

PER UN UMANESIMO DIGITALE

Giovanni Cucci S.I.

Le novità che l'introduzione del digitale nelle sue variegate sfaccettature, non sempre facili da separare con precisione (computer, web, robot, intelligenza artificiale), presenta in ogni campo della vita umana sono sterminate e affascinanti; tutto ciò nello stesso tempo suscita anche interrogativi rilevanti. La rivista si è occupata di queste tematiche anche recentemente¹.

27

Già da tempo i nuovi ritrovati tecnologici sono in grado di svolgere in maniera estremamente efficace diversi compiti della vita umana. È indubbia la velocità e l'efficienza garantite dagli elaboratori elettronici per gestire un numero sempre più grande di operazioni della vita ordinaria che sembrerebbero superare le capacità umane: si pensi all'organizzazione dei collegamenti, alla velocità di calcolo, alla mole di informazioni rese accessibili dai database. La capacità di «memoria» di un computer consente di reperire in breve tempo informazioni che richiederebbero altrimenti complesse e faticose procedure.

Senza smentire tali indubbi vantaggi, vi sono tuttavia altri aspetti che destano preoccupazione nel momento in cui si vorrebbe affidare alle macchine una fetta sempre più ampia di decisioni; non a caso, la gran parte di film e romanzi dedicati all'argomento – come

1. Cfr A. SPADARO – TH. BANCHOFF, «Intelligenza artificiale e persona umana. Prospettive cinesi e occidentali», in *Civ. Catt.* 2019 II 432–443. Uno dei problemi maggiori nella trattazione di queste tematiche (che interpellano saperi diversi come ingegneria, psicologia e filosofia) è quello di fornire definizioni precise. Rifacendosi alla nozione di intelligenza umana, Roberto Cordeschi definisce l'intelligenza artificiale come la capacità di «riprodurre con programmi per calcolatore, o con robot mobili, comportamenti che, se osservati negli esseri umani, o più in generale negli esseri viventi, verrebbero definiti “intelligenti”» (R. CORDESCHI, «Intelligenza artificiale», in *Enciclopedia filosofica*, Milano, Bompiani, 2006, vol. 6, 5734).

ARTICOLI

anche diversi saggi scientifici – mettono in guardia da una possibile «dittatura del digitale» che ridurrebbe l'essere umano a diventare schiavo delle macchine che ha costruito.

Consapevoli della vastità e complessità di tale problematica, ma anche della sua attualità che incrocia saperi molto differenti, in questo contributo prenderemo in considerazione alcuni ambiti specifici della vita umana nei quali l'assenza di controllo può portare a gravi conseguenze sociali; e altri invece dove l'apporto dell'intelligenza artificiale può offrire un contributo, oltre che utile, anche correttivo nei confronti della volubilità umana. Infine cercheremo di illustrare come alcune attività della vita ordinaria mostrino una misteriosa complessità che si pone su un livello qualitativamente differente rispetto all'intelligenza artificiale.

28

Luci e ombre dell'automazione

Nel 2012 diversi disabili dello Stato dell'Idaho (Usa) si videro improvvisamente decurtare le sovvenzioni mediche. Il Dipartimento di Stato aveva deciso di affidare a un *software* la valutazione del *budget*, per stabilire in maniera pianificata chi avesse diritto a un aiuto economico. Quando, di fronte alle proteste degli interessati, si cominciarono a compiere delle verifiche, emersero situazioni di grande arbitrarietà: alcuni avevano ricevuto senza motivo un notevole aumento, altri si erano visti tagliare il sussidio, anche se la loro situazione nel frattempo era peggiorata. Il Dipartimento, in nome della segretezza del brevetto, si rifiutò di dare spiegazioni, limitandosi a confermare quanto elaborato dal *software*. Come si scoprì più tardi, quando la cosa finì in tribunale (dopo 4 anni di battaglie giudiziarie, 4.000 denunce e un'azione legale collettiva), il programma utilizzato era di pessima qualità, pieno di imprecisioni, al punto da essere ufficialmente considerato illegale. Le conseguenze per la salute di molti furono enormi, a motivo di due errori di base: quello di calcolo del *software* e, soprattutto, quello di chi aveva conferito alla macchina una fiducia totale, rifiutandosi di compiere verifiche².

