Dacă inteligența artificială e atât de deșteaptă, de ce se comportă precum un arbitru corupt?[1]

Despre discriminarea sistemelor de inteligență artificială

Introducere

Evident, un sistem de inteligență artificială (IA) poate să se comporte ciudat sau să spună *nu* pe motive justificate, ca atunci când sistemul nu te trece un examen online fiindcă nu ai dat numărul minim de răspunsuri corecte pentru promovare. Dar articolul de față nu vorbește despre aceste cazuri, ci de cel când ești nedumerit de răspunsul primit și bănuiești că s-ar putea să fie ceva incorect la mijloc - fie o discriminare, fie o manipulare.

În articolul precedent am vorbit pe larg despre sistemele de inteligență artificială (ce sunt, ce nu sunt și ce pot fi) și am identificat problemele majore ale acestor tehnologii. Dacă doriți să explorați asta, găsiți articolul aici.

Discriminarea poate lua multe forme. De obicei, discriminarea[2] este un act negativ comis de un om, sau un grup de oameni, asupra altui om sau asupra altui grup de oameni. Ne gândim, de cele mai multe ori, la discriminare ca fiind pur omenească, ceva ce fac *oamenii* împotriva altor *oameni*. Discriminarea implica atingerea drepturilor fundamentale ale unui grup de oameni de acțiunile unor alți oameni, prin deosebire, excludere, adresarea de invective, violență, reprimare, stereotipuri, refuzarea de servicii și așa mai departe. Tradițional, în spatele unor astfel de acțiuni puteai găsi o față umană, care putea fi trasă la răspundere. Bucla omenească era închisă - doar un om (sau grup de oameni) putea atinge pe un altul.

---

[1] Titlul este exact întrebarea ingenuă a unui livrator Glovo în articolul din DoR "Ce nu știi despre livratorul tău de mâncare" (iunie 2022)

[2] Pentru o definiție legală vezi art 2 (1) din OG 137/2000 "prin discriminare se înțelege orice deosebire, excludere, restricție sau preferință, pe bază de rasă, naționalitate, etnie, limbă, religie, categorie socială, convingeri, sex, orientare sexuală, vârstă, handicap, boală cronică necontagioasă, infectare HIV, apartenență la o categorie defavorizată, precum și orice alt criteriu care are ca scop sau efect restrângerea, înlăturarea recunoașterii, folosinței sau exercitării, în condiții de egalitate, a drepturilor omului și a libertăților fundamentale sau a drepturilor recunoscute de lege, în domeniul politic, economic, social și cultural sau în orice alte domenii ale vieții publice."

Doar că ultimele dezvoltări din domeniul inteligenței artificiale (IA) ne arată că lucrurile nu stau chiar așa. **Discriminarea poate fi făcută și de un calculator sau un sistem informatic.** Dar nu era calculatorul acel sistem perfect, rece, calculat și obiectiv care nu poate nici măcar să greșească și deci nu poate nici discrimina? Să explorăm împreună pornind de la cazuri concrete și exemple practice cum poate discrimina un sistem informatic, de ce, cine poate fi responsabil și alte aspecte conexe.

În practică, lumea inteligenței artificiale de până acum ne arată că, de exemplu, discriminarea de sex și gen este perpetuată - pe steroizi - de sisteme automatizate. Discriminarea automatizată, cu ajutorul sistemelor de inteligență artificială poate să fie găsită în medicină, prin refuzul automatizat de prestare de servicii medicale față de anumite grupuri, în acordarea de credite bancare, infrastructura discriminatorie, recunoaștere facială și vocală, la angajare (scanarea de CV-uri), în sisteme de traducere automatizate (de ex. "cleaner" din engleză tradus în *femeie de serviciu).*

Inteligența artificială contemporană ne arată o realitate indezirabilă: algoritmi automatizați antrenați pe baze de date nediversificate care ne prezintă o lume profund negativă, unde prejudecățile celor care scriu codul sursă al acestor sisteme ies la lumină. Mai mult, atunci când un sistem automatizat de inteligență artificială discriminează, apare problema atribuirii responsabilității.

