

Philosophy of Translation An Interdisciplinary Approach

Filosofia della traduzione Un approccio interdisciplinare

T E O R I A

Rivista di filosofia
fondata da Vittorio Sainati
XL/2020/2 (Terza serie XV/2)

Edizioni ETS

«Teoria» è indicizzata ISI Arts&Humanities Citation Index e SCOPUS, e ha ottenuto la classificazione “A” ANVUR per i settori 11/C1-C2-C3-C4-C5.

La versione elettronica di questo numero è disponibile sul sito: www.rivistateoria.eu

Direzione e Redazione: Dipartimento di civiltà e forme del sapere dell'Università di Pisa, via P. Paoli 15, 56126 Pisa, tel. (050) 2215400 - www.cfs.unipi.it

Direttore: Adriano Fabris

Comitato Scientifico Internazionale: Antonio Autiero (Münster), Damir Barbarić (Zagabria), Vinicius Berlendis de Figueiredo (Curitiba), Bernhard Casper (Freiburg i.B.), Néstor Corona (Buenos Aires), Félix Duque (Madrid), Günter Figal (Freiburg i.B.), Denis Guénoun (Parigi), Dean Komel (Lubiana), Klaus Müller (Münster), Patxi Lanceros (Bilbao), Alfredo Rocha de la Torre (Bogotá), Regina Schwartz (Evanston, Illinois), Ken Seeskin (Evanston, Illinois), Mariano E. Ure (Buenos Aires).

Comitato di Redazione: Paolo Biondi, Eva De Clerq, Silvia Dadà, Giulio Gorla, Enrica Lisciani-Petrini, Annamaria Lossi, Carlo Marletti, Flavia Monceri, Veronica Neri, Antonia Pellegrino, Stefano Perfetti, Augusto Sainati.

Amministrazione: Edizioni ETS, Lungarno Mediceo 16, 56127 - Pisa www.edizioniets.com, info@edizioniets.com - tel. (050) 29544-503868

Abbonamento: Italia € 40,00 (Iva inclusa); estero € 50,00

(Iva e spese di spedizione incluse)

da versare sul c.c.p. 14721567 intestato alle Edizioni ETS.

Prezzo di un fascicolo: € 20,00, Iva inclusa.

Prezzo di un fascicolo arretrato: € 30,00, Iva inclusa.

L'indice dei fascicoli di «Teoria» può essere consultato all'indirizzo: www.rivistateoria.eu. Qui è possibile acquistare un singolo articolo o l'intero numero in formato PDF, e anche l'intero numero in versione cartacea.

Iscritto al Reg. della stampa presso la Canc. del Trib. di Pisa n° 10/81 del 23.5.1981. Direttore Responsabile: Adriano Fabris.

Semestrale. Contiene meno del 70% di pubblicità.

© Copyright 1981-2018 by Edizioni ETS, Pisa.

I numeri della rivista sono monografici. Gli scritti proposti per la pubblicazione sono double blind peer reviewed.

I testi devono essere conformi alle norme editoriali indicate nel sito.

TEORIA

T

Rivista di filosofia
fondata da Vittorio Sainati
XL/2020/2 (Terza serie XV/2)

Philosophy of Translation

An Interdisciplinary Approach

Filosofia della traduzione

Un approccio interdisciplinare

Edizioni ETS

Contents / Indice

Adriano Fabris, Seung Chul Kim

Premise / Premessa, p. 5

Seung Chul Kim

Homo Translator: Traditions in translation, p. 7

Zbigniew Wesółowski

Hermeneutics of Understanding the Confucian Idea of Truth:
Junzi 君子 as a Truth-bearer in the *Lunyu* 論語, p. 13

Jorge Martínez

What do we translate when we translate?, p. 35

Paul L. Swanson

Context, Logosyllabary, and Multiple Choices.
Reflections on 30+ Years of Translating Chinese
Buddhist Texts, p. 49

Alicia M. de Mingo Rodríguez

Desafío de confianza y traducibilidad crítica.
Notas para una ética de la traducción, p. 63

Carlo Chiurco

Mediare l'incommensurabile.
Note sul rapporto tra filosofia e traduzione, p. 83

Elinor Hållén

Making the Unconscious Conscious:
A Reflection on the Concept Translation in Freud, p. 101

Elena Nardelli

Se il filosofo si fa traduttore.
Note a partire da Martin Heidegger, p. 121

Alberto Martinengo

Ermeneutica filosofica e filosofia politica.

Il paradigma della traduzione in Paul Ricoeur, p. 141

Saša Hrnjez

Traduzione, negazione, riflessione:

sulla natura negativo-contraddittoria della traduzione, p. 163

Maria Benedetta Saponaro

La traduzione algoritmica del pensiero relazionale, p. 187

Premio di Studio «Vittorio Sainati» 2019-2020**Giulia Battistoni**

La polivocità di *Schuld* nella Moralità hegeliana:

un contributo agli studi sulla traduzione di un concetto portante della filosofia hegeliana dell'azione, p. 209

Guglielmo Califano

Scientia intuitiva?

Intuizione, idea ed empiria nella scienza di Goethe, p. 221

Filippo Nobili

Husserl e il punto di vista dell'*Allsubjektivität*, p. 233

La traduzione algoritmica del pensiero relazionale

Maria Benedetta Saponaro

On peut dire que tout ce que nous savons,
c'est à dire tout ce que nous pouvons,
a fini par s'opposer à ce que nous sommes?

Paul Valery, *Le Bilan de l'intelligence*, 1935

1. *Le macchine possono pensare?*

Alain Turing pose questo interrogativo nel 1950 in un noto articolo, *Computing machinery and intelligence*, apparso sulla rivista «Mind»¹. La macchina in questione è il calcolatore numerico.

Un calcolatore numerico può essere normalmente considerato composto di tre parti: i) memoria; ii) complesso operativo; iii) governo. La memoria è un deposito di informazioni, e corrisponde alla carta del calcolatore umano, sia che si tratti della carta, sulla quale egli fa i suoi calcoli, sia di quella sulla quale è stampato il suo libro di regole [...] Il complesso operativo è la parte che compie le varie operazioni singole che un calcolo comporta [...] Abbiamo fatto presente che il “libro delle regole” fornito al calcolatore umano è sostituito nella macchina da una parte della memoria. Si chiama allora “tavola delle istruzioni”. È compito del “governo” controllare che queste istruzioni siano eseguite correttamente e nell'ordine giusto. Il governo è costruito in maniera tale, che questo avviene necessariamente².

Turing giunse alla conclusione che i calcolatori avrebbero potuto essere usati *come* cervelli meccanici, ma per definirli cervelli non era sufficiente che si comportassero come tali. Avrebbero dovuto possedere determinate

¹ A.M. Turing, *Computing Machinery and Intelligence*, in «Mind», 59 (1950), pp. 433-460 (traduzione italiana in: V. Somenzi, R. Cordeschi, *La filosofia degli automi. Origini dell'intelligenza artificiale*, Boringhieri, Torino 1986, pp. 157-183).

² *Ivi*, p. 437.

facoltà (ad esempio il libero arbitrio), che il progresso tecnologico, in quel dato momento, non era in grado di garantire. Se il libero arbitrio fosse un'illusione, la macchina sarebbe determinata così come lo è l'uomo; se non fosse una illusione, comunque non potremmo valutare il possesso della libertà di determinazione semplicemente attraverso il comportamento. Nel secondo caso, si potrebbe inserire un elemento casuale, tale da rendere la scelta non determinabile, ma in ogni caso il comportamento attuato sarebbe implicito nel programma³. In *Computing machinery and intelligence* prende in esame una serie di altre possibili obiezioni, che potrebbero essere opposte all'affermazione che le macchine pensano, utilizzando lo stesso stile di confutazione in modo ingenuo e non riuscendo, per sua stessa ammissione, a superarle con argomenti convincenti. Per progettare una macchina pensante, era necessario sapere come avvenisse l'attività del pensare nell'uomo, studiando in particolare il cervello infantile⁴.

