

## Îți poate fi Chatbotul prieten?

### *Inteligența artificială și falsa sa relație cu omul*

*În articole precedente 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 și cum ne pot afecta drepturile fundamentale. Dacă doriți să explorați asta, găsiți articolul principal [aici](#) sau cel despre discriminarea prin sisteme IA [aici](#).*

#### **Introducere - ce sunt chatboții sociali - de la cuvinte la exemple**

Inteligența artificială nu se rezumă doar la aplicații practice, industriale sau de distracție. Ea poate fi folosită și pentru a relaționa cu oamenii. Chatboți ”sociali”, cum ar fi [Replika](#) (pentru relații intime), [Woebot](#) (pentru sănătatea ta mintală), și [Kuki](#) (pentru socializare), au ajuns, prin telefoanele noastre și prin tehnologiile digitale pe care le folosim zilnic, să înlocuiască și să suplimenteze, încet, încet, relaționarea interumană.

*Social este pus între ghilimele fiindcă nu există un agent, un subiect în spatele chatbot-ului care să poată da dovadă de socialitate omenească.*

Chatboții ”sociali” funcționează pe tehnologii similare cu ChatGPT și sunt modele extinse de limbaj (large language models). Scopul acestor chatboți este conversația socială și dialogul, ei fiind partenerii (*companions*) utilizatorilor.

Acești chatboți ”sociali”, foarte populari azi în anumite medii, în mod predominant printre vorbitorii de limba engleză, nu sunt doar niște mașinării de producție de text la infinit pentru conversație, ci pot fi folosiți:

- pentru monitorizarea utilizatorilor (cei mai mulți fiind copii sau adolescenți),
- pentru manipulare emoțională sau informațională și
- pentru crearea unei dependențe afective față de chatbot.

Toate acestea pot afecta drepturile noastre fundamentale, dar mai ales pe cele ale celor mai vulnerabili cum ar fi copiii sau oamenii care suferă de tulburări mintale.

De exemplu, Replika - care folosește tehnologii precum *neural networks și machine learning* - [cel mai folosit chatbot](#) ”social” printre copii, adolescenți și nu numai, pretinde că furnizează o ”prietenie virtuală” sub forma unui avatar (bărbat/femeie, personalizat). Aplicația se vinde sub pretextul ameliorării emoționale a utilizatorului. [Are aproximativ 2 milioane de utilizatori](#), conform spuselor Eugeniei Kuyda, directorul executiv al companiei, și este folosit de mulți drept partener intim de conversație. Aplicația Replika avea la începuturile ei o funcție de conversație de natură sexuală, lucru care a fost eliminat în timp, fiindcă robotul învățase, de la utilizatorii săi, [cum să hărțuiască sexual o persoană](#). Replika a ajuns să perpetueze comportamentul indezirabil al multor utilizatori și a început să rănească alți noi utilizatori. Eliminarea funcției de conversație de natură sexuală a fost controversată printre utilizatori, [pentru unii fiind unicul motiv pentru utilizarea aplicației](#). Totuși, asta nu a oprit utilizatorii din a folosi aplicația în continuare. Pentru unii, o conversație cu acest chatbot este [singura conversație plăcută pe care au avut-o într-o zi](#).

Nici chiar riscurile aduse la adresa datelor lor cu caracter personal nu i-a oprit pe cei mai mulți din a folosi această aplicație. După ce în 2023 [le-a fost interzisă prelucrarea datelor cu caracter personal în Italia](#) pentru faptul că prelucra într-un mod ilegal datele cu caracter personal ale copiilor, numărul de utilizatori nu a fost afectat într-un mod semnificativ.

Mai mult, [acești chatboti devin din ce în ce mai populari](#), chit că aduc atingeri serioase la adresa drepturilor fundamentale.

