



Implicazioni etiche della quarta rivoluzione industriale

Articolo di Stefano Zamagni, Presidente Commissione Scientifica AICCON e Paolo Venturi, Direttore AICCON

Questo articolo è estratto dal contributo **“L’impatto economico e la sfida etica delle tecnologie convergenti”** che sarà distribuito in occasione della XVIII edizione delle Giornate di Bertinoro per l’Economia Civile.

È sul fronte dell’etica pubblica che le conseguenze della diffusione nelle nostre società delle tecnologie convergenti vanno ponendo i problemi più delicati, primo fra tutti quello di capire come la digitalizzazione della nostra vita stia riuscendo a modificare anche il modo di percepirla. Eppure, è proprio su tale fronte che si registra una sorta di “fin de non recevoir” da parte dell’alta cultura, scientifica e filosofica. Di due (soli) aspetti particolari desidero qui dire in breve. Il primo concerne la **questione della fiducia**:

può l’intelligenza artificiale creare la fiducia che è necessaria per il corretto funzionamento delle nostre economie di mercato?

Il secondo aspetto chiama in causa il **problema della responsabilità**, di cosa significhi essere responsabili nell’era della digitalizzazione.

Sono le “smart machines” agenti morali e dunque responsabili? Saranno gli algoritmi a governarci, in tutti i casi in cui le persone non sono in grado di comprendere appieno le questioni sulle quali debbono esprimere valutazioni?

Partiamo dal primo aspetto.

Generale è il consenso sul fatto che è **la fiducia uno dei fattori decisivi per assicurare i vantaggi dell’agire collettivo e, per questa via, sostenere il processo di sviluppo**. È agevole darsene conto. Tutti gli scambi che avvengono nel mercato sono incorporati in contratti: espliciti o impliciti; spot o a termine; completi o incompleti; contestabili o no. Eccezion fatta per quelli spot, tutti gli altri tipi di contratto hanno bisogno di un qualche meccanismo per essere resi esecutivi. Sappiamo che l’esecutorietà dei contratti – come fare in modo che i termini e le obbligazioni contrattuali vengano onorate – dipende, in forme e gradi diversi, dalle norme legali, dalle norme sociali di comportamento prevalenti in una determinata comunità e dalla fiducia reciproca. Ebbene, quando i primi due fattori non bastano ad assicurare l’esecutorietà dei contratti, è alla fiducia che si ricorre per far funzionare il mercato. Ciò è specialmente vero ai giorni nostri, dato che globalizzazione e quarta rivoluzione industriale hanno reciso i tradizionali legami (di sangue, di religione, di tradizione) che in passato funzionavano come surrogati, più o meno perfetti, della fiducia.

Si noti il paradosso tipico dell’attuale fase storica. Mentre la fiducia nei confronti delle istituzioni, sia politiche sia economiche, va declinando per una pluralità di ragioni, tra cui l’aumento endemico della corruzione, il mercato globale è conquistato sempre più da imprese e organizzazioni che chiedono ai loro clienti e utenti prove di fiducia, mai viste in passato.



È come se gli individui stessero imparando la lezione della ben nota vicenda della Tosca di Puccini: **la mutua sfiducia genera sempre risultati subottimali**. (Cfr. H. Leibenstein, "On some economic aspects of a fragile input: trust", in R. Feiwel, a cura di, Essays in honour of Kenneth Arrow, MIT, Press, 1987). Si pensi alla concessione dell'uso della propria abitazione a estranei totali (Airbnb) o alla condivisione di viaggi in auto con persone affatto sconosciute (Uber, Blablacar). Sta cioè accadendo che alla diminuzione della fiducia istituzionale di tipo verticale fa riscontro un aumento della fiducia personale, cioè la fiducia orizzontale tra persone. Per Tim Wu, affermato giurista della Columbia University, quello cui stiamo assistendo è un massiccio trasferimento di fiducia sociale: abbandonata la fiducia nelle istituzioni, ci si rivolge alla tecnologia. *"La fiducia – scrive R. Botsman (Di chi possiamo fidarci, Milano, Hoepli, 2017) – è la nuova valuta dell'economia mondiale. E' un vero moltiplicatore di opportunità di guadagno perché consente di far fruttare beni sottoutilizzati"*. Il cuore del paradosso odierno è in ciò che l'economia di mercato contemporanea ha ancora più bisogno di quella del passato di fiducia reciproca per poter funzionare al meglio.