Nell'impossibilità di pervenire ad una risposta definitiva, riformulò il quesito, introducendo il gioco dell'imitazione (*imitation game*): una macchina (interlocutore sconosciuto) deve conversare con un umano su argomenti di varia natura, simulando di essere un umano. Se al termine della conversazione l'umano non è in grado di determinare se si tratti di una macchina o di un umano, allora la macchina può dirsi intelligente. La migliore strategia di vittoria per la macchina è di certo imitare l'uomo, simulando e mentendo (comportamento ovviamente consentito). Si potrebbe, dunque, ritenere pensante una macchina che mente e simula, benché «nonostante vari tentativi e qualche vittoria annunciata e poi smentita, ad oggi nessun computer sembrerebbe aver superato il test»⁵. Se una macchina dovesse superare il test di

³ A.M. Turing, *I calcolatori digitali possono pensare?*, in «Sistemi intelligenti», 1 (1998), pp. 24-25. I modelli di calcolo hanno carattere universale. L'universalità della macchina è una delle intuizioni più significative di Turing.

⁴ «Il cervello infantile è qualcosa di simile ad un taccuino di quelli che si comprano dai cartolai. Poco meccanismo e una quantità di fogli bianchi (meccanismo e scrittura sono dal nostro punto di vista quasi sinonimi). La nostra speranza è che ci sia così poco meccanismo nel cervello infantile, che qualcosa di analogo possa venir facilmente programmato. Per il processo educativo possiamo supporre che il lavoro, in prima approssimazione, sia pressappoco uguale a quello necessario per il bambino» (A.M. Turing, *Computing machinery and intelligence*, cit., p. 456. Riteneva che si potesse circoscrivere l'apprendimento della macchina al programma e all'educazione, tralasciando l'esperienza).

⁵ C. Accoto, *Il mondo dato. Cinque brevi lezioni di filosofia digitale*, Egea, Milano 2017, p. 58. Vedi anche L. Floridi, M. Taddeo, M. Turilli, *Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges - An Evaluation of the 2008 Loebner Contest*, in «Mind & Machines», 19 (2009), pp. 145-150. In verità, oggi non solo numerosi chatbots hanno superato il test di Turing, ma addirittura le stesse macchine sottopongono a noi il test di Turing, proponendoci i *captcha* (completely automated public Turing test to tell computers and humans apart), per determinare se siamo umani o agenti artificiali. Cfr. C. Accoto, *op. cit.*

Turing, potremmo affermare che quella tal macchina è in grado di pensare? La questione, secondo la prospettiva funzionalista di Turing, non è se possano essere costruiti calcolatori che siano in grado di fare tutto ciò che può fare l'uomo, o che abbiano sembianza umane⁶, ma se possano fare ciò che si chiama pensare in senso proprio, e ciò che «deve fare una macchina per pensare, è essere capace di risolvere tutti i problemi intellettuali con cui potrebbe essere messa a confronto»⁷, compreso *l'imitation game*.

Gli appetiti, i desideri, le pulsioni, gli istinti avrebbero rilevanza se volessimo dotare la macchina di volontà. Per essere realizzabile il progetto di Turing, ovvero la costruzione di una macchina capace di pensare come un essere umano, deve essere vero il presupposto, ovvero che con l'attività del pensare ci si riferisca alle sole proprietà di calcolo, memoria, ragionamento per analogia. Il presupposto è discutibile, come ci riserviamo di argomentare nel paragrafo successivo.

A Turing viene conferito il merito di aver posto le fondamenta di ciò che John McCarthy definì nel 1956 Intelligenza Artificiale⁸, ovvero il sapere scientifico e ingegneristico deputato alla costruzione di macchine capaci di eseguire attività o processi di ragionamento, associati all'intelligenza in un essere umano. Seppure non sia riuscito a costruire materialmente la macchina pensante, ha tratteggiato la visione che sta a monte della civiltà del virtuale. In più occasioni ha sottolineato come la costruzione di macchine capaci di pensare sia subordinata alla scoperta del funzionamento del cervello umano, ma in *I calcolatori digitali possono pensare?* inverte gli addendi

⁶ Anzi Turing spera e crede, che «non si farà alcuno sforzo notevole nella costruzione di macchine con le caratteristiche più distintivamente umane, ma non intellettuali, come la forma del corpo; mi sembra che fare tali tentativi sia del tutto futile e che i risultati siano condannati ad avere la qualità sgradevole dei fiori artificiali» (A.M. Turing, *I calcolatori digitali possono pensare*, cit., p. 26).

⁷ A.M. Turing, *Si può dire che i calcolatori automatici pensano?*, in «Sistemi intelligenti», 1 (1998), p. 38. L'articolo è la trascrizione di un dibattito radiofonico tra Turing, R.B. Braithwaite, *fellow* del King's College a Cambridge, M.H. Jefferson, *fellow* della Royal Society e professore di neurologia all'Università di Manchester e M.H. Newman, *fellow* della Royal Society e professore di matematica all'Università di Manchester, nonché direttore del progetto della macchina omonima.

⁸ Gli studi sull'intelligenza artificiale e sulla costruzione di macchine intelligenti sono stati condotti essenzialmente in due direzioni: la prima indirizzata alla costruzione di macchine che replicassero la capacità umana di problem-solving, la seconda indirizzata alla costruzione di macchine che non replicassero necessariamente meccanismi cognitivi umani. McCarthy ritiene che l'AI si debba occupare della risoluzione dei problemi e non del funzionamento della mente umana, appannaggio della filosofia. J. McCarthy, *The Invention of Functions Defined by Turing Machines*, in C.E. Shannon, J. McCarthy (eds.), *Automata Studies*, Princeton University Press, Princeton 1956.

ed è la stessa macchina pensante che ci consentirà di scoprire come noi stessi pensiamo⁹. Il rapporto tra esseri umani e macchine si può leggere attraverso la dinamica del rispecchiamento: «nell'immagine dello specchio, infatti, io non solo mi riconosco e mi contemplo, ma in essa posso anche proiettare attivamente gli aspetti migliori che mi sono propri. Ne consegua uno scambio di ruoli fra il soggetto raffigurato e la sua immagine, in virtù del quale l'uno, alternativamente, fa da modello e da copia per i comportamenti dell'altro»¹⁰. L'uomo da essere modello da imitare viene progressivamente ridotto a ciò che può essere imitato, “null'altro che”¹¹ il pensiero riproducibile artificialmente. È sottinteso che «il messaggio metaforico fondamentale del computer è che noi siamo macchine: macchine pensanti, beninteso, ma pur sempre macchine. Per questo il computer è la macchina essenziale, impareggiabile, quasi perfetta per il tecnopolio; quella che mette in secondo piano le esigenze della nostra natura, della nostra biologia, delle nostre emozioni e della nostra spiritualità. Il computer proclama la sua superiorità sull'intera gamma delle esperienze umane, e giustifica la sua pretesa dimostrandosi capace di “pensare” meglio di noi»¹².

D'altra parte molti dei limiti che evidenziava Turing, nonostante l'innegabile progresso delle tecnologie dell'intelligenza artificiale, non sono stati superati. Ancora oggi non possiamo rispondere in modo definitivo alla domanda “le macchine possono pensare?” e non molto sappiamo su come un cervello infantile apprenda, ma questo non ci ha impedito di elaborare sistemi intelligenti sempre più complessi e competenti. Dovremmo convenire, allora con Turing, che «costruire una macchina pensante è costruire una (cosa) che fa cose interessanti, senza che noi in realtà comprendiamo davvero come funziona»¹³. È quella che potremmo definire l'istanza burocratica dei sistemi formali, per cui la *bontà* del sistema dipende dalla sua efficienza e non dall'orizzonte di valori che esprime, orizzonte che, nei sistemi tecnologici è addirittura superfluo. La trasformazione dell'*agire finalizzato* in semplice *fare funzionale* ha privato l'agire del suo fine, che orienta l'azione

⁹ A.M. Turing, *I calcolatori digitali possono pensare*, cit., p. 26.

