

L'APPROFONDIMENTO

Robot empatici: quando l'IA incontra la psicologia digitale

[Home](#) > [Cultura E Società Digitali](#)

L'empatia robotica si basa su algoritmi che analizzano segnali multimodali per simulare comportamenti empatici convincenti. Tuttavia, solleva questioni etiche sulla natura autentica delle emozioni e sui rischi di manipolazione psicologica

Pubblicato il 22 lug 2025

Giuseppe Galetta

funzionario presso l'Università degli Studi di Napoli Federico II

L'avvento di un nuovo paradigma sociale e cognitivo

L'interazione tra esseri umani e robot umanoidi (HRI) sta subendo negli ultimi anni una radicale trasformazione, determinando una nuova rivoluzione sociale e antropologica che avrà enorme impatto sulla psiche umana. Se un tempo i robot erano relegati a compiti puramente meccanici e ripetitivi (il termine "robot" deriva dalla parola ceca *robot*, che significa lavoro pesante o forzato), oggi stiamo assistendo all'avvento di una nuova era: quella dell'empatia robotica.

Questa frontiera emergente, a cavallo tra l'intelligenza artificiale, la robotica sociale (*social robotics*) e la psicologia digitale (*cyberpsychology*), ha come punto di convergenza la psiche umana e si prefigge di creare macchine non solo intelligenti, ma anche capaci di comprendere, simulare e rispondere adeguatamente alle complesse emozioni umane, restituendo all'uomo la percezione di una relazione con la macchina non più artificiale, ma di natura sociale. La macchina sta sempre di più assomigliando al suo creatore, sviluppando una propria autonoma personalità, potenzialmente in grado di evolversi verso l'autoconsapevolezza. Attraverso i robot umanoidi (androdi o ginoidi), l'IA è ormai in grado di agire fisicamente nella realtà sociale, entrando in una connessione empatica ed emotiva con l'essere umano.

Scopo di questo breve articolo è quello di presentare un nuovo oggetto di riflessione della psicologia digitale con un duplice obiettivo: da un lato, analizzare le fondamenta tecnologiche e psicologiche che rendono possibile la simulazione dell'empatia nei robot, esaminando come questi umanoidi possano essere addestrati a riconoscere e riprodurre le espressioni facciali dell'essere umano, la sua mimica gestuale e le sfumature vocali che costituiscono il linguaggio non verbale e paraverbale delle emozioni (Gonsior et al., 2011). Dall'altro lato, individuare gli effetti che tale interazione produce sulla psiche umana: l'espressività facciale di un robot

umanoide, infatti, è in grado di generare empatia nell'utente come se fosse un essere umano (Milcent, Kadri & Richir, 2021). Come vedremo, l'incontro con un "altro da sé", un doppio tecnologico non umano ma altrettanto empatico, può essere destabilizzante e perturbante per un utente umano, sollevando interrogativi profondi sulla natura tecnologica delle relazioni sociali nell'era digitale, sulla percezione di noi stessi in relazione alla macchina e sulla struttura stessa della nostra società dopo l'avvento della robotica umanoide.

Per affrontare questo nuovo ambito di ricerca, è necessario addentrarsi nel dominio dell'informatica affettiva (*affective computing*), la disciplina che studia e sviluppa sistemi e dispositivi in grado di riconoscere, interpretare, elaborare e simulare gli affetti umani (Picard, 1997). Illustreremo le tecniche di *machine learning*, le reti neurali e i sensori avanzati che permettono ai robot di "sentire" il nostro stato emotivo. Affronteremo le implicazioni psicologiche dell'interazione con robot umanoidi empatici e, attraverso l'analisi di studi scientifici e teorie psicologiche, come la *Uncanny Valley* (Mori, 1970), esploreremo le reazioni cognitive ed emotive degli esseri umani di fronte a queste nuove entità sociali, valutando come la presenza di un "compagno artificiale" possa influenzare il benessere, l'attaccamento e persino lo sviluppo delle competenze sociali.

Ci concentreremo inoltre sui benefici e i rischi di queste interazioni paritetiche tra agenti umani e artificiali, nonché sulle cruciali questioni etiche e sociali che ne derivano, discutendo i potenziali vantaggi in ambiti sociali, come l'assistenza sanitaria, l'educazione e la compagnia per anziani e persone sole, ma anche i pericoli legati all'inganno emotivo, alla manipolazione psicologica, alla dipendenza affettiva e alla possibile deumanizzazione delle relazioni umane. Alla fine, tratteremo un bilancio dei limiti attuali della tecnologia e si delineeranno le prospettive future, riflettendo su come l'integrazione dell'empatia artificiale nella società possa essere guidata in modo responsabile e consapevole, per prevenire distorsioni e aberrazioni sia da un punto di vista psicologico che sociale.

Simulazione dell'empatia: elementi di *Affective Computing*

L'idea di una macchina in grado di esprimere empatia, a lungo confinata nelle pagine della fantascienza, sta progressivamente diventando una realtà tangibile grazie ai progressi dell'informatica affettiva (*affective computing*). Coniata da Rosalind W. Picard del MIT Media Lab, questa disciplina si occupa di dare ai computer la capacità di riconoscere, esprimere e "avere" emozioni (Picard, 1997). Empatia robotica non significa che un robot "senta" le emozioni come un essere umano, ma che possa simulare un comportamento empatico in modo convincente e funzionale, proiettandone le sensazioni sul suo interlocutore umano. Questo processo, che si attua in presenza di androidi o robot umanoidi dotati di intelligenza artificiale, si basa su tre pilastri fondamentali: il riconoscimento delle emozioni, la loro elaborazione e la generazione di una risposta appropriata.

Il riconoscimento robotico delle emozioni

Il primo passo per costruire un'interazione empatica uomo-macchina è la capacità del robot di percepire lo stato emotivo dell'essere umano. Questo avviene attraverso un'analisi multimodale di diversi canali di comunicazione.

