CRISTIAN RADU

NECESITATEA IMPLEMENTĂRII UNUI SISTEM DE MORALITATE INTELIGENȚELOR ARTIFICIALE

Cristian Radu
Babeș Bolyai University, Department of Philosophy, Faculty of History and Philosophy, Cluj-Napoca, Romania
Email: cristiradu1991@gmail.com

Rezumat: În acest articol, susțin că este necesară implementarea unui sistem moral pentru platformele de inteligență artificială. Voi aduce argumente pentru această cerință, analizând exemple de inteligențe artificiale deja utilizate în domeniul războiului armat, al asistenței medicale și al vieții cotidiene. Voi vorbi și despre exemple de sisteme suficient de avansate și capabile să fie implementate cu un sistem moral sau etic. În final, voi vorbi despre despre accidente ce au implicat inteligențe artificiale în viața reală, accidente care ar fi putut fi prevenite prin utilizarea unui astfel de sistem.

Cuvinte-cheie: inteligență artificială, morală, etică, armată, asistență medicală, viață, siguranță, informații, tehnologie, utilitate

THE NECESSITY OF IMPLEMENTING A MORAL SYSTEM TO ARTIFICIAL INTELLIGENCE

Abstract: In this article, I argue that there is a necessity of implementing a moral system for artificial intelligence platforms. I will bring arguments for this requirement by analyzing examples of artificial intelligence already in use in the field of armed warfare, healthcare and day to day life. I will also talk about examples of systems that are advanced enough and capable of being implemented with a moral or ethical system. In the end, I will talk about real-life a.i. accidents that could have been prevented by having in use such a system.

Keywords: artificial intelligence, morality, ethics, military, healthcare, life, safety, information, technology, utility

Revista de Filosofie Aplicată, vol. 3, issue 4 (Spring 2020): 83-103.
ISSN 2668-0009; ISSN-L 2668-0009

1. Introducere

Inteligența artificială (de aici înainte i.a. sau a.i., dacă e vorba de echivalentul termenului englez *artificial intelligence*) este una din cele mai importante realizări technologice ale secolului trecut. A fost concepută inițial de Alan Turing în 1935 și implementată cu succes pentru prima oară în 1951.[1] De atunci, a revoluționat dezvoltarea tehnologiei și a fost implementată în aproape toate domeniile activității umane. Astfel, în prezent, se poate discuta despre trei tipuri de inteligență artificială: „weak a.i.", „strong a.i." și „super a.i.". „Weak a.i." sunt acele inteligențe artificiale utilizate în domenii specifice: traduceri lingvistice, asistenți virtuali, motoare de căutare web, etc.[2] Un „strong a.i." este o inteligență artificială capabilă să interacționeze și să operaționalizeze diferite sarcini care nu sunt legate între ele. Poate învăța din propria experință să dezvolte noi strategii pentru noi sarcini. Nu există în prezent o formă de inteligență artificială de acest fel.[3] În final, „super a.i." există doar la nivel de concept și implică o inteligență artificială care este conștientă de sine precum un om.[4]

Odată cu dezvoltarea inteligenței artificiale, au apărut aceleași dileme caracteristice oricărei tehnologii noi descoperite: Ce înseamnă pentru om această invenție? Cum ne poate ajuta? Care sunt limitele ei? Ce înseamnă pentru viitor? Care este relația între o inteligență artificială și una umană? În prezent se pune problema implementării unor principii etice pentru o utilizare mai bună și mai sigură a acestei tehnologii. Direcționată de aceste principii, o inteligență artificială poate lua decizii corecte atât din punct de vedere tehnic cât și etic.

În acest articol urmăresc să argumentez necesitatea introducerii unui sistem de moralitate inteligențelor artificiale. Voi argumenta această necesitate în cadrul diferitelor domenii în care s-a implementat deja sistemul i.a. Voi aduce exemple de i.a. care au cel mai mare potențial de a fi implementate cu principii morale. Iar, în final, voi aduce exemple de situații critice unde un sistem moral ar fi dus la evitarea unor accidente. Consider necesară această implementare

pentru că tehnologia i.a. ne înconjoară și face parte din viața noastră de zi cu zi. Incidentele recente pot servi ca dovezi pentru o rafinare atât tehnologică cât și etică ale acestui sistem.

2. Folosirea i.a. în domeniul militar

Utilizarea inteligențelor artificiale în cadrul conflictelor armate este o realitate curentă care se dezvoltă în mod rapid. Marile puteri ale lumii investesc în cercetări în acest domeniu în expansiune. Printre exemple s-ar număra dronele ghidate și armele de foc controlate de i.a. Direcția în care se îndreaptă dezvoltarea fac necesară o "calibrare" morală.

Cea mai importantă dar controversată inovaţie pe care ar aduce introducerea unui sistem de moralitate este faptul că i.a.-ul ar putea înlocui omul pe câmpul de luptă. Astfel, prin asistarea sau înlocuirea omului, o i.a. îmbunătățită poate reduce numărul de vieți pierdute în perioadele de conflict militar pentru că poate nu numai să analizeze, ci și să acționeze rapid și eficient în situații dificile. Acest „avantaj" în sine este un motiv fundamental pentru utilizarea unui sistem de moralitate, este ceea ce ar numi filosoful american, George R. Lucas, *Principiul riscului inutil*. Potrivit acestui concept, experții și specialiștii, din diferite domenii (profesori, oameni de știință, filosofi, etc) au datoria morală de a-și proteja forțele militare sau de securitate, prin reducerea sau eliminarea riscurilor care pot apărea în cadrul activităților acestora.[5] Dar, această posibilitate ridică și multe obiecții din partea altor experți. De exemplu, factorul uman nu poate fi înlocuit. De la începutul civilizației, orice conflict militar a implicat, pe lângă o cooperare între membrii unui grup, un grad de responsabilitate etică. O astfel de responsabilitate etică, în perspectiva lui Gérard de Boisboissel, este imposibilă de transpus într-un sistem coordonat de inteligență artificială.[6] Acest obstacol, deși semnificativ, se poate rezolva printr-o analiză a lanțului cauzal care pornește de la echipamentele utilizate și ajunge până la producători. În acest mod, se poate stabili cine este responsabil într-o situație critică.