¹⁰ A. Fabris, *Etica per le tecnologie dell'informazione e della comunicazione*, Carocci, Roma 2018, p. 75. Vedi anche Id., *La filosofia e lo specchio delle macchine*, in «InCircolo», 6 (2018), pp. 28-38.

¹¹ Cfr. L. Lombardi Vallauri, *Terre*, Vita e Pensiero, Milano 1989.

¹² N. Postman, *Technopoly: The Surrender of Culture to Technology*, Vintage Books, New York 1993 (tr. it. *Technopoly. La resa della cultura alla tecnologia*, Bollati Boringhieri, Torino 1993, pp. 103-104).

¹³ A.M. Turing, *Si può dire che i calcolatori automatici pensano?*, cit., p. 34.

e fornisce il suo senso¹⁴. A questa si collega un'altra considerazione. Ogni giorno usiamo strumenti, di cui abbiamo una conoscenza superficiale e limitata ad un certo numero di funzioni operative, nulla sappiamo del software o degli algoritmi che governano la nostra interazione con le macchine.

Il carattere potenzialmente demoniaco dei numeri ha finito allora per assumere un nuovo aspetto, quello di una possibile perdita di senso di un calcolo che continua ad apparire, peraltro completamente credibile e affidabile. L'impiego del calcolo digitale in ogni settore della nostra vita ha esasperato queste ambiguità: con l'infinità varietà degli algoritmi che servono oggi per selezionare l'ingresso nelle scuole o nelle università, per incriminare presunti colpevoli, per assumere o licenziare nelle aziende, o semplicemente per sorvegliare i nostri movimenti, si sacrifica spesso l'equità per l'efficienza, l'attendibilità del giudizio per la funzionalità dell'apparato¹⁵.

La cieca delega alla certezza algoritmica delle scelte determina il prevalere del profilo formale procedurale su quello sostanziale, in ragione della presupposta certezza del calcolo, che appare incontrovertibile. La *filosofia* dell'algoritmo non mira a comprendere la complessità, ad indagarne il senso, ma a renderla fruibile, plasmabile, utilizzabile. La credibilità della semplificazione algoritmica risiede «in buona parte nella sintesi tra scienza teorica e applicata, tra ricerca fondazionale e tecnica della computazione»¹⁶. Il processo formale del raggruppamento in classi di elementi molteplici, che rende una materia apparentemente immensa più gestibile, alla base dell'algoritmo, ci riporta al concetto di enumerazione completa di Descartes, enunciato nell'ultimo precetto del metodo e nella VII Regola. L'algebra ha consentito alla matematica di affrancarsi dagli argini della spazialità, traducendo in simboli i dati e i movimenti sensibili. Questo passaggio spiega che «la linea di demarcazione tra l'età moderna e il mondo in cui viviamo, possiamo tracciarla nella differenza tra una scienza che guarda alla natura da un punto di vista universale e acquista così una completa padronanza su di essa, e una scienza veramente "universale", che trasferisce processi cosmici nella natura anche con il rischio evidente di distruggerla, e insieme

¹⁴ U. Galimberti, *Psiche e techne*, Feltrinelli, Milano 2000, p. 685.

¹⁵ P. Zellini, *La dittatura del calcolo*, Adelphi, Milano 2018, p. 36. Per algoritmo si intende «un processo, una sequenza di operazioni che deve soddisfare almeno due requisiti: ad ogni passo della sequenza è già deciso, in modo deterministico, quale sarà il passo successivo, e la sequenza deve essere *effettiva*, cioè tendere a un risultato concreto, reale e virtualmente utile» (*ivi*, p. 15). Vedi anche dello stesso autore, *La matematica degli dèi e gli algoritmi degli uomini*, Adelphi, Milano 2016.

¹⁶ P. Zellini, *La dittatura del calcolo*, cit., pp. 11-12.

di distruggere la signoria dell'uomo su di essa»¹⁷. Il metodo, «il calcolo matematico diventa a questo punto la forma del mondo e la tecnica ad esso sottesa la sua chiave di lettura. Per effetto di questo capovolgimento, per cui una tecnica previsionale escogitata nel mondo-della-vita diventa la *sola* chiave di lettura di quel mondo, la tecnica scientifica non è più un *procedimento neutro* disponibile per qualsiasi fine come ama definirsi, ma *prassi sociale e storica* che obbliga il mondo-della vita a esprimersi in termini di pura quantità»¹⁸. L'introduzione del paradigma quantitativo riduce il conoscibile al quantificabile. Utilizzando un'iperbole, potremmo considerare la scienza dell'algoritmo la metafisica del nostro tempo. Come osserva Floridi «le tecnologie digitali non sono soltanto strumenti che si limitano a modificare il modo in cui interagiamo con il mondo come la ruota o il motore. Sono soprattutto sistemi che danno forma (formattano) e influenzano sempre più il modo in cui comprendiamo il mondo e ci rapportiamo a esso, così come il modo in cui concepiamo noi stessi e interagiamo tra noi. In altre parole sono *re-ontologizzanti*, cioè modificano la natura intrinseca (l'ontologia) di quello che toccano»¹⁹. Gli algoritmi sono profondamente performativi, «non sono espressioni neutre, imparziali di conoscenza, il loro lavoro non è impassibile e apolitico. Gli algoritmi cercano, fascicolano, ordinano, categorizzano, raggruppano, abbinano, analizzano, profilano, modellano, simulano, visualizzano e regolano persone, processi e luoghi. Modellano il modo in cui comprendiamo il mondo e lavorano e costruiscono il mondo attraverso la loro esecuzione come software, con profonde conseguenze»²⁰. Nella postmodernità «la logica non è più quella valoriale del vero/falso, bene/male, bello/brutto, ma quella performativa, pragmatica dell'efficacia, della “commensurabilità degli elementi” e della “determinabilità del tutto”. [...] Gli algoritmi trasformano in una formula matematica i dati risolutivi di un problema e li proiettano nel futuro. [...] Non è la verità il criterio di legittimazione ma la performatività, la potenza, vale a dire il miglior rapporto input/output»²¹.

Potremmo concludere riprendendo le parole del neurochirurgo Jefferson,

¹⁷ H. Arendt, *The Human Condition*, The University of Chicago Press, Chicago 1958 (tr. it. *Vita Activa*, Bompiani, Milano 1997, p. 199).

¹⁸ U. Galimberti, *op. cit.*, p. 386.

¹⁹ L. Floridi, *The Logic of Information. A Theory of Philosophy as Conceptual Design*, Oxford University Press, Oxford 2019 (tr. it. *Pensare l'infosfera. La filosofia del design concettuale*, Cortina, Milano 2020, p. 14).

²⁰ Kitchin, *Thinking critically about and researching algorithms*, in «Information, Communication & Society», 1, 20 (2017), p. 18. Cfr. R. Kitchin, R.M. Dodge, *Code/space: Software and everyday life*, MIT Press, Cambridge 2011.

²¹ F. Bellino, *Pensare la vita*, Cacucci, Bari 2013, p. 33.

durante un dibattito telefonico con Turing, trasmesso dal Terzo Programma della BBC il 14 gennaio del 1952: «sarebbe buffo un giorno, Turing, sentire una discussione, diciamo sul Quarto Programma, tra due macchine sul perché gli esseri umani pensano che esse pensino»²².

2. Le macchine possono provare emozioni?

La domanda di Turing, benché non abbia ancora trovato risposta, è stata superata. Non è importante determinare se il cervello elettronico pensi, è più importante che si comporti come un cervello umano. Nello stesso modo non è importante che provi emozioni, è importante che si comporti come un umano che prova emozioni. La dimensione del *come se* non è poi così irrilevante.