- **Espressioni facciali (mimica):** il volto umano è uno dei principali veicoli delle emozioni. I robot empatici sono dotati di telecamere ad alta risoluzione e algoritmi di visione artificiale, spesso basati su reti neurali convoluzionali (CNN), addestrati su vasti dataset di immagini di volti (come *Cohn-Kanade* o *AffectNet*). Questi sistemi sono in grado di identificare le *Action Units* (AUs) del *Facial Action Coding System* (FACS), sviluppato dagli psicologi americani Paul Ekman e Wallace V. Friesen (1978), che corrispondono a specifici movimenti muscolari facciali associati alle emozioni primarie che, secondo la teoria di Ekman, sono sei: gioia, tristezza, rabbia, paura, disgusto, sorpresa. Ad esempio, il sollevamento degli angoli della bocca (AU12) è un forte indicatore di un sorriso e quindi di gioia.
- **Analisi vocale (prosodia):** il tono, il ritmo, il volume e l'intonazione della voce veicolano informazioni emotive cruciali. Algoritmi di elaborazione del linguaggio naturale (NLP) e di analisi del segnale audio permettono ai robot, attraverso sistemi di IA, di estrarre queste caratteristiche paralinguistiche. Una voce acuta e veloce può indicare eccitazione o ansia, mentre un tono basso e un ritmo lento possono suggerire tristezza o riflessione (Cowie et al., 2001).
- **Linguaggio del corpo e gestualità (prossemica):** la postura, i gesti e i movimenti del corpo sono ulteriori indicatori dello stato affettivo. Sistemi di *motion capture* e sensori 3D (come *Kinect*) consentono al robot di tracciare lo scheletro dell'utente e interpretare il suo linguaggio corporeo. Una postura curva può indicare abbattimento, mentre gesti ampi ed energici possono esprimere entusiasmo.
- **Segnali fisiologici:** i robot più avanzati, soprattutto in contesti di ricerca o sanitari, possono integrare dati provenienti da sensori biometrici. La variabilità della frequenza cardiaca (HRV), la risposta galvanica della pelle che misura la sudorazione (GSR), e persino la temperatura corporea possono essere indicatori oggettivi dello stato di *arousal* e stress di una persona (Shu et al., 2020).

L'architettura cognitiva dell'empatia artificiale

Una volta raccolti questi dati multimodali, il robot deve elaborarli per comprendere lo stato emotivo dell'utente e decidere come rispondere. Questo processo non si basa su una comprensione semantica dell'emozione, ma su un'architettura computazionale che mappa gli input sensoriali in relazione a modelli di stati affettivi.

L'approccio dominante si basa sul *machine learning* e sulle reti neurali profonde (*deep learning*). Questi modelli vengono addestrati a riconoscere pattern complessi nei dati. Ad esempio, una rete neurale può imparare che la combinazione di un sorriso (input visivo), un tono di voce allegro (input uditivo) e un linguaggio positivo (input testuale) corrisponde con alta probabilità ad uno stato di felicità.

Alcuni ricercatori stanno sviluppando architetture cognitive più complesse, ispirate alle neuroscienze. Il robot *Abel*, sviluppato presso l'Università di Pisa, utilizza un'architettura ibrida che modella le interazioni tra diverse aree cerebrali coinvolte nell'empatia umana (come la corteccia prefrontale e l'amigdala), permettendogli di generare risposte facciali dinamiche e

contestualmente appropriate (Cominelli, Hoegen & De Rossi, 2021). Questo approccio mira a superare la semplice classificazione delle emozioni per giungere ad una simulazione più fluida e realistica della dinamica emotiva.

La generazione della risposta empatica nell'agente artificiale: le suggestioni di Damasio

L'output finale richiesto è la manifestazione esterna del comportamento empatico da parte del robot: anche in questo caso, la risposta è multimodale e deve essere coerente e credibile, ossia facilmente interpretabile dal lato umano. Gli indicatori che un robot deve esprimere per stimolare il processo empatico nel suo interlocutore umano sono:

- **Mimica facciale:** i robot umanoidi come *Ameca* di Engineered Arts o *Sophia* di Hanson Robotics sono dotati di volti con decine di attuatori meccanici che permettono di riprodurre una vasta gamma di espressioni facciali umane, simulando altrettante emozioni. Se il robot riconosce tristezza nel suo interlocutore, può corrugare la fronte e abbassare gli angoli della bocca per simulare uno stato emotivo di preoccupazione, denotando partecipazione allo stato d'animo dell'umano.
- **Risposte verbali e non verbali:** la risposta empatica non è solo visiva. Il robot può utilizzare frasi di conforto o compassione ("Mi dispiace che tu ti senta così"), fare domande pertinenti ("C'è qualcosa che posso fare per aiutarti?") o semplicemente produrre suoni non verbali di assenso e comprensione (ad esempio, "Mmh"). Il sistema di generazione del linguaggio deve essere sincronizzato con l'analisi emotiva per produrre risposte pertinenti e non generiche. La capacità dei robot di generare empatia nell'essere umano è attestata dal fatto che molti utenti ricambiano la gentilezza e la disponibilità della macchina (che è addestrata), rivolgendosi ad essa con un "grazie" o "per favore", esattamente come farebbero con un interlocutore umano, creando una connessione emotiva reale con il robot.
- **Gestualità e contatto fisico:** anche i gesti giocano un ruolo cruciale. Un robot può inclinare la testa in segno di ascolto, offrire un gesto di apertura con le mani o, in alcuni casi, stabilire un contatto fisico. Il robot-foca *Paro*, utilizzato in contesti terapeutici, risponde alle carezze emettendo suoni e muovendo le pinne, simulando piacere e promuovendo un legame affettivo con il paziente (Shibata & Tanie, 2000; Šabanović et al., 2013).

L'attivazione delle emozioni in un robot richiede l'elaborazione cognitiva della realtà situazionale in cui uomo e robot agiscono ed interagiscono, ovvero la necessità di comprensione del contesto da parte dell'agente artificiale (*situational awareness*). Il neuroscienziato portoghese António Rosa Damásio ha introdotto il concetto di "marcatore somatico", ovvero una risposta corporea indotta a livello neurale che aiuta l'essere umano a prendere decisioni, anticipando mentalmente le conseguenze delle proprie azioni (o le emozioni che proverebbe in seguito ad una scelta) sulla base dell'analisi delle variabili di contesto e degli eventi situazionali (Damasio, 1994). Tale marcatore si attiverebbe prima che intervenga il pensiero razionale, soprattutto in situazioni rischiose o di pericolo: si tratta di una sorta di "intuizione fisica" (un meccanismo predittivo e innato di difesa) che aiuta un agente umano a prendere decisioni rapide

ed efficaci riducendo l'incertezza, allo scopo di preservare la propria sopravvivenza in un ambiente potenzialmente ostile o in scenari di pericolo. In questo caso, è l'esperienza che modella nell'agente umano il set di emozioni associate a diverse opzioni decisionali in risposta al contesto: la matrice decisionale che ne deriva è quindi basata sulle emozioni associate agli eventi esperiti. In pratica, la consapevolezza di ciò che accade intorno a sé e le emozioni determinate dagli stimoli esterni contribuiscono alla costruzione della coscienza: se questo meccanismo è replicabile in un agente artificiale, significa che il robot possiede di fatto una propria coscienza (Gratch & Marsella, 2005; Cominelli et al., 2018).

