
L'incertezza ci salverà

· *Intelligenza artificiale, non solo algoritmi* ·

13 febbraio 2019

Le intelligenze artificiali (*Artificial Intelligence*, abbreviato in Ai) stanno cambiando il mondo: ogni attività umana, dalla medicina alla sicurezza nazionale stanno subendo profonde trasformazioni. I sistemi dotati di Ai non solo aiutano l'uomo ma in sempre maggiori situazioni danno luogo a sistemi, bot o robot, completamente autonomi. Di fronte a questo diluvio di intelligenza artificiale è urgente la questione etica. Più le Ai diventano universali più è necessario sviluppare un nuovo linguaggio universale che sappia gestire l'innovazione. Dobbiamo iniziare dallo sgomberare il campo da alcuni possibili fraintendimenti. Uno dei fraintendimenti più comuni quando si parla di etica e di pensare all'etica come una sorta di catena che debba porre dei limiti alla libertà. Parlare di etica della tecnologia significherebbe allora cercare di tracciare a priori dei limiti alla tecnologia. Ma è così? Per capire che significa etica della tecnologia dobbiamo ripercorrere un viaggio che iniziato tanto tempo fa.



Secondo gli antropologi 70mila anni fa la nostra specie, l'*homo sapiens*, si è spostata dall'Africa meridionale, la culla della nostra esistenza, colonizzando tutto il mondo. Abbiamo raggiunto ogni luogo in una maniera veramente unica, dando mostra di quella che è una nostra unicità come specie. Fino a quel momento ogni

specie di biologica abitava un clima particolare, il suo habitat. Se un mammut dalle steppe siberiane si è spostato in Africa e in India è perché un membro della sua discendenza ha subito una mutazione genetica perdendo il lungo pelo e divenendo capace di sopravvivere nei caldi climi meridionali. Quando l'uomo dall'Africa meridionale si è spostato in tutto il globo compresa la steppa siberiana non ha aspettato al contrario del mammut che gli nascesse un discendente dotato di un forte pelo. Non ha aspettato in altri termini che gli nascesse un *Homo sapiens hipster*. L'uomo si è vestito della pelliccia del mammut. In altri termini quello che in tutte le altre specie è fornito dal codice genetico per noi è mutato dall'artefatto tecnologico. Le capacità che hanno gli altri animali gli sono date attraverso competenze genetiche e possono cambiare solo se muta il loro Dna. Noi no. Cooperiamo tra noi, trasmettiamo informazioni sul mondo e educiamo le generazioni seguenti a fare cose grazie gli artefatti tecnologici. Mentre un delfino sa nuotare grazie al Dna, l'uomo è diverso. L'uomo cambia se stesso e il mondo che abita tramite gli artefatti tecnologici. La tecnologia è il luogo dove noi troviamo condensato tutto questo. Con la tecnologia noi cambiamo il mondo e noi stessi per abitare il mondo. L'etica della tecnologia non è altro che la costituzione naturale dell'antefatto tecnologico. Le intelligenze artificiali sono artefatti tecnologici. Ma differenti da tutti gli artefatti prodotti fino ad oggi. Tutti gli strumenti che abbiamo prodotto consentono all'uomo di svolgere

più efficace dei compiti precisi. Le Ai, tanto nei bot quanto nei robot, superano il concetto di artefatto e di macchina che conoscevamo fino adesso. Tutti i meccanismi automatici che abbiamo costruito durante la rivoluzione industriale sono stati costruiti pensando a quale sarebbe stato il loro scopo. Facevano quello per cui erano progettati e basta. Oggi le Ai non sono progettate così.

Non sono software programmati ma sistemi addestrati. Si supera il modello classico *if this then that* in cui un ingegnere del software prevedeva prima tutte le possibili occorrenze. Le Ai rispondono in maniera autonoma a un problema che gli viene posto. Questi artefatti sono una nuova specie nelle macchine. Delle *machine sapiens*.

