

Menu

Commenta

> Anno 2018 > n°2 - 2018 > **L'intelligenza Artificiale tra funzionalità ed etica**

L'intelligenza Artificiale tra funzionalità ed etica



Antonio Santangelo

(/tit/it/notiziariotecnico/autori/santangelo.html)

Antonio Vetrò

(/tit/it/notiziariotecnico/autori/vetrò.html)

Juan Carlos De Martin

(/tit/it/notiziariotecnico/autori/cd-e/juan-carlos-de-martin.html)

L'intelligenza Artificiale tra funzionalità ed etica

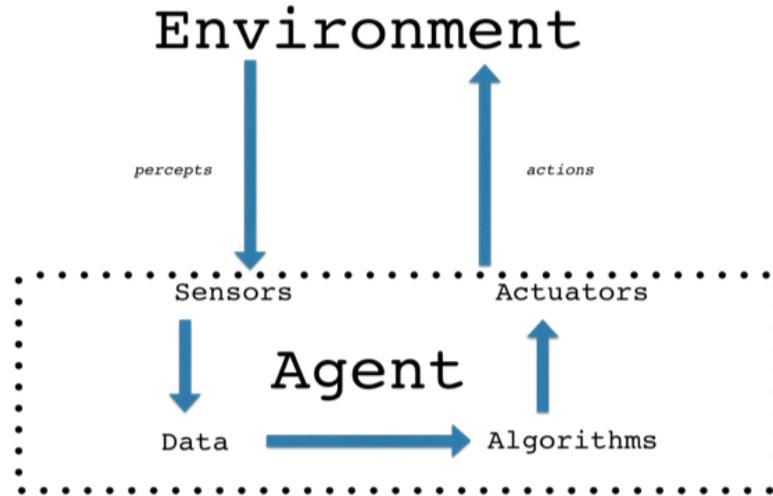
L'avvento dell'Intelligenza Artificiale cambierà il nostro modo di usare la tecnologia: basti pensare che già oggi essa è comandabile tramite il parlato e che con il passare degli anni sempre più mansioni cognitive potranno essere eseguite da un software. Tutto questo sta già modificando il nostro modo di vivere presto certe trasformazioni diventeranno molto diffuse. È necessario, dunque, essere consapevoli di ciò che stiamo realizzando, per progettare collettivamente tecnologie che siano allo stesso tempo intelligenti ed eque. Non basta, infatti, che l'Intelligenza Artificiale sappia svolgere al meglio le sue funzioni: è importante che contribuisca anche a costruire una società più giusta.

Che tipo di razionalità, per la tecnologia che ci renderà "padroni del mondo"?

La locuzione "Intelligenza Artificiale" va di moda e se ne comprende la ragione, visto che si sta investendo in un campo di ricerca che si presenta come una platea di sedicimila studiosi russi, che chi svilupperà le soluzioni migliori, in questo ambito, «diventerà il padrone del mondo» [nota 1 (#note)]. Una convinzione che sembra condivisa da molti, a giudicare dalle ingenti somme di denaro investite da governi e grandi imprese, per assicurarsi una posizione predominante nel settore [nota 2 (#note)].