2. H. FRY, *Hello World. Essere umani nell'era delle macchine*, Torino, Bollati Boringhieri, 2019, 25 s.

PER UN UMANESIMO DIGITALE

Il secondo esempio mostra invece come il giudizio umano necessiti di una correzione che può giungere efficacemente da un elaboratore. I casi dibattuti in tribunale sono affidati a un giudice, il cui responso è legato a una serie di elementi estremamente variegati, che non sono riconducibili unicamente alla competenza, all'esperienza e alla conoscenza dei dettagli della vicenda, ma anche a parametri meramente soggettivi, come l'impressione generale sull'accusato, la valutazione delle sue caratteristiche personali (livello di istruzione, *status* sociale, etnia, sesso) e anche, più semplicemente, l'umore del momento o l'orario della giornata. Una ricerca compiuta sulle sentenze nel Regno Unito mostrava che «la probabilità di ottenere la libertà su cauzione era maggiore subito dopo una pausa e minore poco prima che il giudice facesse uno spuntino»³.

Al di là delle differenze temperamentali, la valutazione di un caso è sempre esposta a una serie di variabili aleatorie. La vicenda di due imputati (Nicholas Robinson e Richard Johnson) processati in seguito ai disordini scoppiati a Londra nel 2011 è, a questo riguardo, significativa. Nicholas Robinson era uno studente di ingegneria: la sera degli incidenti stava rientrando a casa in auto, aveva sete e, vedendo un supermercato con le vetrine sfondate, aveva preso una bottiglia d'acqua senza rendersi conto che i poliziotti erano dietro di lui. Richard Johnson, quando venne a sapere delle devastazioni nei grandi magazzini, si era coperto il volto per non essere riconosciuto ed era uscito per fare incetta di videogiochi. Entrambi furono arrestati e processati, con esiti molto diversi. Nonostante l'esiguità del furto e il pieno riconoscimento di aver fatto qualcosa di cui vergognarsi, Nicholas, tra l'altro incensurato, venne condannato a sei mesi di carcere. Johnson, che aveva premeditato il reato e contribuito ai disordini, rimase in libertà: gli venne chiesto di svolgere lavori socialmente utili per circa tre settimane⁴.

3. Cfr S. DANZIGER - J. LEVAV - L. AVNAIM-PESSE - D. KAHNEMAN, «Extraneous factors in judicial decisions», in *Proceedings of the National Academy of Sciences of the United States of America* 108 (2011) 688-692; K. WEINSHALL-MARGEL - J. SHAPARD, «Overlooked factors in the analysis of parole decisions», *ivi* 108 (2011) E 833; H. FRY, *Hello World...*, cit., 77.

4. Cfr J. ROJAS, «London riots: Lidl water thief jailed for six months», in *The Telegraph*, 11 agosto 2011; D. MILLS, «Paul and Richard Johnson avoid prison over riots», in *News Shopper*, 13 gennaio 2012; H. FRY, *Hello World...*, cit., 55 s.

ARTICOLI

La variabilità imprevedibile di questi esiti rivela una problematica tra le più complesse e controverse della giustizia penale. In base a quali parametri si commina una sentenza? Il tema del ragionamento giudiziario è da molto tempo studiato e analizzato con rigore⁵. Eppure, nonostante la mole di ricerche in proposito, si è ben lontani dall'ovviare alle molteplici bizzarrie proprie della soggettività umana. I risultati di alcune ricerche in proposito sono piuttosto sconcertanti. A 47 giudici dello Stato della Virginia (Usa) venne sottoposto il medesimo caso, chiedendo una possibile sentenza. Si trattava di una ragazza di 18 anni, incensurata, trovata in possesso di una quantità di droga sufficiente allo spaccio. La varietà delle risposte fu enorme: 29 giudici ritennero che dovesse essere assolta, 8 applicarono la condizionale, 4 scelsero la multa, 3 la multa e la condizionale e 3 chiesero l'incarcerazione.