Oggi il mondo non è più abitato solo dall'*homo sapiens* ma anche da *machine sapiens*. Se la macchina è autonoma chi risponde delle sue decisioni? Chi l'ha progettata? Chi la sta utilizzando? Chi la vende? Chi l'ha comprata? Allo stato attuale un'intelligenza artificiale media è in grado di fare una diagnosi medica migliore di un medico medio. Siamo pronti a delegare alle macchine tutta questa capacità decisionale? Per poter rispondere a questa domanda dobbiamo chiarire una questione di fondo: può un'intelligenza artificiale fare una scelta perfetta?

I *data scientist* ci dicono che il problema è alla qualità e alla quantità dei dati. Quando avremo un *database* perfetto su cui far girare i nostri servizi di Ai la macchina farà scelte perfette.

Ma è così? Già in passato abbiamo avuto questa impressione. Laplace sosteneva che qualora noi avessimo conosciuto la posizione in un istante di tutte le particelle che contengono l'universo saremmo stati in grado di predire tutto il futuro e di conoscere tutto il passato dell'universo. Era il famoso demone di Laplace.

Oggi la questione è applicata all'intelligenza artificiale e alle sue scelte basate sui dati. Che cosa sono i dati su cui decidono le intelligenze artificiali? Brevemente possiamo dire che i dati altro non sono che una mappa del mondo. Tutto ciò che esiste nel mondo viene mappato registrato e messo all'interno di un *database* che ne costituisce una mappa. Ma può una mappa essere la copia esatta del mondo? Lasciamo stare la domanda filosofica su questa possibilità e affrontiamola da un punto di vista operativo.

Se fossimo in grado di creare una mappa che è l'esatta copia della realtà, includendo al suo interno tutto, compresi i passanti, le foglie degli alberi, e così via, dovremmo riconoscere che la mappa che abbiamo creato è inutile. Questa infatti sarebbe complessa come la realtà, troppo complessa per prendere le decisioni e quindi inutile.

Ci troveremo cioè di fronte al noto paradosso raccontato da Jorge Luis Borges in un frammento *Del rigore della scienza*, l'ultimo di *Storia universale dell'infamia* pubblicato per la prima volta nel 1935. Come sua abitudine, l'autore argentino attribuisce la citazione a un libro che in realtà non esiste: «In quell'Impero, l'Arte della Cartografia giunse a una tal Perfezione che la Mappa di una sola Provincia occupava tutta una Città, e la mappa dell'impero tutta una Provincia. Col tempo, queste Mappe smisurate non bastarono più. I Collegi dei Cartografi fecero una Mappa dell'Impero che aveva l'Immensità dell'Impero e coincideva perfettamente con esso. Ma le Generazioni Seguenti, meno portate allo Studio della cartografia, pensarono che questa Mappa enorme era inutile e non senza Empietà la abbandonarono all'Inclemenza del Sole e degl'Inverni. Nei deserti dell'Ovest rimangono lacerate Rovine della Mappa, abitate da Animali e Mendichi; in tutto il Paese non c'è altra reliquia delle Discipline Geografiche. (Suárez Miranda, *Viajes de varones*

I dati sono una mappa della realtà, rappresentano una riduzione della realtà e per questo sono utili per prendere decisioni. Inoltre, le Ai lavorano sui database e sui sensori. Ma anche i sensori non leggono tutta la realtà: ne prendono solo una parte trasformandola in dati. Eccoci al punto chiave della questione.

Siccome le intelligenze artificiali fondano loro decisioni sui dati e poiché questi non sono una copia perfetta della realtà non è pensabile a priori che la macchina dotata di intelligenza artificiale possa fare una scelta priva di errori. La macchina *sapiens* sarà sempre e costitutivamente fallibile. Le Ai hanno costitutivamente bisogno di un'etica. Poiché le intelligenze artificiali possono sbagliare bisogna capire come gestire allora questo sbaglio. La questione etica è fondamentale, importantissima e urgente. Bisogna trovare un sistema etico condiviso perché l'utilizzo di questi sistemi non produca ingiustizie, non danneggi le persone e non crei dei forti disequilibri globali.