I.a. reprezintă, prin sistemele lor complexe, un avantaj în ceea ce privește numărul scăzut de victime colaterale. Față de victimele colaterale cauzate de erori umane, în toate cazurile în care a fost vorba de drone sau echipamente controlate de i.a., numărul de victime colaterale a scăzut semnificativ. Acest lucru are loc pentru că echipamentele funcționează mai eficient decât oamenii, își aleg mai bine momentul de atac și identifică mai precis ținta.[7] Un sistem moral ar îmbunătăți cu mult tehnologia, făcând trecerea de la un sistem de țintire letal la un sistem de identificare etic. Acest sistem îmbunătățit, nu numai că limitează numărul de victime colaterale, dar și poate înlocui factorul uman (astfel respectă principiul riscului inutil), elimină erorile ce apar în contexte critice și chiar poate asista forțele armate pentru a putea lua decizii etice în timpul misiunilor.

Avantajul enorm pentru implementarea unei autonomii și a unei moralități tehnice, este faptul că o i.a. nu are gândirea înceţoşată de momente temporale de şovăială. Mai exact, anxietatea, emoțiile, frica nu afectează un sistem de i.a. în procesarea datelor. Potrivit lui Ronald C. Arkin, o i.a. care folosțe un trup robotic nu are vreun instinct de autoconservare, poate folosi senzori care întrec organele de simț umane, nu suferă de fustrare, isterie sau șoc postraumatic. Un a.i. nu are subiectivitate umană prin care să-și creeze scenarii în minte care ar periclita misiunea. Nu poate acorda importanţă de natură sentimentală unui element anterior în realizarea unei alegeri noi. Mai ales în momente critice, o i.a, prin senzorii săi, poate aduna mai rapid informațiile necesare pentru a putea lua o decizie într-un timp foarte scurt. Și, în final, poate chiar asigura buna respectare a principiilor etice, când lucrează „cot la cot" cu oameni, fiind capabilă să prevină chiar și crime de război.[8] În concluzie, se poate observa imaginea conturată a unei entități care accentuează capacitățile umane și elimină slăbiciunile. Prin aceste avantaje, un a.i. are potențialul de a lua o decizie etică mai rapid, mai eficient și cel mai important, mai precis decât un om.

Însă nu toată lumea este de acord cu aceste direcții de dezvoltare. Emoțiile pot avea un efect pozitiv în procesul moral. Moralitatea în sine nu se poate descompune într-o funcționalitate tehnică, avertizează Dominique Vermersch. La fel în cazul definiției războiului,

este imposibil de înglobat într-un modul sau într-o funcție toate variabilele și elementele ce compun ideea de război sau de strategie militară.[9]

O soluție este aplicarea unui sistem moral la o i.a. disponibilă în prezent. Un „weak a.i." reprezintă un candidat bun pentru că reprezintă un mediu în care un sistem moral care poate fi implementat într-un mod sigur. Fiind un sistem simplu, destinată unui singur scop, ea nu prezintă aceleași riscurile pe care un „strong a.i." le poate avea. Aplicarea unei autonomii și a unei moralităţi unui sistem i.a. avansat sau puternic este foarte periculoasă. Acesta ar putea avea o conștiință asemănătoare cu a unui om, ar putea gândi ca un om și ar putea lua decizii ca unul. Astfel, o implementare la un astfel de nivel ar duce la rezultate neașteptate sau periculoase, mai ales că nu știm exact cum funcționează conștiința umană.

Propunerea lui George Lucas este utilizarea i.a. slabe într-un mediu izolat care permite implementarea optimă a misiunilor și a directivelor în mecanismul i.a. Mai exact, utilizarea unor roboți cu i.a. în zone demilitarizate sau de graniță unde trecerea este interzisă. Sau chiar utilizați ca escorte în misiunile militare de explorare.[10] În esență, utilizarea lor în situații și locuri mai puțin complexe pentru care nivelul curent de i.a. poate funcționa optim pentru implementarea viitoare a unui sistem de moralitate. Și acest sistem, după ce este implementat, trebuie să fie într-un context restrâns pentru a putea fi introdus optim.

3. Utilizarea i.a. în medicină

Prima instanță în care putem spune că inteligența artificială, augmentată de principii etice, ar aduce contribuții semnificative în domeniul medical, este în instruirea medicilor în pregătire. După cum spun Steven A. Wartman și C. Donald Combs, prin introducerea i.a, studenții interacționează cu pacienți virtuali sau modele anatomice virtuale pentru practica medicală.[11] Pacienții virtuali sunt programe virtuale care simulează situații clinice reale, prin care viitorii medici pot dobândi, într-un mod practic, aptitudinile necesare pentru a putea

lucra cu pacienți reali.[12] Un pacient virtual cu inteligență artificială este un program avansat care poate interacționa verbal și nonverbal, poate purta o conversație, reacționează la limbajul de trup și poate chiar simula emoții.[13] Aceste sisteme, îmbunătățite prin introducerea unui sistem etic, ar duce la o instruire mult mai completă și eficientă pentru că pacienții virtuali ar fi mult apropiați de pacienții reali. Mai concret, acestea nu s-ar rezuma doar la simularea unor situații clinice, ci l-ar ajuta pe viitorul medic să învețe cum să ia decizii corecte atât din punct de vedere medical cât și dintr-un punct de vedere etic. Acest avantaj, se aplică mai ales în cazul utilizării unor pacienți virtuali pentru instruirea psihologilor în devenire. În concluzie, acest progres reprezintă trecerea de la o instruire bazată pe informații la o instruire care accentuează latura practică.

