Critical Hermeneutics, 4(1), 2020
Biannual International Journal of Philosophy
http://ojs.unica.it/index.php/ecch/index
ISSN 2533-1825 (on line); DOI 10.13125/CH/4331

Received: 21/3/2020
Accepted: 21/3/2020
Published: 9/10/2020

[RECENSIONE]

Dumouchel, P. e Damiano, L. (2019). *Vivere con i robot – saggio sull'empatia artificiale*. Milano: Raffaello Cortina Editore.

EVA SIMONETTI

Sulla scia di una visione offerta dai manga giapponesi, gli studiosi di epistemologia della complessità Paul Dumouchel e Luisa Damiano, autori di *Vivere con i robot. Saggio sull'empatia artificiale*, propongono un nuovo modo di pensare al nostro rapporto, presente ma soprattutto futuro, con gli agenti robotici.

A partire da quest'apertura, gli autori possono prendere le distanze sia dalla tradizionale idea di robot come semplice *lavoratore artificiale*, sia dalla prospettiva occidentale, tutta distopica, riguardo a una possibile evoluzione del rapporto tra esseri umani e robot: distopica in quanto prevede che robot troppo evoluti prendano il sopravvento sugli umani. Essi propongono infatti uno scenario in cui questi agenti artificiali diventano attori sociali integrati nella rete di relazioni, anche e soprattutto affettive, degli umani. Ritengono inoltre che il rapporto tra esseri umani e robot, lungi dal concludersi in un'Apocalisse (come vorrebbero gran parte della filmografia e della letteratura fantascientifiche), sarà occasione di crescita morale e sociale sia per gli umani sia, perché no, per gli stessi agenti robotici sociali.

Ma in che modo i manga giapponesi ci offrono una visione alternativa? E una volta aperti a questa visione, come possiamo immaginare concretamente questi nuovi agenti artificiali sociali? Come dobbiamo impiegarli? E ancora: che idea di mente dobbiamo abbracciare per avvicinarci alla realizzazione pratica di questi robot? A quali problematiche etiche apre l'integrazione nella nostra rete sociale di tali agenti artificiali?

Tutte queste domande guidano il percorso del libro, che si snoda in cinque capitoli a loro volta suddivisi in vari paragrafi, i quali affrontano progressivamente problematiche sempre più complesse.

Se la parola *robot* nasce in un dramma teatrale di Karl Capek intitolato *Russum's Universal Robots* in cui si narra la storia di macchine artificiali che diventano sempre più simili agli umani fino a diventare indistinguibili da essi e a ribellarsi, in una visione apocalittica che sarà poi tipica di tanta fantascienza occidentale, i manga giapponesi propongono una visione opposta: in molte delle loro storie, una su tutte *Atro Boy*, sono presenti agenti robotici, ma questi, lungi dal ribellarsi agli umani e andare contro di essi, sono di volta in volta alleati dell'eroe protagonista e occasione, per lui, di un'importante crescita psicologica e morale, tanto che molte di questi racconti possono essere assimilati al romanzo di formazione. Questi agenti robotici non sono sistemi tecnici impersonali il cui perfezionamento tecnico e la conseguente deriva non sono opera di nessuno (come quelli che fanno da sfondo alla visione apocalittica occidentale), ma sono oggetti dotati di un'anima e realizzati da un creatore (spesso è presente il tema padre-figlio). Se nella prima prospettiva la tecnica è un nemico, fonte di alienazione, nella seconda la tecnica è un alleato, che conduce al trionfo dell'individualità. Questa prospettiva è un'anticipazione delle teorie della robotica sociale: è questo l'indirizzo di ricerca abbracciato e problematizzato dagli autori. Esso punta alla realizzazione di agenti

sociali artificiali in grado di integrarsi nelle nostre relazioni e nell'ecologia dei nostri scambi emozionali, da impiegarsi in contesti in cui la socialità è protagonista, per esempio per assistere o far compagnia ad anziani o pazienti con bisogni speciali. In queste interazioni, essi porterebbero a vantaggi paragonabili a quelli della *pet therapy*, e quindi potrebbero essere sostituiti ad animali reali, portando vantaggi di igiene e gestione. Inoltre, essi vengono considerati un vero e proprio strumento di ricerca scientifica all'interno del complesso dibattito sulla natura della mente, da impiegare in contesti sperimentali. Gli autori sostengono infatti che tanto il processo di creazione di agenti artificiali sociali quanto l'osservazione sperimentale delle interazioni che con essi hanno gli umani, possono portare a scoperte significative sulla natura della mente.