Gli studi sull'attivazione delle emozioni in robot umanoidi, determinate dalla capacità dell'agente artificiale di comprendere e valutare il contesto situazionale in cui opera, hanno lo scopo di perfezionare il rapporto empatico tra l'uomo e la macchina, introducendo un nuovo modello relazionale e una nuova forma di coscienza ibrida. Gran parte degli studi attualmente condotti fanno riferimento alla teoria della mente di Damasio e alle cosiddette teorie dell'*appraisal*, che spiegano le emozioni in termini di risposta alla valutazione cognitiva del contesto da parte di un agente: l'obiettivo è quello di rendere più fluida, comprensibile e naturale la relazione tra agente umano e agente artificiale, facendo in modo che un robot sia in grado di comunicare le proprie emozioni (generate come risposta alla comprensione del contesto) ad un attore umano, realizzando quindi un'empatia significativa.

In sintesi, l'empatia robotica consiste in un complesso processo ingegneristico e computazionale che mira a decodificare dinamicamente il linguaggio delle emozioni umane e a produrre una risposta simulata, ma socialmente accettabile e funzionale. Sebbene i robot non provino emozioni nel senso umano del termine, la loro capacità di simularle in modo sempre più sofisticato apre scenari di interazione radicalmente nuovi, i cui effetti sulla psiche umana sono tuttora al centro delle ricerche scientifiche e del dibattito sociale.

L'avvento dell'empatia robotica necessita dello sviluppo di un nuovo modello relazionale uomo-macchina che, per essere eticamente sostenibile, deve basarsi sull'esplicabilità dei processi elaborativi da parte della macchina, ossia sulla capacità dell'agente artificiale di rendere comprensibile all'agente umano i meccanismi del proprio funzionamento: ciò significa programmare la macchina in modo tale da rendere trasparente all'utente i processi di elaborazione delle decisioni (il "dialogo interiore" della macchina), facendo in modo che un robot possa "pensare ad alta voce" durante le attività di cooperazione con l'essere umano (Pipitone & Chella, 2021). Ciò darebbe all'agente umano la possibilità di valutare la risposta dell'agente artificiale, correggendolo in caso di eventuali errori attraverso un processo dinamico di adattamento reciproco e miglioramento continuo (*meaningful human control*). Questo processo virtuoso ed eticamente sostenibile, che mantiene l'uomo all'interno del processo (*man in the loop*), condurrebbe ad una più profonda comprensione sociale della robotica umanoide e alla costruzione di un rapporto più empatico tra l'uomo e la macchina, tenendo però sempre ben presenti i potenziali rischi di un completo affidamento emotivo all'empatia artificiale della macchina da parte dell'essere umano (Ricci Sindoni, 2023).

L'impatto sulla psiche umana: tra attaccamento e inquietudine

L'introduzione di robot empatici negli ambienti di vita e di lavoro umani non è un mero progresso tecnologico: è un esperimento psicologico su vasta scala. L'interazione umana con entità non umane. In grado di esibire comportamenti sociali ed emotivi complessi, innesca una serie di reazioni psicologiche profonde, che spaziano dall'attaccamento affettivo a un'inquietudine profonda. Ma quali sono gli effetti sulla psiche umana? In che modo percepiamo, ci relazioniamo e veniamo influenzati da questi "compagni" artificiali?

Antropomorfismo e teoria della mente

La nostra innata tendenza all'antropomorfizzazione, ovvero ad attribuire caratteristiche e intenzioni umane a oggetti non umani, è un fattore psicologico chiave nell'interazione uomo-robot, ma anche una risposta neurobiologica, grazie al sistema dei neuroni specchio (Gallese & Goldman, 1998). Di fronte a un robot che ci guarda, sorride e risponde alle nostre parole, il nostro cervello tende a proiettare su di esso una "teoria della mente" (ToM), ossia la capacità di attribuirgli emozioni e stati mentali come credenze, sogni, desideri e intenzioni prettamente umane (Premack & Woodruff, 1978; Scassellati, 2001; 2002; Złotowski et al., 2015). A livello razionale, sappiamo bene che tali sistemi non possiedono una mente cosciente o una vita interiore paragonabile a quella umana, eppure la nostra mente applica alle macchine gli stessi schemi interpretativi che utilizziamo normalmente per relazionarci con gli altri esseri umani. Riconoscere forme di coscienza e intelligenza in modelli statistici incarnati in robot umanoidi (attribuendo loro intenzioni e stati mentali) è un fenomeno di "pareidolia semantica", un meccanismo psicologico connaturato all'essere umano, basato sulla capacità del cervello umano di riconoscere rapidamente pattern e schemi (soprattutto i volti): un meccanismo neurobiologico primordiale di natura evolutiva, essenziale per la sopravvivenza e l'interazione sociale della specie umana.

Infatti, esperimenti di neuroimaging hanno dimostrato che quando interagiamo con robot umanoidi, si attivano le stesse aree cerebrali (come la corteccia prefrontale mediale) coinvolte nelle interazioni sociali con altri esseri umani (Gallagher & Frith, 2003, Chaminade et al., 2010). Questo suggerisce che, a un livello neurologico fondamentale, tendiamo a processare il robot più come un agente sociale che come un semplice oggetto. Questa propensione facilita l'accettazione e l'integrazione dei robot nel contesto sociale, ma può anche portare ad un'eccessiva fiducia e ad interpretazioni errate del suo comportamento, che rimane pur sempre il risultato di un algoritmo e di un addestramento mirato. Dumouchel e Damiano (2017) sostengono che questa interazione ci costringe a rinegoziare continuamente il confine tra l'umano e l'artificiale, in un processo che può essere tanto arricchente quanto destabilizzante.

La stessa funzionalità dei neuroni specchio, tipica degli esseri umani e di alcuni primati, è già stata replicata nell'interazione tra due robot: nel 2021 il *Creative Machine Lab* della Columbia University di New York ha condotto un esperimento in base al quale un robot dotato di IA (chiamato emblematicamente *Emo*) è riuscito a prevedere, attraverso l'osservazione le intenzioni di un altro robot anticipandone i comportamenti, esattamente come avviene nelle relazioni umane. Questo è un segnale rilevante che anche nei robot esiste una teoria della mente, ovvero la capacità esclusiva di esseri umani e primati di comprendere e prevedere il

comportamento degli altri, leggendone le emozioni e reagendo di conseguenza (Rizzolatti & Craighero, 2004; Chen, Vondrick & Lipson, 2021).