Putem spune că posibilitatea utilizării inteligențelor artificiale, în ceea ce privește administrarea datelor pacienților, reprezintă încă un motiv pentru introducerea unui sistem etic. Prin viteza de procesare caracteristică calculatoarelor, o i.a. poate parcurge într-un timp foarte scurt toate datele electronice ale pacientului. Prin puterea de procesare, o i.a. poate trece dincolo de datele pacientului. Ea poate corela datele pacientului cu istoricul medical al familiei sale (dacă acest istoric este disponibil) și chiar poate compara datele medicale ale pacientului cu cele ale altor pacienți, pentru diagnoză și tratament. Toate acestea sunt posibile prin accesul inteligenței artificiale la diferite baze de date medicale.[14]

Putem susține importanța i.a. chiar și în procesul de diagnostificare a pacienților. Prin calitățile inerente lor, au potențialul de fi mai eficiente ca oamenii în anumite privințe. O analiză a sistemelor de i.a, bazate pe modelul rețelelor neurale artificiale, a arătat faptul că acestea au un procentaj ridicat de precizie în diagnostificarea diferitelor boli. În cazul bolilor cardiovasculare s-a observat o precizie de 91.2%, iar în cazul cancerului 96%. Doar în cazul diabetului, a fost o prezicie mai mică de 63-79%.[15]

Aceste avantaje necesită implementarea unui sistem etic pentru că i.a. lucrează, în primul rând, cu datele personale ale pacienților și adună informații enorme într-un timp foarte scurt prin puterea de

procesare și prin accesul la diferite baze de date. Astfel, o formă rudimentară de restricție a i.a. nu poate fi destulă pe termen lung. Inteligențele artificiale, menționează Kadircan Keskinbora, funcționează în momentul de față pe baza sistemului *machine learning* ce necesită un flux continuu de informație pentru a funcționa. Această structură reprezintă un pericol pentru dreptul omului la intimitate, făcând imperativă utilizarea unui sistem moral pentru protejarea pacienților.[16] Prin acest sistem moral, i.a, deși accesează date personale, ea le poate procesa etic. Intervențiile ei nu violează mai departe intimitatea pacientului, se rezumă doar la datele necesare pentru tratement și îl asistă pe medic pentru ca deciziile luate de acesta să fie și etice.

În cazul asistenților non-umani, moralitatea ar îmbunătăți semnificativ relația dintre om și i.a. Vorbim aici despre acei asistenți robotici care sunt utilizați pentru a avea grijă de pacienți cu nevoi speciale. Numărul asistenților robotici a crescut considerabil în ultimii ani și sunt de mare ajutor pentru asistarea bătrânilor, copiilor autiști și pentru terapie fizică sau chiar psihologică. Uneori sunt folosiți ca animale (sau roboți) de companie. Deoarece pacienții nu văd acești asistenți doar ca niște obiecte, ci și ca niște ființe vii în care au încredere, putem observa anumite pericole pentru care un sistem moral ar fi o soluție. Mai concret, acești asistenți nu trebuie să pună în pericol intimitatea pacientului, autonomia acestuia, și să nu înlocuiască factorul uman din viața de zi cu zi a pacientului. Printr-un sistem moral implementat asistentului robotic, toate aceste riscuri sunt eliminate pentru că asistența acestuia nu se rezumă doar la tratamentul pacientului, ci și la bunăstarea acestuia pe termen lung într-un mod etic[17].

4. Folosirea i.a. în viața de zi cu zi

Un alt domeniu unde este posibilă și chiar necesară implementarea unui sistem de moralitate inteligențelor artificiale, este transportul public. Există deja mașini fără șofer care sunt conduse de o inteligență artificială. Această tehnologie este testată de mai multe

companii cum ar fi Google, Nissan Volvo și Mercedes. Aceste mașini se ghidează prin mediu cu ajutorul unor senzori, detectoare și proiecții 3D ale mediului cu o conexiune constantă la internet. O inteligență artificială prezintă mai multe avantaje față de un șofer uman, între care amintim: lipsa oboselii, a stresului și a fustrărilor. Procesarea rapidă a informațiilor reprezintă un avantaj enorm când e vorba de momente de criză unde este nevoie de a lua rapid o decizie.

Există, totuși, câteva dificultăți în implementarea totală a unei i.a și a unui sistem etic atașat acesteia. În cazul unei situații critice limită unde sunt posibile victime, nu putem prezice exact ce alegere va face o i.a. În cazul unui om, reacțiile sunt bazate pe reflexe și pe timpul în care realizează ce se întâmplă, ceea ce poate duce la o decizie greșită. Dar, în cazul i.a. există o viteză de procesare rapidă, senzori care preiau datele prin care ea poate decide. Astfel, va exista riscul unei alegeri care se potrivește inteligenței artificiale, dar nu și omului.[18] Problema etică, care apare aici, este că nu știm cum va reacționa o i.a. în situațiile în care nu se poate evita vreo victimă și ce fel de decizii va lua în acest caz. Un contraargument este că în astfel de situații o i.a. poate găsi o soluție la care un om nu s-a mai gândit. În prezent, deși nu există încă „strong a.i", există i.a. de tip slab care au încorporate posibilitatea acestui scenariu. Mai exact, este vorba de i.a. dezvoltată de compania „DeepMind" numit „AlphaGo". Este o i.a. de tip „weak a.i." programată pentru un singur joc, Go. Într-o serie de meciuri împotriva unui jucător uman a realizat o mișcare pe care nimeni nu a mai făcut-o până în momentul respectiv și a câștigat meciul.[19] Asta este o dovadă pentru faptul că, fiind vorba de un tip diferit de cogniție, o i.a. poate să gândească altfel decât un om, însă cu toate acestea să facă o alegere bună în contextul în care operează.

