Etica in Inteligenta Artificiala

Andrei Theodor IORDACHE

Universitatea Româno-Americană, București, România

Abstract

Inteligența artificială (AI) reprezintă utilizarea mașinilor pentru a face lucruri care ar necesita în mod normal inteligența umană. În multe domenii ale vieții umane, AI a afectat rapid și semnificativ societatea umană și modul în care interacționăm unul cu celălalt. Va continua să facă acest lucru. Pe parcurs, AI a prezentat provocări etice și socio-politice substanțiale care necesită o analiză filozofică și etică aprofundată. Impactul său social ar trebui studiat astfel încât să se evite orice repercusiune negativă. Sistemele AI devin din ce în ce mai autonome, aparent raționale și inteligente. Această dezvoltare cuprinzătoare dă naștere la numeroase probleme. În plus față de potențialul prejudiciu și impactul tehnologiilor AI asupra vieții private, alte preocupări includ statutul lor moral și juridic (inclusiv drepturile morale și legale), posibila lor agenție morală și răbdarea și problemele legate de posibila lor personalitate și chiar demnitate.

Cuvinte Cheie: Artificial Intelligence (Inteligenta Artificiala), Autonom, Etica

Introducere

AI poate însemna mai multe lucruri și este definit în multe moduri diferite. Când Alan Turing a introdus așa-numitul test Turing (pe care îl numea „joc de imitație") în faimosul său eseu din 1950 despre mașinile care pot gândi, termenul „inteligență artificială" nu fusese încă introdus. Turing a analizat dacă mașinile pot gândi și a sugerat că ar fi mai clar să înlocuim această întrebare cu întrebarea dacă ar fi posibil să se construiască mașini care să poată imita oamenii atât de convingător încât oamenilor le-ar fi greu să spună dacă, de exemplu, mesajul scris vine de la un computer sau de la un om (Turing 1950).

La începutul secolului al XXI-lea, scopul final al multor specialiști și ingineri în calculatoare a fost de a construi un sistem robust AI care să nu difere de inteligența umană în niciun alt aspect decât originea mașinii sale. Faptul că acest lucru este posibil este o chestiune de dezbatere de câteva decenii. Proeminentul filozof american John Searle a introdus așa-numitul argument al camerei chinezești pentru a susține că un AI puternic sau general adică construirea unor sisteme AI care ar putea face față multor sarcini diferite și complexe care necesită inteligență asemănătoare omului - este în principiu imposibil. Astfel, a declanșat o dezbatere generală de lungă durată cu privire la posibilitate. Sistemele actuale de AI sunt focalizate în mod restrâns (adică slabe) și nu pot rezolva decât o singură sarcină anume, cum ar fi jocul de șah sau jocul chinezesc Go. Teza generală a lui Searle a fost că, indiferent cât de complexă și sofisticată este o mașină, ea nu va avea totuși nici „conștiință" sau „minte", ceea ce este o condiție prealabilă pentru capacitatea de a înțelege, spre deosebire de capacitatea de a calcula.

Provocările etice majore pentru societățile umane ale AI sunt prezentate bine în introducerile excelente de Vincent Müller (2020), Mark Coeckelbergh (2020), Janina Loh (2019), Catrin Misselhorn (2018) și David Gunkel (2012). Indiferent de posibilitatea de a interpreta AGI, sistemele autonome de IA ridică deja probleme etice substanțiale: de exemplu, părtinirea mașinii în legislație, luarea deciziilor de angajare prin intermediul algoritmilor inteligenți, chatbot rasist și sexist sau traduceri de limbă. Însăși ideea unei mașini care „imită" inteligența umană - care este o definiție comună a AI - dă naștere la îngrijorări cu privire la înșelăciune, mai ales dacă AI este încorporată în roboți concepuți să arate sau să acționeze ca niște ființe umane (Boden 2017; Nyholm și Frank 2019). Mai mult, Rosalind Picard susține pe bună dreptate că „cu cât este mai mare libertatea unei mașini, cu atât va avea nevoie de standarde morale" (1997). Acest lucru fundamentează afirmația că toate interacțiunile dintre sistemele AI și ființele umane implică în mod necesar o dimensiune etică, de exemplu, în contextul transportului autonom.

Ideea implementării eticii în cadrul unei mașini este unul dintre principalele obiective de cercetare în domeniul eticii mașinilor (de exemplu, Lin și colab. 2012; Anderson și Anderson 2011; Wallach și Allen 2009). Din ce în ce mai multă responsabilitate a fost transferată de la ființe umane la sisteme autonome de AI care sunt capabile să funcționeze mult mai repede decât ființele umane fără a face pauze și fără a fi nevoie de supraveghere constantă, așa cum este ilustrat de performanța excelentă a multor sisteme (odată ce au reușit a trecut faza de depanare).

S-a sugerat că existența viitoare a umanității poate depinde de implementarea unor standarde morale solide în sistemele de AI, având în vedere posibilitatea ca aceste sisteme, la un moment dat, să se potrivească sau să înlocuiască capacitățile umane. Acest moment a fost

numit „singularitate tehnologică” de Vernon Vinge în 1983. Celebrul dramaturg Karl Čapek (1920), renumitul astrofizician Stephen Hawking și influentul filosof Nick Bostrom (2016, 2018) au avertizat cu toții despre posibilele pericole ale singularității tehnologice în cazul în care mașinile inteligente se vor întoarce împotriva creatorilor lor, adică a ființelor umane. Prin urmare, potrivit lui Nick Bostrom, este extrem de important să construim AI prietenoase.