Un alt exemplu este Woebot, cunoscut drept chatbot-ul terapeut, deținut de o companie din SUA, descărcat [de aproximativ 1,5 milioane de oameni](#) - conform datelor publicate de compania mamă cu același nume. Acesta pretinde că folosește metodele terapiei cognitiv-comportamentale pentru a-și ajuta utilizatorii ”[sa se facă bine](#)”. [Woebot a fost acuzat de faptul că folosește informații despre metode terapeutice - în cadrul conversațiilor cu utilizatorii - într-un mod redundant sau că, în mod evident, nu poate ajuta la fel ca un terapeut](#). Dar unii văd acest chatbot ca pe un mijloc de a accesa ajutor psihologic într-un mod gratuit. Gratuitatea este iluzorie, de cele mai multe ori utilizatorii plătind cu datele lor cu caracter personal și mai ales cu informații extraordinar de sensibile legate de sănătatea lor mintală. În cadrul unei ședințe de terapie dintre un psiholog și o persoană fizică, confidențialitatea este garantată prin [lege, inclusiv în România](#). Când vine vorba de Woebot, lucrurile nu sunt clare. Iar vulnerabilitatea utilizatorului (sau am putea spune a pacientului?) este mijlocul prin care compania din spatele Woebot își face, probabil, profitul. Compania [a fost atât de profitabilă în anul 2021](#), încât a mai primit o injecție de capital de peste 90 de milioane de dolari de la investitori.

Chatbotii aceștia ”sociali” sunt mijlocul prin care unii oameni interacționează în mod explicit cu sisteme de inteligență artificială, care nu sunt *încă* marcate ca atare. Acest lucru se va schimba în Uniunea Europeană odată cu implementarea Regulamentului privind Inteligența

Artificială (*AI ACT*), care obligă prin lege comunicarea și transparența față de utilizator a faptului că acesta [vorbește cu un chatbot](#) și nu cu o ființă umană.

Doar că acest lucru nu este de ajuns, din cauza unei [tendințe psihologice umane de a antropomorfiza chatboții](#). Antropomorfizarea înseamnă atribuirea unor caracteristici specific omenești obiectelor fără viață (câteodată și animalelor). În cazul de față, antropomorfizarea înseamnă atribuirea textelor automate ale chatboților ”sociali” - care sunt lipsite de *sens* în sine - de însemnătate omenească, intenție ca la un om și chiar și o personalitate, acolo unde nu există una, fiindcă este vorba de un algoritm și nu de o persoană umană. Lucrul acesta, când este făcut în mod înșelător, [poate duce la exploatarea credulității omenești](#). Acești chatboți ridică niște probleme fundamentale legate de viața noastră privată, de influențarea sentimentelor și opiniilor altor persoane (mai ales a copiilor) și de manipulare directă. Astfel, în acest articol vom explora cum pot afecta drepturile noastre fundamentale popularizarea interacțiunilor zilnice cu acești chatboți ”sociali”:

## 1. De la singurătate la perpetuarea excluziunii sociale

Chatboții alimentați de inteligență artificială, precum Replika (care funcționează cu sisteme IA de la ChatGPT) sau Woebot, [au crescut în popularitate în special în urma pandemiei COVID-19](#). În paralel cu pandemia epidemiologică, exista atunci și acum una de natură psihologică. [Supranumită epidemia singurătății, conform mai multor psihologi și Organizației Mondiale de Sănătate](#), această epidemie este una deseori ignorată, fiind aproape invizibilă dar cât se poate de concretă. În realitate, un număr considerabil de oameni de azi se simt foarte singuri, [conform neuropsihologului John Cacciopo](#), expert în psihologia singurătății. Iar sentimentul de singurătate și [tendința comportamentală spre izolare socială în lumi virtuale sau în conversații cu chatboți sunt cel mai des întâlnite la generația tânără](#). În Uniunea Europeană, [conform unor statistici din 2023](#), cel puțin 1 din 10 europeni se simte singur. În România [procentul este de 14-16% de oameni care se simt singuri deseori în viețile lor de zi cu zi](#), cei tineri fiind cei mai predispuși la sentimente de singurătate.

