

ROBOTICA SOCIALE E LE TECNOLOGIE PERSUASIVE

di Fabio Fossa *

Spesso si pensa che la filosofia e l'etica della tecnologia abbiano solo il compito di piantare paletti e tracciare limiti per la ricerca tecnologica. Mi sembra una visione molto limitata. Certamente limiti e paletti sono utili quando le circostanze li richiedono. Ma gli studi sull'etica della tecnologia digitale e robotica non possono essere ridotti a questo. C'è molto di più. L'analisi etica è propositiva, cerca l'integrazione con i saperi dell'ingegneria e del design. Si tratta di capire come le tecnologie che costruiamo interagiscano con noi per svilupparle meglio, cioè in modo da minimizzare i rischi e promuovere il benessere sociale e individuale. Ora come ora sono impegnato ad approfondire varie tematiche legate all'etica della guida autonoma presso il Dipartimento di Ingegneria Meccanica del Politecnico di Milano. Una parte delle mie ricerche passate, a cui continuo a lavorare, ha però a che vedere con la robotica sociale. In particolare mi interessa il fatto che tramite specifiche scelte di design sia possibile influenzare la percezione e il modello mentale che gli utenti si formano di un dato sistema tecnologico. Cosa che si può mettere in pratica, ad esempio, al fine di promuovere l'accettazione o la fiducia nella competenza di un robot. La domanda è: cosa è lecito fare in questo senso? Qual è la linea tra influenza legittima e manipolazione, tra supporto e inganno? E quali possono essere le conseguenze?

Sfruttare i bias in sede di design: qual è il problema?

Porre queste domande è particolarmente importante perché uno dei modi per rendere i robot più familiari e di facile utilizzo consiste nello sfruttare i cosiddetti *bias*, cioè gli stereotipi attraverso i quali semplifichiamo la realtà e ci orientiamo nel mondo sociale.

I bias sono associazioni mentali profondamente radicate nel modo in cui interpretiamo il mondo sociale. Ci aiutano a gestirne la complessità. Per loro tramite ci formiamo aspettative circa i ruoli, le competenze e le caratteristiche meno evidenti dei nostri simili. Sono associazioni che ci troviamo ad avere in testa, il che non dice nulla sulla loro effettiva consistenza. Capita che siano del tutto campate per aria. Funzionano un po' come scorciatoie. Spesso le scorciatoie ci fanno arrivare prima e con più facilità dove vogliamo andare. Altre volte ci portano fuori strada e si finisce per fare o farsi male.

Tra i più diffusi ci sono i bias legati al genere sessuale. Per esempio, associamo capacità di cura e comprensione al sesso femminile, autorità e competenza al sesso maschile. Insieme alla dott.ssa Irene Sucameli dell'Università di Pisa ci siamo chiesti: cosa succede quando si

usano questi bias per ottimizzare le relazioni uomo-macchina, per esempio nel caso di agenti conversazionali?

Non si tratta di una ipotesi. Sfruttare i bias che riguardano la nostra percezione del mondo per promuovere l'accettazione dei robot è una tecnica impiegata in robotica sociale, anche esplicitamente. In questo modo si vuole facilitare l'introduzione di robot nei vari contesti sociali, cioè far sì che l'utente umano non percepisca uno scarto troppo ampio tra l'interazione cui è abituato – quella con altri umani – e quella con il sistema artificiale.

Se si riesce a mantenere la continuità tra questi due mondi, l'utente sarà in grado di interagire con il robot più o meno nello stesso modo in cui si aspetta di interagire con le persone. Non sentirà troppo lo scarto, accetterà il sistema e lo userà come anticipato dai progettisti. Un esempio semplicistico ma efficace potrebbe essere questo: nel progettare un robot sociale per l'assistenza casalinga, dove cura e comprensione sono fondamentali, è consigliabile adottare forme, tono di voce e fattezze usualmente collegate al genere femminile. In questo modo le aspettative dell'utente, per quanto basate su pregiudizi, saranno soddisfatte e l'interazione sarà un successo.

