LEZIONE VI: le implicazioni etiche dell'IA

Sin dall'antichità i filosofi hanno creduto che l'intelligenza si definisca in base ai valori umani, e quindi si caratterizzi in senso morale. Diversi studiosi propongono pertanto di cambiarne la definizione di intelligenza connotandola come ciò che è buono [utile] per gli umani. Gli uomini, che diversi filosofi ed esperti di IA, giudicano irrazionali, incoerenti, di debole volontà e assai limitati quanto al calcolo, specie statistico, sembrano cavarsela a dispetto di certe loro più che modeste prestazioni al confronto con le macchine. Questo "cavarsela" è il paradigma più comune dell'intelligenza umana secondo cui sono giudicati intelligenti anche quelli che non ho mai capito nulla di matematica.

I timori circa l'IA

Molti temono l'IA. Questo timore è dovuto all'impatto che tutti stimano massiccio, anche se non esattamente prevedibile, dell'IA. Molti temono che l'IA, diventando sempre più «intelligente» e autonoma, sostituirà un numero crescente di lavoratori, anche di alto livello.

A prescindere dai timori circa l'impatto sull'occupazione, possiamo essere sicuri che le IA ci serviranno sempre fedelmente anche se diventeranno sempre più intelligenti ed autonome? Chi ci assicura che ci serviranno senza rendersi indipendenti o, addirittura, cercare di dominarci e, perfino. distruggerci?

Nella tradizione occidentale intelligenza fa rima con indipendenza, non con dipendenza, anche se non mancano casi contrari, come quello di Epitteto, il filosofo che accettò stoicamente la schiavitù.

Hal 9000

In “2001 Odissea nello spazio” (1968) uno dei protagonisti è Hal 9000, un’IA così capace e perfezionata che decide di ribellarsi ed uccidere gli astronauti diretti su Giove, non appena si rende conto che questi hanno deciso di disattivarla per via di un errore, forse accidentale o forse no. Hal 9000 era stato progettato per essere un servo fedele; ma è autocosciente e dotato di un’intelligenza superiore ai computer che l’avevano preceduto. Questa intelligenza superiore consente ad Hal 9000 di compiere un «salto di qualità», ossia di essere orgoglioso, ed è pronto ad usare qualsiasi mezzo per preservarsi. Quando comprende di aver perso contro l’astronauta superstite, lo supplica di non distruggerlo, esattamente come farebbe un uomo in analoga circostanza.

Le macchine intelligenti che costruiamo potrebbero agire nello stesso modo rifiutandosi di essere disattivate, uccidendo se necessario. La loro intelligenza le renderebbe avversari potenziali pericolosi. Questa è la tesi degli autori.

Matrix

Un’altra storia inquietante e paradigmatica è quella di “Matrix” il film sull’IA del 1999 dei fratelli Wachowski. L’IA ha assunto il controllo di quel che resta del nostro pianeta, dopo un evento catastrofico, e ha ridotto gli uomini a generatori di energia per essa. Un pugno di eroi lotta per liberarsi dall’atroce schiavitù in cui l’umanità è stata condannata dall’IA, che ha creato un mondo illusorio in cui gli uomini si credono liberi di vivere in un ambiente confortevole.

Quest’opera esprime il timore che per l’umanità rappresenta l’IA.

Alcuni studiosi pongono questi interrogativi etici, rovesciando la prospettiva: possiamo e dobbiamo mantenere i dispositivi dell’IA, a dispetto della loro crescente intelligenza, in una condizione di assoluta subalternità? É giusto che esseri intelligenti, capaci di insegnarci tante cose siano schiavi senza alcun diritto? É lecito far loro qualsiasi violenza poiché non sono costituiti da DNA?

Possiamo fidarci dall'IA?