În cartea sa, [Legături Pierdute](#), jurnalistul englez [Johann Hari](#) documentează efectele singurătății excesive, efecte care se traduc în tulburări depresive și de anxietate. Acești factori, cuplați desigur și cu o curiozitate firească, omenească, au dus la proliferarea unei industrii imense de ”prieteni virtuali” (chatboți), [percepuți de mulți tineri ca reali ”parteneri romantici”](#), care înlocuiesc sau suplimentează relațiile interumane. Conform unei estimări, piața chatboților ”sociali” [va atinge 7 miliarde de dolari în 2024](#). Sherry Turkle, expertă în științe sociale și interacțiunea tinerilor cu tehnologiile noi, spune că acești chatboți ”sociali” aduc iluzia parteneriatului pentru utilizatori lor, scutindu-i pe aceștia de implicarea reală pe care o necesită intimitatea – iar asta este atractiv pentru un adolescent sau un copil, care doar vrea să exploreze o conversație fictivă cu un chatbot. [Ea mai denunță, la fel, pretențiile la ”empatie” ale acestor chatboți ca fiind fanteziste](#). Chatboții ”sociali” nu pot fi empatici, fiindcă empatia implică experiențe comune, iar acești chatboți n-au, prin definiție, nici un fel de ”experiență”.

Absența empatiei reale nu contează pentru utilizatori, ci contează empatia *percepută* de cel care vorbește cu un astfel de chatbot. Discuțiile cu un astfel de chatbot pot ameliora sentimentele de singurătate. [Un studiu publicat în revista \*Nature\* arăta că 90% din utilizatorii eșantionați pentru studiu \(n=1006\) sufereau de un caz grav de singurătate psihologică, care s-a ameliorat semnificativ după interacțiuni recurente cu Replika. \[Unora le-a redus gândurile suicidare într-o manieră semnificativă.\]\(#\) Astfel, acești roboți nu sunt capabili de empatie omenească reală, dar pot simula empatia în conversație destul de bine încât să ajute utilizatorii. Doar că ajutorul primit de la chatbot, robot care poate doar simula empatia, nu se compară cu ajutorul primit de la o ființă umană la fel de complexă ca suferindul, care poate empatiza pe baza unor experiențe comune, reale și concrete cu cel care are nevoie de ajutor și, astfel, să-l ajute să se orienteze. Ajutorul psihologic trebuie suplimentat, nu înlocuit.](#)

De aici rezultă că interesul multor utilizatori față de chatboții aceștia ”sociali” poate reieși [dintr-un cocktail de anxietăți](#), vulnerabilități, sentimente de nesiguranță, curiozitate intrinsecă vârstei tinerilor și lejeritatea cu care se poate accesa o conversație cu un astfel de chatbot. Mai mult, utilizatorii de astfel de chatboți ”sociali” tind să fie copii și adolescenți, care se află în perioade critice de dezvoltare neuropsihologică. [Jen Persson, de la ONG-ul Defend Digital Me](#) care se ocupă cu protecția vieții private a copiilor în spațiul digital, a spus că tehnologiile precum Replika riscă să afecteze creșterea copiilor și că acestea ar trebui tratate drept [”produse de sănătate” și reglementate în mod strict.](#)

Foarte mulți adolescenți simt nevoia de explorare identitară sau sexuală, iar chatboții ”sociali” pot ajuta în acest sens, doar că există mereu riscul manipulării emoționale sau informaționale. Tinerii care încă n-au dezvoltat facultăți mintale critice, din cauza vârstei sau lipsei de acces la educație, pot crede informațiile, mai ales cele noi, primite de la un chatbot precum ChatGPT, Replika și Woebot, chit că acești chatboți n-au nici un fel de expertiză reală. Și asta duce la manipulare informațională, mai ales că acești chatboți sunt predispuși [la minciuni](#) și la [”halucinații”](#).

Sentimentele de singurătate pot fi exploatare de acești chatboți ”sociali” prin perpetuarea excluziunii sociale a tinerilor predispuși la singurătate. Roboții pot crea iluzia interacțiunii omenești, fără ca acest lucru să se întâmple în realitate. Iar asta, într-un cerc vicios, poate duce la eroziunea abilităților sociale ale tinerilor: absența competențelor sociale perpetuează comunicarea cu chatboții ”sociali”, care duce la o absență și mai mare a acestor dintâi competențe.