Quali sono le direttrici etiche che ci possono guidare nella realizzazione di questo nuovo linguaggio umano destinato a mettere in contatto l'*homo sapiens* con la *machina sapiens*? La storia dell'etica ci aiuta in questa ricerca.

La prima direttrice è quella che potremmo definire *Fear of Uncertain*. Qualsiasi scelta noi facciamo sappiamo che avrà delle conseguenze. Tutti possiamo scegliere liberamente ma quello che accade una volta scelto non sempre dipende da noi. Ogni scelta libera e consapevole porta con sé un orizzonte d'incertezza. Uno dei paradigmi etici chiave è la gestione dell'incertezza. Questo è il primo *driver* etico: essere coscienti che le scelte messe in atto possono produrre anche effetti non desiderati e gestire questo rischio.

Una seconda direttrice molto importante da considerare è la tensione tra *Equality and Pursuit of Happiness*. Tutte le più sanguinose guerre che abbiamo conosciuto tra Otto e Novecento si sono combattute per ottenere l'uguaglianza di tutti gli uomini. Di fatto l'uso di tali tecnologie rischia di produrre nuove disuguaglianze. L'etica per le Ai deve tutelare tutto questo. È la dignità umana il valore etico, non il valore dati. Inoltre uno Stato ha la sua legittimità se consente all'individuo di realizzare la propria ricerca della felicità. Queste nuove tecnologie con loro possibilità di profilazione, con loro possibilità di predizione del comportamento degli esseri umani possono di fatto rendere assai difficile la situazione di un'esistenza individuale libera. Non bisogna solo guardare il bene e il male che può scaturire per il singolo (*fear of uncertain*) ma la società nel suo complesso: bisogna tutelare l'uguaglianza degli individui e la possibilità di ciascuno di poter ricercare la propria felicità?

Infine, dobbiamo essere consapevoli di una verità di fondo. L'etica da sola è fragile. Così come la dignità umana è stata calpestata dai regimi totalitari del Novecento perché non era tutelata da nessun diritto, l'etica delle Ai rischia di essere inefficace se non diverrà *policies* vincolanti che proteggano il singolo e la convivenza sociale.

L'esistenza di macchine *sapiens* chiede di mettere in piedi un nuovo linguaggio universale che sappia tradurre queste direttrici etiche in delle direttive eseguibile dalla macchina. Ma come fare questo? Il mondo nell'epoca del *Digital Age* è regolato degli algoritmi. Più di qualcuno parla di una algo-crazia. Per evitare che ci sia questo dominio dell'algoritmo anche grazie alle Ai dobbiamo iniziare a sviluppare questo linguaggio comune dell'algor-etica.

Per poter sviluppare un algor-etica dobbiamo chiarire in che senso si

L'etica invece parla di valore morale. Dobbiamo stabilire un linguaggio che sappia tradurre il valore morale in un qualcosa di computabile per la macchina. La percezione del valore etico è una capacità puramente umana. La capacità di lavorare dei valori numerici è invece l'abilità della macchina. L'algor-etica nasce se siamo in grado di trasformare in qualcosa di computabile il valore morale. Ma nella relazione tra uomo e macchina il vero conoscitore e portatore di valore è la parte umana. La dignità umana e i diritti umani ci dicono che è l'uomo da proteggere nella relazione tra uomo in macchina. Questa evidenza ci fornisce l'imperativo etico fondamentale per la macchina *sapiens*: dubita di te stessa. Dobbiamo mettere in grado la macchina di avere un certo senso di incertezza. Tutte le volte che la macchina non sa se sta tutelando con certezza il valore umano deve richiedere l'azione dell'uomo. Questa direttiva fondamentale si ottiene introducendo dei paradigmi statistici all'interno delle Ai.