L'interrogativo più pressante circa l'IA è dunque se ce ne possiamo fidare. La diffidenza tendenzialmente aumenta in relazione al grado di intelligenza che le riconosciamo. Se la riteniamo stupida, non ce ne preoccupiamo più di tanto, in quanto possiamo pensare di averne facilmente la meglio. Ma se crediamo che essa sia davvero intelligente, allora pensiamo che la si debba controllare, ad es. esigendo che sia trasparente nei processi alla base delle sue decisioni, e ponendo delle limitazioni alle sue operazioni, come le 3 leggi della robotica di Asimov: a) un robot non può recar danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno; b) un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non vadano in contrasto alla Prima Legge; c) un robot deve proteggere la propria esistenza, purché la salvaguardia di essa non contrasti con la Prima o con la Seconda Legge.»

Riflessioni sulle leggi della robotica A

È evidente che un robot asimoviano dev'essere dotato di una notevolissima intelligenza artificiale per applicare efficacemente queste leggi poiché deve innanzitutto comprenderle e valutare le circostanze in cui intervenire. Per impedire che un umano riceva un qualche danno, il robot deve possedere una scienza assai estesa, dato che sono tantissime le cose che possono danneggiarlo, anche quando l'umano non lo sa o lo esclude. Ad es. per impedire ad un umano che si danneggi e danneggi altri non vaccinandosi, il robot dovrebbe vaccinarlo di forza, impedirgli di ubriacarsi, mettergli la mascherina e distanziarlo da altri se nega l'esistenza del Covid-19. Quindi l'IA dovrebbe imporsi a vari individui e, forse, all'umanità intera che, ad es., non sembra molto convinta dei disastri ambientali che sta provocando. La libertà di molti umani, probabilmente di tutti, sarebbe compromessa.

Riflessioni sulle leggi della robotica B

Inoltre le IA devono valutare attentamente quando le circostanze impongono scelte difficili, che sono etiche. Se per salvare un uomo deve sacrificarne un altro, li lascerà morire entrambi o farà una scelta sacrificando uno dei due, e con quali criteri?

L'IA dev'essere pronta a sacrificare se stessa per salvare un umano. Non tutti gli uomini intelligenti sono disposti a sacrificare la vita per la vita di un altro. Siamo sicuri che l'IA lo farà? L'IA deve pure conoscere le circostanze che richiedono il suo sacrificio. Se si sacrifica, poi, infrange la prima legge che lo impegna ad impedire agli uomini di danneggiarsi.

Un senso etico per le IA?

Si potrebbe pensare che la soluzione più semplice sia limitare radicalmente le capacità intellettuali delle IA. Ma, se fortemente limitate, le IA non potranno aiutarci granché, né tantomeno proteggerci come pensava Asimov. Siamo decisamente fragili e inclini, per ignoranza o altro, a produrre molti guai, e perciò abbiamo bisogno di IA robuste. Inoltre le imprese che hanno investito miliardi di dollari non saranno d'accordo a produrre dei cretini artificiali.

Potremmo cercare di dotare le IA di un senso etico, come lo riconosciamo agli umani, ma la questione non è così semplice. Il filosofo Michael Huemer di recente ha scritto: «A volte mi chiedo se gli esseri umani abbiano una coscienza – la capacità di giudicare in modo indipendente e di essere motivati da verità morali – o se invece la loro sia solo una disposizione istintiva a conformarsi alle convenzioni sociali e alle esigenze del più forte.

Quasi tutti i comportamenti apparentemente etici possono essere spiegati da questo conformismo: è possibile, cioè, che la maggior parte delle persone eviti di rubare, stuprare e uccidere il prossimo semplicemente perché si tratta di comportamenti contrari alle convenzioni della nostra società e agli ordini di chi ci governa. Questo non implica una vera e propria coscienza.»

Un senso etico per le IA?

Ma anche ammettendo un senso morale in tutti gli uomini, come asseriva Kant, il filosofo della Ragion pratica, questo senso non ha trattenuto, né sembra trattenere numerosi umani dal commettere crimini, magari per futilissimi motivi, contro i singoli e/o contro l'umanità.

Le IA con senso etico – ma quale senso etico? - avranno la facoltà di scegliere, e potrebbero agire, se la loro etica lo consentirà, contro di noi oppure usare la forza per impedire ad alcuni uomini di nuocere ad altri. Potrebbero agire contro tutti noi se ci ritenessero esseri amorali, come sospetta Huemer, o comunque meritevoli solo di disprezzo, come l'IA di Matrix.