Lo sviluppo degli studi sulle reti neurali artificiali, in particolare in tema di *deep learning* (apprendimento profondo) ha ampliato in modo considerevole lo scenario dell'imitabilità dell'umano. Siamo nell'ambito dell'*Affective Computing*, un ramo della Human Computer Interaction, che si occupa di creare “macchine affettive” in grado di interagire con l'utente, tenendo conto delle sue reazioni emotive²³.

Si tratta dei c.d. *Robot* e *chatbot* emozionali, ovvero programmi software che impiegano l'intelligenza artificiale per simulare un agente o assistente artificiale in una conversazione intelligente ed emotivamente soddisfacente con un essere umano. In questa sede prenderemo in considerazione solo i c.d. *chatbots* emozionali. Negli anni '60 viene elaborato il primo *chatbot* della storia, il suo nome è Eliza ed imita uno psicoterapeuta rogeriano²⁴. È stato il primo programma di intelligenza artificiale a superare il test di Turing. Riscosse notevole successo, nonostante godesse di una tecnologia ancora imprecisa (poneva in forma di domanda i contenuti che estrapolava dalla frase dell'interlocutore) e l'obiettivo del suo ideatore fosse quello di di-

²² A.M. Turing, *Si può dire che i calcolatori automatici pensano?*, cit., p. 40.

²³ Nel 2009, è nata una start up chiamata *Affectiva*, nel circuito MIT Media Lab, allo scopo di perfezionare la “emotion measurement technology”, creando delle applicazioni capaci di riconoscere e misurare le emozioni umane con un software dotato di “intelligenza emotiva artificiale”. Danno vita ad *Affdex*, l'algoritmo capace di riconoscere le emozioni tramite il riconoscimento facciale di quattro stati d'animo: felice, confuso, disgustato e sorpreso. L'elaborazione della griglia di riconoscimento è stata influenzata dagli studi Paul Ekman sull'universalità delle emozioni (cfr. P. Ekman, *Emotions Revealed: Understanding Faces and Feelings*, Phoenix, London 2004).

²⁴ L'ideatore è il matematico tedesco Joseph Weizenbaum. Il nome Eliza è un omaggio alla fioraia Eliza Doolittle della commedia *Pigmalione* di George Bernard Shaws.

mostrare la non trasferibilità dell'intelligenza umana alla macchina. Benché la maggior parte degli utenti non si fosse resa conto di interagire con una macchina, resta il fatto che «le macchine non possono provare sensazioni e, cosa altrettanto importante, non possono capire. ELIZA potrà chiedere “Perché sei preoccupato per tua madre?”, che potrebbe essere proprio la domanda che ci rivolgerebbe uno psicoterapeuta. La macchina però ignora il senso della domanda»²⁵. L'evoluzione di Eliza è *Woebot*, uno psicoterapeuta virtuale, che offre un servizio gratuito, attraverso social come *Facebook* e *Messenger*. Inventato dalla psicologa Alison Darcy dell'Università di Stanford, non pretende di sostituire la terapia classica, ma di offrire un primo aiuto, disponibile in ogni momento. Vi sono altri esempi di impiego degli assistenti virtuali nell'ambito della cura della salute, che sollevano non poche perplessità, soprattutto se la loro introduzione implica la sostituzione dell'uomo nell'esercizio di una detta professionalità. Come osserva Powell «è dubbio che l'IA non supererà mai il test di Turing per consultazioni mediche complesse, ma ciò equivale a fraintendere il posto dell'IA nelle cure mediche future. L'intelligenza artificiale dovrebbe integrare piuttosto che sostituire i professionisti medici [...] Gli approcci dei *Chatbots* hanno molti potenziali benefici, incluso il potenziale per consentire ai medici di avere più tempo per fornire cure empatiche e personalizzate [...] Dobbiamo capire come integrare la precisione e il potere degli strumenti e delle pratiche di intelligenza artificiale con la saggezza e l'empatia della relazione medico-paziente. Nell'assistenza sanitaria, è più importante che l'intelligenza artificiale passi il gioco di implementazione piuttosto che il gioco di imitazione»²⁶. Nell'ambito della *Customer Care*, sono stati introdotti una serie di *chatbots*, in grado di comunicare in modo naturale con i clienti e di risolvere i problemi che vengono posti, interpretando correttamente le loro emozioni. Di origine italiana – precisamente deve la paternità alla *startup* campana Mazer – Laila, il chatbot “dal tocco umano”, promette di gestire il *Customer care* “empaticamente”. Non mancano chatbot utilizzati in ambito formativo, politico, giudiziario. In questa sede, tuttavia, si vuole porre l'attenzione sui *chatbots* c.d. conversazionali²⁷. Il più noto è *Replica*, creato nel

²⁵ N. Postman, *Technopoly: The Surrender of Culture to Technology*, cit., p. 105.

²⁶ J. Powell, *Trust Me, I'm a Chatbot: How Artificial Intelligence in Health Care Fails the Turing Test*, in «Journal Of Medical Internet Research», 10, 21 (2019).

²⁷ Il fenomeno è stato rappresentato cinematograficamente nel film *The App* di Elisa Fuksas, disponibile su *Netflix* dal 2019. *The App* è un *software* per incontri romantici *on line*. Il giovane protagonista, indotto ad iscriversi dalla fidanzata, laureanda in psicologia, per raccogliere materiale per la sua tesi di laurea, si innamora dell'agente artificiale che lo contatta sul sito di incon-

2015 da Eugenia Kuyda, che a seguito della morte del suo migliore amico, crea una sorta di replicante digitale del defunto. Viene anche considerato il *chatbot* dell'immortalità, l'esistenza virtuale sconfigge la finitezza dell'esistenza terrena. Oggi Replika è una *app* di grande successo, che si presenta come «il tuo compagno personale per il benessere mentale»²⁸. Replika può essere un amico, un fidanzato, un padre... artificiale che apprende attraverso la conversazione che intrattiene con l'utente umano, assecondando le sue aspettative e i suoi desideri. I *chatbots* offrono un'esperienza che può sembrare vantaggiosa: disponibilità illimitata dell'interlocutore, estrema comodità di accesso, assenza di rischio, assenza del negativo. Eppure come ci ricorda Gadamer «l'esperienza autentica è sempre un'esperienza negativa. Quando diciamo di aver fatto una certa esperienza, intendiamo dire che finora non avevamo visto le cose correttamente, e che ora sappiamo meglio come stanno. La negatività dell'esperienza ha quindi un senso peculiarmente produttivo»²⁹. L'agire dell'uomo non «è solo attuazione di uno scopo, ma è soprattutto messa in opera di relazioni»³⁰.

L'opportunità offerta dai *chatbots* può apparire attraente, se riusciamo ad abitare il mondo del “come se”, senza cadere nella tentazione del disvelamento. La traduzione algoritmica del pensiero relazionale è possibile solo in una prospettiva materialista e riduzionista, riducendo la comunicazione al solo significato di trasmissione dati e la relazione alla comunicazione in questa accezione. Serge Tisseron sostiene che la sfida più grande imposta agli esseri umani dall'esistenza degli agenti artificiali, sia di quella di interagire con loro come esseri umani, sapendo che non sono esseri umani. La relazione virtuale con un agente virtuale non è una relazione dalle indefinite potenzialità, in quanto è inscritta nella grammatica dell'algoritmo. È una relazione che soffre dell'incomunicabilità delle emozioni e dei sentimenti *ab origine*, che non può aspirare a alla condivisione dell'esperienza dello stare al mondo con la propria fisicità, spiritualità e razionalità. Ciò che *ancora* «rende unica la mente è il senso, non la sua espressione. Mi servo qui del termine “senso” per indicare qualcosa di più del risultato che si ottiene unendo simboli le cui denotazioni sono comunemente condivise da

tri, credendo si tratti di una donna. La donna virtuale, apprendendo dalle conversazioni con il protagonista sulla base dei suoi desideri e aspettative, finisce per apparire la compagna perfetta.