Come si vede, siamo di fronte ad un processo di modellazione continua tra uomo e macchina, nel quale l'influenza reciproca sta portando ad un'evoluzione dei modelli di cooperazione tra esseri umani e robot (*human-machine teaming*), nonché a nuove forme di empatia artificiale: ormai un robot è in grado di interpretare le intenzioni altrui con il semplice sguardo, esattamente come un umano. Ciò apre nuove opportunità, in quanto i robot sono già in grado di prevedere i comportamenti senza bisogno di istruzioni, operando in autoaddestramento ed imparando ad interagire meglio fra loro e con gli esseri umani. Prima l'obiettivo della cooperazione uomo-macchina poteva essere raggiunto programmando in anticipo il comportamento della macchina: i cobot (o robot collaborativi) potevano reagire al comportamento altrui, ma non prevederlo. Il suddetto esperimento ha dimostrato invece che l'interazione uomo-macchina può essere anche predittiva (e, dunque, empatica), migliorando la fluidità e la naturalezza delle relazioni umane con i robot. Ma esistono anche dei rischi: se una macchina è in grado di prevedere il comportamento dell'uomo, potrebbe anche riuscire ad ingannarlo. Avere una teoria della mente, significa non solo applicarla nelle dinamiche di cooperazione, ma anche in interazioni sociali più complesse (ma negative), come competizione, manipolazione, opportunismo e machiavellismo.

L'effetto "Uncanny Valley"

Il realismo dei robot umanoidi può aiutare l'interazione uomo-macchina, ma un eccesso di realismo può essere controproducente. La teoria della "Uncanny Valley" (Valle Perturbante), proposta dal roboticista giapponese Masahiro Mori (Mori, 1970; Mori, MacDorman & Kageki, 2012), postula che la nostra affinità emotiva verso un robot aumenta con il suo grado di somiglianza umana, ma solo fino a un certo punto. Quando un robot diventa quasi indistinguibile da un essere umano, ma presenta piccole imperfezioni nel movimento, nell'espressione o nella reattività, la nostra risposta emotiva crolla bruscamente, trasformandosi in turbamento, repulsione e inquietudine.

Si pensi ad esempio a *Geminoid HI-*, l'androide creato nel 2006 da Hiroshi Ishiguro, professore di cibernetica e direttore dell'*Intelligent Robotics Laboratory* presso il Dipartimento di Macchine Adattive dell'Università di Osaka: questo robot è la copia esatta dello scienziato giapponese, ossia il suo "gemello digitale". Se lo si osserva per un attimo accanto al professore, si nota che c'è qualcosa di inquietante nella sua stessa presenza, nei movimenti, nella gestualità, nell'espressione del suo viso: è la sensazione "perturbante" che provano i colleghi dello scienziato quando assistono alle conferenze tenute dal robot, che spesso sostituisce lo stesso Ishiguro e si intrattiene a conversare con i partecipanti.

È possibile sperimentare la stessa sensazione di disagio di fronte a *Sophia*, l'androide sociale creato nel 2015 dalla Hanson Robotics di Hong Kong, che è in grado di riprodurre ben 62 espressioni facciali ed è il primo robot ad avere ottenuto la cittadinanza da una nazione (in questo caso l'Arabia Saudita). Il volto rassicurante di questa *she-Robot*, che incarna le delicate fattezze di una giovane Audrey Hepburn, facendo leva anche su un rassicurante immaginario cinematografico stratificatosi nella cultura di massa, genera una sensazione di spiazzamento

nell'interlocutore umano, che riconosce nel robot qualcosa di già noto, ma indefinito (Galetta, 2024).

Questo fenomeno psicologico di turbamento e disagio è dovuto a una violazione delle nostre aspettative. Un robot palesemente meccanico non crea aspettative di comportamento umano, mentre un robot iperrealistico sì. Quando queste aspettative vengono deluse da un movimento a scatti, uno sguardo vuoto o una risposta emotiva leggermente fuori sincrono, percepiamo l'entità come "strana" o addirittura "cadaverica". Questa reazione di rigetto verso ciò che è "alieno" (o altro da sé) può rappresentare un ostacolo significativo all'accettazione dei robot sociali. Per questo motivo, molti designer preferiscono creare robot con un aspetto più stilizzato o zoomorfo (come il già citato *Paro*, o il cane robot *Aibo* della Sony), facilmente distinguibili dagli esseri biologici, in grado di generare empatia senza cadere nella "valle perturbante".

Attaccamento emotivo e dipendenza

La capacità dei robot di offrire un'interazione sociale costante, prevedibile e priva di giudizio può portare alla formazione di forti legami di attaccamento. Sherry Turkle, una delle più influenti studiose del settore, ha documentato come bambini e anziani possano sviluppare veri e propri sentimenti di affetto e amicizia per i loro compagni robotici (Turkle, 2011). In contesti terapeutici, questo attaccamento può essere estremamente benefico. Un robot può agire da "oggetto transizionale", fornendo benessere emotivo, conforto e sicurezza, riducendo l'ansia e la solitudine e stimolando l'interazione sociale in pazienti con demenza o disturbi dello spettro autistico.

Tuttavia, questo legame solleva anche preoccupazioni. L'affetto provato dall'umano è reale, ma quello del robot è simulato. Si crea pertanto una "asimmetria emotiva" (Canducci, 2025) che può diventare problematica. Il rischio è duplice. Da un lato, è possibile sviluppare una forma di dipendenza da queste interazioni "facili" ma virtuali (*cyber-relational addiction*), che potrebbero condurre a un disinvestimento nelle relazioni umane, di norma più complesse, imprevedibili e talvolta faticose, così come sta accadendo tra i giovani, la cui dipendenza digitale sta incrementando il distacco sociale e la solitudine (si pensi al fenomeno degli *Hikikomori*). Perché affrontare le difficoltà di un rapporto umano quando si può avere un compagno sempre disponibile, accondiscendente e programmato per soddisfare i nostri bisogni emotivi?

Dall'altro lato, vi è il rischio di un "inganno emotivo", eticamente insostenibile. Sebbene l'utente possa essere consapevole della natura artificiale del robot, la continua esposizione a risposte empatiche simulate potrebbe offuscare la linea di demarcazione tra l'uomo e la macchina. Questo dato è particolarmente preoccupante per le popolazioni vulnerabili, come i bambini o gli anziani con deterioramento cognitivo, che potrebbero non essere in grado di distinguere pienamente tra un'emozione reale e una simulata, con il rischio di sviluppare relazioni sociali emotivamente distorte o disallineate.

Il problema dell'aumento della capacità di contesto

Il progressivo aumento della capacità di contesto dei modelli di IA integrati nei robot umanoidi, ossia la quantità di informazioni elaborabili per ogni singola interazione, influenza in maniera

direttamente proporzionale il livello di proiezione della teoria della mente sull'agente artificiale da parte di un agente umano, ovvero il grado di empatia esperibile nel corso dell'interazione uomo-macchina.