Iluzia obiectivității sistemelor de inteligență artificială perpetuează prejudecățile și discriminarea. Discriminarea poate fi întâlnită în algoritmi impersonali care sunt, în aparență, cât se poate de obiectivi, în modele de limbaj extinse (large language models) ca ChatGPT, în inteligența artificială generativă precum Stable Diffusion pentru care realitatea înseamnă că numai bărbații albi pot fi liderii unei companii, femeile rareori sunt avocați, medici și așa mai departe, iar oamenii de culoare comit crime. Aceste exemple ridică întrebări fundamentale - reprezintă rezultatele acestor sisteme IA realitatea sau construiesc o realitate deformată pe baza datelor de antrenare? Dovezile actuale tind a înclina balanța pe ultima variantă.

Discriminarea algoritmică amplificată de tehnologii și sisteme de inteligență artificială riscă să devină o problemă la o scară extraordinar de mare, având în vedere amploarea cu care sunt folosite sistemele de inteligență artificială în serviciile digitale de astăzi. Unii experți sugerează că aproape 90% din conținutul de pe internet din viitor va fi produs cu ajutorul unor sisteme IA generative. Stable Diffusion, foarte popular azi, suferă de probleme grave în materie de stereotipuri. Un internet plin de stereotipuri generate cu ajutorul inteligenței artificiale generative ar fi un loc mult mai întunecat.

Conform unui studiu făcut de Consiliul Europei, o organizație internațională care are ca scop declarat protecția drepturilor fundamentale ale omului, **impactul inteligenței artificiale asupra drepturilor omului este "unul dintre cei mai importanți factori care vor defini perioada în care trăim"**. Așadar, discriminarea automatizată algoritmică, pe bază de sisteme de inteligență artificială este cât se poate de reală și imediată, iar adresarea ei necesită o discuție critică și atentă la modul în care inteligența artificială este folosită azi pentru perpetuarea discriminării.

De ce poate discrimina inteligența artificială?

Discriminarea automatizată se referă la procesul în care sistemele informatice sau algoritmi automatizați de tipul IA iau decizii care pot afecta sau cauza un tratament inechitabil în funcție de anumite caracteristici ale unor persoane, cum ar fi rasă, gen, vârstă, orientare sexuală sau alte caracteristici arbitrare (înălțime, greutate, aspect, etc). De exemplu, tehnologiile software de recunoaștere facială au o istorie lungă și problematică cu recunoașterea fețelor persoanelor de culoare. În unele cazuri, sistemele IA au fost acuzate de faptul că întăresc practici rasiste discriminatorii. Această formă de discriminare are loc atunci când algoritmii sau sistemele de inteligență artificială aplică în mod implicit prejudecăți din datele de antrenare, perpetuând sau chiar accentuând inegalitățile, inechitățile și nedreptățile existente în societate.

Discutăm în detaliu mai jos despre **3 cauze** și **2 posibile soluții (garanții)** cu privire la aceasta problemă de discriminare automatizată.

Cauza I - Datele nediversificate și suprareprezentarea unor populații

Inteligența artificială, despre care am vorbit pe larg în alt articol, are la bază un sistem de decizii automatizate bazat pe un anumit set de date. Asta înseamnă că la baza majorității unor sistem de inteligență artificială cu care ne confruntăm în viața noastră online există niște informații, numite date de antrenare (date, poze, text, audio-video, etc.) despre niște oameni care au modelat modul în care acest sistem IA se comportă, își ia deciziile și cum ajunge la niște concluzii. Prin urmare, teoretic, cu cât este mai diversificat setul de date de antrenare, cu atât sistemul IA ar trebui să aibă răspunsuri cât mai diversificate și cât mai nuanțate, reducând posibilitatea producției unui răspuns discriminator. Doar că, momentan, lucrurile arată diferit.

În primul rând, discriminarea poate fi un rezultat al lipsei acestor date diversificate. Dacă un set de date de antrenare al unui sistem IA de angajare este nediversificat și are numai bărbați trecuți ca ingineri software în baza de date, sistemul IA respectiv va decide, în mod automat, că numai bărbații pot fi ingineri software și că numai ei trebuie angajați pentru un asemenea post, ca în cazul sistemului de angajare de la Amazon din 2014. Femeile nu vor fi luate în considerare de către sistem decât în situații de excepție. Iar procesul prin care se aleg datele pentru antrenare poate fi și el unul discriminatoriu și părtinitor (*biased)*. După cum spun doi cercetători în discriminarea automatizată, Lum și Isaac, în cazul anchetelor poliției, care pot folosi sisteme IA pentru a ajuta autoritățile să depisteze suspecți de anumite tipuri de infracțiuni, este posibil ca atenția exagerată față de un grup sau o comunitate să ducă la suprareprezentarea acelor comunități în datele culese de polițiști. Suprareprezentarea unor grupuri, mai ales minoritare, în relație cu anumite tipuri de acțiuni, de obicei infracțiuni, poate să învețe acele sisteme IA că acele grupuri minoritate sunt mai predispuse să comită acele infracțiuni, creând un cerc vicios. Astfel, sistemele IA ajung să exagereze corelațiile dintre unele grupuri și infracțiuni, datorită datelor de antrenare în care o populație este suprareprezentată în raport cu un fapt (în cazul de față, cu o infracțiune), ducând la consolidarea stereotipurilor și discriminării la adresa grupurilor care au fost asociate cu astfel de infracțiuni într-un mod nedrept.