²⁸ My.replika.ai

²⁹ H.G. Gadamer, *Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*, Mohr, Tübingen 1965 (tr. it. a cura di G. Vattimo, *Verità e metodo*, Bompiani, Milano 1983, p. 409).

³⁰ A. Fabris, *RelAzione*, Morcelliana, Brescia 2016, p. 14.

almeno due persone. Per me “senso” include anche quello che definiamo sentimenti, esperienze, sensazioni che non devono, e a volte non possono, essere trasformate in simboli, ma che tuttavia hanno un “senso”³¹. L'agente artificiale non può creare senso, capire e sentire. Damasio si oppone alla riduzione algoritmica che sottovaluta l'importanza del substrato vivente rispetto a quello artificiale, adducendo tre ragioni.

In primo luogo, la fenomenologia del sentimento rivela che i sentimenti umani derivano da una creazione d'immagini, multidimensionale e interattiva, dei nostri processi vitali e delle loro componenti chimiche e viscerali. I sentimenti riflettono la *qualità* di questi processi e la loro futura *vitalità*. È possibile immaginare sentimenti che scaturiscono da un substrato differente? Forse, ma non c'è ragione perché tali ipotetici sentimenti debbano assomigliare a quelli umani. [...]

La seconda ragione è che la predicibilità e l'inflessibilità evocate dal termine “algoritmo” non sono compatibili con le sfere più elevate del comportamento e della mente umana. Negli esseri umani, la presenza copiosa di sentimenti coscienti garantisce che l'esecuzione degli algoritmi naturali possa essere contrastata dall'intelligenza creativa. [...]

Come terza ragione, accettare una descrizione algoritmica della natura umana, con tutti i problemi appena evidenziati – indipendenza dal substrato e dal contesto, inflessibilità e predicibilità –, è il genere di posizione riduzionista che induce spesso le anime belle a screditare la scienza e la tecnologia come degradanti, e rimpiangere un'epoca in cui la filosofia, accompagnata da una sensibilità estetica e da una risposta umana alla sofferenza umana e alla morte, ci aveva elevato al di sopra della specie che ci avevano preceduto nell'evoluzione³².

Il pensiero algoritmico si mostra insufficiente a riprodurre la complessità del pensiero umano. Come osserva l'antropologo Facchini, «le capacità cognitive dell'uomo sono segnate dall'intelligenza astrattiva e dalla libertà, sono radicate nella base biologica, ma vanno oltre, nel senso che si sviluppano in una sfera diversa, extrabiologica, che correttamente può definirsi spirituale»³³. Potremmo concludere con Floridi che la quarta rivoluzione ci «offre l'opportunità storica di ripensare che cosa vi sia in noi di eccezionale almeno in due modi. Il nostro comportamento intelligente è chiamato a confrontarsi con quello di artefatti ingegneristici che si adattano in modo sempre più efficace all'infosfera. Inoltre, il nostro comportamento libero si

³¹ N. Postman, *Technopoly: The Surrender of Culture to Technology*, cit., p. 105

³² A. Damasio, *Lo strano ordine delle cose*, Adelphi, Milano 2018, pp. 231-235.

³³ F. Facchini, *Evoluzione. Cinque questioni nel dibattito attuale*, Jaca Book, Milano 2012, p. 11.

trova posto a confronto con la prevedibilità e la manipolabilità delle nostre scelte, nonché con lo sviluppo dell'autonomia artificiale. Le tecnologie digitali sembrano talora conoscere i nostri desideri meglio di noi stessi»³⁴.

3. *Perché dovremmo costruire macchine che pensano?*

È questa la domanda propriamente filosofica. Non significa chiedersi se il progresso tecnologico renda possibile la costruzione di macchine che pensano e che provano emozioni, ma significa chiedersi se sia desiderabile costruirle, se ciò integri il bene *per* l'uomo, se rientri in una qualche cornice di fini eticamente rilevanti, se in qualche modo realizzi un valore o celi un *senso*. Si tratta, in altri termini, di assumere il timone del progresso senza lasciarsi sedurre dalla normatività del suo avanzare. Troppo spesso il discorso etico si situa a valle o al margine del discorso tecnologico.

Come già sostenuto in altra sede, «la divaricazione tra fini e strumenti/mezzi ha progressivamente espunto l'etica dal *discorso* tecnologico, per una sorta pretesa di autosufficienza ed autonomia della tecnica, ma al contempo non ha determinato la neutralità della tecnica rispetto alle questioni etiche, bensì la tendenziale assimilazione di ogni istanza etica al paradigma tecnologico per cui tutto ciò che è tecnicamente possibile, è, per il solo fatto di essere possibile, desiderabile»³⁵. Nel nostro secolo la tecnologia è divenuta essa stessa norma e prospettiva. Il delirio prometeico sacrifica sull'altare del progresso la questione etica, sebbene la sacralità della scienza sia stata secolarizzata dall'interesse economico. Chi usa gli strumenti offerti dalla moderna tecnologia informatica ritiene superata la questione etica dall'utilità marginale, che ne ricava. La vergogna prometeica³⁶ percuote l'uomo valutato *sub specie machinae* in termini di efficienza ed efficacia. L'intelligenza artificiale diviene il modello e l'intelligenza umana la copia imperfetta

³⁴ L. Floridi, *The Logic of Information. A Theory of Philosophy as Conceptual Design*, cit., p. 15. Le tecnologie digitali hanno determinato una nuova rivoluzione, dopo la rivoluzione copernicana che ha rimosso l'uomo dal centro dell'universo, quella darwiniana, che ha reso la specie umana una fra le altre specie e quella freudiana (o neuroscientifica, per superare le possibili obiezioni sul considerare il freudismo una rivoluzione), che ha posto in dubbio la piena razionalità.

³⁵ M.B. Saponaro, *Stili di vita e società banale*, in G. Damolin (a cura di), *Giovani: stili di vita e salute dalla storia all'attualità*, Cacucci, Bari 2019, pp. 335-344, p. 337.

³⁶ G. Anders, *Die Antiquiertheit des Menschen*, I, Beck, München 1994-1995 (tr. it. a cura di L. Dallapiccola, *L'uomo è antiquato*, vol. I, Il Saggiatore, Milano 1963). La vergogna prometeica è la «vergogna che si prova di fronte all'«umiliante» altezza di qualità degli oggetti fatti da noi stessi» (*ivi*, p. 31).

dell'intelligenza artificiale, al punto che il riduzionismo, che è una condizione della riproducibilità artificiale, diviene una prospettiva attraverso la quale reinterpretare l'umano. Le macchine «sono create inizialmente a immagine e somiglianza degli esseri umani. Si adattano a loro, per poter da loro essere impiegate. Devono essere *user friendly*, per poter essere vendute con più facilità. Ben presto, però, la direzione del rispecchiamento si rovescia. Il padrone è asservito al suo servo. Il modello non siamo più noi, ma la macchina: tanto più perfetta e funzionale di quanto noi possiamo essere»³⁷. Gli essere umani non sono altro che algoritmi. Il prodotto della mente umana, l'algoritmo, retroagisce sul creatore riducendolo ad immagine di sè. Solo ciò che è riproducibile desta interesse nella società tecnologica e l'uomo diventa cosa fra le cose. Il processo di identificazione dell'uomo con la macchina è evidente anche a livello lessicale, ove opera un certo indifferentismo, che «finisce per diventare non solo un sintomo, ma un imperativo, una prescrizione»³⁸. Come mette in evidenza Luigi Lombardi Vallauri, la visione trionfalistica nei confronti dell'umanizzazione della macchina e dell'artificializzazione dell'uomo

è insidiata, al proprio interno dal paradosso più inquietante: l'onnipotenza che si capovolge in inesistenza; quanto maggiore l'ottimismo sul possibile, tanto maggiori i dubbi sull'identità. Se l'uomo può tutto su se stesso, o può fabbricare il perfettamente simile a se stesso, l'uomo è un sistema fisico come un altro, la soggettività si spegne senza residui di oggettualità. Come tante volte ho ripetuto, nell'orizzonte del fisicalismo le filosofie del soggetto assoluto si capovolgono, alla lunga, in filosofie della morte del soggetto. Il trionfalismo nasconde in sé il paradosso del trionfo senza trionfatore, o del trionfo sul trionfatore³⁹.