Al tempo stesso, però, gli straordinari livelli di efficienza finora raggiunti dai nostri sistemi economici sembrano far dimenticare che è necessario rinforzare le reti fiduciarie tra persone perchè il mercato mentre "consuma" sempre più fiducia non riesce, stante l'attuale assetto istituzionale, a produrne a sufficienza. Di qui l'inquietante dilemma sociale: chiediamo sempre più efficienza per accrescere il benessere materiale, la ricchezza, la sicurezza, ma per conseguire un tale obiettivo decumuliamo irresponsabilmente il patrimonio di fiducia che abbiamo ereditato dalle generazioni passate. (Patrimonio è parola che viene da "patres munus": il dono dei padri). Si badi che un'economia di comando può ben fare a meno della fiducia per assicurare il proprio funzionamento, non così una economia di mercato, come si è detto sopra. Ai tempi dell'URSS ("La fiducia è bene, il controllo è meglio", era solito dire Lenin) non si avvertiva alcun bisogno di investire sulla fiducia interpersonale; bastava e avanzava quella istituzionale. (L. De Biase, "L'algoritmo della fiducia, "Vita, Genn. 2018).

Che fare per sciogliere questo dilemma? È nota la proposta di **David Hume**. Per il fondatore dell'empirismo filosofico (e iniziatore del non cognitivismo etico) la disposizione ad accordare fiducia, e a ripagare la fiducia concessa, trova il proprio fondamento nei vantaggi personali che scaturiscono da una buona reputazione. *"Possiamo soddisfare i nostri appetiti meglio in un modo indiretto e artificiale. E' così che imparo a prestare un servizio ad un altro senza provare per lui una vera benevolenza. Infatti io prevedo che egli mi renderà il servizio attendendosene un altro dello stesso tipo, per conservare la medesima reciprocità di buoni uffici con me o con altri"* (Trattato sulla natura umana [1740] 1971, pp. 552-3).

È quasi incredibile che un grande filosofo come Hume sia potuto cadere in una così patente svista concettuale, quella di confondere la reciprocità con una sequenza di scambi auto-interessati.

La reciprocità, a differenza dello scambio di equivalenti, è un insieme di relazioni di dono tra loro interrelate.

La cosa ancora più strana è che il pensiero filosofico successivo mai abbia rilevato la contraddizione pragmatica in cui cade Hume quando, poche righe sopra il brano riportato, fa l'esempio dei due coltivatori di grano che finiscono in disgrazia per mancanza di garanzie reciproche. Si legge sempre nel Trattato: *"Il tuo grano è maturo, oggi, il mio lo sarà domani. E' utile per entrambi se io oggi lavoro con te e tu domani dai una mano a me. Ma io non provo nessun particolare sentimento di benevolenza"*



nei tuoi confronti e so che neppure tu lo provi per me. Perciò io oggi non lavorerò per te perché non ho alcuna garanzia che domani tu mostrerai gratitudine nei miei confronti. Così ti lascio lavorare da solo oggi e tu ti comporterai allo stesso modo domani. Ma il maltempo sopravviene e così entrambi finiamo per perdere i nostri raccolti per mancanza di fiducia reciproca e di una garanzia". (Ivi).

Anche la soluzione dell'imperativo categorico kantiano non ci è di grande aiuto agli scopi presenti. "Segui la regola che, se ognuno la seguisse, tu potresti volerne il risultato". È questo un principio di eguaglianza del dovere. Tuttavia, la teoria di Kant soffre di una evidente aporia quando si cercasse di porla in pratica. Infatti, l'individuo kantiano sceglie la regola (la massima) che va ad applicare assumendo che anche tutti gli altri la applichino. Ma poiché persone diverse, in generale, hanno preferenze diverse circa il risultato finale, anche le regole kantiane da esse preferite saranno a priori diverse. Ne consegue che ciascuno seguirà la sua regola preferita (da cui la sua azione) assumendo che gli altri agiscano nel modo in cui in realtà essi non agiranno affatto. Ciò significa che il **principio kantiano** non può applicarsi a se stesso; non può validare se stesso: davvero una seria incongruenza logica per una dottrina morale che ambisce ad essere universale. Solamente se tutti gli individui fossero tra loro identici quanto al loro sistema preferenziale l'aporia in questione scomparirebbe. Ma è evidente che se così fosse il principio kantiano perderebbe tutta la sua rilevanza pratica.