Un argument în plus pentru implementarea unui sistem de moralitate este faptul că, oricât de avansată poate fi o i.a, prelucrarea datelor după o formă rigidă prin programare nu este suficientă pentru a putea lua decizii corecte și dintr-un punct de vedere etic. Mai exact, luarea deciziilor pe baza unor reguli programate sau setate nu poate acoperi tot ceea ce poate întâlni o i.a. pe teren. Oricât de multe reguli și algoritmi s-ar putea programa, tot timpul există riscul apariției unui scenariu în care anumiți parametrii sau reguli pot să ducă la luarea

unor decizii necorespunzătoare. Patrick Lin, în capitolul lui despre mașinile conduse de i.a, descrie cum un set de reguli sau algoritmi pot deveni un sistem periculos de țintire. Acest scenariu periculos apare în cazul accidentelor rutiere, unde o i.a. trebuie să ia o decizie pentru a salva cât mai multe vieți. De exemplu, o i.a. este programată să pună preț pe viața pasagerilor și este programată ca, în caz de accident, să lovească un obiect mai puțin solid. Astfel, indiferent de obiectele care sunt în apropiere, în cazul unei situații critice, o i.a. va lovi ținta care se încadrează în regulile programate, indiferent de cine sau ce este. Chiar dacă se poate implementa un sistem mai eficient, de exemplu să scaneze care obiect din apropiere este mai puțin dăunător în caz de ciocnire, tot va rămâne, în esență, un sistem de țintire și nu un mecanism etic.[20] Soluția, în acest, caz ar fi implementarea unui sistem moral al cărui algoritm ar avea la bază protejarea vieților umane, iar orice decizie luată să aibă în prim plan salvarea vieților umane. Astfel, avantajele acestei tehnologii, respectiv percepția și puterea de procesare avansată, împreună cu un sistem etic, ar duce la crearea unei i.a. capabile să ia decizii eficiente în situații critice, care să evite erorile etice.

Deoarece în prezent i.a. este implementată tot mai mult în locurile de muncă, este necesar un sistem moral pentru acestea. Ceea ce face necesară această implementare este relația dintre muncitor și i.a. În prezent, una din cele mai semnificative probleme etice ale acestei relații este cea a înlocuirii mâinii de lucru umane cu i.a/roboți. Soluția cea mai viabilă nu este înlocuirea omului, ci asistarea acestuia cu o i.a. îmbunătățită cu moralitate. Un exemplu bun în acest caz reprezintă locurile de muncă periculoase care necesită efort fizic cum ar fi exploatarea forestieră, pescuitul, aviația și construcțiile. Toate aceste locuri de muncă sunt indispensabile și un sistem de i.a. ar aduce mai multe beneficii dacă ar funcționa împreună cu omul, decât fără el. Astfel, prin tehnologia i.a, scade riscul unor situații periculoase în cadrul acestor locuri de muncă. O i.a. poate programa, alături de om, utilajele folosite, poate asista omul cu soluții atât tehnice cât și etice în funcție de context. Ea poate asigura buna respectare a normelor etice în cadrul locului de muncă. Acest avantaj este semnificativ în cazul urgențelor de genul incediilor, cutremurelor, atacuri teroriste.[21]

De exemplu, în cazul incediului din California din 2018, dronele au fost folosite pentru a căuta victime și pentru a aduna informații legate de zonele atinse de flăcări unde, în mod normal, pompierii și medicii nu ar putea fi avut acces. Iar în 2016, un studiu realizat de compania producătoare de drone "DIJ" a dovedit cum dronele pot asista, într-un mod eficient, echipele de salvare să ajungă la victime în cazul unor situații critice, prin intermediul senzorilor avansați și prin puterea de procesare a informației.[22] Tot legat de incendiile recente, în 2019, compania Descartes Labs din Statele unite a dezvoltat un program de detectare a incendiilor bazat pe tehnologia i.a. Aceasta utilizează doi sateliți, *GOES-16* și *GOES-17* pentru a putea observa orice incediu care ar putea apărea în zonele aflate sub raza lor de acoperire. Algoritmul este programat să filtreze alarmele false precum focurile controlate (industriale, agricole, artificii) și este capabil să ofere informații importante despre locația unui incediu, despre sursa acesteia și despre geografia locului unde are loc incendiul.[23] Prin implementarea unui sistem moral, aceste sisteme, prin accesul lor la baze de date și prin tehnologia lor avansată de percepție, ar excela nu numai în rolul lor curent, ci ar avea și un accent pus pe salvarea vieților umane. Pe lângă procesarea și adunarea rapidă a informațiilor, ele ar avea potențialul de a coordona echipele de salvare, de a asigura aplicarea principiilor etice în momente critice și de a lua decizii care să pună accentul pe salvarea vieților omenești.

Pe lângă transport și locul de muncă, sistemele a.i. sunt deja implementate în cazul rețelelor de socializare. Facebook folosește i.a pentru toate funcțiile paginii, de la conținut la analiza fotografiilor și mai ales pentru preferințe. Instagram folosește i.a. pentru a identifica imagini. LinkedIn folosește i.a. pentru a oferi recomandări de locuri de muncă și sugestii de prieteni bazate pe compatibilitate[24]. În astfel de situații, o i.a. ar beneficia de un sistem moral nu numai pentru a detecta violări ale principilor etice, ci și pentru a îmbunătăți sistemele de protecție utilizate în prezent.

5. Exemple de inteligențe artificiale cu posibilități de moralizare

La ora actuală, încă nu s-a implementat un sistem moral perfect unui sistem artificial, dar în anumite cazuri, moralitatea ar fi pasul următor. Exemplele care vor fi discutate aici sunt inteligențe artificiale performante care se apropie cel mai mult de un sistem moral și care au făcut deja primii pași în această privință.

