Etica dell'intelligenza artificiale

Adriano Fabris

1. *Introduzione.* Nell'ambito della nostra tavola rotonda (su *Sviluppi e culture dell'intelligenza artificiale*) il mio contributo s'incentra sull'aspetto etico della questione. L'approccio etico – la *motivazione* etica – come sappiamo è essenziale nel contesto educativo. Senza questa motivazione non ci sarebbe insegnamento: né in presenza, né a distanza. "Etica" significa far emergere i criteri e i principi condivisi delle nostre scelte consapevoli. L'insegnante, per esempio, ha il compito di richiamare anche questo e mostrare e fare in modo che certi criteri e principi condivisi, quelli relativi ai comportamenti relazionali di ciascuno secondo regole sociali partecipate, siano assunti e praticati.

Oggi però la situazione è ben particolare. Oggi anche l'educazione – in un mondo dominato da dispositivi dotati di intelligenza artificiale – deve tenere conto del fatto che viviamo in ambienti molteplici e paralleli. Non ci sono infatti solo quelli quotidiani, offline, ma anche quelli – molteplici – online. E con tutti dobbiamo essere in grado di confrontarci

Si tratta di ambienti governati da algoritmi. Si tratta di processi che rimandano a ciò che chiamiamo "intelligenza artificiale" (AI). Con essa, e con le sue applicazioni, dobbiamo fare i conti, anche nella nostra attività quotidiana: soprattutto perché, proprio grazie all'AI, certi dispositivi sembrano davvero in grado di "comunicare".

Tutti questi temi meritano di essere approfonditi. Lo farò, seppure nei limiti di tempo a me consentiti, partendo da un chiarimento di certi concetti: "AI", "robot", "comunicazione". Poi cercherò di approfondirne alcune caratteristiche importanti per la nostra interazione con essi. Infine cercherò di dare alcune indicazioni di fondo, etiche, al fine di governare tale interazione.

2. *Intelligenza artificiale e comunicazione*. Anzitutto dobbiamo chiarire che cos'è l'intelligenza artificiale. Fra le sue varie definizioni vi propongo quella elaborata dalla EU, a supporto del lavoro svoltosi tra il 2018 e il 2019 da un High Level Group (HLEG), che aveva l'obbiettivo di redigere delle "Linee guida etiche per un'Intelligenza Artificiale

degna di fiducia" (*Ethic Guidelines for Trustworthy Artificial Intelligence*). Proprio in relazione a questo progetto è stato elaborato e messo a disposizione, dallo stesso gruppo di esperti, un testo parallelo, in cui è stata fornita appunto una definizione di AI (*A Definition of AI: Main Capabilities and Disciplines*: consultabile al link https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top). Leggiamola:

Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).

Sembra forse una definizione che rimanda a qualcosa di complicato, di sofisticato? Prendiamo il nostro smartphone. Apriamo la app delle mappe. Accendiamo il microfono e dettiamo un indirizzo. L'assistente vocale ci risponde e ci guida alla nostra destinazione.

Ecco: i dispositivi tecnologici comunicano. Non solo servono per comunicare, non solo ampliano le nostre possibilità di comunicazione, non solo dischiudono nuovi ambienti in cui la nostra capacità comunicativa si può ulteriormente sviluppare, ma anch'essi, anche i dispositivi tecnologici, sono in grado a loro volta di comunicare.

Non è necessario fare riferimento a film recenti – come ad esempio *Her*, che narra la vicenda di un uomo, Theodore Twombly, che s'innamora della voce femminile del proprio assistente vocale, di nome Samantha, interagendo con questo programma, ovvero con "lei" – per sottolineare come la comunicazione tra un essere umano e una macchina, o un sistema, possa assumere le sembianze di un dialogo. Pare anzi che qui, almeno dal punto di vista dell'essere umano, questo dialogo riesca, e riesca in maniera soddisfacente. Da una ricerca condotta dall'agenzia pubblicitaria statunitense J. Walther Thompson già nell'aprile 2017 risultava che il 37% degli intervistati affermava di essersi a tal punto invaghito delle varie Cortana, Alexa o Siri – i nomi dati alle voci dei vari assistenti virtuali – da immaginare o desiderare che esse corrispondessero a persone reali.

