de inițiativă ridică o mulțime de dificultăți – de fapt, în cele mai multe cazuri ar fi imposibil de pus în practică. Așadar, cum planificăm în linii mari ceva ce necesită adaptări fine?

O abordare ar fi să controlăm și să limităm inferențele pe care calculatoarele le fac utilizând informațiile pe care le-au colectat deja. În acest sens, am putea „interzice" algoritmului judiciar să folosească rasa ca o variabilă de prezicere sau, un alt exemplu, să excludem sexul ca variabilă din analiza unor potențiali candidați pentru locuri vacante de muncă. Însă această abordare nu este lipsită de probleme.

În primul rând, limitarea informațiilor pe care marile companii le pot folosi este o inițiativă costisitoare și dificil de pus în practică din punct de vedere tehnic. Ar necesita nu doar ca aceste companii să ofere acces liber la sursa algoritmilor folosiți, ci și o activitate de audit permanent din partea marilor agenții guvernamentale. În plus, odată ce segmente mari de informații au fost colectate, sunt multe feluri indirecte de deducere a „informațiilor interzise" acestora. Să presupunem că utilizarea sexului ca variabilă pentru prezicerea succesului academic ar fi declarată ilegală. Ar fi foarte ușor pentru aceste companii să folosească variabile precum „mașina deținută" și „genul de muzică preferat" ca substitute pentru sex, efectuând mai apoi o inferență de gradul doi și făcând prezicerea pornind de la substitutele folosite pentru sex. Inferențele despre sex ar putea fi integrate într-un algoritm în ciuda bunelor intenții ale unei companii. Aceste inferențe de gradul doi fac activitatea de audit mult mai descurajantă. Cu cât mai multe variabile sunt incluse în analiză, cu atât mai mari sunt șansele să apară inferențe de gradul doi.

O abordare mai radicală – și probabil mai eficientă – ar fi să prevenim mai întâi colectarea datelor. Spre exemplu, într-o inițiativă de pionierat în domeniul legislativ, Germania a adoptat în anul 2017 o serie de legi care interzic mașinilor autonome să identifice rasa, vârsta și sexul oamenilor de pe stradă. Asta se traduce prin faptul că mașina nu va putea niciodată să integreze în deciziile sale – și în special în deciziile care trebuie luate când un accident este inevitabil – informații din aceste categorii.

Consecventă cu acest mod de a gândi, Uniunea Europeană a prevăzut prin Regulamentul General pentru Protecția Datelor (GDPR), în vigoare din mai 2018, că utilizatorilor europeni li se poate stoca de către companii un minim de informații personale, doar cu consimțământul lor și în raport cu un serviciu specific oferit de către companiile care colectează astfel de informații. O astfel de restricție ar putea reduce semnificativ inferențele de gradul doi. O limitare importantă a abordării GDPR reiese din faptul că unele companii pot să-și fixeze obiective foarte generale. Un exemplu este dat de fosta (inexistentă, în prezent) Cambridge Analytica a cărei obiectiv clar era să evalueze personalitatea, astfel că așa-zisa „controversată" campanie de colectare a

datelor personale ale utilizatorilor Facebook satisfăcea normele GDPR. În mod asemănător, accentul GDPR pe raportul dintre informații și un serviciu specific nu exclude posibilitatea oferirii unor informații cu caracter imoral, nici nu interzice companiilor să cumpere informații ilegale de la un broker de informații, atât timp cât utilizatorul consimte în acest sens – și nu puțini oameni refuză să ofere informații chiar și pentru niște beneficii derizorii. Cercetătorii au aflat că niște studenți de la Institutul de Tehnologie din Massachusetts ar oferi informațiile de contact ale prietenilor lor pentru o simplă felie de pizza.[5] Este cât se poate de evident că mai multe limitări sunt necesare, dar cât de multe?

Activistul și programatorul american Richard Stallman a dat următorul răspuns: „Sunt atât de multe feluri în care informațiile pot fi folosite pentru a face rău oamenilor, încât singura bază de date sigură ar fi cea în care nu s-ar stoca nimic, niciodată.” Dar restrângerea radicală a procesului de colectare ar îngreuna progresul tehnologic și ne-ar putea priva de beneficiile folosirii AI. Cine ar trebui să decidă în aceste cazuri? Ei bine, noi toți ar trebui să o facem.

