**Moralitatea si inteligența artificială**

**Nume:** Nedelcu Radu – Ioan

**Grupa:** 152

Introducere:

În acest eseu voi aborda o temă de actualitate, cea a inteligenței artificiale și a modului în care societatea noastră integrează schimbările impuse de aceasta. Inteligența artificială apare treptat în tot mai multe aspecte ale vieții moderne: de la aplicații menite să ne ușureze viața, precum mașini care se conduc singure, până la cele care reușesc deja să salveze vieți, precum computerele inteligente folosite în spitale (Rohaidi, 2016). Este însă adevărat că aceasta tehnologie nu funcționează întotdeauna corect, iar aceste erori sunt de obicei foarte costisitoare, în special atunci când le este alocat un rol important, precum tranzacționarea la bursă (Salmon, 2010) sau conducerea unei mașini.

La începuturile sale, inteligența artificială a fost folosită pentru a rezolva probleme de matematică sau chimie, pentru care există reguli simple si clare. Dar ce se întâmplă atunci când domeniul în care este aplicată această tehnologie nu are astfel de reguli? Moralitatea este un subiect complex, asupra căruia nici oamenii nu se pot pune de acord.

Inteligența artificială, (abreviată IA), „este știința si ingineria de a face mașinării inteligente, în special programe de calculator inteligente. Este legată de obiectivul similar de a utiliza calculatoare pentru a înțelege inteligența umană, dar IA nu trebuie să fie constrânsă la metode observabile biologic.” (McCarthy, 2007, 2). Astfel, prin inteligentă artificială înțelegem multitudinea de aplicații în care programe de calculator sunt create pentru a servi anumite scopuri mai bine decât o poate face un om.

Personal, consider că obținerea unui astfel de sistem, bazat pe inteligența artificială care să fie capabil nu doar să înțeleagă dar să si aplice conștiința si moralitatea umană constituie o mare realizare tehnologică. Având în vedere tendința actuală de a obține sisteme care să fie mai performante decât omul, nu este greu să ne imaginăm un viitor în care tehnologiile din jurul nostru vor fi mai inteligente decât noi. Fără îndoială astfel de sisteme vor face decizii pentru noi (Spre exemplu: modul în care aplicații precum Google Maps deja decid pentru noi ruta până la destinație), adesea decizii cu implicații morale, motiv pentru care este important ca un calculator să poată distinge între ce este etic si ce nu.

Pentru a avea un program inteligent capabil de a face decizii morale este nevoie în primul rând ca acel sistem să înțeleagă ce este moral. Astfel, primul pas ar fi stabilirea unui set de reguli pe care un calculator să le poată utiliza atunci când are de luat o decizie. Problema este că moralitatea nu poate fi restrânsă la un număr finit de enunțuri / reguli de tipul: „Să nu minți!” sau „Să nu furi!”, pe care un program inteligent să le urmeze, deoarece lumea reală este plină de excepții de la aceste reguli. Spre exemplu, dacă naziștii bat la ușa ta, iar tu o adăpostești pe Anne Frank, majoritatea dintre oamenilor ar fi de acord că este moral să minți și să spui că nu este acolo, deși încalcă una dintre regulile anterior menționate. Totuși, această problemă se poate rezolva folosind o ramură modernă a inteligenței artificiale, anume învățarea automată (machine learning). Aceasta folosește cantități mari de date și algoritmi cu scopul de a imita modul în care oamenii învață, prin îmbunătățiri treptate (IBM, 2020). Această abordare s-a dovedit de succes în domenii precum șahul, unde programele inteligente sunt capabile să învingă cu ușurință chiar și pe cei mai buni dintre jucători. Se pleacă de la un set simplu de reguli, cele ale jocului de șah, iar prin încercări repetate de a juca, programul înțelege ce tipuri de mutări și în ce situații îl duc mai aproape de victorie. Spre exemplu, programul știe inițial ca nu este bine să își piardă tura, dar pe parcurs, pe măsură ce se lovește de astfel de situații, programul învață că există situații în care este bine să îți pierzi tura pentru a câștiga un alt avantaj.

Diferența între computerele de neînfrânt la jocul de șah și mașinile autonome ce folosesc inteligența artificială este că într-o situație critică doar cea de-a doua are de făcut o alegere ce poate afecta viața cuiva. Multe dintre accidentele din trafic se produc în fracțiuni de secundă, dese ori mult prea rapid pentru un om sa acționeze, dar îndeajuns pentru un calculator modern sa efectueze zeci de mii de calcule și să încerce să minimizeze daunele. De aici apare si nevoia ca mașinile autonome să fie capabile de a lua decizii morale.