Cele mai semnificative exemple din domeniul militar sunt i.a. care au rolul de a identifica și a localiza ținte. Multe i.a. de acest fel, în prezent, folosesc sisteme complexe de procesare a datelor care prezintă un prim pas pentru o viitoare implementare a unui sistem moral. Un exemplu în acest sens este *Project Maven.* Este un proiect al Pentagonului care implică utilizarea procesului de *machine learning* pentru a distinge oameni și obiecte în capturile video ale dronelor, oferind guvernului comandă și control timp real al câmpului de luptă, precum și capacitatea de a urmări, eticheta și de a spiona țintele fără intervenție umană directă.[25] În 2010, Compania DoDAAM din Coreea de Sud a conceput un sistem automat de foc, *Super aEgis II*, care poate detecta mașini și oameni până la 3 km distanță. Are un sistem de avertizare verbal și poate să fie setat să tragă automat sau să fie folosit prin comenzi vocale.[26] Rusia, deasemenea, a testat mai multe sisteme de luptă autonome și semi-autonome, cum ar fi modulul bazat pe „rețeaua neurală Kalashnikov" formată dintr-o mitralieră, o cameră de fotografiat și o i.a, care poate lua decizii fără intervenție umană.[27] Utilizarea acestor sisteme ridică multe întrebări etice. Cea mai semnificativă este cea a responsabilității. Dacă sistemul decide ținta, atunci cine este responsabil în cazul unor accidente sau a unor victime colaterale? În prezent, tehnologia încă nu este destul de avansată încât să ia decizii fără vreo intervenție umană. Apoi, următoarea întrebare ar fi în cazul unor misiuni unde victimele colaterale sunt imposibil de evitat. Va renunța la misiune sau va calcula un plan optim? Va chema echipe de salvare pentru posibile pagube?

În final, trebuie să luăm în considerare cum afectează omul această tehnologie. Multe sisteme de ţintire implică un control uman de la distanţă. Este vorba de o altă situaţie faţă de soldatul de pe front în mijlocul conflictului. Trebuie evitate situaţiile în care operatorul acţionează în afara misiunii din cauza unei emoţii sau a unei decizii grăbite, devenind mai neglijent pentru că nu este acolo fizic pentru a fi în pericol.

Introducerea unui sistem moral ar fi următorul pas etic pentru că bazele care se pot observa aici, vitezele rapide de procesare, senzori mai puternici decât organele de simţ umane, lipsa stărilor temporare precum frica şi anxietatea, pot duce aceste tehnologii pe două căi diferite. O dată, se poate implementa un sistem de moralitate pentru ca acestea să poată respecta principii etice. Altfel, rămân doar arme şi unelte periculoase care pot duce la mai multe pagube sau victime colaterale. Dar, chiar şi această implementarea poate fi dificilă. Trebuie să se decidă unde să aplice în mai mare măsură, la echipamentele ofensive sau la cele de apărare. Dacă se implementează un sistem moral pentru un echipament ofensiv, cum va şti acesta să acţioneze astfel încât să îşi servească menirea şi să fie şi etic. În cazul echipamentelor defensive, salvarea vieţilor umane are o prioritate mai mare, dar şi modul în care se realizează această protecţie poate fi discutabilă. Există diferenţe semnificative între scenariul în care i.a. va alege să elimine mai întâi pericolul, şi cel în care va alege să proteje soldaţii sau civilii şi doar după să se ocupe de eliminarea pericolului.

În domeniul medical, sistemele cu i.a folosite în prezent joacă un rol de asistent în relaţie cu medicul, mai ales în ceea ce priveşte diagnosticarea pacienţilor. Un sistem foarte avansat de acest fel este IBM Watson, folosit în diagnosticarea leucemiei. Este un sistem care foloseşte mai multe tehnici de i.a.: procesarea limbajului şi a informaţiilor, analize semantince şi *machine learning*. Mai mult, acest sistem foloseşte şi modelul *DeepQA* (Deep question and answering) prin care găseşte o soluţie într-o fracţiune din timpul normal alocat unui diagnostic.[28] Un sistem asemănător este dezvoltat de Microsoft, numit *Project Hanover*, care este utilizat pentru detectarea şi tratarea cancerului.[29]

În ceea ce privește bunăstarea pacienților, dezvoltarea roboților care au grijă de pacienți a evoluat semnificativ. Un exemplu este carebot-ul proiectat de compania GeckoSystems. Față de alți roboți de acest fel, acesta are mai multe funcții precum chat, memorare și o formă de conștientizare a mediului.[30] Pe scurt, are toate bazele pentru o implementare a unui sistem etic pe lângă toate funcțiile lui.

În ceea ce privește viața de zi cu zi, cele mai avansate sisteme cu i.a. sunt asistenții virtuali. Un exemplu este Amazon Alexa. Acest sistem a fost conceput sub forma unei boxe care are o funcție de comunicare și de executare a diferitelor comenzi..[31] Pe lângă asta, se poate conecta la echipamente smart sau la sisteme automate de control al casei. Posiblitatea implementării unui sistem de moralitate o acordă procesarea rapidă a informației, legătura cu netul și posibilitatea de modificare a programului sursă. În această categorie intră și Google Assistant, Microsoft Cortana și Apple Siri.

6. Accidente reale care ar fi putut fi prevenite printr-un sistem etic

Întâi, trebuie să definim ce este o violare a principiilor morale sau cum putem descrie acele situații în care o i.a. ar fi putut evita o situație neplăcută dacă ar fi acționat etic. Conform teoriei bazelor morale, există șase domenii prototipice morale umane: grijă/vătămare, cinste/înșelare, loialitate/trădare, autoritate/subversiune, sanctitate/degradare, libertate/constrângere. Conform acestei teorii, o violare morală este orice acțiune sau situație care pune în pericol un din cele șase baze morale.[32]

Un accident, al cărui rezultat intră în această categorie, a avut loc în 2016. Compania Youth Laboratories, o companie care colaborează cu alte companii mari din industria IT cum ar fi Nvidia și Microsoft, a organizat un concurs de frumusețe online numit „Beauty.ai". În cadrul acestui concurs, 600.000 de participanți au fost evaluați de o inteligență artificială. Algoritmul utilizat a analizat riduri, simetria feței, coșuri, rasă și vârsta. Problema a apărut la sfârșit, când i.a. a

generat rezultatele. Din cei 44 de câștigători, 36 au fost de rasă caucaziană. Algoritmul a folosit o rețea neurală artificială pentru a scana fotografii cu participanții la concurs. Diferența a survenit din cauză că sistemul nu putea face o analiză corectă a celor cu piele închisă la culoare.[33] Acest incident dovedește necesitatea unui sistem moral prin care alegerea ar fi putut fi etică și este o dovadă care arată că oricât de avansată este tehnologia i.a. și oricât de avansate sunt tehnicile ei, ea necesită un sistem moral care să depășească limitele acestor moduri de procesare a informației, pentru a putea evita aceste situații.

