

Fiorella Battaglia

Robotiche emozioni

I robot e le sfide della morale artificiale

visualizza la scheda del libro sul sito www.edizioniets.com



Edizioni ETS



www.edizioniets.com

*Volume finanziato dall'Unione Europea - Next Generation EU,
Missione 4 Componente 1 Investimento 1.1 CUP F53D23010740001*



**Ministero
dell'Università
e della Ricerca**



© Copyright 2025

EDIZIONI ETS

Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa

info@edizioniets.com

www.edizioniets.com

Distribuzione

Messaggerie Libri SPA

Sede legale: via G. Verdi 8 - 20090 Assago (MI)

Promozione

PDE PROMOZIONE SRL

via Zago 2/2 - 40128 Bologna

ISBN 978-884675623-7

Indice

1. Introduzione	7
2. Le emozioni umane rispetto ai robot	13
2.1. Non meri esecutori ma fidi maestri sostituiti?	20
2.1.1. Le emozioni e la ragione pratica	24
2.1.2. Il corpo e l' <i>embodiment</i>	27
2.2. Effetti trasformativi	33
2.2.1. <i>Human Enhancement</i>	36
2.2.2. Epistemologia morale	39
2.2.3. Colonizzazione del mondo della vita	44
2.3. Riassunto	45
3. La morale artificiale	47
3.1. Coscienza	49
3.2. Agenti morali autonomi	51
3.2.1. Responsabilità	56
3.3. Apprendimento automatizzato e opacità epistemica	65
3.3.1. Decisioni automatizzate	68
3.3.2. Errori e discriminazioni	70
3.3.3. Questioni etiche ed epistemiche	75
3.3.4. Alcune riflessioni metaetiche	75
3.3.5. Epistemologia dell'apprendimento automatico	80
3.3.6. Potenza umana e algoritmica di calcolo	84
3.4. Riassunto	85

4. Imparare a essere umani	89
4.1. Ragioni pro e contro l'utilitarismo	89
4.2. Etica e interculturalità	94
4.3. Consenso o verità?	95
4.4. Indicazioni filosofiche	98
4.5. Conclusioni	100
5. Bibliografia	105

Introduzione

Quando Kevin Warwick mi ha raccontato per la prima volta del collegamento invasivo di elementi macchinici con il suo corpo, mi ha colpito la descrizione che ha dato del suo attaccamento ai sistemi robotici che per un periodo limitato gli avevano consentito di superare i vincoli spaziali della condizione umana. Con il chip impiantato nel suo braccio, Kevin era stato capace di attivare un arto robotico artificiale a più di 5500 chilometri di distanza e di accarezzare da New York sua moglie che si trovava in Gran Bretagna (Warwick *et al.* 2003). Nel momento in cui l'artefatto gli fu rimosso sviluppò un vero e proprio sentimento profondo del tipo che si prova per la scomparsa di una persona cara. Kevin cercò di spiegare il suo lutto come un'elaborazione rispetto a quello che egli stesso avrebbe potuto essere – ed era stato – per un breve periodo. Egli non stava solo soffrendo per la semplice perdita dell'oggetto, ma piuttosto per il tramontare di quell'esperienza che l'impianto gli aveva consentito di fare. Insomma, l'artefatto posto nel suo corpo era diventato parte del sé e in tal modo aveva parassitariamente avuto accesso allo status di personalità. Kevin si rammaricava per la perdita di una possibilità della sua vita che svaniva con l'eliminazione dell'artefatto. È interessante osservare che il livello fenomenologico dell'esperienza di Kevin trova un complemento nell'argomentazione giuridica. Nella regolamentazione della protesi, queste perdono il loro status di oggetti per attingere allo stesso status del loro portatore. Diventano simbioticamente parte del soggetto. Vent'anni più tardi il pionieristico collegamento invasivo tra sistema nervoso periferico da un lato e sistemi robotici e informatici dall'altro sperimentato da Kevin Warwick, una nuova ricerca clinica, oltre a fornire evidenze robuste di vantaggi funzionali e di adattamento reciproco, conferma alcuni dei risultati più speculativi dell'esperimento di Warwick (Gilbert, Ienca