3. *I robot antropomorfi.* Ma perché ciò può accadere? Un sistema dotato di intelligenza artificiale è in grado davvero di "comunicare"? Più concretamente, che cosa sono, o magari "chi" sono, i robot in grado di comunicare?

Quando si pensa a un robot, di solito viene in mente un dispositivo dalle sembianze per lo più umane, o comunque animali, dotato di un sistema e di programmi che lo rendono capace di muoversi nello spazio, a differenza del computer (che invece sta dove lo si appoggia), e dotato di qualcosa come una faccia. È così, d'altronde, che ce lo raffigurano i romanzi di fantascienza, i fumetti, i film. In realtà il termine oggi ha vari significati ed è funzionale a molteplici applicazioni. Esso indica un meccanismo inizialmente progettato per svolgere quei compiti che l'essere umano non è in grado di fare o che comporterebbero determinati pericoli. In quest'accezione il robot ha un impiego soprattutto in ambito industriale e militare, o viene usato in operazioni di salvataggio condotte in condizioni particolarmente critiche.

Ma quest'attenzione alla forma esteriore è solo uno degli aspetti di cui tenere conto. Per svolgere infatti le sue funzioni, ad esempio in contesti in cui è richiesta un'alta precisione o una grande velocità di elaborazione dei dati, il robot dev'essere dotato di programmi specifici. Si tratta di programmi che non solo analizzano e processano informazioni in modo metodico e automatizzato, ma che sono fatti in modo da elaborare le risposte fornite dai sensori esterni di cui il robot è dotato e di adattarle ai compiti a esso assegnati. C'è dunque, in questo caso, un'interazione con l'ambiente in cui il sistema automatizzato opera e la capacità per tale sistema di raggiungere un obbiettivo seguendo procedure diverse a seconda delle situazioni.

Tenendo conto di quest'ultimo punto, l'aspetto, cioè il sembiante antropomorfo del robot non è soltanto il frutto di un'idea romanzesca, ma è giustificato da una serie di motivi. Il robot è stato prodotto per affiancare l'essere umano nelle sue attività e pertanto, per assolvere ai suoi scopi, deve essere fatto in un certo modo: deve avere ad esempio la capacità di afferrare oggetti attraverso qualcosa che funzioni come una mano. Ben presto, però, da sistema di *affiancamento* esso si trasforma in un sistema di *accompagnamento* dell'essere umano stesso. Il robot, cioè, viene costruito non solo per coadiuvare e perfezionare l'essere umano in certe sue azioni (come le operazioni chirurgiche), ma

anche per *interagire* con lui o con lei, e per assisterli in certe loro esigenze. È il caso del robot badante o del robot cagnolino. Anche per questo, per facilitare un'interazione, la macchina deve avere una "faccia".

Tutto questo ha contribuito però a sviluppare un'altra idea: l'idea che il robot possa non solo affiancare e accompagnare, ma addirittura *sostituire* l'essere umano in certe sue attività, dato che in alcuni casi le può svolgere meglio. E questo ha provocato una serie di reazioni contrastanti. Da una parte si è verificato un rigetto psicologico degli esseri umani nei confronti dei robot, che sta alla base anche di quei timori che essi suscitano. Dall'altra parte si è invece consolidata una specie di attrazione nei confronti di tali entità, che ha prodotto una dinamica di vero e proprio rispecchiamento fra esseri umani e robot. Non solo i robot, cioè, sono fatti in modo da assomigliarci, provocando una serie di problemi, ma noi stessi possiamo essere indotti a prenderli a modello: come accade nella prospettiva del transumanismo e del post-umano.

Per chiarire meglio il primo punto, è utile fare riferimento alle ricerche di Mori Masahiro, note con il nome di teoria della "uncanny valley", ossia della "curva perturbante". Esse mostrano come la sensazione di familiarità che proviamo nei confronti di un robot antropomorfo cresca fino al punto in cui la sua troppa somiglianza con noi non ci provoca un rigetto emotivo, segnalato da una brusca flessione – la "perturbazione" cui fa riferimento la teoria – nella curva che rappresenta, in un ipotetico grafico, il nostro atteggiamento rispetto a tali entità artificiali. Ciò significa dunque che il robot, per avere a che fare con noi senza suscitare disagio, dev'essere riconoscibile come tale, deve cioè assomigliarci ma senza essere scambiato per uno di noi. Forse è questo il motivo per cui ai replicanti di *Blade Runner*, il film di Ridley Scott del 1982, viene data la caccia: perché sembrano, e forse sono, troppo umani.