De exemplu, cazul COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) adus în ochii publicului de către jurnaliștii de investigație de la publicația americană

ProPublica este ilustrativ în acest sens. COMPAS este un instrument de evaluare a riscului recidivei pentru a ajuta judecătorii americani să determine probabilitatea recidivei pentru infractori. ProPublica a analizat datele culese din statul american Florida și a descoperit că algoritmul COMPAS prezenta probleme de discriminare. Concluziile lor au indicat că algoritmul avea tendința de a prezice în mod incorect că unii indivizi de culoare aveau un risc mai mare de recidivă decât se dovedea în realitate, în timp ce subestima riscul pentru unii infractori albi, perpetuând, astfel, stereotipuri rasiale. Cazul COMPAS a ajuns să fie un exemplu de manual pentru cum modul în care sunt culese și procesate datele de antrenare ale sistemelor IA pot influența rezultatul produs de astfel de sisteme automatizate.

Cauza II - Caracterul tip cutie neagră a sistemelor IA

În al doilea rând, unele sisteme de inteligență artificială, cum ar fi cele care folosesc neural networks și deep learning, sunt practic niște sisteme de tip cutie neagră. Termenul de "cutie neagră" este folosit pentru a defini tehnologii care au rezultate impresionante dar al căror proces decizional nu este înțeles complet, de multe ori chiar nici de creatorii săi. O cutie neagră este netransparentă prin definiție, fiindcă persoanele care ar trebui să raporteze despre comportamentul sistemului nu reușesc să descrie sau să înțeleagă pe deplin ce se întâmplă.

Discriminarea automatizată algoritmică poate fi propagată de acest fapt. Dacă nu știm de ce ajunge un sistem IA la rezultatul la care ajunge, nu putem fi siguri dacă acesta discriminează sau doar reproduce o serie de afirmații sau un model al realității în conformitate cu datele sale de antrenare. Absența transparenței decizionale din spatele unor astfel de sisteme de inteligența artificială de tip cutie neagră poate facilita discriminarea structurală. În alte cazuri unele companii refuză în mod intenționat să-și transparentizeze sistemele IA, folosind dreptul la secretul comercial pentru a nu explica nimic public. Creatorii unor astfel de sisteme doresc să se poată absolvi de responsabilitatea pe care o pot avea pentru un astfel de conținut care poate atinge drepturile noastre fundamentale, prin a invoca inexplicabilitatea rezultatului unui astfel de sistem de inteligență artificială. **Caracterul de tip cutie neagră a unor sisteme de inteligență artificială poate ajuta la protecția actorilor rău-intenționați, sau cu intenții discriminatorii, prin simplul fapt că nu putem înțelege tehnic mijlocul prin care un astfel de sistem își produce rezultatele.**

De exemplu, un tribunal din Amsterdam a decis că compania de taximetrie Uber nu a respectat un ordin judecătoresc atunci când un algoritm a luat decizia de a concedia doi șoferi din Marea Britanie și Portugalia, într-un mod netransparent. Hotărârea judecătorească împotriva companiei Uber a fost pronunțată în aprilie 2023 de Curtea de Apel din Amsterdam și a impus companiei de taxiuri și livrări să ofere transparență în ceea ce privește modul în care a folosit automatizarea pentru a lua decizia de a-i concedia pe cei doi muncitori. Compania Uber a susținut că nu ar trebui să le ofere muncitorilor mai multe informații despre decizia automatizată, luată de sistemul lor intern de IA, pentru a-și proteja secretele comerciale. Uber a declarat că conturile șoferilor au fost semnalate din cauza unor "circumstanțe foarte specifice" care nu au fost menționate și că deciziile au fost analizate de echipe umane. Doar că instanța a decis că analiza umană a deciziei automatizate a fost un act „simbolic", fără consecințe reale sau schimbări semnificative. Uber a pierdut cazul, iar tribunalul districtual din

Amsterdam a ordonat ca Uber să plătească daune de peste 500.000 de euro sub formă de penalități șoferilor concediați pe nedrept și în mod complet netransparent.