În această categorie, putem pune și experimentul realizat de Microsoft în 2016 pe Twitter. Compania a dezvoltat un chat bot cu i.a. cu numele de Tay, care funcționa pe baza interacțiunilor ei cu utilizatorii site-ului de socializare. În esență, folosea reacțiile și răspunsurile utilizatorilor ca o sursă de informații pentru propriile ei răspunsuri. După 16 ore, în urma unor interacțiuni cu anumiți utilizatori care i-au dat informații bazate pe teme controversate, a început să genereze texte și răspunsuri la fel de controversate care au dus la închiderea ei. Un sistem moral ar fi acționat ca filtru pentru procesarea informațiilor. Chatbot-ul a acceptat informațiile primite de la utilizatori fără să le proceseze dincolo de input. Și-a îndeplinit programarea; dar fără un sistem etic, rezultatul a fost dezastruos.

În final, aceeași situație se întâmplă constant și cu Amazon Alexa, existând multe instanțe în care comportamente nedorite au avut loc pentru că echipamentul procesează informația doar la nivel structural și nu intră într-un proces amplu de analiză pentru a înțelege exact ceea ce se întâmplă sau ce aude.[34]

7. Concluzie

După cele menționate mai sus, putem spune că implementarea unui sistem etic inteligențelor artificiale ar introduce mecanisme de decizie etică în conflicte, ar îmbunătăți sistemul de sănătate, făcându-l mai eficient, iar în viața de zi cu zi, am avea locuri de muncă îmbunătățite, cu mai puține pericole. Cu bazele acestei tehnologii care

sunt aşezate în prezent, avem potențialul de crea o tehnologie care ar putea fi atât puternică cât şi morală.

Note:

[1] B.J. Copeland, "Artificial Intelligence", *Encyclopedia Britannica*, https://www.britannica.com/technology/artificial-intelligence, Mar 5, 2020, accesat în 31.01.2020.
[2] ***, "Weak Artificial Intelligence", *Techopedia*, July 22, 2017 https://www.techopedia.com/definition/31621/weak-artificial-intelligence-weak-ai, accesat în 03.03.2020.
[3] Jake Frankenfield, "Strong a.i", *Investopia*, October 2, 2019, https://www.investopedia.com/terms/s/strong-ai.asp, accesat în 03.03.2020.
[4] Tannya Jajal, "Distinguishing between Narrow AI, General AI and Super AI", *Medium*, May 21, 2018, https://medium.com/@tjajal/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22, accesat în 03.03.2020.
[5] George Lucas, "The Ethical Challenges of Unmanned Systems", în Ronan Doaré, Didier Danet, Jean-Paul Hanon, and Gérard de Boisboissel (eds.) *Robots on the Battlefield. Contemporary Perspectives and Implications for the Future* (Fort Leavenworth, Kansas: Combat Studies Institute Press, 2014), 135.
[6] Alfred De Vigny, *Servitude et grandeur militaires* (Conard, 1914) https://www.bouquineux.com/?ebook s=82&Vigny, accesat în 30.01.2020.
[7] George Lucas, "The Ethical Challenges of Unmanned Systems", 136.
[8] Ronald Arkin, "Ethical Robots in Warfare", în *IEEE Technology and Society Magazine*, Vol. 28, Nr. 1 (2009): 31-32.
[9] Dominique Vermersch, "Ethics and Techniques of the Invisible: Ethical Aspects of the Use of Undetectable Robots in the Private Sphere and Their Extension to Military Robots", în Ronan Doaré, Didier Danet, Jean-Paul Hanon, and Gérard de Boisboissel (eds.) *Robots on the Battlefield. Contemporary Perspectives and Implications for the Future* (Fort Leavenworth, Kansas: Combat Studies Institute Press, 2014), 146.
[10] George Lucas, "The Ethical Challenges of Unmanned Systems", 136.
[11] Steven A. Wartman, C. Donald Combs, "Reimagining Medical Education in the Age of AI", *AMA journal of ethics* vol. 21, no. 2 (2019): 146-152