Ora, se pensiamo che i robot sociali possano portare importanti benefici alla società, saremo motivati a ricorrere a questa strategia che ne facilita l'accettazione. Ma quali effetti rischiamo di causare?

Il sospetto che si stia giocando con il fuoco mi sembra lecito. Molti stereotipi si diffondono a livello sociale pur non avendo alcuna base razionale. Li troviamo già pronti e li riproduciamo non solo (e forse non tanto) per esplicita convinzione, ma anche in modo implicito, come un automatismo. Certo, alcuni di questi stereotipi possono essere innocui. Altri però possono veicolare credenze ingiuste, offensive e discriminatorie nei confronti di alcuni gruppi sociali. Pensiamo ad esempio al pregiudizio secondo cui competenza e autorevolezza sono caratteri maschili: un retaggio storico-culturale che continua a pesare sulla condizione delle donne nella nostra società. Non dovremmo almeno ammettere la possibilità che assecondare questi stereotipi attraverso il design dei robot sociali non finisca per consolidarli e istituzionalizzarli, quando sarebbe molto meglio disfarsene?

Cosa fare dei bias?

Certo, il dubbio che ho appena sollevato si basa su un assunto che non tutti accettano. La domanda ha senso solo se si ammette la possibilità che dinamiche pregiudizievoli possano trasferirsi dalle interazioni con robot a quelle tra umani. Alcuni ricercatori sostengono che sia un errore ammettere questa permeabilità tra le due dimensioni. Altri fanno notare però

che l'idea di utilizzare i bias come risorse di design ha senso solo se si ammette fin dall'inizio che la separazione tra le due forme di interazione non sia poi così netta. D'altronde, se le interazioni con i robot non avessero nulla a che spartire con le interazioni tra umani, l'impiego dei pregiudizi non avrebbe alcun senso. Il ricorso stesso ai bias nella progettazione dei robot sociali mette già in discussione che ci sia una divisione netta tra le due forme di interazione. Se ha senso sfruttare i bias per facilitare le interazioni tra umani e robot, allora bisogna ammettere la possibilità che gli stessi pregiudizi si possano trasferire nuovamente alle relazioni tra umani.

Consideriamo un esempio ipotetico. Un team di programmatori deve sviluppare un agente conversazionale che gestisca le patologie più comuni per diminuire il carico di lavoro dei medici di base. Poniamo che, per scegliere i caratteri dell'avatar, il team faccia una ricerca empirica da cui emerge che il pubblico associa all'immagine del medico competente e autorevole quella di un uomo bianco, sulla sessantina, con capelli bianchi e barba. Di conseguenza, il team adegua l'avatar alle aspettative degli utenti, così da facilitarne l'accettazione e semplificare l'interazione. Così facendo, però, il team consolida le aspettative pregiudizievoli degli utenti, rafforzando l'associazione tra un certo genere sessuale, un certo colore della pelle, una certa età, e qualità professionali come l'autorevolezza e la competenza. Un'associazione che difficilmente reggerà a uno scrutinio approfondito. Che impatto avrà questa tecnologia sulla fiducia dei pazienti nei confronti, ad esempio, di giovani dottoresse di colore agli inizi della professione? Il team avrà contribuito a trasmettere e consolidare aspettative pregiudizievoli e discriminatorie?

Se il dubbio ci sembra fondato, è naturale chiedersi come sia meglio comportarsi. Una prima possibilità potrebbe essere di eliminare completamente l'utilizzo di bias dalla progettazione dei robot. Si tratta però di una opzione estremamente difficile da perseguire: ci sono buone ragioni per dubitare che sia possibile, data la pervasività dei pregiudizi. E il rischio che il prodotto sia rifiutato dagli utenti è alto. Potremmo allora propendere per rendere off-limits i bias discriminatori e permettere l'utilizzo di quelli più innocui. Ma chi dovrebbe distinguere tra bias discriminatori e innocui? Le aziende, i progettisti, i designer, il legislatore, i decisori politici? In quest'ultimo caso, come assicurare che tempi lunghi dell'amministrazione statale non inficino la ricerca e lo sviluppo dei robot?