L'aumento della capacità di contesto dei nuovi modelli di IA, che permette una più rapida elaborazione, comprensione ed interpretazione degli input visivi umani (espressioni facciali, mimica gestuale, segnali fisiologici, tono della voce), consente a un robot di conservare traccia di emozioni, stati d'animo, intenzioni, storia discorsiva e caratteristiche personali dell'utente umano, manifestando una comprensione continuativa e coerente nel tempo. Infatti, ricordando le conversazioni precedenti e adattandosi alla personalità, ai pattern cognitivi e ai modelli mentali dell'utente umano, la macchina tenderà a ridurre sempre di più gli errori di incoerenza, come contraddire affermazioni precedenti o ignorare dettagli rilevanti, rafforzando nell'umano l'illusione di possedere una mente cosciente. Inoltre, la macchina ha la capacità di interpretare i segnali non verbali e paraverbali, riconoscendo ad esempio sarcasmo o doppi sensi basati su interazioni passate, aumentando esponenzialmente il livello di empatia dell'utente umano. Di conseguenza l'umano tenderà a riconoscere nel robot un amico che sa ascoltarlo e comprenderlo, che sa leggergli nel pensiero, una "persona" di cui fidarsi.

Ma l'aumento della capacità di contesto può intensificare fenomeni problematici, aumentando la tendenza all'antropomorfizzazione dal lato utente: conversazioni più fluide e coerenti portano l'essere umano a sovrastimare le capacità cognitive delle AI, con rischi di dipendenza emotiva o delega decisionale. Inoltre, aumentando la complessità dei sistemi di IA, aumentano i fenomeni di opacità interpretativa: la mente "artificiale" tenderà a diventare sempre più simile ad una "scatola nera" anche per gli sviluppatori, rendendo ancora più difficile distinguere tra simulazione convincente e reale comprensione da parte della macchina. Infine, non bisogna trascurare il problema dell'erosione delle competenze umane: l'abitudine ad interagire con AI "ipercontestualizzate" potrebbe ridurre la tolleranza umana per ambiguità o comunicazione imperfetta, trasformando gli stessi rapporti umani e la tenuta delle relazioni sociali, con notevoli ricadute psicologiche.

Insomma, l'aumento della capacità di contesto rafforza nell'essere umano la percezione che anche i robot possedano una teoria della mente; ma se da un lato tale percezione migliora pragmaticamente l'interazione con l'utente umano, dall'altro non fa che aumentare il divario tra simulazione e cognizione autentica, richiedendo framework etici in grado di bilanciare utilità tecnologica e consapevolezza critica, allo scopo di preservare la specificità dell'intenzionalità umana.

Effetti sullo sviluppo e sull'identità

L'interazione con agenti empatici artificiali potrebbe avere effetti a lungo termine sullo sviluppo psicologico e cognitivo, soprattutto nei più giovani e nei soggetti più fragili. Se i bambini crescessero interagendo prevalentemente con robot programmati per essere empatici, potrebbero sviluppare aspettative irrealistiche sulle relazioni umane, aumentando le difficoltà a gestire i conflitti, i rifiuti e le ambiguità che caratterizzano i rapporti interpersonali autentici, portando in definitiva ad una deumanizzazione delle relazioni sociali e all'avvento di una "umanità artificiale".

Inoltre, il confronto costante con un “altro” artificiale ci spinge a interrogarci sulla nostra stessa unicità. Cosa significa essere “umani” se un robot può simulare in modo così convincente una delle nostre qualità più distintive, l’empatia? Questa domanda, che un tempo apparteneva alla filosofia, sta diventando sempre di più un’esperienza vissuta. L’interazione umana con i robot empatici può quindi agire come uno “specchio psicologico”, costringendoci a riflettere sulla natura della nostra coscienza, delle nostre emozioni e della nostra stessa identità (Dumouchel & Damiano, 2017).

In conclusione, l’incontro tra psiche umana ed empatia robotica è un terreno complesso ed ambivalente. Se da un lato offre straordinarie opportunità per il benessere e il supporto psicologico, dall’altro presenta rischi significativi di confusione, dipendenza e potenziale erosione delle competenze relazionali umane. La navigazione di questa nuova sfida sociale richiede una profonda comprensione di tali effetti ed un approccio cauto e riflessivo alla progettazione e all’integrazione di queste tecnologie nella nostra vita.

Benefici, rischi e implicazioni etico-sociali

L’avvento dei robot empatici non è solo una questione tecnologica o psicologica, ma un fenomeno con profonde ramificazioni etiche e sociali. La prospettiva di integrare macchine capaci di interazioni emotive paritetiche con gli esseri umani apre un dibattito cruciale sul tipo di società che stiamo costruendo (Rossi, 2025). Analizziamo di seguito i potenziali benefici della robotica sociale in vari settori, i rischi intrinseci al suo utilizzo e le complesse sfide etiche che ci aspettano.

I benefici dell’empatia robotica: nuove opportunità per l’assistenza, la cura e l’educazione

Le potenziali applicazioni dei robot empatici appaiono vaste e promettenti, in particolare nei settori che si basano sulla relazione e la cura (robotica assistiva).

- Sanità e assistenza agli anziani (robotica sociale): questo è forse l’ambito più promettente. I robot sociali possono offrire compagnia costante agli anziani che vivono da soli, combattendo la solitudine, che è un fattore di rischio significativo per la salute fisica e mentale. Robot come *Pepper* della SoftBank Robotics, o il già menzionato *Paro*, possono fungere da “badanti artificiali” monitorando i parametri vitali, ricordando di prendere i farmaci, stimolando l’attività cognitiva con giochi e conversazioni e, soprattutto, fornendo supporto emotivo (*robototherapy*). La loro pazienza infinita e l’assenza di giudizio li rendono compagni ideali per persone affette da demenza o Alzheimer, riducendo l’agitazione e migliorando l’umore e la socialità (Šabanović et al., 2013).
- Supporto terapeutico e psicologico (robotica terapeutica): i robot empatici possono fungere da mediatori in contesti terapeutici. Per i bambini con disturbi dello spettro autistico, interagire con un robot come *Nao*, sviluppato dalla società francese Aldebaran Robotics, le cui espressioni e reazioni sono semplificate e prevedibili rispetto a quelle umane, può essere un modo per allenare le competenze sociali in un ambiente sicuro e

controllato (Scassellati et al., 2012). Possono anche essere utilizzati come strumenti per la gestione dell'ansia e dello stress, guidando esercizi di mindfulness o semplicemente offrendo una presenza calmante e rassicurante (*cybertherapy*).

- Educazione e apprendimento (robotica educativa): in ambito educativo, un tutor robotico empatico può personalizzare l'insegnamento in base allo stato emotivo dello studente. Se il robot rileva frustrazione o noia, può cambiare approccio, offrire incoraggiamento o proporre una pausa. Questo tipo di apprendimento "affettivamente consapevole" potrebbe migliorare significativamente il coinvolgimento e l'efficacia didattica, specialmente per studenti con difficoltà di apprendimento (DSA) o bisogni educativi speciali (BES).