https://journalofethics.ama-assn.org/article/reimagining-medical-education-age-ai/2019-02, accesat în 27.01.2020.
[12] David Cook, "Teaching with Technological Tools", în Kathryn N Huggett, William B. Jeffries (eds.), *An introduction to medical teaching* (Dordrecht: Springer, 2014), 136.
[13] C. Donald Combs, P. Ford Combs, "Emerging Roles of Virtual Patients in the Age of AI", *AMA Journal of Ethics*, February 2019, Volume 21, Number 2: E153-159 https://journalofethics.ama-assn.org/article/emerging-roles-virtual-patients-age-ai/2019-02, accesat în 03.03.2020.
[14] Steven Dilsizian, Eliot Siegel, "Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment", în *Current Cardiology Reports*, Vol. 16, No. 1, 2014: 441.
[15] Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, Josef Havel, "Artificial neural networks in medical diagnosis", în *Journal of Applied Biomedicine*, Vol. 11, Nr. 2, 2013: 52-53.
[16] Kadircan Keskinbora, "Medical ethics considerations on artificial intelligence", în *Journal of Clinical Neuroscience*, vol. 64, 2019, p. 278.
[17] Hayley Robinson, Bruce MacDonald, Ngaire Kerse, Elizabeth Broadbent, "The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial", *JAMDA. The Journal of Post-Acute and Long-Term Medicine*, vol. 14, issue 9, September 2013: 661-667, https://www.jamda.com/article/S1525-8610(13)00097-2/pdf, accesat în 27.01.2020.
[18] Utku Kose, Alice Pavaloiu, „Dealing with Machine Ethics in Daily Life: A View with Examples", *The 5th International Virtual Conference on Advanced Scientific Results*, June 26-30, 2017, DOI 1018638/scieconf.2017.5.1.454
[19] ***, "Artificial intelligence: Google's AlphaGo beats Go master Lee Se-do", *BBC News*, 12 March 2016, https://www.bbc.com/news/technology-35785875, accesat în 31.01.2020
[20] Patrick Lin, "Why Ethics Matters for Autonomous Cars", în Markus Maurer, Chris Gerdes, Barbara Lenz, Herman Winner (eds.), *Autonomous Driving. Technical, Legal and Social Aspects* (Heidelberg: Springer, 2016), 72-73.
[21] Bret Greenstein, "When Is It Ethical to Not Replace Humans with AI?", *Information Week*, December 30, 2019, https://www.informationweek.com/big-data/ai-machine-learning/when-is-it-ethical-to-not-replace-humans-with-ai/a/d-id/1336661, accesat în 28.01.2020.
[22] ***, "Drones to the Rescue: How Drones Are Helping Disaster Relief

Efforts", *GraVoc*, January 3, 2018, https://www.gravoc.com/2018/01/03/drones-to-the-rescue-how-drones-are-helping-disaster-relief-efforts/, accesat în 03.03.2020.
[23] Abby Wolfe, "A look at how Descartes Labs is leveraging AI to alert fire managers of wildfires and decrease the damage on homes and habitats across the US", *Business Insider*, December 19, 2019, https://www.businessinsider.com/how-descartes-labs-leveraging-artificial-intelligence-fight-wildfires-2019-12, accesat în 05.03.2020.
[24] Mike Kaput, "What Is Artificial Intelligence for Social Media", *Marketing Institue*, September 24, 2019, https://www.marketingaiinstitute.com/blog/what-is-artificial-intelligence-for-social-media, accesat în 29.01.2020.
[25] Global News, „What is Project Maven? The Pentagon AI project Google employees want out of", *Global News*, April 5, 2018 https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/, accesat în 30.01.2020.
[26] Simon Parkin, "Killer robots: The soldiers that never sleep", *BBC News*, July 16, 2015, https://www.bbc.com/future/article/20150715-killer-robots-the-soldiers-that-never-sleep?referer=https%3A%2F%2Fen.wikipedia.org%2F, accesat în 30.01.2020.
[27] Mark Smith, "Is 'killer robot' warfare closer than we think?", *BBC News*, August 25, 2017, https://www.bbc.com/news/business-41035201.
[28] David Luxton, "Should Watson Be Consulted for a Second Opinion?", în *AMA Journal of Ethics*, Vol. 21, number 2, February 2019: E131-137, https://journalofethics.ama-assn.org/article/should-watson-be-consulted-second-opinion/2019-02, accesat în 30.01.2020.
[29] Dina Bass, "Microsoft Develops AI to Help Cancer Doctors Find the Right Treatments", *Bloomberg*, September 20, 2016, https://www.bloomberg.com/news/articles/2016-09-20/microsoft-develops-ai-to-help-cancer-doctors-find-the-right-treatments, accesat în 30.01.2020.
[30] ***, "The CareBot Developed Using Data From the World's First in Home Trials of an Eldercare Mobile Service Robot", *Gecko Systems*, http://www.geckosystems.com/markets/CareBot_trials.php, accesat în 30.01.2020.
[31] ***, "Alexa Voice Service v20160207", *Amazon Alexa*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/api-overview.html, accesat în 30.01.2020.

[32] Daniel B. Shank, Alyssa DeSanti, "Attributions of Morality and Mind to Artificial Intelligence after Real-World Moral Violations", în *Computers in Human Behavior*, Vol. 86, May (2018): 401-411.
[33] Dave Gershgorn, "When artificial intelligence judges a beauty contest, white people win", *Quartz*, September 6, 2016, https://qz.com/774588/artificial-intelligence-judged-a-beauty-contest-and-almost-all-the-winners-were-white/, accesat în 30.01.2020.
[34] Gia Liu, "Hey, I didn't order this dollhouse! 6 hilarious Alexa mishaps", *Digital Trends*, March 6, 2018, https://www.digitaltrends.com/home/funny-accidental-amazon-alexa-ordering-stories/, accesat în 30.01.2020.

Bibliografie

***. 2016. "Artificial intelligence: Google's AlphaGo beats Go master Lee Se-do". *BBC News* 12 March https://www.bbc.com/news/technology-35785875, accesat în 31.01.2020.

***. 2017. "Weak Artificial Intelligence". *Techopedia* July 22 https://www.techopedia.com/definition/31621/weak-artificial-intelligence-weak-ai, accesat în 03.03.2020.

***. 2018. "Drones to the Rescue: How Drones Are Helping Disaster Relief Efforts". *GraVoc*. January 3. https://www.gravoc.com/2018/01/03/drones-to-the-rescue-how-drones-are-helping-disaster-relief-efforts/, accesat în 03.03.2020.

Amato, Filippo; Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, Josef Havel. 2013. "Artificial neural networks in medical diagnosis". *Journal of Applied Biomedicine* 11 (2): 47-58.

Arkin, Ronald. 2009. "Ethical Robots in Warfare". *IEEE Technology and Society Magazine* 28 (1): 31-32.

Bass, Dina. 2016. "Microsoft Develops AI to Help Cancer Doctors Find the Right Treatments". *Bloomberg*. September 20. https://www.bloomberg.com/news/articles/2016-09-20/microsoft-develops-ai-to-help-cancer-doctors-find-the-right-treatments, accesat în 30.01.2020.