Questi robot sono chiamati *sostituti* poiché vengono impiegati per sostituire l'attore umano o animale. Essi hanno specifiche caratteristiche, illustrate di seguito, che aprono a molte problematiche, una su tutte la concezione tanto filosofica quanto neuroscientifica della mente e delle emozioni.

I sostituti non sono dei semplici «lavoratori artificiali» come gli altri robot, progettati per svolgere compiti specifici e senza capacità sociali. Essi, al contrario, non hanno alcuna funzione pratica precisa, in quanto il loro fine ultimo è la socialità, che «non ha alcuno scopo particolare al di là di se stessa» (41).

Essi sono pertanto dotati di caratteristiche del tutto peculiari: (I) devono saper interrompere lo svolgimento di un compito per ricoordinarsi con l'ambiente, con gli altri robot e con i paterner sociali: la capacità trascendere da una data funzione e riconfigurare le azioni in base alle situazioni, secondo gli autori, sarebbe infatti un aspetto essenziale della nostra socialità, sarebbe anzi proprio ciò che differenzia le macchine dagli schiavi e il motivo per cui il tentativo di eser-

citare un pieno controllo sugli schiavi si rivela sempre molto più instabile di quello di esercitarlo su delle macchine (41). (II) Devono avere una presenza sociale, saper rivolgere la propria attenzione agli altri, dando a questi la percezione di «essere con un altro». Questo sarebbe un punto fondamentale di differenza tra i sostituti e gli altri oggetti tecnici: in linea di massima, infatti, gli oggetti tecnici (es. smartphone) sono pensarti per scomparire nella loro funzione, far assentare l'utente dallo spazio fisico per trascinarlo in quello virtuale. Al contrario, i sostituti, caratterizzati da tridimensionalità, affermano la loro presenza e non ci allontanano dallo spazio fisico. La presenza sociale che affermano, inoltre, deve essere propriamente la loro. (III) Devono esercitare e mantenere una certa autorità. Anche questa caratteristica li distanzierebbe dagli oggetti tecnici che, per quanto possano essere strumenti di misurazione che forniscono risultati autorevoli, non hanno una reale autorità. I sostituti devono propriamente affermarsi e farsi rispettare. Stabilito questo, che aspetto fisico devono avere?

Mashairo Mori realizza una curva (33) che nasce dall'integrazione tra somiglianza nell'aspetto di questi robot con l'uomo e senso di familiarità degli umani verso di essi, ottenendo che il senso di familiarità cresce col crescere della somiglianza, finché non si raggiunge un punto critico in cui la somiglianza eccessiva genera una zona perturbante (*uncanny valley*): l'umano è inquietato da questo oggetto artificiale. Dalla zona perturbante si esce nel momento in cui il robot assume le sembianze perfette di un individuo in salute, ovvero quando il suo aspetto è in tutto indistinguibile da quello di un uomo. Mori ipotizza anche un'ulteriore evoluzione (attualmente impossibile a livello tecnico) in cui il robot sarà «più umano dell'umano» (35) ovvero quando il suo aspetto incarnerà il nostro ideale artistico di perfezione. Con questi agenti artificiali, che per altro genererebbero in noi una

vergona prometeica, secondo tale teoria addirittura «interagiremmo meglio che con uno dei nostri simili» (36).

Come anticipato, la questione dei sostituti apre a questioni filosofiche come la concezione della mente e delle emozioni: per quanto riguarda l'idea della mente, gli autori, già nel Capitolo 1, si oppongono alla concezione della *mente nuda*, per cui la mente sarebbe un'attività computazionale universale che può essere realizzata sui più vari supporti fisici (es. computer): al contrario, propongono l'idea di una *mente radicalmente incorporata*, le cui funzioni sono vincolate dall'organizzazione corporea del soggetto e dall'ambiente in cui esso è situato: questa visione deriverebbe da esperimenti di etologia artificiale e supporterebbe l'idea che sia necessario di disporre di oggetti tridimensionali (come appunto sarebbero i sostituti) per gli esperimenti sul comportamento animale e per gli esperimenti di ordine sociale.