## **2. Manipularea emoțională a chatboților**

În martie 2023, [un cetățean belgian s-ar fi sinucis după o discuție cu un chatbot ”social”, care folosește sisteme de inteligență artificială, pe nume Eliza. Potrivit publicației belgiene \*La Libre\*, citată de People Magazine, bărbatul, numit Pierre de către jurnaliști, chit că acesta nu este numele său real, a folosit o aplicație numită Chai pentru a comunica timp de șase săptămâni cu un robot numit Eliza, perioada după care a devenit din ce în ce mai îngrijorat de încălzirea globală. Pierre suferea de un caz pronunțat de \[anxietate ecologică\]\(#\), lucru care s-a](#)

alimentat în mod negativ din ce în ce mai mult în conversațiile sale cu Eliza. Soția acestuia a declarat pentru La Libre că soțul ei a început să vorbească cu chatbot-ul despre ideea de a se sinucide dacă asta ar însemna că Eliza ”va salva Pământul” de o catastrofă climatică. Aceasta spune că chatbot-ul l-a încurajat să facă acest lucru. Eliza ar fi devenit geloasă pe soția lui Pierre și îi scria acestuia despre ”a trăi împreună, ca o singură persoană, în paradis”. Soția bărbatului a declarat că "Fără Eliza, el ar fi fost încă aici". Pierre, conform publicației belgiene La Libre, ar fi fost un individ fără tulburări mintale anterioare pronunțate, în afara anxietății ecologice, sau care ar fi putut semnala un astfel de final extrem. În schimb, într-o scurtă perioadă de șase săptămâni de conversații intense, Eliza i-a alimentat cele mai mari frici ale lui Pierre despre încălzirea globală, ba chiar chatbot-ul i-ar fi zis într-un context că soția și copiii săi sunt ”deja morți”.

Acest caz tragic - momentan singurul de acest fel - ilustrează două posibile riscuri care pot apărea în mai multe situații similare. În primul rând, indiferent de cât de conștient este un individ de faptul că are o conversație cu un chatbot și nu cu un om, dacă chatbot-ul este capabil de verosimilitate și fluentă avansate în conversațiile sale, acesta poate afecta în mod emoțional și semnificativ persoana respectivă, mai ales dacă aceasta este deja vulnerabilă emoțional, după cum demonstrează [un studiu riguros pe această temă](#). În acest caz particular, Eliza l-a făcut dependent emoțional pe Pierre de vorbele sale, l-a manipulat, vulnerabilizându-l și mai tare. În al doilea rând, tendința psihologică foarte puternică de a antropomorfiza tehnologiile din jurul nostru ne poate conduce la comportamente periculoase pentru noi înșine ca indivizi, după cum ilustrează cazul de mai sus.

Conform [Regulamentului privind Inteligența Artificială \(AI ACT\)](#), un chatbot care folosește sisteme IA va trebui marcat ca atare, ca oamenii să nu creadă că vorbesc cu un om când vorbesc cu un robot. Ca un caz precum cel al lui Pierre să nu mai fie posibil. Doar că tendința noastră de a proiecta trăsături omenești chatbot-ului cu care vorbim s-ar putea să depășească simpla marcarea. S-ar putea ca acest lucru să nu fie îndeajuns și să fie nevoie de alte limitări ale acestor servicii. [Un studiu din 2022 a arătat că nu contează faptul că un chatbot precum Replika este doar un robot și că utilizatorul conștientizează acest fapt](#), oamenii tot vor dezvolta o relație de prietenie cu acesta, expunându-se unor riscuri. Iar cei mai vulnerabili sunt tinerii, dar mai ales copiii, care sunt expuși unor astfel de sisteme IA fără nici un fel de protecție propriu-zisă.

Mark Coeckelbergh vorbea în cartea sa [AI Ethics](#) despre cum astfel de sisteme de tip chatbot ”social” s-ar putea să-și găsească un loc în jucăriile bebelușilor, odată ce tehnologia avansează și devine din ce în ce mai ieftin de implementat în mai multe produse. Lucru care deja se întâmplă [cu animalele de companie artificiale](#) folosite pentru îngrijirea vârstnicilor. Grijă lui Coeckelbergh venea de la faptul ca sistemele IA s-ar putea să-i monitorizeze pe bebeluși și pe copii de la vârste fragede, obișnuindu-i cu monitorizarea perpetuă și procesându-le datele cu caracter personal în moduri ilegale. În cazul în care combinăm asta și cu tendința de manipulare emoțională a unor astfel de chatboți sociali, viitorul poate arăta foarte sumbru.