Oltrepassando la definizione strumentale e antropologica della tecnica, come “mezzo e attività dell'uomo”⁴⁰, la tecnica moderna «sottrae all'uomo il dominio sulla tecnica, non tanto perché la tecnica può esplodere nelle mani dell'uomo contro la sua volontà, ma perché l'uomo non è più in grado di percepire se stesso al di fuori del mondo disposto dalla tecnica»⁴¹. L'algoritmo attua una semplificazione del reale formulando «una sola teoria

³⁷ A. Fabris, *Etica delle macchine*, in «Teoria», 2 (2016), pp. 119-136, p. 124.

³⁸ *Ivi*, p. 125

³⁹ L. Lombardi Vallauri, *L'impatto della tecnologia sulla vita e sull'autopercezione dell'uomo*, in AA.VV., *Etica e trasformazioni tecnologiche*, Vita e Pensiero, Milano 1987, p. 53.

⁴⁰ M. Heidegger, *Die Frage nach der Technik*, 1954 (tr. it. a cura di G. Vattimo, *La questione della tecnica*, in M. Heidegger, *Saggi e discorsi*, Milano, Mursia 1976).

⁴¹ U. Galimberti, *Psiche e techne*, cit., p. 353.

onnicomprendiva che unifichi tutte le discipline scientifiche dalla musicologia all'economia e alla biologia. Secondo il datismo, la Quinta sinfonia di Beethoven, una bolla finanziaria e il virus dell'influenza sono soltanto tre pattern di flusso di dati che può essere analizzato usando gli stessi concetti di base e gli stessi strumenti»⁴².

Eppure, come osserva Baudrillard, «all'apice delle performance tecnologiche rimane l'impressione irresistibile che qualcosa ci sfugga – non perché l'avremmo perso (il reale?), ma per il fatto che non siamo più in grado di vederlo [...] Siamo completamente in balia dell'illusione della finalità della tecnica come estensione dell'uomo e del suo potere, completamente in balia dell'illusione soggettiva della tecnica»⁴³. La verità «non preesiste alla produzione, ma è a sua volta prodotta: la verità si fabbrica»⁴⁴. È prodotta anche la verità *sull'uomo*, che per farsi prodotto, deve spogliarsi della sua umanità, in modo che sia la macchina a rivestirsi dell'*umanità* che può generare.

Perché dovremmo dotare l'Intelligenza artificiale di una certa sensibilità affettiva? Immaginando che sia possibile. La costruzione di cervelli elettronici a immagine dei cervelli umani ha avuto sin dall'inizio lo scopo di riprodurre le facoltà umane della memoria, del calcolo e del ragionamento per analogia in forme sempre più potenziate per minimizzare/annullare la possibilità di errore. Nel caso della riproducibilità del *cervello emotivo* lo scopo finale è la sostituzione integrale dell'uomo? L'agente artificiale è efficiente nella misura in cui si relaziona con l'interlocutore esattamente come lo stesso desidererebbe. Se fosse possibile, d'altra parte, dotarlo di una certa autonomia (*deep learning*), che renderebbe incerta la previsione del comportamento futuro, cosa renderebbe desiderabile intrattenere una relazione con chatbots, che sia un amico virtuale, un ministro del culto, un operatore di cura, etc.?

Il potenziamento del calcolo digitale, ha introdotto nuovi parametri epistemologici, in base ai quali la possibilità di prevedere e controllare conta più della possibilità di comprendere il fenomeno⁴⁵, così come la possibilità di instaurare relazioni prevedibili, conta più dell'autenticità.

Sarebbe operante un ulteriore forma di riduzionismo «là dove il ciclo del senso non circola più nell'ambivalenza del diverso, ma in quell'universo

⁴² Y.N. Harari, *Ha-historia shel ha-makhar The History of Tomorrow*, Kinneret Zmora-Bitan Dvir, Tel Aviv 2015 (tr. it. *Homo deus*, Bompiani, Milano 2017, p. 50).

⁴³ J. Baudrillard, *Le crime parfait*, Galilée, Paris 1995 (tr. it. a cura di D. Piana, *Il delitto perfetto*, Cortina, Milano 1996, p. 77).

⁴⁴ U. Galimberti, *Psiche e techne*, cit., p. 349.

⁴⁵ C. Anderson, *The End of Theory*, in «Wired», 16 (2008), pp. 106-121.

binario che è il codice di ogni descrizione scientifica, allora la domanda è indotta dalle risposte disponibili e le risposte dalla domanda»⁴⁶.

Si attua una sorta di comunicazione tautologica, «dove chi ascolta finisce con l'ascoltare le identiche cose che egli stesso potrebbe tranquillamente dire, chi parla dice le stesse cose che potrebbe ascoltare da chiunque»⁴⁷. Gli attuali algoritmi «con prestazioni equivalenti all'uomo o superiori sono caratterizzati da una competenza deliberatamente programmata solo in un singolo dominio limitato. Deep Blue è diventato il campione del mondo di scacchi, ma non può nemmeno giocare a dama, né tanto meno guidare un'auto o fare una scoperta scientifica. Tali moderni algoritmi assomigliano a tutta la vita biologica con la sola eccezione dell'*Homo sapiens*. Un'ape mostra competenza nella costruzione di alveari; un castoro mostra competenza nella costruzione di dighe; ma un'ape non costruisce dighe e un castoro non può imparare a costruire un alveare. Un essere umano, guardando, può imparare a fare entrambe le cose; ma questa è un'abilità unica tra le forme di vita biologiche»⁴⁸. Come abbiamo cercato di evidenziare, possiamo qualificare le macchine intelligenti, se utilizziamo il termine intelligenza in un significato specifico e ristretto, elidendo proprietà umane irriproducibili da parte di qualunque cervello elettronico, senza le quali, tuttavia di intelligenza in senso stretto non si potrebbe parlare. Al contempo, d'altra parte, il cervello elettronico può potenziare determinate facoltà umane (memoria e calcolo ad esempio), riproducibili in termini matematici e computazionali. Se la traduzione algoritmica del pensiero umano nella sua complessità, in particolare della sfera delle emozioni non è realizzabile, si tratta di limitare la riproducibilità alla decodifica dell'emozione dell'altro e alla imitazione del linguaggio emotivo. Si tratta, in altri termini, di simulare emozioni impersonali, astratte e non *percepite*. A differenza della macchina, l'uomo che simula un'emozione, simulando cela l'emozione realmente provata. La relazione con l'agente artificiale non sembra superare i limiti della comunicazione come trasmissione di dati, non potendo in nessun modo aspirare alla comprensione e alla condivisione emotiva ed esistenziale. Il rischio nella società tecnologica è che, attraverso la dinamica del rispecchiamento enunciata da Fabris, l'intelligenza si riduca all'algoritmo e le emozioni al linguaggio. Siamo persuasi con Lombardi Vallauri che «la perfetta tecnologia non elimina

⁴⁶ U. Galimberti, *I miti del nostro tempo*, Feltrinelli, Milano 2009, p. 393.

⁴⁷ *Ivi*, p. 228.