Riguardo al secondo punto, invece, emerge una dinamica diversa. Io mi specchio nell'apparato che ho costruito, mi riconosco, magari, e mi contemplo. Ma su di esso posso anche proiettare gli aspetti migliori che mi sono propri. E dunque, vista la sua maggiore funzionalità (all'interno di una prospettiva in cui si assume implicitamente il principio etico dell'utile), il dispositivo artificiale, antropomorfo o meno, diventa il modello al quale devo aspirare. Si diffondono allora i

tentativi, compiuti da esseri umani, di ibridarsi con gli automi, o di uniformare i propri comportamenti a procedure standard.

In questo quadro non stupisce se un ruolo importante viene giocato dalla comunicazione. Lo abbiamo visto con l'esempio del nostro assistente vocale: anche i sistemi dotati d'intelligenza artificiale "comunicano". Con tali sistemi, addirittura, sembra che riusciamo a "dialogare". Ma che cosa vuol dire, propriamente, "comunicare" nel caso di un robot? In quali forme un sistema automatizzato è in grado d'intrattenersi con un essere umano, oppure con un altro sistema automatizzato? Per rispondere a queste domande dobbiamo prima di tutto approfondire un tema-chiave, che non può essere lasciato sullo sfondo: l'idea che il robot possieda una sorta di "autonomia".

4. *Autonomia, responsabilità, comunicazione.* Condizione infatti perché vi possa essere un robot in grado di comunicare e capace, anche per questo aspetto, di compiere determinate "scelte", è che esso sia considerato qualcosa di autonomo. Ma che cosa significa qui, propriamente, "autonomia"? Ho già parlato degli apparati programmati per "apprendere", modificando il proprio agire a seconda di come si presenta una determinate situazione. Il famoso caso di AlphaGo, il programma che è stato elaborato da Google per giocare le partite di questo antico gioco cinese, il Go, e che nel 2016 ha avuto ragione del campione in carica Lee Sedol per 4-1, è certamente l'esempio più noto di questa capacità che caratterizza un sistema dotato di "intelligenza artificiale".

In generale, tuttavia, per il robot la situazione è molto più complessa di quanto non lo sia per un programma. Nel robot il programma è "*embedded*", "incorporato". Nel suo operare esso è preso in un vero e proprio reticolo di azioni, retroazioni e interazioni, che solo in parte sono dipendenti dal suo funzionamento. Esse dipendono invece da azioni precedenti che hanno prodotto il robot e il suo agire. Ognuna di esse, a sua volta, si trova orientata e governata dai criteri e principî in base ai quali si realizza, e implica vari livelli di responsabilità. Posso provare a indicarli qui solo schematicamente.

Vi sono anzitutto alcuni principî generali di riferimento a partire dai quali progettista, costruttore e programmatore regolano la propria attività lavorativa e, più in generale, la loro stessa vita, e che incidono nelle scelte che portano alla creazione di una determinata entità

artificiale. Pensiamo per esempio al caso di un ingegnere pacifista, che certo non vorrebbe essere coinvolto nella costruzione di droni da combattimento. Vi sono poi i criteri che motivano tutti costoro a progettare, a costruire e a programmare proprio quella determinata macchina. Tali criteri debbono tener conto della portata e dei limiti dello sviluppo tecnologico di una data epoca. Essi, inoltre, sono vincolati alla struttura stessa del robot e alla portata della sua azione, come nel caso di quei valori "incorporati" nella macchina (o in un programma: ad esempio nel modo in cui funziona un Social). Infine i principî di cui parlo sono anche quelli che rendono possibile i programmi d'interazione di un agente artificiale con il suo ambiente e con gli altri esseri che vi si trovano: ciascuno dei quali può agire in modi diversi e sulla base di criteri differenti.