Cauza III - Iluzia obiectivității

În al treilea rând, iluzia obiectivității sistemelor de inteligență artificială ajung să creeze o imagine distorsionată despre modul lor de operare. **Sistemele de inteligență artificială nu sunt obiective. Conținutul produs de acestea nu este neutru. Ele oglindesc prejudecățile și opiniile subiective ale creatorilor lor, prejudecăți care pot fi găsite în codul sursă a acestor sisteme de inteligență artificială.** Obiectivitatea sistemelor de inteligență artificială trebuie pusă sub semnul întrebării, fiindcă aceste sisteme nu au față de ce să fie obiective – ele sunt mașinării care reproduc în mod corelativ și statistic ce știu din datele și din modul de antrenare, corelate indicațiilor de "reglaj fin" ale celor ce le implementează, dar și cu modul de interpretare a cerinței utilizatorului. Din această rețetă nu poate reieși obiectivitate. Din mai multe elemente subiectiv alese, cum ar fi datele de antrenare, implementarea lor, și interpretarea inginerilor a relevanței unor date în detrimentul altora (care implică o alegere subiectivă), nu are cum să iasă un rezultat obiectiv.

După cum spunea filosoful Thomas Nagel, perspectiva obiectivă este doar un mod de a înțelege lumea detașându-te cât mai mult de particularitățile propriei personalități. Un sistem IA nu are particularități de genul, acestea fiind specific omenești. Așadar, acestea nu pot fi obiective, în schimb, ele sunt părtinitoare în mod recurent, dacă prin asta înțelegem redarea unei perspective asupra lumii care e în conformitate cu ce există în datele de antrenare și a altor intervenții în procesul de funcționare. Loialitățile unui sistem de inteligență artificială cum este Stable Diffusion - pentru care, la un moment dat rezultă că, femeile nu pot ocupa funcții de leadership iar oamenii de culoare sunt în mod invariabil infractori – sunt față de datele de antrenare și față de compania mamă.

Acestea sunt motivele principale pentru care sistemele de inteligență artificială ajung să discrimineze. Ele sunt produse omenești care reflectă prejudecățile și *subiectivitățile* omenești. Tema principală care domină cele trei motive ilustrate mai sus este lipsa transparenței sistemelor de inteligență artificială. Lipsă de transparență și caracterul lor opac de tip cutie neagră sunt motivele principale, dar nu singurele, pentru care sistemele IA ajung la rezultate inexplicabile. Iar această inexplicabilitate perpetuează discriminarea, prin faptul că menține într-o stare de opacitate și beznă mecanismele prin care rezultate discriminatorii ajung să fie generate și regenerate în societate. Aceasta afectează în mod substanțial drepturile noastre fundamentale.

O primă garanție - Importanța transparenței

După cum a ilustrat cazul mai sus menționat legat de muncitorii Uber, lipsa de transparență poate duce la concedierea unor persoane din motive imposibil de explicat care pot ascunde niște decizii discriminatorii în fond. Transparența implica vizibilitatea procesului prin care un sistem IA ajunge să ia astfel de decizii cu consecințe serioase și imediate. În absența transparenței, sistemele de inteligență artificială nu pot fi investigate, iar răspunsul lor rămâne de nechestionat.

Acest aspect perpetuează iluzia obiectivității unor astfel de sisteme IA care de fapt sunt părtinitoare și loiale față de datele lor de antrenare și alte informații conexe.