În majoritatea cazurilor, vorbim despre informațiile noastre personale. Am fost nepăsători oferindu-le unor aplicații atrăgătoare fără să ne gândim la posibilele consecințe. De fapt, ne oferim informațiile personale de atâta timp încât am uitat că ne aparțin. Redobândirea lor ne-ar permite să decidem fiecare dacă există lucruri pe care vrem să le știm și altele pe care nu vrem să le aflăm. Restaurarea informațiilor către proprietarii de drept – în acest caz fiecare dintre noi – reprezintă soluția ideală la numeroasele probleme discutate mai sus. Nu necesită dezvoltarea unei legi universale care să reglementeze utilizarea informațiilor. În schimb, milioanele de utilizatori își vor ghida alegerile ce privesc folosirea datelor cu caracter personal conform propriilor convingeri. Cu toții am putea reacționa în funcție de evoluția politicilor companiilor, pedepsindu-le sau răsplătindu-le în raport cu modul în care tratează informațiile utilizatorilor.

Programatorul și filosoful în știința calculatoarelor Jaron Lanier a propus un argument în plus, de factură economică, în favoarea redobândirii de către indivizi a informațiilor oferite diverselor servicii și companii. Lanier consideră că am putea cu toții să profităm de pe urma informațiilor noastre prin vânzarea lor către marile companii. Această abordare ridică însă probleme sub două aspecte. Un prim aspect constă în confuzia creată în ce privește etica utilizării informațiilor și a proprietății. Intenția de a oferi date personale cu titlu gratuit reprezintă testul de turnesol pentru integritatea etică a întrebărilor la care astfel de informații vor fi folosite pentru a răspunde. Câți indivizi dintr-o minoritate și-ar oferi informațiile personale gratuit pentru a crea o aplicație de recunoaștere facială precum gaydar? Câți dintre aceștia ar accepta să fie plătiți pentru asta? Pe de altă parte, o parte importantă din populație ar contribui pro bono la găsirea unui leac pentru cancer. Un al doilea aspect constă în faptul că valoarea ridicată a

datelor personale ar conduce la crearea unui privilegiu pentru cei bogați, deoarece doar aceștia și-ar putea permite să le cumpere.

Totuși, nu putem afirma cu rigiditate că acțiunea individuală ar fi suficientă. Acțiunea colectivă prin instituțiile statului este și ea necesară. Chiar dacă o mică parte a populației își oferă informațiile personale, rezultatul ar putea avea o mai mare acuratețe decât în cazul oferirii informațiilor de către majoritatea populației. Nu toți suntem conștienți de asta. Pentru a preveni situații nedorite sunt necesare legi și dezbateri publice.

The Economist a scris că cea mai valoroasă resursă a lumii nu mai este petrolul, ci informațiile, iar acestea sunt foarte diferite de petrol. Pe lângă faptul că reprezintă o resursă nelimitată, ele sunt deținute de persoane și este cel mai bine să nu atașăm valoare economică schimbului lor. Lipsirea petrolului de potențialul profit ar închide piața de petrol. Ca un prim pas, lipsirea informațiilor de valoare economică creează mediul necesar menținerii unor standarde etice care pot supraviețui avântului inteligenței artificiale, și deschid drumul către gestionarea ignoranței colective. Cu alte cuvinte, dacă informațiile devin unele dintre cele mai utile bunuri ale lumii moderne, ele trebuie să devină și cele mai ieftine.

NOTE

1. Wang, Y. & Kosinski, M. “Deep neural networks are more accurate than humans at detecting sexual orientation from facial images”. *Journal of Personality and Social Psychology* 114, 246-257 (2018). ↑
2. Dressel, J. & Farid, H. “The accuracy, fairness, and limits of predicting recidivism”. *Science Advances* 4, eaao5580 (2018). ↑
3. Hertwig, R. & Engel, C. “Homo ignorans: Deliberately choosing not to know”. *Perspectives on Psychological Science* 11, 359-372 (2016). ↑
4. Gigerenzer, G. & Garcia-Retamero, R. “Cassandra’s regret: The psychology of not wanting to know”. *Psychological Review* 124, 179-196 (2017). ↑
5. Athey, S. Catalini, C., & Tucker, C.E. “The digital privacy paradox: Small money, small costs, small talk.” *Stanford University Graduate School of Business Research Paper* No. 17-14 (2018). ↑

Imagine: Unplash