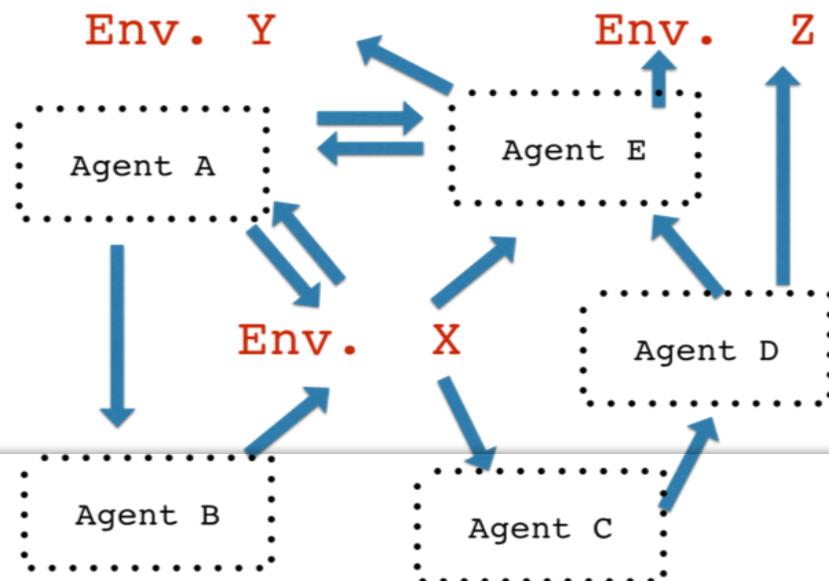
Tuttavia, come spesso accade, tutta questa attenzione ha generato confusione, anche tra gli addetti ai lavori, che si riferiscono all'Intelligenza Artificiale per parlare di cose molto diverse. Cercando di mettere un po' d'ordine, per Russell e Norvig [nota 3 (#note)] (2010) si tratta dell' «ambito di studi in cui si progettano e si costruiscono agenti intelligenti» [nota 4 (#note)] (p.34), dove per "agente" si intende «qualunque cosa possa dimostrare la capacità di percepire il proprio ambiente attraverso sensori e di agire su quest'ultimo attraverso degli attuatori» [nota 5 (#note)] (p.34), mentre per "intelligente" si definisce un agente "razionale", che «per ogni possibile sequenza di percetti [...] è in grado di selezionare, tra le varie possibili azioni in risposta, quella che si presuppone massimizzerà la misura della sua performance, sulla base dei dati derivanti dalla sequenza degli stessi percetti e della forma di conoscenza inserita al suo interno» [nota 6 (#note)] (p.36).

Figura 1 - Un agente interagisce con l'ambiente attraverso sensori e attuatori.



La Figura 1 schematizza questa definizione, che costituisce il blocco elementare attraverso cui si determinano le tassonomie [nota 7 (#note)] che guidano il design di sistemi tecnologici a complessità variabile (es. da un "semplice" traduttore vocale sino a un veicolo autonomo): basti pensare che la sola aggiunta della connettività allo schema base rende molto più complesso lo scenario, come mostrato in Figura 2 .

Figura 2 - La connettività abilita azioni ed effetti tra agenti e ambienti diversi



Questa definizione di Intelligenza Artificiale si basa su una visione molto precisa, e in un certo senso ristretta, del concetto di intelligenza, molto concentrata sulla funzionalità delle macchine, sulla loro capacità di rispondere in maniera efficace agli stimoli dell'ambiente in cui operano. Ma è chiaro che la parte più rilevante è giocata dalla "forma di conoscenza" con cui si decide di misurare le loro azioni. In questo senso, il mondo scientifico sta muovendo passi rilevanti [nota 8 (#note)], sostenendo la necessità di mettere l'etica al centro della programmazione dell'Intelligenza Artificiale, al fine di non rischiare di dare origine a un mondo distopico, in cui, per l'appunto, i padroni possano diventare i possessori di algoritmi sicuramente efficaci e molto performanti, ma allo stesso tempo decisamente ingiusti. Esistono già parecchi esempi, a questo proposito, dagli strumenti di Intelligenza Artificiale che aiutano gli istituti finanziari a decidere a chi erogare un prestito, e che si basano sull'idea che sia meglio favorire i cittadini bianchi, istruiti, residenti in certe zone specifiche delle città, a quelli con cui viene stabilito quali pratiche portare avanti, per concedere la libertà vigilata ai detenuti, che ancora una volta privilegiano individui appartenenti a certi gruppi etnici. Oppure, ancora, si parla di quei programmi che indirizzano offerte di lavoro economicamente più vantaggiose agli uomini anziché alle donne [nota 9 (#note)], e così via. Evidentemente, le "forme di conoscenza" su cui si appoggiano gli algoritmi che "animano" queste macchine sono il frutto di basi di dati (o, nel più semplice dei casi, di indagini statistiche), che possono anche essere – anche se non è detto che lo siano – accurate, ma che rappresentano certe storture della nostra società.

Si veda ad esempio la Tabella 1: essa mostra la frequenza dei valori dell'attributo "gruppo etnico" nel database utilizzato dall'algoritmo COMPAS [nota 10 (#note)] nel sistema giudiziario americano. I numeri in tabella mostrano una grossa disparità di rappresentanza: infatti, il 33,22% delle osservazioni del dataset si riferisce ai bianchi, mentre il 53,45% ai neri. Ne consegue che vi è una sovrastima dell'attributo gruppo etnico – principalmente contro la parte di popolazione afro-americana - che distorce la stima della probabilità di recidiva fornita dall'algoritmo. Dal momento che l'algoritmo fornisce supporto alle decisioni dei giudici, questi sono a loro volta condizionati da un *bias*. Se si pensa che in alcuni ambiti i sistemi di intelligenza artificiale prendono numerose decisioni in autonomia, allora il *bias* ha un impatto potenzialmente scalabile all'intero insieme delle persone su cui la decisione automatica viene presa.