L'interrogativo posto ci introduce al **tema della responsabilità** – il secondo aspetto di cui abbiamo scritto all'inizio del paragrafo – intrigante quanto pochi. Come sappiamo, la responsabilità possiede significati diversi. Si può dire responsabilità per significare una libertà che ha il senso della responsabilità. Ma si può dire responsabilità in senso molto diverso quando si è incaricati di un compito di cui si deve rispondere (è il concetto americano di "accountability"). Infine, si può dire responsabilità per indicare che si è colpevoli di un'attività portata a compimento. In tal senso, "io sono responsabile" significa che sono colpevole di qualcosa. Responsabilità e libertà risultano pertanto fortemente correlate, anche se, in tempi recenti, sull'onda degli avanzamenti registrati sul fronte delle neuroscienze, si tende ad allentare il nesso tra libertà e responsabilità. Si considerino gli interventi di potenziamento cui ho fatto cenno dianzi. Il soggetto potenziato prenderebbe le sue decisioni non sulle ragioni pro e contro, ma in seguito all'influsso causale esercitato sul suo cervello dai mezzi di manipolazione biotecnologica. Quanto a dire che per migliorare la performance degli esseri umani li si priva della loro autonomia morale, che è il bene più prezioso. (P. Donati, "Globalization of markets, distant harms and the need for a relational ethics", Rivista Internazionale di Scienze Sociali, 1, 2017).

Mentre sembra relativamente facile identificare la responsabilità diretta degli agenti – come quando il proprietario di uno sweatshop sfrutta il lavoro minorile per trarne un vantaggio – che dire dell'azione economica che è intrapresa con l'intenzione di non svantaggiare nessuno e tuttavia provoca effetti negativi in capo ad altri? Ad esempio, di chi è la responsabilità della disoccupazione, della povertà, delle disuguaglianze, etc?

Le risposte tradizionali, in economia, consistono nel sostenere che si tratta di conseguenze non volute delle azioni intenzionali (le "unintended consequences of intentional actions" di cui ha parlato la Scuola dei moralisti scozzesi del 18° secolo). E dunque l'unica cosa da fare è di attribuire alla società il compito di porre rimedio (o di alleviare) le conseguenze negative. E infatti il **welfare state** è sorto e si è sviluppato precisamente per rendere collettiva e impersonale la responsabilità dei singoli.



Ma è veramente così? Siamo sicuri che i meccanismi del libero mercato siano inevitabili e che gli effetti che ne derivano siano inattesi, come si tende a far credere?

Valga un solo esempio. **Albert Carr** nel saggio "Is business bluffing ethical?" (Harvard Business Review, 1968) – il saggio più citato di sempre in teoria della finanza – scrive: *"La finanza, dove si cerca di fare agli altri ciò che non si desidera gli altri facciano a noi (sic!) dovrebbe essere guidata da un insieme di standard etici diversi da quelli della morale comune o della religione: gli standard etici del gioco. Se un'azione non è strettamente illegale, e può dare un profitto, allora compierla è un obbligo dell'uomo d'affari"*. E' questo modo di pensare – fondato sulla tesi della doppia moralità – che è all'origine di tutti i grandi scandali finanziari, tra cui quelli dell'ultimo ventennio. Come, tra i primi, aveva notato **Z. Bauman**, l'organizzazione sociale della seconda modernità è stata pensata e disegnata per neutralizzare la responsabilità diretta e indiretta degli agenti.

La strategia adottata – di grande raffinatezza intellettuale – è stata quella, per un verso, di allungare la distanza (spaziale e temporale) tra l'azione e le sue conseguenze e, per l'altro verso, di realizzare una grossa concentrazione di attività economica senza una centralizzazione di potere. E' in ciò il carattere specifico dell'**impresa adiaforica**, una figura di impresa ignota alle epoche precedenti la seconda guerra mondiale e il cui fine è quello di annullare la questione della responsabilità morale dell'azione organizzata. Adiaforica è la responsabilità "tecnica" che non può essere giudicata in termini morali di bene/male.