30

Varietà altrettanto composita si è registrata per un altro caso ipotetico presentato a 81 giudici del Regno Unito, con un'aggiunta significativa. A molti di loro venne presentato un caso che essi avevano già giudicato in passato, cambiandone soltanto il nome e il sesso. Quasi nessuno emise la medesima sentenza. La percentuale finale era quasi quella della casualità (50%): in pratica, tirando una moneta, si sarebbe arrivati alla stessa conclusione. L'esercizio della giustizia è dunque solo questione di fortuna (o sfortuna)⁶?

L'utilizzo di un algoritmo (cioè di un metodo di calcolo) consentirebbe di avere un'idea più chiara della situazione, e di come si è agito in circostanze simili, fornendo una sorta di «Digesto elettronico». Si pensi alla decisione di concedere o meno a un detenuto una riduzione della pena o la libertà condizionale. È un dilemma mostrato drammaticamente da quanto accaduto a Londra il 29 novembre 2019: quel giorno un uomo armato di coltello ha ucciso due persone e ne ha ferite altre tre prima di essere bloccato da alcuni passanti. L'attentatore

5. Cfr, ad esempio, P. GARBOLINO, *Probabilità e logica della prova*, Milano, Giuffrè, 2014; A. COSTANZO, *L'argomentazione giuridica*, ivi, 2003; ID., *Logica giudiziaria*, Roma, Aracne, 2015.

6. Cfr W. AUSTIN - TH. A. WILLIAMS III, «A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity», in *Journal of Criminal Law and Criminology* 68 (1977/2) 306-310; M. K. DHAMI, «Bailing and jailing the fast and frugal way», in *Journal of Behavioral Decision Making* 14 (2001/2) 141-168; H. FRY, *Hello World...*, cit., 56 s.

PER UN UMANESIMO DIGITALE

era in libertà vigilata e aveva scontato meno della metà della condanna per terrorismo. Uno degli «eroi» che lo hanno disarmato era anch'egli in libertà vigilata: era stato condannato per l'omicidio di un ragazza disabile, suscitando le proteste dei familiari per il suo rilascio⁷.

Uno stesso episodio mostra in maniera emblematica le variabili del caso: in base a quali criteri concedere la libertà vigilata? L'utilizzo di algoritmi potrebbe aiutare a stabilire con maggiore precisione la possibilità di recidive? O a stabilire di tenere in carcere una persona che ha davvero deciso di cambiare vita? È evidente che la decisione ultima spetterà sempre a un giudice, ma il calcolo delle probabilità e la memoria storica di casi simili può fornire indicazioni preziose e un fondamento meno volubile alla decisione.

Tuttavia rimangono delle proprietà prettamente umane, che una macchina, anche la più sofisticata che si possa immaginare, non sembra essere in grado di riprodurre, neppure in futuro.

31

Il linguaggio

Il linguaggio umano ha una complessità enorme e misteriosa, perché non è riconducibile a una teoria, a un programma, o a una serie di informazioni accumulate. Esso fa inoltre riferimento a dimensioni «impalpabili», come gli affetti, l'immaginazione, la fantasia, i simboli, le sfumature della pronuncia, le relazioni. Nello stesso tempo, viene appreso con facilità fin dai primissimi anni di vita. Il bambino non inizia a parlare in forza di continue e ripetute istruzioni, ma perché possiede le sue strutture basilari: «La complessità della competenza linguistica di cui è provvisto un bambino di 5 anni richiederebbe un apprendimento che duri più di 100 anni senza interruzione»⁸. Cento anni di apprendimento non basterebbero, eppure in pochi anni ogni bambino è in grado di parlare.