Tabella 1 - Distribuzione dei gruppi etnici all'interno del database COMPAS, per livello di rischio di recidiva

Ethnic group	High	Low	Medium	N/A	TOT ROW
African-American	3400	3369	3010	12	9791
Asian	9	50	12	0	71
Caucasian	943	3554	1579	10	6086
Hispanic	191	945	315	0	1451
Native American	15	26	16	0	57
Other	56	653	150	1	860
TOT COLUMN	4614	8597	5082	23	18316

Il problema, dunque, è che se questi *bias* del mondo in cui viviamo non sono identificati e vengono inseriti all'interno degli strumenti di cui ci serviamo, e di cui sempre più ci serviremo, certe ingiustizie non potranno che perpetuarsi e acuirsi.

La domanda che ci dobbiamo porre, dunque, è: che tipo di razionalità vogliamo imporre all'Intelligenza Artificiale?

Per un'Intelligenza Artificiale come "agente etico"

Gli esempi che abbiamo riportato, ci fanno comprendere che accanto alla definizione di Intelligenza Artificiale come "agente razionale", dovremmo portarne avanti un'altra in parallelo: quella di "agente etico".

In questi anni, si è sollevato nella comunità scientifica [nota 11 (#note)] e nella società civile [nota 12 (#note)] un grande dibattito, a questo proposito, man mano che l'utilizzo dei cosiddetti "big data" ha dimostrato tutte le sue potenzialità in diversi ambiti, tra cui, appunto, quello dell'Intelligenza Artificiale. Come abbiamo visto in precedenza, in questo campo uno dei problemi più rilevanti è quello dei data set sbilanciati, che «sovrastimano o sottostimano il peso di alcune variabili — spesso, ancora una volta, di genere o legate all'appartenenza degli individui ad alcune minoranze [n.d.a.] — nella ricostruzione della relazione causa-effetto necessaria per spiegare certi eventi e, soprattutto, per prevederli» [nota 13 (#note)], come è avvenuto con alcuni algoritmi utilizzati dalla polizia americana per prevenire i crimini [nota 14 (#note)]. Inoltre, vi sono casi in cui i *bias* dei dati (o quelli incorporati dagli algoritmi) non sono solo un semplice riflesso del mondo in cui viviamo, ma possono essere inseriti durante il processo di training degli agenti. Si pensi, ad esempio, che nelle tecniche di *supervised learning*, al momento tra le più diffuse, le macchine devono essere istruite, per svolgere i loro calcoli e hanno bisogno di dati "annotati" da esseri umani, che possono interpretarli in

maniera scorretta, inducendole in errore. Per esempio, si parla molto delle differenze di genere, o di provenienza etnica e sociale, che possono produrre diversi *bias* di valutazione del significato di un'immagine o di un concetto [nota 15 (#note)]. È necessario, in pratica, che i dati su cui poggia l'Intelligenza Artificiale non inducano quest'ultima a compiere valutazioni o azioni che varino a seconda della popolazione o che possano discriminare i vari gruppi di cui la popolazione stessa è costituita.