L'azione adiaforica va valutata in termini solamente funzionali, sulla base del principio che tutto ciò che è possibile per gli agenti sia anche eticamente lecito, senza che si possa giudicare eticamente il sistema, come Luhmann ha insegnato.

Ebbene, la responsabilità adiaforica ha ricevuto, in tempi recenti, nuovo impulso proprio dalla 4a rivoluzione industriale, la quale va producendo "mezzi" che sono alla ricerca di "domande" o di problemi da risolvere. Esattamente il contrario di quanto era accaduto con le precedenti rivoluzioni industriali. Invero, cosa ne è del principio di responsabilità nella società degli algoritmi? Dalle **nuove tecnologie industriali** alla diagnostica medica, dai social networks ai voli degli aerei, dai big data ai motori di ricerca: ci affidiamo a complesse procedure cui deleghiamo la buona riuscita di operazioni che gli esseri umani, da soli, non saprebbero eseguire. Eppure, gli algoritmi sono irresponsabili, pur non essendo neutrali, né oggettivi, come invece erroneamente si crede. Quando un programma commette un errore non ne paga le conseguenze, perché si ritiene che la matematica resti al di fuori della moralità. Ma non è così, perché gli algoritmi non sono pura matematica; sono opinioni umane incastonate in linguaggio matematico. E dunque discriminano, al pari dei decisori umani. Ad esempio, il processo delle assunzioni di lavoro si va sempre più automatizzando, perché si pensa di rendere obiettivo il reclutamento del personale, eliminando pregiudizi. Ma le dinamiche discriminatorie, anziché diminuire, stanno aumentando nelle nostre società. (P. Zellini, "Saranno gli algoritmi a governarci?", Repubblica, 12 marzo, 2017).

Generalizzando un istante, il vero problema delle **smart machines** comincia nel momento in cui queste compiono azioni che coinvolgono la necessità di scegliere oppure di decidere. Il soldato-robot, l'automobile-robot, la scopa-robot potrebbero compiere scelte esiziali per vite non robotiche. Di chi è la responsabilità in questi casi? Quale il messaggio subliminale della recente provocazione di Bill Gates di tassare i robot, che andrebbero dotati di personalità elettronica, al pari delle



corporations che sono dotate di personalità giuridica? Come ha lucidamente spiegato Gunther Anders, (L'uomo è antiquato. Sulla distribuzione della vita nell'epoca della terza rivoluzione industriale, Torino, Bollati Boringhieri, 2002) – il XXI secolo ha inaugurato l'era dell'irresponsabilità umana, immunizzando i soggetti dalle loro relazioni. Le "smart machines" (quelle dotate di intelligenza artificiale) sono in grado di prendere decisioni autonome, che hanno implicazioni sia sociali sia morali. (Si veda il caso dell'auto senza pilota Tesla, creata da Elon Musk, che nel maggio 2016 uccise un passeggero).

Come assicurare, allora, che le decisioni prese da tali oggetti siano eticamente accettabili? Dato che queste macchine possono causare danni di ogni sorta, come fare in modo che esse siano poste in grado di differenziare tra decisioni corrette" e "sbagliate" ?

E nel caso in cui un qualche danno non possa essere evitato – si pensi al caso dell'auto senza conducente che deve scegliere se gettarsi contro un altro veicolo uccidendone i passeggeri oppure investire dei bambini che attraversano la strada -, come istruire (nel senso di programmare) tali macchine a scegliere il danno minore? Gli esempi in letteratura sono ormai schiera. E tutti concordano sulla necessità di dotare l'IA di un qualche canone etico, per sciogliere dilemmi morali del tipo "guida autonoma". (Rinvio all'accurato lavoro di L. Palazzani, Dalla bioetica alla tecnoetica: nuove sfide al diritto, Torino, Giappichelli, 2017).