Printre problemele regăsite in argumentul precedent, care susține avantajele creării unor mașinării inteligente, capabile de a lua decizii morale, se află punctul slab al companiilor din spatele acestor algoritmi. Cum am menționat anterior, pregătirea algoritmilor de învățare automată necesită cantități mari de date pe care acesta să le prelucreze, lucru care este de obicei accesibil doar giganților tehnologici precum Google sau Meta (Ex Facebook) și companiilor in care aceștia investesc. Este cunoscut faptul ca multe dintre acestea nu au un istoric bun în ceea ce privește respectarea confidențialității si libertăților utilizatorilor (Ma și Gilbert, 2019) și totodată sunt caracterizate de lipsa lor de transparență în ceea ce privește produsele lor. Din acest motiv, chiar dacă tehnologia permite crearea unei mașinării cu inteligență artificială capabilă de a lua decizii morale, nu te putem asigura că aceasta va respecta într-adevăr principii etice sau doar interesele corporațiilor din spate. Spre exemplu, Christoph von Hugo, managerul de departamentului „Active Safety” al producătorului auto Mercedes Benz a declarat faptul că mașinile lor autonome vor prioritiza salvarea vieții persoanelor din mașină (Taylor, 2016).

Această idee contrazice valorile morale regăsite de experimentul „[The Moral Machine](https://www.moralmachine.net/)” derulat de o echipă de cercetători de la MIT, în care participanții din întreaga lume au avut de ales cum ar trebui sa acționeze o mașină autonomă în cazul unei situații limită. Informațiile obținute în urma experimentului au arătat că preferința etică a oamenilor este de a salva cât mai multe vieți și de a salva vieți cât mai tinere, indiferent dacă acestea se află in postura pasagerilor sau a pietonilor (Awad și altii, 2018).

Problema ridicată de contra argument este una destul de dificilă pentru ca nu are o rezolvare simplă. În schimb, cunoaștem exemplele istorice ale altor tehnologii si modul în care acestea au evoluat. Primele calculatoare comerciale au fost vândute în anii ’50 și ocupau întregi încăperi iar prețul era pe măsură, dar cu timpul, datorită interesului pentru aceste produsele, a îmbunătățirilor treptate făcute asupra lor și implicit a proceselor de fabricație, calculatoarele se află astăzi în majoritatea gospodăriilor. Interesul asupra inteligenței artificiale este în continuă creștere, tot mai multe companii investesc în astfel de sisteme și dezvoltă moduri cât mai eficiente pentru a le opera.

Pe de altă parte, este de așteptat ca o tehnologie modernă, nouă, să fie exploatată atât pentru a face bine cât si pentru a face rău. Spre exemplu, cercetarea în domeniul tehnologiei nucleare a început în încercarea de a valorifica energia produsă de fisiunea nucleară (World Nuclear Association, 2020). Datorită ei avem astăzi unul dintre cele mai eficiente moduri de a produce energie, reactoarele nucleare, dar și cea mai puternică armă de distrugere, bomba nucleară.

Concluzie:

În concluzie, inteligența artificială este o unealtă tot mai des folosită și constant îmbunătățită, iar dezvoltarea unui astfel de sistem, capabil să înțeleagă si să aplice moralitatea umană este complet necesară într-o lume ce se bazează tot mai tare pe deciziile luate de calculatoarele din jurul nostru.

Bibliografie:

Awad Edmond, Dsouza Sohan, Kim Richard, Schulz Jonathan, Henrich Joseph, Shariff Azim, Bonnefon Jean-François, Iyad Rahwan, ”The Moral Machine Experiment”, Massachusetts Institute of Technology, 2018 <https://core.ac.uk/download/pdf/231922494.pdf>

IBM Cloud Education, ”Machine Learning”, IBM, 15.07.2020 <https://www.ibm.com/cloud/learn/machine-learning>

McCarthy John, ”What is artificial intelligence?”, Stanford, Stanford University, 2007

Ma Alexandra, Gilbert Ben, ”What Is Cambridge Analytica?”, Business Insider, 23.08.2019 <https://www.businessinsider.com/cambridge-analytica-a-guide-to-the-trump-linked-data-firm-that-harvested-50-million-facebook-profiles-2018-3>

Rohaidi Nurfilzah, ”IBM`s Watson Detected Rare Leukemia In Just 10 Minutes”, AsianScientist, 15.08.2016

<https://www.asianscientist.com/2016/08/topnews/ibm-watson-rare-leukemia-university-tokyo-artificial-intelligence/>

Salmon Andy, ”Algorithms Take Control of Wall Street”, Wired, 27.12.2010, <https://www.wired.com/2010/12/ff-ai-flashtrading/>

Taylor Michael, ”Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians”, Car And Driver, 7.10.2016 <https://www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/>

# TEDx Talks, ”Morality and Artificial Intelligence: The Science and Beyond | Devin Gonier | TEDxAustinCollege”, 16:24, 30.10.2018 <https://www.youtube.com/watch?v=2_8kgV-EfTA>

World Nuclear Association, ”Outline History of Nuclear Energy”, 11.2020

<https://world-nuclear.org/information-library/current-and-future-generation/outline-history-of-nuclear-energy.aspx>