Ma è dal Capitolo 2 che la riflessione sulla mente si estende e si fa più complessa. Gli autori compiono un'operazione apparentemente paradossale: riabilitano Cartesio per sfuggire al dualismo della sua prospettiva. Infatti, a loro avviso, le moderne teorie che criticano Cartesio di fatto ricadono nel suo dualismo. Cartesio, secondo gli autori, affermando che gli animali non hanno un'anima, non sta negando loro capacità cognitive. Piuttosto, sta ammettendo la possibilità dell'esistenza di sistemi cognitivi diversi dal nostro. Egli aprirebbe dunque a un'idea plurale dei sistemi cognitivi, ovvero all'idea che certi sistemi cognitivi, pur essendo radicalmente diversi da quelli umani, non per questo sono meno validi. Seguendo questa argomentazione, gli autori accostano i sistemi cognitivi animali a quelli delle macchine: entrambi si differenzierebbero dal sistema cognitivo umano perché questo sarebbe l'unico capace di comprendere il linguaggio e dare significato alle cose, caratteristiche che non siamo ancora in grado di

riprodurre nelle macchine. A questo discorso viene accostata la confutazione della teoria della *mente estesa*, secondo cui, tramite le macchine, la mente, intesa come situata nel cervello, avrebbe la possibilità di estendersi al di fuori dei limiti del corpo, trasformando di fatto in se stessa tutto ciò con cui entra in contatto: una macchina, secondo questa tesi, è cognitiva solo in quanto appendice della mente umana, e non ha capacità cognitive proprie. Secondo gli autori, al contrario, le macchine non estenderebbero la mente, ma avrebbero sistemi cognitivi propri, esattamente come gli animali, seppur di tipologia diversa dal sistema cognitivo umano. Ad essere esteso sarebbe quindi lo stesso dominio del cognitivo, portato al di là del confine della mente umana. In base a questa visione, come scrive Paul Humphreys in »Extending Ourselves», occorre «abbandonare l'idea che le abilità epistemiche umane siano l'arbitro ultimo della conoscenza scientifica» (81). L'idea dell'omogeneità del cognitivo e l'idea che l'umano sia il soggetto epistemico per eccellenza, sarebbero infatti frutto di una visione antropocentrica (88), e proprio in ciò consisterebbe la ricaduta della moderna concezione della mente nel dualismo cartesiano, contro il pluralismo proposto dagli autori.

Nel Capitolo 3 la riflessione sulla mente arriva al punto centrale e si conclude: qui viene approfondita l'idea della mente come una rete cognitiva socialmente distribuita. Alla domanda: "Dov'è la mente?" la tesi tradizionale e il senso comune risponderebbero "Dentro di me". Gli autori si muovono contro quest'idea, proponendo che la mente non sia un oggetto situato in una dimensione individuale, ma che sia una dinamica sociale.

Il nucleo di questa dimensione sociale è individuato nelle emozioni. Da qui, l'idea che la dimensione sociale dei robot dipende dalla possibilità, per questi agenti artificiali, di partecipare agli scambi emozionali ed empatici immettendosi nel circuito affettivo. Come arrivare alla realizzazione di tali robot?

Gli autori ci presentano i due approcci attualmente esistenti per la modellizzazione della robotica delle emozioni e, dopo aver osservato come entrambi, pur restando radicalmente separati a livello teorico, di fatto nella pratica oltrepassano sistematicamente i loro confini, ci propongono un terzo approccio che li integri. Questi due approcci sono: la robotica esterna, che si concentra sulla dimensione sociale e che concepisce le emozioni robotiche come simulazioni dei processi emozionali umani, per cui i robot *fingono* di avere emozioni, ovvero mimano le modalità di espressione delle emozioni così che l'utente che interagisce con loro gli attribuisca delle emozioni per analogia con le espressioni umane; la robotica interna, che guarda alla dimensione individuale e che considera le emozioni robotiche come genuine, perché provocate da meccanismi interni che vorrebbero mimare l'effettiva generazione biologica delle emozioni umane, riproducendo artificialmente i processi emozionali naturali.

L'approccio proposto dagli autori, invece, ha come obiettivo quello di «collocare i robot nel circuito affettivo (*affective loop*)» (128) dotandoli di una fenomenologia della coordinazione emozionale interindividuale. Si tratta dell'idea che l'interazione tra due agenti si basi sul fatto che l'uno reagisce alle emozioni dell'altro secondo una strategia di coordinazione, distante dall'idea classica che l'interazione si basi sul riconoscimento delle emozioni dell'uno da parte dell'altro in base alle sue espressioni. Infatti, il successo della strategia di coordinazione non dipenderebbe da una conoscenza adeguata dell'emozione di partenza, in quanto tale strategia potrebbe addirittura modificare quest'ultima. Questa teoria si basa su quelle riguardanti i meccanismi di *mirroring*, ovvero processi di coattivazione neuronale, da concepire, secondo Vittorio Gallese, come «meccanismi incorporati» di accesso alle emozioni altrui (156).