e Cook 2023). In quest'articolo Frederic Gilbert, Marcello Ienca e Mark Cook riportano i risultati di uno studio che indaga la fenomenologia della simbiosi umano-macchinica durante un esperimento con una interfaccia neurale (*Brain Computer Interfaces*), progettata per prevedere le crisi epilettiche e perciò realizzare vantaggi funzionali. Per esplorare il livello fenomenologico, i ricercatori hanno utilizzato interviste qualitative semi-strutturate. Di là dai vantaggi funzionali è interessante confrontare l'esperienza di questo paziente con quanto già narrato da Kevin Warwick (Tamburrini 2013). Per quanto riguarda i vantaggi funzionali: dopo l'impianto dell'interfaccia, il paziente ha riportato esperienze di aumento della capacità e della continuità agenziale; dopo l'espianto del dispositivo, il paziente ha riportato danni traumatici persistenti legati alla discontinuità agenziale. Per quanto invece riguarda l'esperienza, il paziente stesso racconta che:

[L'interfaccia neurale] all'inizio era come un alieno, poi ci si abitua gradualmente e diventa parte di tutti i giorni, [...], diventa parte di te. Perché è questo che ha fatto, era me, è diventato me, [...] con questo dispositivo mi sono ritrovato [...] Ha cambiato chi era quella persona allora e mi sono ritrovato a cambiare ... a crescere, suppongo [...] (Gilbert, Ienca e Cook 2023, 785).

Entrambe le esperienze si prestano a essere interpretate nei canoni della cultura occidentale che sono stati formulati mirabilmente da Pico della Mirandola, uno dei primi umanisti, e cioè che gli esseri umani non intendono soggiacere né ai condizionamenti naturali né a quelli sociali, nel momento della costruzione di sé, siccome essi si percepiscono come autrici e autori della propria esistenza (1486). Questo tratto dell'autorialità sovrana trova nello sviluppo della tecnica un alleato potente. La robotica e tutte le tecnologie di intelligenza artificiale che sussumiamo sotto questo termine allora sembrano concorrere per rinforzare questo programma di sviluppo autonomo degli esseri umani. Anche se proprio nel suo esercizio la tecnica paradossalmente può rovesciarsi nel suo contrario, dove l'umano diventa piuttosto soggetto alla tecnica. Le interfacce neurali costituiscono "l'ultima fortezza" (Farahany 2023; Tesink *et al.* 2024; Ienca e Andorno 2017) di quella che Thomas Nagel considerava

la cifra della natura umana: “Il confine tra ciò che riveliamo e ciò che non riveliamo, e un certo controllo su tale confine, sono tra gli attributi più importanti della nostra umanità” (Nagel 1998). I modelli d’intelligenza artificiale hanno bruciato una nuova tappa. È oggi possibile elaborare e decodificare dati anche non neurali e attraverso un processo noto come “inferenza inversa” esplorare i contenuti mentali altrui così come altri oggetti d’indagine empirica. È chiaro che questi avanzamenti spostano il dibattito etico sulla lettura della mente su un terreno concettuale inesplorato finora. Misure individuali sia delle strutture cerebrali e della loro attività sia anche delle loro attività indirette – come, per esempio, registrazione di voce o testo – vengono elaborate attraverso un procedimento di *reverse inference*.

L’inferenza inversa a “processi psicologici” da “modelli di attivazione” rivelati dalla risonanza magnetica funzionale o da altre tecniche di scansione (Glymour e Hanson 2016) conduce a un insieme di dati che presi assieme indicano un significato logico e rivelano con precisione aspetti di stati mentali come pensieri, memorie ed emozioni. Lo scenario si è ampliato. È così possibile arrivare a conoscere pensieri, ricordi, intenzioni ed emozioni di una persona a partire sia da dati cerebrali (Bellier *et al.* 2023) sia da dati non cerebrali. Le interfacce neurali sono il prodotto tecnologico più prossimo alla nostra integrità personale. Di conseguenza sono anche oggetto di forti reazioni emotive. Questo prologo aiuta a capire perché in questo libro ho scelto di focalizzare la riflessione filosofica sulla robotica iniziando dalle emozioni.