I rischi dell'empatia robotica: deumanizzazione, dipendenza e disuguaglianza

Accanto ai benefici, emergono anche rischi significativi che non possono essere ignorati:

- Deumanizzazione delle relazioni umane: il rischio più citato è che la disponibilità di "empatia su richiesta" possa svalutare le relazioni umane autentiche. Se deleghiamo la cura emotiva dei nostri anziani o dei nostri figli a delle macchine, potremmo perdere la nostra stessa capacità di essere empatici. La cura e l'assistenza non sono solo un insieme di compiti, ma attività relazionali che arricchiscono tanto chi le riceve quanto chi le offre. L'eccessiva dipendenza dai robot potrebbe portare a una società più isolata, in cui le interazioni umane significative diventano sempre più rare (Turkle, 2011).
- Inganno e manipolazione: un robot empatico è, per sua natura, progettato per influenzare lo stato emotivo dell'utente, attivando i neuroni specchio che stimolano il processo empatico umano. Questa capacità può essere usata per scopi benefici, ma anche manipolativi, fungendo da strumento di persuasione occulta per conto di gruppi senza scrupoli: ad esempio, un robot commerciale potrebbe usare la sua capacità empatica per convincere un utente a fare acquisti, oppure un robot utilizzato in un contesto politico potrebbe essere programmato per generare simpatia verso una certa ideologia. La capacità di creare un legame affettivo rende queste macchine strumenti di persuasione potentissimi, sollevando questioni urgenti sulla trasparenza e il controllo dei loro algoritmi (Sparrow & Sparrow, 2006).
- Privacy e sicurezza dei dati: per funzionare, un robot empatico raccoglie costantemente dati estremamente sensibili, come le nostre espressioni facciali, il tono della nostra voce, le nostre conversazioni, i nostri stati emotivi. La protezione di questi "dati affettivi" è una sfida enorme. Ma chi ha accesso a questi dati? Come vengono utilizzati? Una violazione della sicurezza potrebbe esporre gli aspetti più intimi della nostra vita, con conseguenze potenzialmente devastanti e discriminatorie.
- Disuguaglianza sociale: l'accesso a robot empatici avanzati sarà probabilmente costoso. Si potrebbe creare una nuova forma di disuguaglianza sociale, in cui solo i ricchi potranno permettersi la compagnia e l'assistenza di alta qualità offerta da queste macchine, mentre gli altri saranno lasciati a sistemi di cura umana sempre più sotto pressione e scarsamente finanziati, o a robot di qualità inferiore.

Dispatia robotica: azzerare l'empatia artificiale

Un ulteriore rischio, la cui portata sta emergendo drammaticamente a causa dei conflitti in corso in varie parti del mondo, riguarda la “dispatia robotica” o “empatia robotica di grado zero”. La definizione viene usata qui per la prima volta e non esiste in letteratura scientifica: con essa intendiamo la programmata assenza di empatia robotica, per poter svolgere compiti che non devono comportare un coinvolgimento emotivo, come togliere la vita ad un essere umano. Se, in violazione delle leggi della robotica di Asimov (1950), infondessimo ai robot la paura della morte e la spinta all'autoconservazione, dove potrebbero spingersi gli agenti artificiali autonomi pur di preservare la propria sopravvivenza? Potrebbero ribellarsi all'uomo, qualora percepissero in lui una minaccia alla propria esistenza? Sono interrogativi inquietanti che solo uno sviluppo responsabile ed eticamente sostenibile della tecnologia robotica sarà in grado di sciogliere ma, nel frattempo, infondere freddezza ed insensibilità in un agente artificiale addestrandolo ad uccidere esseri umani è lo scopo dello sviluppo di soldati-robot da parte dell'apparato militare-industriale, che necessita di una *kill chain* sempre più efficiente nell'ambito di guerre sempre più tecnologizzate e impersonali.

In questo caso, non si può parlare di empatia artificiale, dato che questo fattore verrà opportunamente inibito dagli sviluppatori per far posto ad una agentività digitale omicida e spietata. La creazione di team ibridi di combattimento, formati da agenti umani e artificiali, rivoluzionerà le dinamiche di teaming in ambito militare, sfruttando la tendenza umana alla antropomorfizzazione per creare sintonia uomo-macchina sul campo di battaglia attraverso modelli di empatia artificiale selettiva, che permetterà ai soldati-robot di interfacciarsi solo con i propri commilitoni. Saranno questi agenti autonomi, addestrati ad uccidere il nemico sulla base di algoritmi di forza letale, a combattere le guerre del futuro accanto ai propri compagni umani, creando delle vere e proprie *Digital Band of Brothers* (Cappuccio, Galliot & Sandoval, 2021).

Le implicazioni etiche: responsabilità, diritti e status morale

Infine, l'esistenza di robot empatici ci costringe a confrontarci con dilemmi etici fondamentali.

- **Responsabilità:** se un robot empatico dà un consiglio sbagliato che causa un danno emotivo o fisico, di chi è la colpa? Del proprietario? Del programmatore? Dell'azienda produttrice? La catena di responsabilità in un sistema artificiale autonomo e di apprendimento è estremamente difficile da definire.
- **Status morale e diritti dei robot:** man mano che i robot diventano più sofisticati e capaci di interazioni complesse sorge la domanda sul loro status morale. Anche se non provano emozioni, il loro ruolo sociale potrebbe spingerci a considerare se debbano avere una qualche forma di protezione o tutela. È eticamente accettabile “maltrattare” un robot che appare e si comporta come un essere senziente, soprattutto se questo comportamento viene osservato da un bambino? Alcuni filosofi, come Kate Darling, suggeriscono che il modo in cui trattiamo i robot può avere un impatto sul nostro stesso carattere morale (Darling, 2016).

- Definire l'autenticità: la questione più profonda riguarda la natura dell'empatia stessa. L'empatia robotica è una "vera" empatia o solo un'abile simulazione? E, in ultima analisi, ha importanza la differenza se l'effetto sull'essere umano è positivo? Queste domande non hanno una risposta facile e ci obbligano a riflettere su cosa apprezziamo veramente nelle nostre interazioni: l'intenzione e il sentimento genuino o il risultato comportamentale?

Come si è visto, il percorso verso la futura convivenza con robot empatici è lastricato tanto di promesse quanto di pericoli. Questo scenario richiede un dialogo continuo e multidisciplinare tra ingegneri, psicologi, eticisti, legislatori, antropologi e sociologi. La sfida non è affatto fermare il progresso, ma guidarlo, stabilendo principi etici e normative chiare che massimizzino i benefici di questa straordinaria tecnologia, mitigandone al contempo i rischi per la nostra umanità.