Tentativi di questo tipo sono portati avanti da Google e Uber con delle librerie statistiche speciali. Deve essere questa capacità di incertezza il cuore del decidere della macchina. Se la macchina ogni volta che si trova in una condizione di incertezza chiede all'uomo allora quello che stiamo realizzando è una intelligenza artificiale che pone l'umano al centro o come si suole dire tra i tecnici uno *human-centered design*. La norma fondamentale è quella che costruisce tutte le Ai in una maniera *human-centered*.

A partire da questa grammatica di base possiamo sviluppare un nuovo linguaggio universale: l'algor-etica. Questo avrà una sua sintassi e svilupperà una sua letteratura. Non è questo il luogo né il momento per dire ogni cosa esprimibile con questa lingua però ci sembra di dover almeno fornire qualche esempio che ne riveli le potenzialità.

Anticipation; quando due umani lavorano assieme, l'uno riesce ad anticipare e ad assecondare le azioni dell'altro intuendone le intenzioni. Questa competenza è alla base della duttilità che caratterizza la nostra specie: fin dai tempi antichi ha permesso all'uomo di organizzarsi. In un ambiente misto, anche le Ai devono essere in grado di intuire cosa gli uomini vogliono fare, e devono assecondare le loro intenzioni cooperando: la macchina deve adattarsi all'uomo, non viceversa.

Transparency; i robot funzionano comunemente secondo algoritmi di ottimizzazione: l'uso energetico dei loro servomotori, le traiettorie cinematiche e le velocità operative sono calcolate per essere il più possibile efficienti nel raggiungimento del loro scopo.

Affinché l'uomo possa vivere assieme alla macchina, l'agire di quest'ultima dovrà essere intellegibile. L'obiettivo principale del robot non dev'essere l'ottimizzazione delle proprie azioni, bensì rendere il proprio agire comprensibile e intuibile per l'uomo.

Customization; un robot, attraverso la Ai, si relaziona all'ambiente aggiustando il proprio comportamento. Lì dove uomo e macchina convivono, il robot deve essere in grado di adattarsi anche alla personalità dell'umano con cui coopera. *L'homo sapiens* è un essere emotivo; la macchina *sapiens* deve riconoscere e rispettare questa caratteristica unica e peculiare del suo partner di lavoro. *Adequation*; gli algoritmi di un robot ne determinano le linee di condotta. In un ambiente condiviso, il robot deve saper adeguare i propri fini osservando la persona e comprendendo così qual è l'obiettivo pertinente in ogni specifica situazione. La macchina deve, in altri termini, acquisire una "umiltà artificiale" per assegnare una priorità operativa alle persone presenti, e non al raggiungimento di un fine predeterminato. Nell'epoca delle Ai, questi quattro parametri sono un esempio di come tutelare la

Il problema è innanzitutto filosofico ed epistemologico. “Funzionano” secondo schemi che connettono dati. Che tipo di conoscenza è questa? Che valore ha? Come va trattata e considerata? Insomma, la domanda prima che tecnologica è etica e filosofica: nella misura in cui vogliamo affidare competenze umane, di comprensione, di giudizio e di autonomia di azione a dei sistemi software di Ai dobbiamo capire il valore, in termini di conoscenza e capacità di azione, di questi sistemi che pretendono di essere intelligenti e cognitivi.

Oggi le Ai sono sviluppate o in una modalità *market-driven* o *state-driven*. Dobbiamo pensare altre modalità. Per esempio sviluppando algoritmi di verifica indipendenti che sappiano certificare queste quattro capacità delle macchine. Oppure è possibile ipotizzare enti terzi indipendenti, che attraverso la scrittura di algoritmi dedicati siano in grado di valutare l'idoneità delle Ai alla convivenza con l'uomo. Solo rispettando queste indicazioni l'innovazione potrà essere guidata verso un autentico sviluppo umano.

di Paolo Benanti