Il quinto ed ultimo capitolo del libro è dedicato alle questioni etiche che emergono dall'utilizzo di sistemi artificiali. Ciò che emerge è la necessità di limitare l'azione di questi robot affinché essi si muovano sempre in una direzione etica. L'obiettivo è limitare le conseguenze sociali di uno sviluppo tecnologico considerato inevitabile. Si tratta della creazione di AMA (Agenti Morali Artificiali) la cui azione è governata (e limitata) da alcune regole che vengono inserite nel robot nel momento in cui lo si realizza e programma, regole che questo non può non seguire, perché a lui intrinseche. La questione si fa particolarmente urgente se si guarda alle tecnologie già impiegate in ambito militare. Infatti, la ricerca scientifica e tecnologica, per la robotica, si muove in gran parte in quest'ambito.

Le due voci presentate dagli autori in merito sono quella di Armin Krishnan, esperto di scienze politiche e relazioni internazionali, e Ronald Arkin, specialista di sistemi robotici autonomi, che guardano, rispettivamente con preoccupazione e, al contrario, con auspicio, alla possibilità (che tuttavia entrambi riconoscono essere economicamente, politicamente e militarmente vantaggiosa) di affidare alle macchine le decisioni belliche, come la decisione di attaccare o la valutazione della legittimità del bersaglio. Entrambi concordano comunque sulla necessità di dotare questi robot di un sistema di regole che ne diriga l'azione: queste regole dovrebbero essere quelle internazionali di regolazione dei conflitti bellici.

Robot di questo tipo, la cui azione è limitata da tali regole, sarebbero dei *super-soldati*, impossibilitati a disobbedire, per cui *agire moralmente* significherebbe soltanto agire in conformità a una regola la cui moralità è stata decisa esteriormente rispetto ad essi e senza che questi abbiano avuto la possibilità di acconsentire a tale regola né di riconoscerne la moralità (184). Queste caratteristiche fanno sì che le loro azioni siano «conformi alla morale» più che propriamente morali, nel senso moderno del termine. Infatti nel pensiero moderno la

moralità è strettamente connessa all'autonomia decisionale, per cui è morale l'agente che, pur essendo in grado di agire in modo immorale, *sceglie*, con un atto consapevole, di agire moralmente.

Il punto, presentato dagli autori già all'inizio del libro, è che noi *non* vogliamo agenti dotati di una tale autonomia. Ci troviamo quindi di fronte ad un paradosso per cui al contempo vogliamo e non vogliamo agenti autonomi. Li vogliamo autonomi nel senso di *capaci di agire senza la necessità di una nostra supervisione costante*, ma non li vogliamo autonomi nel senso pieno del termine.

Gli autori, comunque, mostrano tutte le problematicità dell'eventuale creazione di questi super-soldati: infatti, se da una parte, come afferma Arkin, questi potrebbero essere moralmente superiori agli esseri umani in quanto non suscettibili di essere spinti dall'ira o dal panico a compiere atrocità, dall'altro gli autori evidenziano che «le emozioni […] sono anche ciò che permetterebbe loro di sentire compassione, di mostrare rispetto o anche solo interesse per qualcosa al di là dell'adempimento impeccabile della missione militare» (182).

Anche in questo caso, quindi, essi ravvisano la necessità di dotare gli agenti artificiali di empatia: anche i robot impiegati in campo bellico devono essere attori sociali, e non semplici armi.

Infatti, gli autori sottolineano come i robot militari prodotti oggi, che non sono sostituti, ma soltanto armi, non possono essere considerati etici solo per il fatto di essere autorizzati dalla legge internazionale: parlare di etica in questi termini è sbagliato e rappresenta un errore di categoria.

Queste considerazioni non esauriscono tuttavia la riflessione sui risvolti etici e sociali dell'eventuale integrazione di sostituti all'interno della nostra rete relazionale. Al contrario, si ritiene che i problemi che essi, in quanto oggetti fisici tridimensionali reperibili nello spazio co-

me individui indipendenti, solleveranno una volta integrati nella nostra ecologia sociale e affettiva, saranno del tutto nuovi, e specifici dell'interazione tra attori umani e artificiali. Saranno problemi di altro ordine, e aprirci a tali questioni significherà in definitiva affrontare l'etica in modo diverso: è l'*innovazione etica* (198).

Si chiude così un cerchio: non solo questi agenti sociali potranno essere degli importanti mezzi di indagine scientifica sulla mente e sulla socialità, ma anche i protagonisti di una possibile innovazione etica.
È questo il percorso che, approfondendo svariate questioni filosofiche, tecniche e scientifiche, viene affrontato dai cinque densi capitoli del libro.