Ma i problemi etici sollevati dal funzionamento dell'Intelligenza Artificiale vanno ben oltre la composizione delle sue basi di dati. Tutti gli esempi che abbiamo riportato mettono infatti in evidenza il grande interrogativo legato all'utilizzo di questo genere di strumenti nell'attività decisionale, sia come assistenti degli esseri umani, sia come "soggetti" autonomi. In tutti e due i casi, queste tecnologie producono degli effetti sulla vita delle persone, a proposito dei quali è necessario potere sempre individuare un *responsabile* in carne e ossa, o quantomeno istituzionale. Il punto, però, è capire di chi si tratti, se del produttore [nota 16 (#note)] o del possessore [nota 17 (#note)] dell'Intelligenza Artificiale, oppure ancora del suo utente finale: è giusto incolpare il Ministro di un Governo, oppure intentare una causa a uno Stato, per aver attuato certe politiche, sulla base del supporto di algoritmi che si appoggiano su dati affetti dai *bias* di cui si è scritto? Se un robot – per esempio un'auto senza conducente, oppure un drone pilotato da un computer – fa del male a qualcuno, chi deve essere incriminato? Da un punto di vista etico, poi, l'Intelligenza Artificiale pone problemi di *trasparenza* e di *apertura*, poiché spesso non è possibile determinare né quali sono i dati su cui essa basa il suo funzionamento, né l'architettura dei suoi algoritmi, che sono coperti dal segreto industriale. Questo può rivelarsi pericoloso in molti ambiti. Per esempio, nel mondo del lavoro, cominciano a sollevarsi perplessità sull'utilizzo di strumenti di Intelligenza Artificiale nella selezione e nella gestione del personale, di cui né gli impiegati, né i corpi intermedi conoscono i meccanismi. Ma si pensi anche agli scenari distopici dell'adozione di macchine "opache" da parte dello Stato, che amministrerebbe il suo potere senza consentire ai cittadini di poterne controllare l'operato. Per questa ragione, in Paesi come la Francia, si cerca di perseguire una politica legata alla promozione degli *open data* e dell'*open code*, anche nel settore dell'Intelligenza Artificiale [nota 18 (#note)].

Ovviamente, il contraltare di quanto appena scritto è la necessità di tutelare la *privacy* degli individui, un problema che induce a porre dei paletti oltre i quali la trasparenza non può essere perseguita. Uno dei nodi tipici, nel campo dell'Intelligenza Artificiale e non solo è, per esempio, quello del cosiddetto "effetto mosaico", legato all'utilizzo secondario di certi dati, molto frequente nelle ricerche in ambito sanitario, che non è facile prevedere sin dall'inizio e che, per questo, rende di complessa attuazione il consenso informato. Un consenso che deve essere chiamato in causa anche per stabilire in quali situazioni ognuno di noi possa rifiutarsi di essere sottoposto a "trattamento" mediante strumenti di Intelligenza Artificiale.

Conclusioni

Per affrontare queste sfide poste dalla progettazione di un'Intelligenza Artificiale animata da una solida "razionalità etica", può essere utile seguire alcuni principi generali, che citiamo dal libro bianco *Intelligenza Artificiale al servizio del cittadino* (2018), che due degli autori di questo articolo hanno contribuito a redigere, insieme all'Agenzia per l'Italia Digitale (Agid). Tra questi, innanzitutto, va menzionato quello che si potrebbe definire "umanistico" o "antropocentrico", secondo cui l'Intelligenza Artificiale deve essere sempre messa al servizio dell'uomo e non viceversa [nota 19 (#note)]. Ci sono, poi, «principi di equità, come quello procedurale (non arbitrarietà delle procedure), formale (uguale trattamento per individui o gruppi uguali) e sostanziale (rimozione effettiva degli ostacoli di natura economico-sociale)» [nota 20 (#note)], il soddisfacimento di alcuni bisogni di base universali, come il rispetto dei diritti di libertà e di rivendicazione di questi ultimi.

Qualcuno, oggi, animato da una visione distopica dell'Intelligenza Artificiale, teme che essa prenderà il sopravvento sugli uomini, deciderà per loro, ruberà loro il lavoro, li discriminerà, ne violerà la *privacy*, li controllerà di nascosto e, in contesti estremi come quelli di guerra, li ucciderà. Altri, invece, più utopisti, sognano un mondo più equo, in cui i governanti saranno supportati da potenti strumenti di calcolo che processeranno e sapranno interpretare nella maniera migliore grandi moli di dati, in cui i lavoratori saranno sollevati dai compiti più gravosi e ripetitivi, i prodotti e i servizi costeranno di meno, le aziende aumenteranno i profitti, gli apparati burocratici si snelleranno e le pratiche si velocizzeranno, i crimini diminuiranno e le malattie verranno studiate meglio, fino a debellarle. Noi, naturalmente, vorremmo che tutto ciò si realizzasse, ma proponiamo una via critica verso l'utopia, vale a dire uno studio attento delle problematiche etico-sociali nascoste dietro ai meccanismi di design e sviluppo degli agenti intelligenti, affinché questi ultimi vengano progettati in maniera responsabile e inclusiva. Non per diventare, come suggerisce la metafora di Vladimir Putin che abbiamo riportato all'inizio dell'articolo, padroni del mondo, ma per essere, più semplicemente, padroni del nostro destino e, soprattutto, architetti di un futuro migliore, anche grazie all'Intelligenza Artificiale.