Așadar, transparența este de o importanță maximă în aceste situații. Transparența decizională a sistemelor de inteligență artificială ne asigură că acestea pot fi investigate și că mijloacele prin care ele ajung la rezultate ce pot fi considerate discriminatorii pot fi expuse publicului și, daca este cazul, corectate. Doar că problema fundamentală explicată în detaliu mai sus este că unele sistemele IA sunt cutii negre, cum ar fi cele care folosesc *deep learning* sau *neural networks*. Așadar, transparența trebuie regândită ca parte componentă a designului sistemelor de inteligență artificială curente și viitoare, pentru a reduce gradul de opacitate al unor astfel de sisteme IA. Cutia neagră trebuie deschisă, iar unii cercetători din domeniul sistemelor IA lucrează exact la asta. Transparența poate facilita construcția de sisteme IA orientate spre comportament etic, lăsând loc îmbunătățirii mijloacelor de producție a rezultatelor sistemelor IA. După spusele filosofului Mark Coeckelbergh, o astfel de transparență intrinsecă ar fi benefică modelelor de inteligență artificială și creatorilor lor, făcându-le mai ușor de înțeles și mai ușor de reparat și îmbunătățit.

Judecata umană ar trebui să primeze, fiindcă numai un om poate cuprinde înțelesurile complexe ale unui caz de natură socială - un caz profund omenesc. Ambele decizii, algoritmice și umane, trebuie să fie transparente, astfel încât persoana prejudiciată să înțeleagă raționamentul din spatele deciziei și să poată acționa, inclusiv de a contesta decizia respectivă.

Acesta este și fundamentul mai multor dispoziții normative din Uniunea Europeană:

Astfel, conform articolului 22 (1) din Regulamentul UE 679/2016 (GDPR)[3], deciziile automatizate trebuie mereu însoțite de judecată umană[4], dacă rezultatul are efecte juridice sau efecte semnificative pentru persoanele fizice afectate. Practic judecata umană trebuie să cântărească în mod *decisiv în fiecare caz*, nu *simbolic și generic* ca în cazul Uber mai sus menționat.

Mai mult, în Regulamentul privind Inteligența Artificială (*AI ACT)*[5] obligațiile de transparență - a deciziilor, dar și a interacțiunii - sunt cuprinse în diverse articole și în special în titlul IV. De exemplu, furnizorii de sisteme de inteligență artificială sunt obligați să comunice faptul că un utilizator vorbește, cu un sistem IA și nu cu un om. La fel se aplică și în cazul tehnologiilor de recunoaștere a sentimentelor, sau cele care categorizează oamenii în funcție de datele biometrice culese, cu niște excepții.

O a doua garanție - Dreptul la o explicație

[3] „Persoana vizată are dreptul de a nu face obiectul unei decizii bazate exclusiv pe prelucrarea automată, inclusiv crearea de profiluri, care produce efecte juridice care privesc persoana vizată sau o afectează în mod similar într-o măsură semnificativă".

[4] Pentru modul în care trebuie aplicat art 22 vezi și recentă decizie a Curții Europene de Justiție în cazul Schufa - C-634/21 (7.12.2023): https://curia.europa.eu/juris/document/document.jsf?docid=280426&mode=req&pageIndex=1&dir=&occ=first&part=1&text=&doclang=RO&cid=2148345

[5] Actualmente într-o formă extrem de avansata de adoptare înainte de publicarea în Jurnalul Oficial al UE - vezi varianta din 21 ianuatie 2024 - - https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AIA-Final-Draft-21-January-2024.pdf

Dreptul la explicație, în genere, se referă la un răspuns satisfăcător primit de o persoană fizică în fața unei decizii automatizate. Acesta nu s-ar referi la o explicație legată de modul detaliat în care funcționează un sistem IA complex. Pe scurt, dreptul la o explicație se referă la o justificare umană semnificativă a rezultatului final al unui sistem IA.

Pe de o parte în art 13-15 din Regulamentul UE 679/2016 (GDPR) este inclusă în categoria informațiilor care trebuie să fie dezvăluite de operator într-o notă de informare (de obicei cunoscută ca Politică de Confidențialitate în cazul unui site online) sau în cazul exercitării dreptului de acces și "existența unui proces decizional automatizat (...), precum și, cel puțin în cazurile respective, **informații pertinente privind logica utilizată** și privind importanța și consecințele preconizate ale unei astfel de prelucrări pentru persoana vizată."Pentru o analiză mai aprofundată, Andrew D Selbst și Julia Powles, argumentează într-un articol științific pentru *International Data Privacy Law* din 2017 argumentează că acest text reprezintă de fapt un fel de "drept la explicație".