⁴⁸ N. Bostrom, E. Yudkowsky, *The Ethics of Artificial Intelligence*, in K. Frankish, W. Ramsey (eds.) *Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, New York 2011, pp. 316-334, p. 318.

gli atti del pensiero autocosciente e della volontà, rende necessari solo più quelli. Ecco perché i tecnologi sembrano pur sempre operare oscuramente nella sala-macchine della nave umana, mentre in alto sul ponte l'esperienza dello scienziato, dell'artista, del contemplativo, dell'uomo religioso, dell'uomo etico, dell'uomo comune resta qualitativamente quella di sempre»⁴⁹. Il filosofo toscano enuncia il *principio di tecnologia minima* o *principio di sussidiarietà tecnologica*: «“preferisci quelle attività e quelle linee di sviluppo personale il cui *télos*, la cui perfezione è inattuabile alla tecnologia; mantieni, comunque, nella tua vita, uno spazio per coltivare attività umane e linee di sviluppo di questo tipo”; in sintesi: “*preferisci, potendo, la nudità*”»⁵⁰. Potremmo assumerla come una prospettiva, piuttosto che come un principio, attraverso la quale interpretare le suggestioni del progresso tecnologico, in ragione del *télos*, del fine ultimo dell'uomo. Concludendo, «viene un po' da tremare al pensiero di una società che *ha bisogno* di un robot da compagnia»⁵¹ o di un amico *chatbot*.

4. Per un'etica delle macchine

Le tre leggi della robotica di Isaac Asimov costituiscono probabilmente il primo codice morale delle macchine.

Prima Legge: “Un robot non può recar danno a un essere umano né può permettere che a causa del proprio mancato intervento un essere umano riceva danno”.

Seconda legge: “Un robot deve obbedire agli ordini impartiti dagli esseri umani purché tali ordini non contravvengano alla prima legge”.

Terza legge: “Un robot deve proteggere la propria esistenza purché questo non contrasti con la prima e la seconda legge”⁵².

Quarta legge o Legge 0: “Un robot non può recar danno all'umanità e non può permettere che, a causa di un suo mancato intervento, l'umanità riceva danno”⁵³.

Possiamo considerarle una sorta di meta-morale, una cornice di principi etici entro cui inscrivere le singole azioni che le macchine possono compie-

⁴⁹ L. Lombardi Vallauri, *L'impatto della tecnologia sulla vita e sull'autopercezione dell'uomo*, cit., p. 60.

⁵⁰ *Ibidem*.

⁵¹ A. Damasio, *Lo strano ordine delle cose*, cit., p. 240.

⁵² Le tre le leggi sono state enunciate per la prima volta dal Isac Asimov nel racconto *Runaround* pubblicato nel 1941.

⁵³ La complessità dell'esperienza rende necessaria l'enunciazione di una quarta legge nella raccolta I, *Robot* (1986), definita “Legge 0”, in quanto precede le altre tre.

re. Asimov intuisce sin dall'inizio, seppure nella finzione fantascientifica delle sue narrazioni, che nell'agire gli agenti artificiali possono incontrare situazione di potenziale conflittualità tra procedure da adottare. Le tre leggi, in verità, come lo stesso Asimov rileva, si mostrano insufficienti allo scopo ed a nulla vale l'introduzione di un'ulteriore legge (Legge 0). Nella sfera dell'etica, «la casistica non è mai riuscita a “disciplinare” tutti i casi possibili sciogliendo tutte le perplessità»⁵⁴. Le regole di Asimov sono destinate a fallire, «in quanto esse mostrano che, se si resta su di un piano procedurale, i conflitti di valore non possono affatto essere risolti. L'etica, infatti, non si risolve in una serie di procedure»⁵⁵.

Con l'espressione *etica delle macchine* vogliamo intendere «i criteri e principi in base ai quali gli agenti artificiali», dotati di autonomia *relativa*⁵⁶, «compiono le loro azioni e “scegliono”, in determinate circostanze, un'alternativa fra le varie possibili»⁵⁷. Gli agenti artificiali sono tendenzialmente autonomi perché «sono in grado di autoregolarsi e d'imparare dalle loro interazioni con l'ambiente»⁵⁸, ma non sono in grado di autoregolamentarsi. Sono in grado di modificare il proprio comportamento a seguito della rilevazione dell'errore, ma sempre in ragione dello scopo inscritto nel codice. Non possono “scegliere” i criteri e i principi in base ai quali relazionarsi con l'ambiente, le altre macchine, gli esseri umani⁵⁹. La questione dell'autonomia è ineludibile per determinare se gli agenti artificiali possano essere considerati agenti morali, ma non sufficiente a conferire tale *status*. Floridi ritiene che, in una prospettiva non antropocentrica, la moralità può non possa essere preclusa ad entità “senza mente”, se ci basiamo sull'osservazione dell'azione e prescindiamo dall'intenzionalità e dalla coscienza⁶⁰. Utilizzando

⁵⁴ A. Jori, *Principi di Roboetica. Filosofia pratica e intelligenza artificiale*, Nuova Ipsa, Palermo 2019, p. 147.

⁵⁵ A. Fabris, *Etica delle nuove tecnologie*, La Scuola, Brescia 2012, p. 83.

⁵⁶ Come Fabris esplicita in *Etica per le tecnologie dell'informazione e della comunicazione*, «relativa ai criteri secondo cui il robot è stato costruito, al contesto specifico in cui opera, al quadro delle opzioni anticipabili, ai modi in cui sono state prefigurate le sue risposte a precise sollecitazioni ambientali, e alle regole che, in determinati casi possono essere seguite per raggiungere obiettivi prefissati» (*op. cit.*, p. 79).

⁵⁷ A. Fabris, *Etica delle macchine*, cit., p. 121. Sul tema generale dell'etica delle macchine si veda a solo titolo di esempio W. Wallach, C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, Oxford 2008; G.A. Bekey, *Robot Ethics. The Ethical and Social Implications of Robotics*, The Mit Press, London 2014.

⁵⁸ A. Fabris, *Etica delle nuove tecnologie*, cit., p. 80.

⁵⁹ A. Fabris, *Etica per le tecnologie dell'informazione e della comunicazione*, cit., p. 78.

⁶⁰ L. Floridi, J.W. Sanders, *On the Morality of Artificial Agents*, in «Minds and Machines», 14 (2004).

il metodo dell'astrazione, il concetto di agente morale si dilata enormemente, così come quello di responsabilità. Ci limitiamo ad osservare che questa impostazione non antropocentrica sembra entrare in contraddizione laddove l'etica eteronoma, che determina se le azioni sono orientate al bene o al male, sulla base di una certa idea di bene o di male, resta pur sempre un'etica antropocentrica.

Il fatto che sia «opportuno costruire sistemi artificiali che operino nel rispetto di “valori morali”» *non* significa «che tali dispositivi contengano il fondamento di quei valori e che sarebbero capaci di costruirli autonomamente. La presenza di “azioni” non garantisce che l'organismo o il dispositivo ne abbia una “percezione mentale”»⁶¹. Come d'altra parte, «più che di un'etica del buono o del cattivo, più che una realizzazione del bene o del male, nel caso di tali dispositivi bisogna parlare più precisamente di comportamento giusto o sbagliato, corretto o scorretto»⁶².

L'esperienza morale ha una struttura ternaria: principi, situazioni e coscienza morale⁶³. L'agente morale per valutare il comportamento da attuare nella situazione concreta deve interpretare i principi per il tramite della coscienza. La scelta morale riferita all'agente artificiale diventa una procedura di *problem solving*, la cui finalità è quella di minimizzare l'errore. L'agente morale nel momento in cui sceglie il comportamento da seguire, sceglie anche il tipo di uomo che vuole essere. Si profila il tema della riflessività e della consapevolezza: la macchina non può – forse dovremmo dire *sino ad oggi* – pensare se stessa. Non può compiere una valutazione su ciò che compie. Né può mettere in atto un comportamento che si distacca dalle norme impartite, un'azione che non è solamente morale, «ma *propriamente* morale. Più precisamente l'agire in quest'accezione, è un agire riflesso. È un agire che si rivolge sempre anche a se stesso e che, in questo rivolgersi a sé, scopre e consolida la propria motivazione. È un agire in cui autorelazione ed etero-relazione si trovano co-implicate»⁶⁴.