Un'altra possibile via è quella di dedicare il design dei robot sociali non soltanto a favorirne l'accettazione, ma anche a contrastare attivamente le discriminazioni sociali, progettando robot che sfidano e mettono in discussione le aspettative pregiudizievoli degli utenti. È un'idea interessante, che però solleva molte questioni di etica sociale: su quali bias sarebbe

permesso agire in questo modo? E chi dovrebbe deciderlo? Non stiamo minando l'autonomia dell'utente, non lo stiamo manipolando a sua insaputa, per quanto a fin di bene?

La robotica sociale e l'inganno

In un certo senso, lo sfruttamento dei bias per facilitare l'accettazione dei robot sociali può essere descritto come una forma di inganno. L'inganno è un tema importante nell'etica della robotica sociale. Ho avuto modo di rendermene meglio conto scrivendo un capitolo proprio su questo tema per il volume *Automi e persone. Introduzione all'etica dell'Intelligenza Artificiale e della robotica*, che ho curato insieme a Viola Schiaffonati e Guglielmo Tamburrini per Carocci ed è stato pubblicato a ottobre.

Alcuni studiosi si sono espressi con decisione contro l'idea di ricorrere ai robot per soddisfare bisogni sociali – per esempio, per alleviare la solitudine di persone anziane e sole. La critica più comune afferma che l'intera pratica consiste nell'ingannare persone vulnerabili, e quindi predisposte a cadere preda di manipolazioni. Il bisogno di contatto umano avvertito dalle persone sole non può essere soddisfatto tramite i robot sociali. Spacciare questi come sostituti di quello è ingannevole e sfrutta in modo irrispettoso condizioni di disagio e vulnerabilità.

Si tratta di una critica molto forte, che getta una luce obliqua sugli studi empirici che sono stati condotti in tal senso. Molti di questi dimostrano che la sensazione di benessere dei soggetti coinvolti è effettivamente migliorata quando è stata concessa la possibilità di interagire con robot sociali anche in modi piuttosto basilari. Una delle prime ricerche in questa direzione è stata realizzata con Paro, un robot a forma di cucciolo di foca che emette versi e muove il corpo a seconda delle carezze che gli vengono somministrate. È stato usato in case di riposo e ospedali per alleviare la solitudine di anziani, in molti casi affetti da Alzheimer o demenza senile.

Il dato empirico sulla percezione che i soggetti hanno del proprio benessere quando provvisti di robot sociali può essere messo in discussione facendo leva sull'estremo bisogno di contatto umano che avvertono. Le persone coinvolte nelle sperimentazioni sono fortemente motivate a lasciarsi ingannare, per cui è almeno lecito dubitare della loro autoconsapevolezza. Insomma, i dati empirici non bastano per risolvere la questione. Chi non è convinto dirà che l'inganno ha funzionato anche troppo bene! Ma rimane da dimostrare che il bisogno reale dei soggetti sia stato davvero soddisfatto, che la loro dignità sia stata rispettata, e che il loro benessere sia stato perseguito come merita. Se poi

aggiungiamo che dietro alla robotica sociale vi sono ovvi interessi commerciali, si capisce che il discorso si complica ulteriormente

Un punto secondo me interessante è che molto spesso si assume che il soggetto vulnerabile venga ingannato in modo inconsapevole, che sia in un certo senso in balia del robot sociale. Ma perché non pensare invece che stia gestendo il proprio mondo di finzioni consapevolmente? L'esperienza dell'autoinganno consapevole non è poi così insolita. Tutti noi ci lasciamo ingannare da finzioni, per esempio quando scene di vita rappresentate su uno schermo o scritte su pagine di libro ci commuovono come se fossero realtà. Oppure quando le dinamiche di un gioco ci coinvolgono del tutto e ci lasciamo trasportare completamente dalla sua logica interna. Anche se si è consapevoli di essere immersi in una finzione, non per questo l'effetto gratificante e benefico sparisce. Al contrario, direi.