Usando un linguaggio allusivo potrei dire che i robot vantano doppi natali. Da un lato sono il prodotto più avanzato e tecnologicamente più sofisticato della manifattura. Dall’altro rappresentano la proiezione delle nostre attese più audaci e delle nostre paure più arcane rispetto al destino umano, all’evoluzione di noi stessi. Nel momento in cui vogliamo tentare di comprendere qualcosa rispetto al nostro rapporto con la robotica, allora prendere le mosse dalle emozioni sembra garantirci un’analisi in grado di fornire un ampio contesto all’interno del quale collocare questo fenomeno. Cosa che invece non potrebbe accadere né se ci limitassimo semplicemente al suo aspetto manifatturiero né se ci concentrassimo sui soli aspet-

ti normativi. È questa, infatti, un'alternativa semplificata che lascia fuori molti elementi. Occorre ricostruire analiticamente le trasformazioni delle pratiche indotte dall'introduzione della robotica e dell'Intelligenza Artificiale (IA) per comprendere il significato profondo dei suoi effetti sulla condizione umana.

La relazione tra emozioni e robotica ci porta a isolare *tre* motivi che costituiscono anche i *tre* capitoli in cui è diviso questo libro che si compone quindi di un *primo* capitolo dedicato alle emozioni che gli umani nutrono nei confronti dei robot. Le emozioni, dalla paura alla speranza, sono un rilevatore del fatto che la tecnica non può essere compresa pienamente, se relegata solo alla sua dimensione strumentale. Ciò non significa condannarla all'altro polo, cioè alla produzione di un tipo di entità che mirano a sostituirci (Harari 2019). Il *secondo* capitolo esplora il ruolo delle emozioni nella "agency" dei robot, introducendo la sfida della morale artificiale. La tesi che cercherò di dimostrare è che i robot e i sistemi autonomi non si limitano all'esecuzione dei compiti, ma producono dei modelli di esecuzione delle attività che con il tempo possono scalzare le modalità, alterare il contesto e assurgere a esemplari di riferimento anche per le pratiche umane. Espungere le emozioni è una delle caratteristiche dei modelli così generati. Questo capitolo fa riferimento non a quando i sistemi robotici vengono usati per influenzare direttamente la vita delle persone, perché in questo caso gli attori morali rimangono le persone, ma quando sono i robot e i sistemi automatici a prendere delle decisioni che hanno un impatto sulla vita delle persone. In questo caso i robot assumono il ruolo di agenti morali e la disciplina in questione non è più l'*etica della robotica*, concentrata su codici deontologici delle professioni o sugli aspetti etici, legali, sociali (ELSA) ma piuttosto l'*etica delle macchine*. A questo campo ci si riferisce qualche volta con le espressioni di "macchine morali" e di "moralità artificiale". Non è per niente una disciplina autonoma, ma si compone piuttosto di un sapere interdisciplinare costituito da robotica, informatica, filosofia, psicologia, scienze cognitive, diritto ed economia. Il capitolo *finale* presenta le conclusioni dei diversi argomenti sviluppati nei due precedenti e tenta di dare delle indicazioni rispetto ai cambiamenti che possono modificare la condizione umana e la sfera emozionale nell'epoca dell'intelligenza artificiale e della robotica.

Dal punto di vista della metodologia, i filosofi morali non disdegnano di fare i conti con i dati di carattere empirico. Come potrebbero fare diversamente? Lo stesso Kant, che programmaticamente esclude la considerazione dei moventi sensibili dalla ragion pratica, riconosce che l'essere umano è spinto a considerare se stesso in un "duplice modo": "dalla coscienza di sé quale oggetto affetto dai sensi" e "dalla coscienza di sé come intelligenza" (GMS, IV 457; trad. it., 100). D'altra parte, se ci s'interroga sulle condizioni richieste per condurre una vita buona, è necessario conoscere il dato di partenza. La tesi generale di questo libro è che tanto le scienze umane quanto la conoscenza necessaria per lo sviluppo delle tecnologie possono contribuire alla comprensione del comportamento umano (Doris e Stich 2005, 115). Tale conoscenza, se sappiamo come analizzarla, ci permette di capire sia il modo in cui gli esseri umani sono, sia ciò di cui gli esseri umani hanno bisogno: un ordinamento etico delle vite che vivono insieme.

Edizioni ETS

Palazzo Rancioni - Lungarno Mediceo, 16, I-56127 Pisa

info@edizioniets.com - www.edizioniets.com

Finito di stampare nel mese di maggio 2025