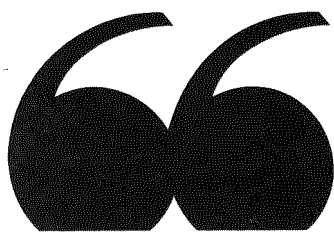


TEMPI PRESENTI

* I dispositivi possono decidere in modo autonomo o va esercitato un controllo sul piano etico e politico?

Le qualità morali delle macchine

Una riflessione a partire da due recenti volumi di ricercatori italiani



«Intelligenza artificiale», a cura di Stefano Quintarelli, per Bollati Boringhieri e «Etica delle macchine», di Guglielmo Tamburrini, per Carocci

TERESA NUMERICO

■ ■ ■ Tutti parlano di intelligenza artificiale come se producesse una trasformazione inevitabile e un progresso indiscutibile. I due volumi di cui discutiamo in questa pagina offrono, invece, una prospettiva originale perché si occupano delle sue conseguenze sociali e politiche e propongono un confronto sull'assunzione di responsabilità e la

regolamentazione dell'esecuzione di attività automatiche per la presa di decisioni che si presumono «intelligenti».

Il primo è *Intelligenza artificiale* a cura di Stefano Quintarelli (Bollati Boringhieri, pp. 138, euro 16), una raccolta di articoli che affrontano il problema da vari punti di vista. I saggi (firmati collettivamente) analizzano il futuro rispetto alla politica del lavoro, alle guerre digitali, all'etica, alla necessaria regolamentazione dell'uso di questi strumenti. Il concetto stesso di efficienza e ottimizzazione attribuito alle macchine deve essere precisato in relazione ai criteri che vengono adottati per misurarle. Sebbene si ritenga che l'intelligenza artificiale introduca parametri oggettivi di misurazione in tanti contesti nei quali regna l'incertezza, nessuno si prende la briga di sottoporre a valutazione queste tecnologie per controllare se davvero portino solo i benefici sperati.

IL VANTAGGIO CENTRALE attribuito agli algoritmi di *machine learning* - al cuore dell'intelligenza artificiale - è la capacità di fare previsioni sul futuro a partire dai dati di eventi passati. Ma una previsione resta solo una stima di quello che accadrà, non la certezza del futuro. Se si ipotizza che all'ottanta per cento domani pioverà, quella previsione è vera anche se non piove, perché verificherebbe l'ipotesi che al venti per cento non avrebbe piovuto.

Non è, quindi, un malfunzionamento. In certi casi, però, la previsione genera una nuova rappresentazione dei fenomeni che può avere impatto sul futuro che dovrebbe limitarsi a indo-

vinare. Mentre nel caso della pioggia possiamo controllare l'affermazione rispetto alla percezione dei fenomeni atmosferici che avvengono in città, non possiamo fare altrettanto in altri casi che poi hanno influenza su decisioni, magari anche delicate, o rilevanti socialmente come la scelta di un candidato per un posto di lavoro, la valutazione del rischio di reiterazione di un reato da parte di un condannato, l'entità di un premio assicurativo, il diritto a prestazioni di welfare o la concessione di un mutuo. In queste circostanze le conseguenze delle ipotesi predisposte sul futuro intervengono a modificarlo e non è più consentito controllarne la veridicità.

COME SUGGERIVA già il teorema sulla scelta sociale dimostrato da Kenneth Arrow nel 1950, non è possibile definire in maniera formale un criterio di ottimizzazione razionale univoca per risolvere una volta per tutte problemi multicriterio, nei quali, cioè, si tratti di mediare nella scelta tra interessi contrapposti, ma tutti legittimi. Negli anni 2000 è stato dimostrato che questo risultato è valido anche per la progettazione di algoritmi per il supporto o l'automazione della decisione.

È UN ESERCIZIO della ragion pratica, che ha bisogno di un criterio di azione operativa, tale da intervenire su alternative incommensurabili tra loro. L'equità è quindi un prodotto di preferenze e valori, e non è razionalmente identificabile. Deve essere politicamente determinata.

Si pone quindi il problema della responsabilità delle decisioni. A chi devono essere attri-

buite, dal momento che non sono valide in assoluto? Il capitolo sull'etica prende una posizione molto netta e condivisibile. Quando il sistema algoritmico di apprendimento per rinforzo AlphaGo ha battuto il campione di Go, il gruppo di ricercatori della DeepMind che lo aveva programmato a sé è attribuito il merito del risultato. Quindi se degli algoritmi per la presa di decisione causassero danni, l'assunzione di responsabilità dovrebbe essere dei progettisti e non degli utenti finali che non hanno possibilità di incidere sul loro funzionamento. La responsabilità non può essere data nemmeno ai dispositivi stessi, perché non sono agenti consapevoli o capaci di autoriflessione per valutare il loro comportamento.