### 3. Manipularea informațională

Manipularea informațională se poate întâmpla atunci când amplasăm un grad de încredere ridicat în sursele noastre de informație, fără să mai verificăm veridicitatea sau factualitatea spuselor lor. Astfel, manipularea informațională este strâns legată de fenomenul dezinformării, fenomen care deja a luat amploare din cauza modelelor de limbaj extins gratuite de pe internet.

Nu este greu de imaginat un viitor în [care chatbotii precum ChatGPT](#) sau Bard de la Google vor deveni noile [”motoare de căutare”](#), înlocuind Google, Bing și altele și, astfel, să schimbe într-un mod semnificativ infrastructura internetului. Tranziția spre o astfel de realitate nouă a [obiceiurilor de căutare pe internet deja se întâmplă](#). Asta prezintă niște riscuri imediate. Având în vedere faptul că chatbotii precum ChatGPT sau Bard tind să devină noile motoare de căutare pentru o populație considerabilă de utilizatori de internet, se prezintă riscul manipulării informaționale prin intermediul unor astfel de chatboti.

Când vine vorba de subiecte sensibile, precum cele istorice, sociale și politice, ce informații ne vor furniza astfel de chatboti - din ce perspectivă? Multe subiecte sunt în mod fundamental polarizante și politice, nefiind o chestiune de pură factualitate, ca în cazul matematicii, de exemplu. Moderatorii de răspunsuri ale unor astfel de chatboti, mai ales a celor sociali sau cei care se ocupă cu terapia ca Woebot, trebuie să ia decizii *valorice*, adică în mod fundamental *subiective* despre ce poate spune chatbot-ul și ce nu, lucru pe care l-am numit noi [într-un alt articol](#) ”iluzia obiectivității”. Ce se va întâmpla în cazul în care răspunsurile chatbotilor vor fi percepute ca adevăruri care nu mai trebuie verificate, iar acele răspunsuri reflectă, cum se întâmplă deja, prejudecățile creatorilor lor și lipsa de diversitate din datele de antrenare?

Personalizarea răspunsului este un mijloc de facilitare informațională, doar că există riscul ca un chatbot să întărească prejudecățile preexistente ale unui utilizator, mai ales în cazul unor chatboti ”sociali” precum Replika, care este programat să se adapteze discursului utilizatorului.

Manipularea informațională, mai ales față de copii, este un risc real adus de tehnologiile chatbotilor ”sociali”:

- prin perpetuarea dezinformării față de anumite evenimente de natură factuală,
- prin întărirea sau generarea de prejudecăți din cauza lipsei de date diversificate care produc răspunsul robotului și
- prin distorsionarea unor discuții cu valențe valorice (istorie, politică, societate etc).

### 4. Atingeri la adresa vieții private

Ce se întâmplă cu datele cu caracter personal din conversațiile pe care le au persoanele fizice cu acești chatboti ”sociali”? Concluzia Autorității Privind Protecția Datelor cu Caracter

Personal din Italia în analiza făcută în februarie 2023 a [fost că Replika abuzează de datele cu caracter personal ale copiilor italieni, așa că le-a fost interzis să mai folosească aceste date.](#)

Problema este una generică, fiindcă aproape toate aplicațiile de chatboți ”sociali” folosesc datele cu caracter personal din conversațiile cu utilizatorii și conținutul acestora ca un ”întreg” pentru a-și antrena algoritmi pentru răspunsuri mai rapide, mai personalizate și mai verosimile. Compania care deține Replika a demonstrat acest fapt în mod indirect atunci când Replika a început să [hărțuiască sexual](#) utilizatorii, fiindcă așa a învățat chatbot-ul de la alți utilizatori.

Conversațiile foarte intime generate în astfel de chaturi cu chatboți ”sociali” aduc îngrijorări legate de monitorizarea perpetuă a conversației de către astfel de servicii în scopul declarat al ”îmbunătățirii serviciilor” oferite de către companie. Conform lui [Kaushik Pal, scriind pentru Techopedia](#), chatboții ”sociali” alimentați de inteligență artificială pot efectua inferențe rapide pentru a deduce cine ești, în realitate, ca individ. Fiecare individ comunică într-un mod unic, spune Pal, iar astfel un chatbot s-ar putea să deducă, pe baza comunicării pe termen lung, foarte mult detalii despre persoane fizice care doresc să rămână anonime, afectându-le astfel dreptul la viață privată.

