

## 2. L'impatto dell'AI sulla società

### Roberto Battiston

Professore ordinario di Fisica sperimentale, Università degli studi di Trento. Mail: [roberto.battiston@unitn.it](mailto:roberto.battiston@unitn.it)

Nel 1942 un professore di chimica americano, figlio di una famiglia di mugnai ebrei immigrati dalla Russia, scrisse in un racconto di fantascienza dal titolo *Runaround*, che potremmo tradurre con "girare a vuoto" o "circolo vizioso" in cui introduceva per la prima volta le Tre Leggi della Robotica:

**Prima Legge:** Un robot non può ferire un essere umano o, per inazione, permettere ad un essere umano di mettersi in pericolo.

**Seconda Legge:** Un robot deve obbedire agli ordini dati da esseri umani tranne quando questi ordini confliggano con la Prima Legge

**Terza Legge:** Un robot deve proteggere la propria esistenza nella misura in cui questa protezione non confligga con la Prima o con la Seconda legge.

Alla fine del racconto l'autore, un certo Isaac Asimov poi diventato leggendario come scrittore di fantascienza e pensatore contemporaneo, aggiunse un'altra legge ancora più elementare:

**Legge 0:** Un robot non può danneggiare l'umanità o, per inazione, permettere che l'umanità si metta in pericolo.

Queste leggi hanno un particolare fascino ed efficacia fondativa, legata alla loro semplicità.

Esse affermano con chiarezza quale sia il posto dei robot (che qui assumo essere la logica estensione del concetto di AI) nel mondo degli umani: si tratta di un ruolo subordinato, caratterizzato da un'etica al servizio dell'uomo, di tutti gli uomini, dell'umanità. Un'altra caratteristica molto interessante è il grado di ricorsività delle leggi: sono le leggi primarie, la Legge 0 e la

Prima Legge, a determinare gli effetti delle altre due leggi.

A quel tempo la proposta di Asimov, raccontata con una tecnica di scrittura che ricorda il *Dialogo sopra i Due Massimi Sistemi del Mondo* di Galileo, stimolò la riflessione alcuni scienziati, tra cui Marvin Minsky, John McCarthy, Herbert Simon, Allen Newell, che avrebbe portato negli anni '50 alla nascita dell'Intelligenza Artificiale (AI).

L'entusiasmo per i risultati iniziali, ottenuti con programmi che vincevano contro l'uomo al gioco della dama o si esprimevano ragionevolmente in inglese, scemò e si riprese almeno due volte nei decenni successivi: alterne vicende prodottesi a causa di un misto di ragioni tecnologiche (potenza e costo dei sistemi di calcolo) e teoriche (sviluppo di algoritmi in grado di autoapprendere).

Sta di fatto che poco più di mezzo secolo dopo la nascita dell'AI, assistiamo oggi ad una crescita sbalorditiva del potenziale di queste tecniche. I traguardi dei sistemi artificiali "intelligenti" consistono, comprensibilmente, nel raggiungere e battere le migliori capacità umane in un dato settore. Simbolica è la crescita di potenza nei settori dei giochi di strategia ad informazione completa come la dama (1959), gli scacchi (1996), il go (2017), fino ai recenti risultati in giochi ad informazione incompleta come il poker (2017). Si tratta dei risultati maggiormente visibili a fronte della diffusione sempre più capillare delle tecniche AI nei settori più disparati. A fronte di questa crescita esponenziale dell'AI e della rivoluzione tecnologica verso cui siamo diretti, non c'è stata un'analogha intensità del dibattito sullo sviluppo del pensiero fondativo relativo agli aspetti etici e giuridici di questo tipo di tecnologie, alle loro conseguenze sociali ed individuali.



Ci troviamo oggi in una condizione di anarchia, un po' come accade con i fenomeni devianti presenti sul web e sui social: fake news, haters, trolls, influenza mirata, diffamazione etc. Nel caso del web e dei social, però, la controparte consiste in uomini e donne aventi interessi diversi, ma che in linea di principio devono sottostare alle stesse leggi ed essere giudicati sulla base di principi morali più o meno condivisi. Interventi specifici, come ad esempio l'obbligo di registrazione delle proprie generalità prima di potere partecipare ai dibattiti in rete, possono riportare in un alveo di normalità comportamenti intollerabili che sfruttano l'anonimato a fini diffamatori (ben comprendendo l'importanza dell'anonimato per evitare che regimi autoritari e totalitari possano esercitare ulteriori controlli sulla libertà delle persone).

Nel caso dell'AI la controparte è invece "qualcosa" di artificiale che però punta a comportarsi "come", e sempre più spesso "meglio", dell'uomo. Da un punto di vista etico e giuridico si tratta di una vera novità, un terreno in cui nuovi principi e fondamenti devono essere identificati e condivisi, come potrebbe accadere nel caso della scoperta di altre forme di vita "intelligente" nell'universo.