In questo quadro articolato parlare di "autonomia" della macchina assume un significato ben preciso. Potremmo dire che la macchina, nel nostro caso il robot, è certamente in grado di *auto-regolare* i propri processi, ma non è capace di *auto-regolamentarli*. In altre parole, la macchina non è in grado di "scegliere" i criteri e i principî in base ai quali essa viene a relazionarsi all'ambiente, alle altre macchine, agli esseri umani. Può solo adottarli. Può, in altre parole, seguire i criteri e i principî in base a cui è stata costruita. Può, anche, modificare il proprio operare a seguito di certi scenari che possono essere anticipatamente individuati. Ma, almeno allo stato attuale del suo sviluppo, non è capace d'intervenire sui principî base da cui dipendono la sua costruzione e il suo funzionamento, nonché le stesse modalità d'interazione con il suo ambiente.

L'autonomia, insomma, riguarda il modo in cui un robot attiva i suoi programmi, anticipa scenari possibili e risponde a essi. È qualcosa che riguarda le sue procedure e i modi in cui esse possono essere seguite. Si tratta dunque di un'autonomia "relativa": relativa ai criteri secondo cui il robot è stato costruito, al contesto specifico in cui opera, al quadro delle opzioni anticipabili, ai modi in cui sono state prefigurate le sue risposte a precise sollecitazioni ambientali, e alle regole che, in determinati casi, possono essere seguite per raggiungere gli obbiettivi prefissati.

A partire da questo chiarimento possiamo rispondere alla nostra domanda iniziale. Che cosa vuol dire, per un robot, "comunicare"? Vuol dire due cose. Allo stato attuale, il robot è capace di comunicare nel

significato dell'elaborazione e della trasmissione di dati. Ciò lo distingue dalle modalità comunicative, più estese, che sono proprie dell'essere umano, che mettono in gioco specifiche funzioni simboliche, metaforiche, interpretative. Al tempo stesso, però, il robot è stato sviluppato anche per elaborare e acquisire, come abbiamo visto, una capacità interattiva nei confronti delle affermazioni, delle risposte, delle emozioni umane. Questo permette a esso di esprimere una sorta di corrispondenza, una capacità di rispondere "a tono", che può essere scambiata per "empatia". Come accade per *Her*, la protagonista del film di cui ho parlato prima.

Da questo punto di vista, però, il lavoro comunicativo maggiore è svolto da noi. In altri termini, la differenza tra umano e non-umano è qualcosa che viene gestita dall'essere umano stesso. Noi ce ne possiamo far carico perché siamo in grado d'immaginare e, immaginando, d'interpretare situazioni differenti dalle nostre, e perciò di comunicare all'interno di esse, dando la parola anche a soggetti non-umani.

Ecco perché è certamente possibile dialogare con un robot. E può anche essere qualcosa di produttivo, nella misura in cui in questa relazione può verificarsi un reale scambio d'informazioni. Ma nel caso sia ricercata un'interazione più profonda, una vera e propria "fusione di orizzonti" (come avrebbe detto Gadamer), che riguardi gli stessi principî e criteri che stanno alla base delle posizioni dei soggetti coinvolti nel dialogo, allora quest'interazione può essere solamente a carico dell'essere umano. Nel momento in cui, poi, l'interazione si trasforma da cooperativa, come in questo caso, in competitiva, allora la potenza di calcolo e di anticipazione del sistema artificiale avrà probabilmente il sopravvento. Come nel caso di AlphaGo.

5. *Etica, comunicazione, intelligenza artificiale.* Riassumo. I dispositivi dotati di AI comunicano. La loro comunicazione, però, si svolge secondo caratteristiche ben precise. È trasmissione di dati; si svolge secondo programmi prestabiliti; è in grado di adattarsi all'interlocutore, di "imparare" da varie situazioni comunicative. Entro certi limiti un dialogo tra l'essere umano e un sistema automatico di comunicazione è possibile. Almeno allo stadio attuale dello sviluppo tecnologico. Almeno in un certo senso della parola "dialogo". Nel caso del robot questa capacità d'interazione, anche comunicativa, è resa possibile da una sorta di autonomia, cioè da quell'autonomia "relativa", come l'abbiamo

chiamata, che lo caratterizza. Ma tale condizione di autonomia è anche quella che rende possibile l'esercizio di azioni che possono essere riconosciute e qualificate, in una certa misura, come "etiche".