Propunerile pentru un sistem de etică a mașinilor. sunt discutate din ce în ce mai mult în legătură cu sistemele autonome a căror funcționare prezintă un risc de a dăuna vieții umane. Cele două exemple cele mai des discutate - care sunt uneori discutate împreună și contrastate și comparate între ele - sunt vehicule autonome (cunoscute și sub numele de mașini cu conducere automată) și sisteme de arme autonome (numite uneori „roboți ucigași”) (Purves și colab. 2015; Danaher 2016; Nyholm 2018).

Mulți oameni cred că utilizarea tehnologiilor inteligente ar pune capăt prejudecăților umane din cauza presupusei „neutralități” a mașinilor. Cu toate acestea, am ajuns să ne dăm seama că mașinile pot menține și chiar demonstra părtinirea umană față de femei, etnii diferite, vârstnici, persoane cu deficiențe medicale sau alte grupuri.

Ideea de a utiliza sisteme de AI pentru a sprijini luarea deciziilor umane este, în general, un obiectiv excelent, având în vedere eficiența, precizia, amploarea și viteza crescuta ale AI în luarea deciziilor și găsirea celor mai bune răspunsuri. Cu toate acestea, părtinirea mașinii poate submina această situație aparent pozitivă în diferite moduri.

AI ca formă de îmbunătățire morală sau consilier moral

Sistemele AI tind să fie utilizate ca „sisteme de recomandare” în cumpărături online, divertisment online (de exemplu, muzică și streaming de filme) și alte tărâmuri. Unii eticieni au discutat despre avantajele și dezavantajele sistemelor de IA, ale căror recomandări ne-ar

putea ajuta să facem alegeri mai bune și mai compatibile cu valorile noastre de bază. Poate că sistemele de AI ar putea chiar, la un moment dat, să ne ajute să ne îmbunătățim valorile. Lucrările la aceste întrebări și întrebări conexe includ Borenstein și Arkin (2016), Giubilini și colab. (2015, 2018), Klincewicz (2016) și O'Neill și colab. (2021).

AI și viitorul muncii

Multe discuții despre AI și viitorul muncii se referă la problema vitală a faptului dacă AI și alte forme de automatizare vor provoca „șomaj tehnologic" pe scară largă prin eliminarea unui număr mare de locuri de muncă umane care ar fi preluate de mașinile automate. Acest lucru este adesea prezentat ca o perspectivă negativă, unde întrebarea este cum și dacă o lume fără muncă ar oferi oamenilor orice perspectivă pentru activități satisfăcătoare și semnificative, deoarece anumite bunuri realizate prin muncă (altele decât venitul) sunt greu de realizat în alte contexte. Cu toate acestea, unii autori au susținut că munca în lumea modernă expune multe persoane la diferite tipuri de rău (Anderson 2017). Danaher (2019a) examinează întrebarea importantă dacă o lume cu mai puțină muncă ar putea fi de fapt preferabilă. Unii susțin că plictiseala existențială ar prolifera dacă ființele umane nu mai pot găsi un scop semnificativ în munca lor (sau chiar în viața lor), deoarece mașinile le-au înlocuit. În contrast, argumentând că plictiseala nu va fi deloc o problemă substanțială. O altă problemă conexă - poate mai relevantă pe termen scurt și mediu - este modul în care putem face ca lucrările din ce în ce mai tehnologizate să rămână semnificative.

AI și viitorul relațiilor personale

Diverse tehnologii bazate pe AI afectează natura prieteniilor, romantismelor și a altor relații interpersonale și le-ar putea afecta și mai mult în viitor. „Prieteniile" online aranjate prin intermediul rețelelor sociale au fost cercetate de filozofi care nu sunt de acord cu privire la faptul dacă relațiile care sunt parțial curate de algoritmi AI ar putea fi adevărate prietenii. Unii filozofi au criticat aspru aplicațiile de întâlnire bazate pe AI, care cred că ar putea întări stereotipurile negative și așteptările negative. În mai multe filozofii asemănătoare științifico-ficțiunii, care ar putea totuși să devină din ce în ce mai prezente în viața reală, s-a discutat, de asemenea, dacă ființele umane ar putea avea adevărate prietenii sau relații romantice cu roboți și alți agenți artificiali echipați cu AI avansată.

Concluzie

Etica AI a devenit unul dintre cele mai vii subiecte din filosofia tehnologiei. AI are potențialul de a ne redefini conceptele morale tradiționale, abordările etice și teoriile morale. Apariția mașinilor inteligente artificiale care se pot potrivi sau înlocui capacitățile umane reprezintă o mare provocare pentru înțelegerea de sine tradițională a umanității ca fiind singurele ființe cu cel mai înalt statut moral din lume. În consecință, viitorul eticii AI este imprevizibil, dar probabil va oferi o emoție și o surpriză considerabile.

Bibliografie

https://www.youtube.com/watch?v=iiAirfn-lBI

https://books.google.ro/books?hl=ro&lr=&id=RYOYAwAAQBAJ&oi=fnd&pg=PA316&dq=Importance+of+ethics+in+Artificial+intelligence&ots=A1VZui8Epx&sig=rAMYIYKIDHaIfMRdORpUQcIkr0E&redir_esc=y

https://link.springer.com/article/10.1007/s10892-017-9252-2

https://journals.sagepub.com/doi/abs/10.1016/j.carj.2019.08.010

https://www.nature.com/news/program-good-ethics-into-artificial-intelligence-1.20821

https://link.springer.com/article/10.1007/s00146-017-0768-6

https://link.springer.com/article/10.1007/s00146-017-0760-1

https://edusoft.ro/brain/index.php/brain/article/view/818

https://link.springer.com/article/10.1007/s43681-020-00022-3

https://link.springer.com/content/pdf/10.1007/s11023-017-9449-y.pdf