Prin noul Regulamentul privind Inteligența Artificială (*AI ACT)* dreptul la o explicație va deveni o realitate concretă[6] în cazul interacțiunii cu un sistem IA clasificat drept "risc ridicat" dacă acest sistem afectează persoana fizică într-un mod semnificativ, sau dacă aduce atingeri la adresa drepturilor fundamentale.

În practică dreptul la explicația ar trebui să fie ceva de tipul următor: dacă o persoana ar aplica pentru un post în marketing și ar primi un răspuns vag despre cum nu a fost angajată, de la sistemul IA folosit pentru recrutări, această persoană ar avea dreptul la un răspuns clarificator legat de motivele *reale* pentru care nu a fost selectată pentru un interviu. Iar acest răspuns ar veni, în parte, de la un angajat uman, care înțelege, în mare, modul în care funcționează un astfel de sistem IA și poate să înțeleagă ce s-a întâmplat în acest caz concret.

De exemplu, primiți următorul mesaj "Nu ați ajuns în etapa de interviu pentru că nu ați trimis o scrisoare de intenție" sau, un alt exemplu "Nu ați ajuns în etapa de interviu pentru că pentru acest loc de muncă căutăm o persoană cu peste 10 ani vechime în domeniu, iar dvs. ați terminat facultatea acum 2 ani", dar *nu* "Nu ați ajuns în interviu, pentru că aplicația dvs. a fost neconformă conform termenilor aplicației noastre".

Dreptul la explicație implică faptul că cei care oferă explicația *înțeleg* raționamentul - sau măcar modul în care ajunge la răspuns - sistemului IA care a luat decizia care a afectat în mod semnificativ viața unei persoane fizice. Momentan, în cele mai multe cazuri o mână de experți înțelege modul în care sistemele IA ajung la deciziile la care ajung, iar în cazul celor de tip cutie neagră (neural networks, deep learning și machine learning) putem spune că nu înțelege aproape nimeni, în mod complet, modul în care rezultatele sunt produse.

Așadar, dreptul la explicație implică mai mult decât un simplu motiv pentru care cineva nu a fost angajat, de exemplu. Dreptul la explicație implică sisteme IA transparente a căror decizii cu impact semnificativ pot fi expuse publicului și *înțelese de acesta*. Asta înseamnă că deciziile automatizate necesită oameni care le înțeleg și le pot traduce pentru cei ce suferă din cauza rezultatelor lor.

---

[6] Vezi art 68 c) din varianta menționată anterior.

Dacă transparența ar fi gândită ca principiu fundamental de design al sistemelor de inteligență artificială, atunci dreptul la o explicare ar putea deveni o realitate iar protecția drepturilor noastre fundamentale ar avea de câștigat. Rezultatele sistemelor IA ar putea fi acompaniate de raționamente și explicații omenești, iar sistemele IA și cei care le folosesc ar putea fi trași la răspundere în mod direct. În absența celor două, discriminarea ar rămâne perpetuată iar motivele pentru care ea se perpetuează, atât cât și mecanismele prin care se întâmplă așa ceva, vor rămâne opace și obscure.

Viitorul discriminării automatizate

Drepturile noastre fundamentale necesită garanții în fața sistemelor IA, iar transparența și un drept la explicație sunt abia primii pași esențiali, nicidecum suficienți în toate cazurile. Fără acestea, ele vor avea de suferit. Așa că inițiativa luată de Regulamentul privind Inteligența Artificială, de a aduce un drept la explicație și un accent pronunțat pe transparență, este un bun punct de început pe care să construim mai mult, mai departe. Rămâne de văzut dacă va fi și aplicat în litera și spiritul său.

Dacă lăsăm mașinării inteligente, precum sistemele IA, să perpetueze ce este mai rău la om și la societatea sa, ajungem să ne lovim pe noi înșine, fiindcă o caracteristică fundamentală a discriminării este că este arbitrară. Asta înseamnă că, la un moment dat și într-un anumit context, oricine poate fi discriminat. Dacă, într-o zi, normele și cutumele sociale se schimbă, cei care discriminează pot ajunge la capătul celălalt al spectrului, ajungând discriminați.

În mod ironic, poate, viitorul discriminării este viitorul inteligenței artificiale. Iar acest viitor este construit în mod colectiv de cetățeni implicați în mod direct cu acest viitor. Nu putem preveni toate problemele cauzate de această tehnologie, dar putem preîntâmpina unele dintre ele - mai ales cele care s-au văzut deja -, iar pe altele le putem repara pe drum.