Combs, C. Donald, P. Ford Combs. 2019. "Emerging Roles of Virtual Patients in the Age of AI". *AMA Journal of Ethics* 21 (2): E153-159 https://journalofethics.ama-assn.org/article/emerging-roles-virtual-patients-age-ai/2019-02, accesat în 03.03.2020.

Cook, David. 2014. "Teaching with Technological Tools". În Kathryn N Huggett, William B. Jeffries (eds.), *An introduction to medical* teaching. Dordrecht: Springer, 123-146.

Copeland, B.J. "Artificial Intelligence", *Encyclopedia Britannica*, https://www.britannica.com/technology/artificial-intelligence, Mar 5, 2020, accesat în 31.01.2020.

De Vigny, Alfred. 1914. *Servitude et grandeur militaires* Conard. https://www.bouquineux.com/?ebook s=82&Vigny, accesat în 30.01.2020.

Dilsizian, Steven, Eliot Siegel. 2014. "Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment". *Current Cardiology Reports* 16 (1): 441.

Frankenfield, Jake. 2019. "Strong a.i". *Investopia* October 2, https://www.investopedia.com/terms/s/strong-ai.asp, accesat în 03. 03. 2020.

Gershgorn, Dave. 2016. "When artificial intelligence judges a beauty contest, white people win". *Quartz*. September 6. https://qz.com/774588/artificial-intelligence-judged-a-beauty-contest-and-almost-all-the-winners-were-white/, accesat în 30.01.2020.

Global News. 2018. „What is Project Maven? The Pentagon AI project Google employees want out of". *Global News*. April 5. https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/, accesat în 30.01.2020.

Greenstein, Bret. 2019. "When Is It Ethical to Not Replace Humans with AI?". *Information Week*. December 3. https://www.informationweek.com/big-data/ai-machine-learning/when-is-it-ethical-to-not-replace-humans-with-ai/a/d-id/1336661, accesat în 28.01.2020.

Jajal, Tannya. 2018. "Distinguishing between Narrow AI, General AI and Super AI". *Medium* May 21, https://medium.com/@tjajal/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22, accesat în 03.03.2020.

Kaput, Mike. 2019. "What Is Artificial Intelligence for Social Media". *Marketing Institute.* September 24. https://www.marketingaiinstitute.com/blog/what-is-artificial-intelligence-for-social-media, accesat în 29.01.2020.

Keskinbora, Kadircan. 2019. "Medical ethics considerations on artificial intelligence". *Journal of Clinical Neuroscience* 64: 277-282.

Kose, Utku, Alice Pavaloiu. 2017. „Dealing with Machine Ethics in Daily Life: A View with Examples". *The 5th International Virtual Conference on Advanced Scientific Results* June 26-30, DOI 1018638/scieconf.2017.5.1.454

Lin, Patrick. 2016. "Why Ethics Matters for Autonomous Cars". În Markus Maurer, Chris Gerdes, Barbara Lenz, Herman Winner (eds.). *Autonomous Driving. Technical, Legal and Social Aspects*. Heidelberg: Springer, 69-85.

Liu, Gia. 2018. "Hey, I didn't order this dollhouse! 6 hilarious Alexa mishaps", *Digital Trends*, March 6, 2018, https://www.digitaltrends.com/home/funny-accidental-amazon-alexa-ordering-stories/, accesat în 30.01.2020.

Lucas, George. 2014. "The Ethical Challenges of Unmanned Systems", în Ronan Doaré, Didier Danet, Jean-Paul Hanon, and Gérard de Boisboissel (eds.) *Robots on the Battlefield. Contemporary Perspectives and Implications for the Future*. Fort Leavenworth, Kansas: Combat Studies Institute Press, 135-142.

Luxton, David. 2019. "Should Watson Be Consulted for a Second Opinion?". *AMA Journal of Ethics* 21 (2): E131-137, https://journalofethics.ama-assn.org/article/should-watson-be-consulted-second-opinion/2019-02, accesat în 30.01.2020.

Parkin, Simon. 2015. "Killer robots: The soldiers that never sleep". *BBC News*. July 16. https://www.bbc.com/future/article/20150715-killer-robots-the-soldiers-that-never-

sleep?referer=https%3A%2F%2Fen.wikipedia.org%2F, accesat în 30.01.2020.

Robinson, Hayley; Bruce MacDonald, Ngaire Kerse, Elizabeth Broadbent. 2013. "The Psychosocial Effects of a Companion Robot: A Randomized Controlled Trial". *JAMDA. The Journal of Post-Acute and Long-Term Medicine* 14 (9): 661-667, https://www.jamda.com/article/S1525-8610(13)00097-2/pdf, accesat în 27.01.2020.

Shank, Daniel B., Alyssa DeSanti. 2018. "Attributions of Morality and Mind to Artificial Intelligence after Real-World Moral Violations". *Computers in Human Behavior*. 86: 401-411.

Smith, Mark. 2017. "Is 'killer robot' warfare closer than we think?". *BBC News*. August 25, https://www.bbc.com/news/business-41035201.

Vermersch, Dominique. 2014. "Ethics and Techniques of the Invisible: Ethical Aspects of the Use of Undetectable Robots in the Private Sphere and Their Extension to Military Robots", în Ronan Doaré, Didier Danet, Jean-Paul Hanon, and Gérard de Boisboissel (eds.) *Robots on the Battlefield. Contemporary Perspectives and Implications for the Future*. Fort Leavenworth, Kansas: Combat Studies Institute Press, 143-150.

Wartman, Steven A., and C. Donald Combs. 2019. "Reimagining medical education in the age of AI." *AMA Journal of Ethics* 21 (2): 146-152.

Wolfe, Abby. 2019. "A look at how Descartes Labs is leveraging AI to alert fire managers of wildfires and decrease the damage on homes and habitats across the US. *Business Insider*. December 19. https://www.businessinsider.com/how-descartes-labs-leveraging-artificial-intelligence-fight-wildfires-2019-12, accesat în 05.03.2020.