Note

1. Corriere della Sera, 4 settembre 2017: https://www.corriere.it/tecnologia/economia-digitale/17_settembre_04/putin-sull-intelligenza-artificiale-chi-sviluppa-migliore-governa-

- mondo-musk-rilancia-l-allarme-c2a46c9c-916f-11e7-8332-148b1c29464d.shtml (ultima consultazione 7 luglio 2018).
2. Techwire Asia, 23 marzo 2018: <https://techwireasia.com/2018/03/investments-cognitive-ai-will-reach-19-1bn-2018/> (ultima consultazione 7 luglio 2018).
The Guardian, 25 aprile 2018:
<https://www.theguardian.com/technology/2018/apr/25/european-commission-ai-artificial-intelligence> (ultima consultazione 7 luglio 2018).
 3. Norvig è il direttore della ricerca di Google, dunque uno dei soggetti più direttamente impegnati nello sviluppo di questo genere di tecnologie.
 4. Traduzione ad opera degli autori di questo articolo.
 5. Traduzione ad opera degli autori di questo articolo.
 6. Traduzione ad opera degli autori di questo articolo.
 7. Ad esempio, un ambiente può essere reale o virtuale, deterministico o stocastico, a tempo continuo o discreto, competitivo o cooperativo, ecc.
 8. Si pensi ad esempio all'iniziativa congiunta del Berkman Klein Center dell'Università di Harvard e del MIT Media Lab, che sono a capo di un programma da 27Mln. \$ per studiare etica e la governance dell'Intelligenza Artificiale (<https://cyber.harvard.edu/research/ai>)
 9. <https://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html>
 10. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) è un algoritmo usato dai giudici negli USA per calcolare la probabilità di recidiva entro due anni da un crimine: <https://www.kaggle.com/danofer/compass>
 11. Si veda, per esempio, Bruno Lepri, Nuria Oliver, Emmanuel Letouz, Alex Pentland, Patrick Vinck, *Fair, transparent and accountable algorithmic decision-making processes. The premise, the proposed solutions, and the open challenges*, Science business media, Springer, 2017.
 12. Si veda, per esempio, Cathy O'Neil, *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia* (2017).
 13. *L'Intelligenza Artificiale al servizio del cittadino* (2018, p. 36), libro bianco a cura di Agid (Agenzia per l'Italia Digitale).
 14. Bruno Lepri, Nuria Oliver, Emmanuel Letouz, Alex Pentland, Patrick Vinck, *ibidem*.
 15. <https://techcrunch.com/2016/09/11/a-cautionary-tale-about-humans-creating-biased-ai-models/>
<https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>
 16. Ma, almeno secondo alcuni, ci sono reti neurali i cui algoritmi di calcolo non sono del tutto ricostruibili nemmeno dai loro programmatori, generando quello che viene definito in gergo "effetto black box". Si veda, su questi temi: <https://arxiv.org/abs/1706.08606>, <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>.
 17. Come avviene attualmente nel campo della robotica.
 18. Si veda ad esempio l'iniziativa Software "Heritage": <https://www.softwareheritage.org/>
 19. Per esempio, le famose leggi della robotica di Asimov vanno in questa direzione: un robot non può recar danno a un essere umano né può permettere che, a causa del proprio mancato intervento, un essere umano riceva danno; un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Prima Legge; un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la Prima o con la Seconda Legge.
 20. *Intelligenza Artificiale al servizio del cittadino* (2018, p. 37).



Partecipa alla discussione...

ENTRA CON

o REGISTRATI SU DISQUS 

Nome



Domenico · 5 giorni fa

Le considerazioni svolte dagli autori sono molto stimolanti, dischiudendo campi di indagine vasti ed intricati. Non dimentichiamo che i timori di quanto paventato, o già oggi realizzato, tramite i metodi dell'AI, non può prescindere dalla medesima realtà, di certo non utopica, nella quale siamo immersi e da cui, necessariamente, l'AI trarrà la sua "natura". Un esempio banalissimo su tutti: la Boston Dynamics studia robot di ausilio all'esercito americano (che finanzia appunto la ditta in questione), mica robot che possano sostituire il lavoro umano, ed infernale, svolto nelle miniere....

^ | ▾ · Rispondi · Condividi ▾