Le divergenze nascono nel momento in cui si deve scegliere il modo (cioè l'approccio) secondo cui procedere: top-down (i principi etici sono programmati nella macchina intelligente: **l'uomo trasferisce all'intelligenza artificiale la sua visione etica del mondo**) oppure bottom-up (**la macchina impara a prendere decisioni eticamente sensibili dall'osservazione del comportamento umano in situazioni reali**). Entrambi gli approcci pongono problemi seri, che non sono tanto di natura tecnica quanto piuttosto concernono la grossa questione se le macchine intelligenti debbano o meno essere considerate agenti morali (cioè moral machines). Siamo appena agli inizi di un dibattito culturale e scientifico che già si preannuncia affascinante e preoccupante ad un tempo. Si veda, ad esempio, la recente presa di posizione di A. Etzioni e O. Etzioni ("Incorporating Ethics into Artificial Intelligence", Journal of Ethics, March, 2017) che negano la possibilità di attribuire lo status di agente morale all'Intelligenza Artificiale e dunque negano ogni fondamento al programma di ricerca della Internet Ethics che studia gli aspetti etici della Internet Communication nelle sue varie articolazioni.

Da ciò essi traggono la conclusione che non vi sarebbe alcun bisogno di insegnare etica alle macchine, anche se ciò potesse essere fattibile. Non è del medesimo parere, ad esempio, il gruppo di ricerca che opera per la **NeuroLink Corporation**, in California, che da qualche tempo sta sviluppando tecnologie digitali per realizzare connessioni tra computer e mente umana e che sta progettando un uomo-cyber con microchip nel cervello. Sulla intricata e delicata questione concernente la possibilità di attribuire "personalità elettronica" ai robot intelligenti e, più in generale, la opportunità di favorire il passaggio dalla selezione naturale darwiniana alla scelta deliberata del processo di selezione mediante la scorciatoia biotecnologica, si vedano: N. Bostrom, "Welcome to a world of exponential change", in P. Miller, J. Wilsdon, a cura di, Better Humans? The Politics of Human Enhancement and Life Extension, Demos, London, 2006 e S.M. Kampowski, D. Moltisanti, a cura di, Migliorare l'uomo? La sfida dell'enhancement, Cantagalli, Siena, 2011. **Una riflessione importante è quella di Adriano Fabris che distingue tra "etica di internet" e "etica in internet".**



Può essere opportuno considerare che, nel dibattito corrente, due diversi modi di concettualizzare l'IA si vanno confrontando. Come chiarisce **L. Palazzani**, il primo concerne software che cercano di ragionare e di prendere decisioni cognitive al modo in cui gli umani lo fanno. Per tale concezione, l'IA aspirerebbe a rimpiazzare l'uomo. (Il famoso test di Turing ha a che vedere con questo tipo di IA).

Il secondo modo mira invece a fornire un'assistenza smart agli attori umani. Si tratta di una IA partner dell'uomo, spesso indicata come "Intelligence Augmentation" ovvero "Cognitive augmentation". Nel concreto, Google si sta muovendo nella prima direzione; l'obiettivo dichiarato è quello di arrivare a fondere l'uomo con la macchina per accrescerne, senza limite, le capacità; IBM, con il suo cognitive computing, nella seconda. Nel 2013, IBM ha lanciato il sistema di Intelligenza Artificiale "Thomas Watson" in omaggio al nome del suo primo presidente. Watson risponde alle domande poste in linguaggio naturale su qualsiasi tematica.

Nella prima pagina del sito dedicato a Watson si legge: "Watson è una tecnologia cognitiva che può pensare come un essere umano". Si tratterà di vedere se la macchina potrà diventare più intelligente dell'uomo. In ogni caso, resta vero che le risposte standardizzate che Watson (o altra macchina) potrà dare non saranno mai più efficaci di quelle che possono dare persone in grado di comprendere i problemi di altre persone. Invero, le macchine, per quanto intelligenti, mai saranno capaci di empatia, perché non dotate di sentimenti morali.

In ogni caso, è oggi più che mai irrisolto il problema di conoscere come e quanto lo sviluppo delle tecnologie potenziative avrà effetti relativamente ai valori della giustizia e dell'eguaglianza sociale. Nella misura in cui queste tecnologie non potranno mai essere fruite da tutti, in quanto assai costose, si pone la questione dell'incremento delle diseguaglianze sociali, già oggi insopportabilmente elevate. Il rischio serio è quello di formare una società parallela costituita dalla fascia di popolazione più ricca che progressivamente si distacca dai più poveri... "non potenziati". (M. Sandel, *Contro la perfezione. L'etica nell'età dell'ingegneria genetica*, Milano, Vita e Pensiero, 2008).