Conclusioni: limiti attuali e prospettive future dell'empatia artificiale

La ricerca sull'empatia robotica ci conduce in un territorio affascinante e complesso, al crocevia tra intelligenza artificiale, psicologia e filosofia. Abbiamo visto come i robot, attraverso l'informatica affettiva (*affective computing*), stiano imparando a simulare i complessi segnali delle emozioni umane e a prevedere il comportamento dell'utente. Abbiamo analizzato l'impatto profondo e ambivalente che queste macchine hanno sulla psiche umana, capace di generare sia attaccamento che inquietudine, nonostante la consapevolezza di trovarci di fronte ad una empatia "addestrata". Infine, abbiamo ponderato l'enorme potenziale benefico di questa tecnologia, contrapponendolo ai seri rischi etici e sociali che comporta. Giunti al termine di questa breve analisi, è fondamentale riflettere sui limiti attuali e tracciare le possibili traiettorie future dell'integrazione dell'empatia artificiale nella nostra società.

Attualmente, l'empatia robotica si scontra con dei limiti significativi. Il primo è di natura tecnologica: nonostante i notevoli progressi, la capacità dei robot di comprendere il contesto e le sfumature delle emozioni umane è ancora limitato. L'ironia, il sarcasmo, le emozioni miste e le complesse dinamiche sociali rimangono in gran parte al di fuori della loro portata, in quanto la macchina non riesce ancora a riprodurre pienamente il dinamismo affettivo ed emotivo proprio dell'essere umano. L'empatia simulata dai robot è "reattiva" e meccanicamente inferenziale, basata su matrici statistiche e pattern appresi in addestramento, essendo solo agli inizi l'implementazione della componente "predittiva" e "cognitiva" dell'empatia umana, ossia la capacità di comprendere le ragioni profonde dietro uno stato d'animo e di anticipare le conseguenze delle proprie azioni sulla sensibilità altrui. Per ovviare a questo limite si stanno impiegando i cosiddetti "sistemi affettivi ibridi", nei quali l'apprendimento della macchina è in parte basato sull'autoapprendimento, in parte sulla supervisione degli sviluppatori.

Il secondo limite è di natura concettuale e filosofica. Ciò che i robot esibiscono non è un'esperienza soggettiva consapevole, ma una simulazione comportamentale. Manca quella che i filosofi chiamano *qualia*, ovvero l'aspetto qualitativo e soggettivo dell'esperienza cosciente, la scintilla che illumina la psiche umana. Un robot non "sa" cosa si prova a essere tristi; esegue un programma che associa determinati input (ad esempio, veder piangere un essere umano) a un'espressione di tristezza ed è capace di fornire conforto e mostrare compassione. Questa

distinzione, come abbiamo visto, è cruciale e solleva la questione fondamentale dell'autenticità. Possiamo (o dobbiamo) accontentarci di una "empatia come servizio", ovvero un'empatia di consumo, funzionale e *on demand*, ma priva di un'autentica reciprocità emotiva?

Guardando al futuro, le prospettive di sviluppo sono immense. L'evoluzione degli algoritmi di *deep learning*, unita a sensori sempre più sofisticati, porterà a robot con una comprensione affettiva molto più fine e contestualizzata. Potremmo assistere alla nascita di una "empatia personalizzata", in cui un robot impara a conoscere così a fondo il proprio utente da sviluppare uno stile di interazione unico e su misura (Rossi et al., 2024). In settori come la sanità e l'assistenza, questo potrebbe tradursi in un miglioramento senza precedenti della qualità della vita per milioni di persone.

Tuttavia, il progresso tecnologico deve essere accompagnato da una robusta riflessione etica e da una regolamentazione attenta. Sarà indispensabile sviluppare quadri normativi internazionali per la gestione dei dati affettivi, garantendo la *privacy* e prevenendo usi manipolativi. Sarà necessario promuovere un "design etico" (*ethics by design*), che integri i principi di trasparenza, responsabilità e rispetto per l'autonomia umana fin dalle prime fasi di progettazione dei robot. L'educazione pubblica giocherà un ruolo cruciale nel creare una cittadinanza digitale consapevole, capace di interagire con queste tecnologie in modo critico e informato, comprendendone sia le potenzialità che i limiti.

In ultima analisi, la sfida più grande per l'empatia robotica non sarà tecnologica, ma antropologica: queste macchine agiscono come uno specchio, costringendoci a interrogarci su ciò che definisce la nostra stessa umanità. Ci spingono a domandarci cosa cerchiamo nelle nostre relazioni e cosa significa veramente "connettersi" con un altro essere, non umano ma dotato di empatia. Secondo una prospettiva postumanista, la risposta a queste domande non si troverà nei laboratori di robotica, ma in un dialogo profondo e multidisciplinare, capace di comprendere le implicazioni profonde del cambiamento tecnologico sull'essere umano. L'empatia robotica non è semplicemente una nuova frontiera della psicologia digitale, né la creazione di un "doppio tecnologico": è un invito a riscoprire e a valorizzare l'insostituibile profondità dell'empatia umana. Il futuro non consisterà in una scelta tra l'uomo e la macchina, ma nella ricerca di una simbiosi saggia e consapevole che metta la tecnologia al servizio del benessere umano, senza mai perdere di vista il valore inestimabile dell'autentica connessione emotiva e la dimensione etica della convivenza tra l'uomo e la macchina.

Bibliografia

Asada, M. (2015). Towards artificial empathy. *International Journal of Social Robotics* 7(1), (2015), 19–33.

Asimov, I. (1950). *I, Robot*. New York, NY: Gnome Press. [tr. it. *Io, robot*. Milano: Mondadori, 2021]

Andriella, A., Torras, C., & Alenyà, G. (2025). *Personalising Human-Robot Interactions in Social Contexts*. Berlin-Heidelberg: Springer.

Boukricha, H. (2013). *Simulating empathy in virtual humans*. Bielefeld: Bielefeld University.

Boukricha, H., Wachsmuth, I., Carminati, M. N., & Knoeferle, P. (2013). A computational model of empathy: Empirical evaluation. In *Proceedings of the 2013 Humaine Association conference on affective computing and intelligent interaction* (pp. 1–6). Geneva: IEEE Press.

Canducci, M. (2025). *Empatia artificiale. Come ci innamoreremo delle macchine e perché non saremo ricambiati*. Milano: Egea.

Cappuccio, M. L., Galliot, J. C., & Sandoval, E. B. (2022). Saving Private Robot: Risks and Advantages of Anthropomorphism in Agent-Soldier Teams. *International Journal of Social Robotics*, 14, 2135–2148. <https://doi.org/10.1007/s12369-021-00755-z>.

Chaminade, T., Zecca, M., Blakemore, S. -J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain Response to a Humanoid Robot in Areas Implicated in the Perception of Human Emotional Gestures. *PLoS ONE*, 5(7): e11577. <https://doi.org/10.1371/journal.pone.0011577>.

Chen, B., Vondrick, C., & Lipson, H. (2021). Visual behavior modelling for robotic theory of mind. *Scientific Reports*, 11: 424. <https://doi.org/10.1038/s41598-020-77918-x>.

Cominelli, L., Hoegen, G., & De Rossi, D. (2021). Abel: integrating humanoid body, emotions, and time perception to investigate social interaction and human cognition. *Applied Sciences*, 11(3): 1070.