Qui il punto cruciale è quello della consapevolezza dell'inganno. A differenza dell'effetto placebo, dove manca la consapevolezza della finzione, nel gioco o in alcune esperienze estetiche sappiamo che la struttura è fittizia, ma più la prendiamo sul serio, più l'esperienza è gratificante. E non mi sembra che in questo caso ci siano da sollevare problemi relativi al rispetto della dignità e dell'autonomia personale. Perché non pensare che lo stesso possa avvenire con la finzione della robotica sociale?

Rimane pur sempre vero, però, che nel nostro dibattito l'attenzione è centrata su soggetti vulnerabili, nel cui caso è complicato stabilire fino a che punto siano effettivamente in grado di gestire esperienze di autoinganno consapevole. Detto ciò, mi sembra comunque sensato indagare la questione anche da questa prospettiva.

Inganno e finzione: il caso dell'antropomorfismo

Si può poi notare che alcuni inganni e finzioni sono profondamente radicati nel modo in cui gestiamo la complessità dell'ambiente in cui ci troviamo a vivere. Se certe finzioni fanno intimamente parte della nostra mente, e alcune di queste influiscono sul modo in cui i robot vengono percepiti e interpretati, forse non è il caso di rifiutarle del tutto in sede di design. Forse è meglio comprendere come fare per riconoscerle, minimizzare i danni che ne possono seguire, e sfruttare i benefici che offrono. Qui la riflessione etica ha evidentemente un ruolo cruciale, che non può essere svolto adeguatamente se ci si limita a tracciare limiti e piantare paletti.

Un esempio in questo caso è l'antropomorfismo. L'antropomorfismo è la tendenza a interpretare fenomeni secondo categorie e schemi che pertengono alla vita umana. È una attitudine così radicata in noi che sembra essere parte costitutiva del nostro modo di esistere

e di interagire con l'ambiente. Nel caso della robotica l'antropomorfismo è tanto naturale che cercare di opporsi con specifiche scelte di design è un tentativo disperante. Bisogna imparare a farci i conti, a gestire questa tendenza in modo che ci aiuti a promuovere i benefici e ridurre i rischi della robotica sociale. Il che significa progettare i robot in modo che sostengano, ma solo fino al punto giusto, la nostra tendenza ad antropomorfizzare.

Come nel caso dei pregiudizi, i benefici dell'antropomorfismo sono legati infatti ad un suo sfruttamento dosato in modo intelligente e responsabile. Cioè, che ci permetta di approcciare il robot non come qualcosa di totalmente estraneo e alieno, ma come un interlocutore tutto sommato familiare per quanto diverso da noi. I rischi sono legati invece a situazioni in cui l'antropomorfismo non viene dosato con misura, per cui l'utente o rifiuta il robot come un corpo estraneo o è portato a perdere consapevolezza delle differenze che lo distinguono dagli esseri umani – fino a trattare uno strumento come se fosse un individuo. Il problema è definire il livello ottimale di antropomorfismo, il che riguarda la gestione ad arte di vari elementi di design del robot. È una questione assai delicata ma cruciale per far sì che l'utente mantenga la consapevolezza della finzione e goda dei benefici connessi.

Resta da chiedersi come fare per definire il livello ottimale di antropomorfismo. È un problema che rivela delle difficoltà serie. Da una parte, la questione sembra soprattutto empirica: bisogna sperimentare. Ma non si può pensare di farlo senza sbagliare mai, e ad ogni errore possono seguire danni morali non indifferenti in questo ambito molto delicato. Si tratta di una questione complessa a cui bisogna mettere mano per evitare che abbia la meglio un approccio più spregiudicato dove a fare le spese di esperimenti sull'antropomorfismo sono gli utenti in generale.

Fabio Fossa è assegnista di ricerca presso il Dipartimento di Ingegneria Meccanica del Politecnico di Milano, dove si occupa di filosofia degli agenti artificiali ed etica dei veicoli autonomi. La sua ricerca verte sui temi di etica applicata, filosofia della tecnologia, etica della robotica e dell'IA e sul pensiero di Hans Jonas. È Direttore della rivista *InCircolo*, *Rivista di filosofia e culture* e membro fondatore del gruppo di ricerca *Zetesis*.

-