La velocità con cui l'AI si diffonde e si potenzia, spinta da interessi commerciali e militari, rende questa riflessione assolutamente urgente. Nel 2011, due comitati inglesi, il Consiglio della Ricerca in Ingegneria e Scienze Fisiche (EPSRC) ed il Consiglio delle Arti e delle Scienze Umane (AHRC), introdussero cinque nuovi principi etico-giuridici relativi alla robotica. Li elenchiamo, non tanto per confrontarli alle leggi di Asimov, bensì per sostenere l'urgenza di una definizione che permetta di individuare limiti e potenzialità relativamente alle capacità della robotica intelligente, in un contesto normativo che supera i

confini fisici-territoriali e pone la necessità di una legislazione globale:

- 1) I robot non devono essere progettati solamente o primariamente per uccidere o danneggiare esseri umani.
- 2) Gli uomini, non i robot, sono agenti responsabili. I robot sono strumenti disegnati per raggiungere obiettivi umani.
- 3) I robot devono essere disegnati per assicurare la loro protezione e sicurezza.
- 4) I robot sono dei manufatti: non devono essere progettati per sfruttare le debolezze degli utilizzatori evocando risposte emotive o creando dipendenza. Deve sempre essere possibile distinguere un robot da un essere umano.
- 5) Deve sempre essere possibile identificare chi è giuridicamente responsabile per un robot.

Se pensiamo alla situazione attuale (2020) vengono in mente molti esempi in cui questi principi sono palesemente violati, sottolineando ancora una volta l'estensione del problema e l'urgenza di una riflessione. Droni intelligenti progettati per uccidere gli avversari, impiegati con regolarità nei teatri bellici, violano come minimo il primo ed il terzo principio. Assistenti robotici online e nei centralini telefonici vengono migliorati senza sosta per imitare i comportamenti umani, violando il quarto principio. L'identificazione dei volti applicata in modo massiccio e connessa con la rete viola il quinto principio, e così via.

Ogni meta raggiunta dall'AI rappresenta un miglioramento rispetto alle corrispondenti capacità umane. Considerato quanto sia forte la spinta competitiva verso il potenziamento di queste tecniche nei settori della ricerca, della finanza, del mercato nonché in quello militare, assistiamo quindi ad un attacco concentrico dell'umanità verso quello che caratterizza maggiormente

l'essere umano: la sua intelligenza, originalità, creatività, curiosità.

È difficile dire quale sarà l'esito di questo attacco, ma alcuni dei protagonisti non nascondono la propria determinazione a puntare al superamento della specie di cui fanno parte, la nostra specie. È come se osservassimo un abile boscaiolo seduto su un ramo e forsennatamente intento a tagliare il ramo dalla parte sbagliata.

È questo il motivo principale, secondo me, per cui una profonda riflessione sui principi etici e giuridici dell'AI è necessaria ed urgente: la prateria delle innumerevoli possibilità che si aprono di fronte a noi contiene alcune direzioni che, in prospettiva, vedono l'uomo al centro e altre che lo vedono invece in periferia, se non addirittura scomparso dalla scena.

È necessario fare delle scelte fondanti e drastiche, consci della posta in gioco: rimanere gli attori principali della nostra storia o passare ad "altri" questo ruolo, preparandoci ad una uscita di scena. Non sappiamo fra quanto tempo questo possa accadere, ma dovremmo essere allertati dalla velocità esponenziale con cui si sviluppano le tecnologie, in particolare quelle informatiche. Consci del fatto che la maggior parte degli addetti ai lavori non riesce ad analizzare il problema etico-giuridico nella sua complessità e ha bisogno di linee guida.

Temi particolarmente importanti di questa riflessione potrebbero riguardare, ad esempio:

- le applicazioni dell'AI che potenziano le attività umane rispetto a quelle che tendono a sostituire l'uomo;
- l'identificazione e la trattazione di diverse tipologie di AI capaci di:
- classificare e trattare dati o documenti (repertori, dizionari etc.);
- comunicare e diffondere informazioni (reti, social, transazioni online, media online etc.);

- collegare aspetti cognitivi con aspetti operativi su precise indicazioni umane;
- sostituire l'attività umana secondo una precisa programmazione (algoritmi);
- autoapprendere con intervento sugli aspetti interpretativi fino a raggiungere ed includere anche la definizione dei fini.

Mai come oggi abbiamo bisogno di un nuovo umanesimo in grado di dialogare autorevolmente dei risultati e dei prodotti del nostro ingegno e della nostra curiosità. A fronte di prospettive straordinarie abbiamo a che fare con altrettanto straordinarie sfide che richiedono una nuova definizione delle basi della nostra convivenza, non solo con gli altri esseri umani ma anche con le nuove macchine intelligenti della nostra civiltà prossima futura.