Con ciò siamo giunti al punto conclusivo del mio discorso. Che cosa significa "etica" in uno scenario in cui sono presenti anche entità dotate di AI? Che cosa cambia per questa disciplina rispetto al passato, in modo tale che – come dicevo all'inizio – è bello, interessante, sorprendente, riflettere su queste tematiche da un punto di vista filosofico?

Nel caso dei sistemi artificiali di comunicazione che ho preso in esame, ripeto, c'è la possibilità di un agire comunicativo per certi aspetti autonomo. Di conseguenza c'è la possibilità che l'etica riguardi non solo i nostri comportamenti, ma anche, sotto un certo rispetto, le azioni che sono proprie delle macchine. Questa è la novità nello scenario filosofico attuale.

In questo scenario, l'espressione "etica delle tecnologie dell'informazione e della comunicazione" (ICTs) (o anche, più nello specifico, "etica dell'AI") ha significato non solo nel caso di una riflessione sui comportamenti umani resi possibili da tali tecnologie (cioè nel senso oggettivo del genitivo), ma anche (nel senso soggettivo del genitivo) a proposito di quell'agire che viene compiuto da certi dispositivi mediante i quali comunichiamo, i quali interagiscono con il nostro comunicare e che, a loro volta, "comunicano" con un certo grado di autonomia. Tutto ciò comporta, indubbiamente, un allargamento dell'etica e nuovi problemi che questa disciplina deve oggi affrontare. Bisogna però stare attenti, di nuovo, a non proiettare caratteristiche umane sulle entità artificiali e ricordare il guadagno delle nostre analisi precedenti.

Qual è, infatti, la portata etica del robot? Il robot è chiamato a *seguire la procedura*, non già a *scegliere di seguirla*. Questo è il motivo per cui, più che un'etica relativa a ciò che è "buono" o "cattivo", più che una realizzazione del "bene" o del "male" mediante un'azione, nel caso di queste macchine bisogna parlare di azioni "corrette" o "scorrette", in un certo senso "giuste" o "sbagliate".

Proprio se teniamo ferme tali distinzioni, però, si presentano nuovi e specifici problemi. Esistono conflitti morali che riguardano l'agire dei robot: non solo nell'ambito di quei comportamenti che una procedura ha il compito di regolamentare, ma anche per quanto riguarda il rapporto fra questa stessa procedura e ciò che essa invece

non prevede, e che bisogna invece cercar di prevedere. Per esempio, un'automobile senza conducente deve sterzare per evitare d'investire alcuni pedoni, oppure no, se il cambio di traiettoria comporta un rischio per altri veicoli o per i suoi stessi passeggeri? Fino adesso, uno dei modi in cui si tenta di rispondere a questa e ad altre domande simili è quello che fa riferimento alle scelte della maggioranza: ma della maggioranza pur sempre degli esseri umani (si veda in proposito il famoso sito www.moralmachine.edu).

Emerge qui comunque, con chiarezza, la differenza fra l'ambito dell'etica umana e quello di una gestione meramente procedurale dei comportamenti che caratterizza l'agire delle macchine. L'essere umano vive la propria vocazione etica proprio mantenendosi su di un duplice piano: per un verso applicando le regole, per altro verso potendole anche mettere in questione, e deliberando magari di scegliere regole differenti. Le macchine, per ottenere lo stesso risultato, hanno bisogno invece di far riferimento a procedure diverse: a procedure che consentano di controllare l'eventualità che si delineino altri scenari, di calcolare ciò che in essi potrà verificarsi e, in tal modo, di evitare che siano prese decisioni arbitrarie.

La questione di fondo da un punto di vista etico, insomma, è che le macchine non sono in grado di andare con le loro azioni al di là di un orizzonte semplicemente procedurale. Ma l'etica non si risolve in una serie di procedure. Il problema è dunque quello di subordinare la regolamentazione procedurale a quei criteri etici che solo l'essere umano è in grado di elaborare e di cui solo l'essere umano è in grado di richiedere l'applicazione. Oggi rischiamo, invece, di subordinare i nostri comportamenti, sempre di più, a quelli di un programma o di una macchina. Questo è il vero pericolo. L'etica, proprio l'etica, è in grado di avvertirci di tutto ciò e d'indirizzarci su un'altra strada.