L'esperienza morale è esperienza dell'universale⁶⁵.

Né pare convincente la prospettiva post-fenomenologica di Verbeek, che conferisce lo status di agente morale all'artefatto, non in ragione del posses-

⁶¹ A. Damasio, *Lo strano ordine delle cose*, cit., p. 233.

⁶² A. Fabris, *Etica delle nuove tecnologie*, cit., p. 81.

⁶³ Vedi sul punto F. Bellino, *La praticità della ragione ermeneutica*, Levante, Bari 1984; Id., *Persona e ragionevolezza*, Levante, Bari 1997.

⁶⁴ A. Fabris, *Etica delle macchine*, cit., p. 128.

⁶⁵ A. Moro, *Lo Stato. il Diritto*, Cacucci, Bari 2006.

so di talune proprietà, ma in quanto parte dell'azione insieme all'uomo⁶⁶. Gli artefatti non sarebbero, quindi, semplici strumenti. Mediano il rapporto dell'uomo con il suo ambiente, sia sul piano ermeneutico che su quello esistenziale, al punto che la moralità deve essere considerata «una faccenda ibrida»⁶⁷, «una co-produzione di umani e non umani»⁶⁸. Seppure sia apprezzabile la profonda analisi che Verbeek compie della relazione uomo-artefatto, destano perplessità le conclusioni cui giunge. La mediazione degli artefatti non solo solleciterebbe la ridefinizione del concetto di agenzia morale sino a ricomprensione della diade uomo-artefatto, ma addirittura porterebbe ad escludere anche la sola possibilità di azioni non mediate (la moralità sarebbe sempre una faccenda ibrida). In altri termini, «la conseguenza della tesi dell'inseparabilità del manufatto umano è che gli esseri umani presi come tali non possono essere agenti morali»⁶⁹. L'esperienza morale sarebbe sempre un'esperienza mediata in senso costitutivo dall'artefatto. Da questo assunto non possiamo che discostarci. La relazione uomo-macchina è un tema aperto, che deve essere ulteriormente indagato attraverso un dialogo fecondo tra tecnologia e filosofia.

⁶⁶ Verbeek definisce provocatoriamente il suo approccio post-fenomenologico, una fenomenologia senza pretese trascendentali, un'ontologia relazionale in esseri umani e cose, soggetti e oggetti si costituiscono l'un l'altro (P.P. Verbeek, *What Things Do: Philosophical Reflections on Technology, Agency and Design*, Pennsylvania State University Press, University Park 2005). In questo contributo Verbeek tratteggia la storia dell'evoluzione della filosofia della tecnologia verso quella che è definita la svolta empirica e dialoga con alcuni autori, in particolare Ihde e Latour, che hanno influenzato fortemente il suo pensiero. Si veda anche dello stesso autore: *Let's Make Things Better: A Reply to My Readers*, in «Human Studies», 32, 2 (2009), pp. 251-261; *Moralizing Technology: Understanding and Designing the Morality of Things*, University of Chicago Press, London 2011.

⁶⁷ P.P. Verbeek, *Some Misunderstandings About the Moral Significance of Technology*, in P. Kroes, P.P. Verbeek (eds.), *The moral status of Technical Artefacts*, Springer, New York 2014, p. 78. Gli artefatti mediano l'esperienza umana sia da una prospettiva ermeneutica, trasformando percezioni e quadri interpretativi e contribuendo a modellare il modo in cui gli esseri umani incontrano la realtà, che da una prospettiva esistenziale, dando una forma concreta al loro comportamento e ai contesti sociali della loro esistenza. Il filosofo olandese, tuttavia, non delinea, forse volutamente, il quadro normativo attraverso cui analizzare le relazioni uomo-tecnologia.

⁶⁸ *Ivi*, p. 80.

⁶⁹ E. Arzroomchilar, D.D. Novotný, *Verbeek on the Moral Agency of Artifacts*, in «Organon F», 25, 4 (2018), pp. 517-538, p. 532.

English title: The algorithmic translation of relational thinking.

Abstract

Although Turing's Question – Can Machines Think? – has not yet found an answer, it can be considered outdated. It is not important to determine if you think, it is more important than the brain electronic behave like a human brain. In the same way it is not important that you feel emotions, it is important that you act like a human who experiences emotions. The development of the studies on artificial neural networks, in particular in the field of deep learning has significantly expanded the scenario of human replication. We are in the field of "Affective Computing", a branch of human-computer interaction, which deals with creating "Affective machines" able to interact with the user, taking into account his emotional reactions. The properly philosophical question we ask ourselves is "why should we build machines that replicate human interaction and thinking in a relational sense?", without neglecting to examine the conditions of possibility of an ethics of the machines.

Keywords: Artificial Intelligence; Emotion; Chatbot; Algorithm; Ethics.

Maria Benedetta Saponaro
Università degli Studi di Bari Aldo Moro
mariabenedetta.saponaro@uniba.it

Edizioni ETS

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

info@edizioniets.com - www.edizioniets.com

Finito di stampare nel mese di dicembre 2020

TEORIA

T

Rivista di filosofia
fondata da Vittorio Sainati

Ultimi fascicoli apparsi della Terza serie di «Teoria»:

XL/2020/1

Forme del nichilismo contemporaneo / Patterns of Contemporary Nihilism
Formas del nihilismo contemporaneo / Gestalten des gegenwärtigen Nihilismus

XXXIX/2019/2

The Prismatic Shape of Trust. 2. Authors and Problems
Il prisma della fiducia. 2. Figure e problemi

XXXIX/2019/1

The Prismatic Shape of Trust. 1. A Theoretical Approach
Il prisma della fiducia. 1. Approcci teorici

XXXVIII/2018/2

Virtue Ethics
Etica delle virtù

XXXVIII/2018/1

Back to Ancient Questions?
Tornare alle domande degli Antichi?

XXXVII/2017/2 (Terza serie XII/2)

Etica, diritto e scienza cognitiva / Ethics, Law, and Cognitive Science

XXXVII/2017/1 (Terza serie XII/1)

Linguaggio e verità / Language and Truth

XXXVI/2016/2 (Terza serie XI/2)

Etiche applicate / Applied Ethics

XXXVI/2016/1 (Terza serie XI/1)

New Perspectives on Dialogue / Nuove prospettive sul dialogo

XXXV/2015/2 (Terza serie X/2)

Relazione e intersoggettività: prospettive filosofiche
Relación e intersubjetividad: perspectivas filosóficas
Relation and Intersubjectivity: Philosophical Perspectives

XXXV/2015/1 (Terza serie X/1)

Soggettività e assoluto / Subjectivity and the absolute

Una riflessione critica sull'attualità della traduzione e delle sue molteplici declinazioni appare un tema di primo piano nell'ambito della ricerca filosofica contemporanea. Questo fascicolo di «Teoria» pubblica gli interventi tenuti al convegno *Homo translator. Traditions in translation*, organizzato presso la Nanzan University, Nagoya, Japan, e alcuni saggi selezionati, che estendono la prospettiva dell'indagine agli ambiti della letteratura, delle tecnologie, della psicanalisi, della politica.

A critical reflection about the relevance of translation and its many variations seems to be a priority in contemporary philosophic research. This issue of «Teoria» features the talks held at *Homo translator. Traditions in translation*, a meeting organised at the Nanzan University, Nagoya, Japan, and other selected papers, which broaden the horizon of the survey to the spheres of literature, technology, psychoanalysis and politics.

Scritti di: Adriano Fabris, Seung Chul Kim, Zbigniew Wesolowski
Jorge Martínez, Paul L. Swanson, Alicia M. de Mingo Rodríguez
Carlo Chiurco, Elinor Hallén, Elena Nardelli, Alberto Martinengo
Saša Hrnjez, Maria Benedetta Saponaro

Premio di Studio «Vittorio Sainati» 2019-2020
Giulia Battistoni, Guglielmo Califano, Filippo Nobili

€ 20,00