LA PROPOSTA DEL GRUPPO di Quintarelli è considerare l'intelligenza artificiale come l'industria farmaceutica: sappiamo che i farmaci possono avere effetti collaterali gravi e condurre persino a morte se non sono dosati adeguatamente, ma di solito sono efficaci per risolvere il problema. La casistica è regolamentata: un farmaco non può essere usato off-label, pena l'incertezza del suo funzionamento. Proseguendo nella metafora dei farmaci, sappiamo che non tutto si cura e che le medicine servono solo se indispensabili, quindi è necessario valutare bene - da una prospettiva politica - per cosa sia davvero utile e benefico usare agenti artificiali e dove invece lo stato di salute del sistema consenta agli esseri umani di decidere autonomamente.

La discussione sulla responsabilità delle macchine che aumenta progressivamente in relazione alla diminuzione del

controllo umano sul loro agire è al centro anche del volume di Guglielmo Tamburrini, *Etica delle macchine* (Carocci, pp. 151, euro 14). È un testo meno divulgativo e più scientifico che si concentra principalmente sull'analisi delle questioni etiche aperte dai veicoli a guida autonoma e dalle armi autonome. Analizzando i veicoli autonomi Tamburrini illustra due diverse prospettive decisionali rispetto alla valutazione morale in caso di collisione inevitabile a seconda se a prevalere sia l'etica dei doveri o quella delle conseguenze. L'etica dei doveri riguarda gli obblighi morali che abbiamo nel giudicare il valore delle azioni compiute ex ante; mentre l'etica delle conseguenze mira a valutare, non tanto la moralità dell'agire, ma le conseguenze moralmente buone o cattive di un'azione avvenuta.

Chi è giusto proteggere nelle collisioni? Al momento la macchina a guida autonoma non è in grado di valutare con molto anti-

cipo gli ostacoli che non riuscirebbe a evitare, ma sebbene il problema posto non sia concreto per l'imaturità della tecnologia, ci troviamo comunque di fronte a dei dilemmi morali di difficile soluzione. La macchina dovrebbe privilegiare il figlio dell'utente o un passante? La difficoltà nasce perché queste decisioni non hanno solo un carattere individuale, ma presentano una valenza collettiva. Chi userebbe una macchina che in caso di pericolo non lo proteggesse? Ma chi consentirebbe l'uso di macchine che in una collisione inevitabile proteggano solo il passeggero a bordo e non il passante, o scelgano di investire il ciclista con casco per limitare il danno della collisione e tutelino quello senza?

Chi è responsabile dei comportamenti dei veicoli? L'attribuzione della responsabilità è ancora più controversa nel caso di armi autonome che per loro costru-

zione possono essere offensive e, quindi, produrre conseguenze dannose. Che succede se l'arma non è in grado di riconoscere un pulmino di studenti e bombardare i danni collaterali? Di chi sarebbe l'errore? La mancanza di responsabilità della macchina renderebbe un crimine di guerra impossibile da perseguire.

LA SOLUZIONE è, per Tamburrini, avere un controllo umano significativo sull'arma, a cui attribuire la responsabilità in caso di conseguenze negative. Ma sia l'autonomia dell'arma, sia il livello di controllo umano significativo sono difficili da precisare razionalmente. Sarebbe necessario allora compiere scelte politiche internazionali per definire la nozione di controllo e quella di autonomia. Anche in questo lavoro si dimostra che la tecnologia non ha in sé le risposte etiche, giuridiche e politiche del suo utilizzo. Ha bisogno di meccanismi di re-

golazione e di individuazione esterna di una soggettività a cui attribuire la facoltà di giustificare l'azione.

I due testi suggeriscono che non sia possibile nessun automatismo nel considerare responsabilità, conseguenze, valutazioni dell'operato delle macchine. La tecnologia, infatti, è intessuta di valori poiché, diversamente dalla scienza che vuole conoscere come stanno le cose nel mondo, si propone come un sistema a sua volta normativo e regolativo dei fenomeni sui quali interviene. Per questo abbiamo bisogno di decidere politicamente - come per i farmaci - a cosa serve e quali siano i suoi legittimi obiettivi. Non possiamo dare per scontato che tutto quello che può essere automatizzato debba esserlo. Non possiamo, cioè, ricadere come società nel pregiudizio dell'automazione che ritiene che ogni compito sia eseguito meglio da una macchina che da un essere umano.

Non dobbiamo ricadere nel pregiudizio dell'automazione che ritiene che l'uomo possa essere rimpiazzato per ogni cosa



«Chain», un'installazione di Cyril Lancelin