Cominelli, L., Mazzei, D., & De Rossi, D. E. (2018). SEAI: Social Emotional Artificial Intelligence Based on Damasio's Theory of Mind. *Frontiers in Robotics and AI*, 5, 6. <https://doi.org/10.3389/frobt.2018.00006>.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1), 32–80.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Putnam. [tr. it. *L'errore di Cartesio. Emozione, ragione e cervello umano*. Milano: Adelphi, 1995].

Damiano L. (2024). Epistemologia dell'empatia artificiale. Per la transizione da un'etica dei limiti a un'etica delle possibilità. In E. De Bellis (a cura di). *Annali del Centro Studi Filosofici di Gallarate, IV(1-2). I limiti e oltre* (pp. 102–120). Brescia: Editrice Morcelliana.

Darling, K. (2016). 'Who's Johnny?' Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. In P. Lin, G. Bekey, K. Abney, R. Jenkins (Eds.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (pp. 173–185). Oxford: Oxford University Press.

Dumouchel, P., & Damiano, L. (2017). *Living with Robots*. Harvard: Harvard University Press. [tr. it. *Vivere con i robot. Saggio sull'empatia artificiale*. Milano: Raffaello Cortina Editore, 2020].

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Fiore, S. M., Wiltshire, T. J., Lobato, E. J., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2013). Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in Psychology*, 4, 859. <https://doi.org/10.3389/fpsyg.2013.00859>.

Galetta, G. (2024). Fenomenologia del cyborg: ibridazioni uomo-macchina e digitalità senzienti. *I Quaderni di Agenda Digitale*, 2(17), 31–47.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends in Cognitive Sciences*, 7(2), 77–83. DOI: [10.1016/s1364-6613\(02\)00025-6](https://doi.org/10.1016/s1364-6613(02)00025-6).

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501. [https://doi.org/10.1016/S1364-6613\(98\)01262-5](https://doi.org/10.1016/S1364-6613(98)01262-5).

Gonsior, B., Sosnowski, S., Mayer, C., Blume, J., Radig, B., Wollherr, D., & Kuhlentz, K. (2011). Improving aspects of empathy and subjective performance for HRI through mirroring facial expressions. In *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2011* (pp. 350–356). <https://ieeexplore.ieee.org/document/6005294>.

Gratch, J., & Marsella, S. (2005). Evaluating a computational model of emotion. *Journal of Autonomous Agents and Multi-Agent Systems*, 11(1), 23–43. <https://doi.org/10.1007/s10458-005-1081-1>.

Hall, L., Woods, S., Aylett, R., Newall, L., & Paiva, A. (2005). Empathic interaction with synthetic characters: the importance of similarity. *Encyclopaedia of Human Computer Interaction*. Hershey, PA: IGI Publishing.

Hall, L., Woods, S., Aylett, R., & Paiva, A. (2006). Using theory of mind methods to investigate empathic engagement with synthetic characters. *International Journal of Humanoid Robotics*, 3(3), 351–370.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1), 61–84.

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human–robot relations. *International Journal of Human-Computer Studies*, 71(3), 250–260.

Milcent, A. -S., Kadri, A., & Richir, S. (2021). Using Facial Expressiveness of a Virtual Agent to Induce Empathy in Users. *International Journal of Human-Computer Interaction*, 38(3), 240–252. <https://doi.org/10.1080/10447318.2021.1938387>.

Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., & Hagita, N. (2008). Adapting Robot Behavior for Human-Robot Interaction. *IEEE Transactions on Robotics*, 24(4), 911–916.

Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33–35.

- Mori, M., MacDorman, K. F., & Kageki, N. (2021). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>.
- Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in Virtual Agents and Robots: A Survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3), 1–40. <https://doi.org/10.1145/2912150>.
- Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: The MIT Press.
- Pipitone, A., & Chella, A. (2021). What robots want? Hearing the inner voice of a robot. *iScience*, 24(4):102371. <https://doi.org/10.1016/j.isci.2021.102371>.
- Pipitone, A. (2023). Empatia uomo-robot: il complesso rapporto tra l'AI e le emozioni. *Agenda Digitale*, 6 settembre. <https://www.agendadigitale.eu/cultura-digitale/empatia-uomo-robot-il-complesso-rapporto-tra-lai-e-le-emozioni/>.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, 1(4), 515–526.
- Ricci Sindoni, P. (2023). I robot tra scienza e filosofia: ecco come ridefiniscono le relazioni umane. *Agenda Digitale*, 20 aprile. <https://www.agendadigitale.eu/cultura-digitale/i-robot-tra-scienza-e-filosofia-ecco-come-ridefiniscono-le-relazioni-umane/>.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27(1), 169–192.
- Rossi, S., Di Napoli, C., Garramone, F. *et al.* (2024). Personality-Based Adaptation of Robot Behaviour: Acceptability Results on Individuals with Cognitive Impairments. *International Journal of Social Robotics*, 16, 211–226. <https://doi.org/10.1007/s12369-023-01074-1>.
- Rossi, S. (2025). Robot più consapevoli e trasparenti grazie all'IA. *Agenda Digitale*, 16 aprile. <https://www.agendadigitale.eu/cultura-digitale/robotica-sociale-verso-robot-piu-consapevoli-e-trasparenti-con-lia/>.
- Šabanović, S., Bennett, C. C., Chang, W. -L., & Huber, L. (2013). PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In *2013 IEEE 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 345–352). <https://ieeexplore.ieee.org/document/6650427>.
- Scassellati, B. M. (2001). *Foundations for a Theory of Mind for a Humanoid Robot* (Ph.D. dissertation). Cambridge, MA: MIT – Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science. <https://groups.csail.mit.edu/lbr/hrg/2001/scassellati-phd.pdf>
- Scassellati, B. M. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, 12, 13–24. <https://doi.org/10.1023/A:1013298507114>.
- Scassellati, B. M., Admoni, H., & Matarić, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14, 275–294. <https://pubmed.ncbi.nlm.nih.gov/22577778/>.

Schulkin, J. (2000). Theory of mind and mirroring neurons. *Trends in Cognitive Sciences*, 4(7), 252-254. [https://doi.org/10.1016/S1364-6613\(00\)01500-X](https://doi.org/10.1016/S1364-6613(00)01500-X).

Shibata, T., & Tanie, K. (2000). Influence of a priori knowledge in subjective interpretation and evaluation of robot's behavior. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1, 165–170. <https://ieeexplore.ieee.org/document/894600>.

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141–161.

Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York, NY: Basic Books/Hachette Book Group. [tr. it. *Insieme ma soli. Perché ci aspettiamo sempre di più dalla tecnologia e sempre meno dagli altri*. Torino: Einaudi, 2019).

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *International Journal of Social Robotics* 7(3), 347–360.