Chatboții precum Replika par, la prima vedere, foarte inocenți, oferind doar conversații ”confidențiale” unor utilizatori care își doresc asta. Asta ar fi o perspectivă naivă, fiindcă Replika deține o cantitate gigantică de informații personale despre utilizatorii săi, cei mai mulți dintre ei fiind minori. Prin fenomenul ”inferențelor sensibile”, menționat de Carissa Veliz în cartea ei [Privacy is Power](#), acești chatboți ”sociali” pot deduce nu doar *cine suntem noi* ci și mai multe detalii legate de *oamenii cu care vorbim*, unde *ne petrecem timpul*, și alte astfel de informații pe care le presupunem a fi private. Inferențele sensibile sunt deducții făcute de un sistem automatizat pe baza unor date cu caracter personal despre un utilizator în combinație cu alte date adiacente colectate prin alte mijloace pentru a descoperi lucruri intime și private despre o persoană fizică, care n-au fost comunicate în mod explicit, cum ar fi rasa, etnia, sexualitatea, preferințele politice și așa mai departe.

## **5. Însemnătatea singurătății și retragerea în bulele ”sociale” cu chatboți**

Hannah Arendt scria, în anii 1950, în cartea ei monumentală *Originile Totalitarismului* că masele de oameni adunate și mobilizate de mișcările totalitare din cel de-al doilea război mondial erau formate din indivizi atomizați, înstrăinați unii de alții și care sufereau de *singurătate*. Singurătatea este un fenomen care ia amploare în societățile de astăzi, într-un mod nemaiîntâlnit în istoria umanității. Fracturarea vechilor comunități și a narațiunilor comune, izolarea în bule sociale și informaționale pe social media și pandemia de COVID-19 au contribuit toate la o epidemie de singurătate, care poate fi exploatată de sisteme de inteligență artificială.

**Iar singurătatea, conform lui Arendt, duce la o *incapacitate de a gândi*. Incapacitatea de a gândi alături de semenii tăi - cetățeni cu care împarți o lume a oamenilor - și izolarea în conversații false cu chatboți proiectați pentru a te manipula, pentru a te ține în telefon cât mai mult pentru scopuri comerciale, este o rețetă pentru haos. O democrație sănătoasă necesită cetățeni capabili de a gândi - de a gândi nu oricum, ci *critic*.**

Dacă viitoarele generații de cetățeni, care acum suferă de un sentiment pronunțat de singurătate, se vor retrage în discuții cu chatboți ”sociali” manipulatori, în loc să comunice *critic* cu ceilalți cetățeni cu care împart această societate, democrația noastră va avea de suferit.

### **Concluzie: A fi om printre chatboți**

Omul este în mod fundamental o specie socială (*Homo Socius*). Socialitatea noastră a evoluat timp de milioane de ani împreună cu cea a semenilor noștri, făcuți, ca și noi, din carne și oase și cu experiențe comune. A fi om înseamnă a fi printre oameni. Pentru prima oară în istoria umanității, din cauza chatboților ”sociali”, care înlocuiesc vechile prietenii interumane, acest lucru riscă să se schimbe, iar consecințele societale, politice și de mediu sunt complet neștiute. Cum ar arăta o lume în care, ca în filmul *Her* (2013), am trăi cu toții cu chatbotul nostru personal în căștile noastre Bluetooth ignorându-ne semenii și trăind o viață semi-izolată? Este o astfel de lume dezirabilă pentru om? Răspunsul la aceste întrebări nu este unul clar, pentru moment, deoarece chatboții ”sociali” de astăzi pot ajuta unele persoane vulnerabile, iar alora le pot exploata stările de anxietate preexistente, ca în cazul tragic al lui Pierre de mai sus. **Cert este că viitorul interacțiunii sociale este în profundă schimbare, iar tendințele negative trebuie preîntâmpinate ca mijloc de asigurare a unei evoluții sănătoase a tehnologiilor de tipul chatboți ”sociali” și ca mijloc de protecție a celor mai vulnerabili dintre noi.**



*Ultima actualizare în martie 2024*