7. «Londra, l'attentatore era in libertà vigilata. L'uomo che lo ha bloccato è un omicida», in <http://www.rainews.it/dl/rainews/articoli/Londra-attentatore-era-28enne-Usman-Khan-in-liberta-vigilata-era-stato-condannato-per-terrorismo-Attacco-anche-Aja-1ff39587-f84e-496f-a4a6-8149e5110959.html>

8. E. RIGOTTI, «La linguistica strutturale», in S. VANNI ROVIGHI (ed.), *Storia della filosofia contemporanea*, Brescia, La Scuola, 1980, 647 s; cfr N. CHOMSKY - G. MILLER, «Modelli finiti di utenti linguistici», in N. CHOMSKY, *Saggi linguistici. I. L'analisi formale del linguaggio*, Torino, Boringhieri, 1969, 288-375.

ARTICOLI

L'influsso della *corporeità* è in particolare rilevante, non solo per la complessità dell'apparato labiale e dei movimenti della lingua che consentono la varietà dei suoni e le sfumature della pronuncia. Un bambino non impara a parlare davanti a uno schermo, o mostrandogli programmi registrati, ma solo nella relazione, a contatto con una persona, osservandone le sfumature del volto, il movimento delle labbra e ascoltandone i suoni⁹. Come nota Daniel Siegel, la mente umana si modifica e si sviluppa nella relazione affettiva, generando nuove connessioni sinaptiche e imparando a cogliere ulteriori significati da queste esperienze¹⁰.

Linguaggio e salute mentale

32

Quanto il linguaggio e le sue proprietà siano indice di sanità mentale è mostrato da uno studio compiuto dall'università del Kentucky su 678 religiose (appartenenti alla Congregazione *Scuole scolastiche di Nostra Signora*) di età compresa tra 75 e 107 anni. Scopo della ricerca era di comparare le caratteristiche dello stile di vita (orario della giornata, presenza di valori etici e religiosi, alimentazione, attività condivise) con lo stato di salute mentale. I termini dell'accordo prevedevano anche che le religiose, alla loro morte, avrebbero donato il proprio cervello perché fosse analizzato per individuare possibili segni di malattie, in particolare il morbo di Alzheimer.

Ma i ricercatori poterono soprattutto contare su un altro parametro prezioso: gli scritti delle religiose. Al momento dell'ingres-

9. Cfr P. K. KUHL - F. M. TSAO - H. LIU, «Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning», in *Proceedings of the National Academy of Sciences* 100 (2003) 9096-9101; M. SPITZER, *Demenza digitale*, Milano, Corbaccio, 2016, 191 s.

10. «Le esperienze relazionali hanno un'influenza fondamentale sul cervello, perché i circuiti che mediano le percezioni sociali sono strettamente correlati a quelli responsabili dell'integrazione dei processi che controllano l'attribuzione di significati, la regolazione delle funzioni dell'organismo, la modulazione delle emozioni, l'organizzazione della memoria e le capacità di comunicazione. Le esperienze interpersonali svolgono un ruolo organizzativo cruciale nel determinare lo sviluppo delle strutture cerebrali nei primi anni di vita e nel plasmare le attività del cervello durante tutta la nostra esistenza» (D. J. SIEGEL, *La mente relazionale. Neurobiologia dell'esperienza interpersonale*, Milano, Raffaello Cortina, 2013, 32).

PER UN UMANESIMO DIGITALE

so nella Congregazione, le autorità religiose chiedevano a tutte le candidate di scrivere la loro biografia. Fu così possibile operare una comparazione tra le capacità espressive delle novizie e l'anamnesi del loro stato di salute mentale così come sarebbe emerso nei decenni successivi. I risultati mostrarono convergenze significative tra la qualità linguistica di quegli scritti (indipendentemente dal grado di istruzione ricevuto) e la possibilità di contrarre malattie mentali, costituendo un sorprendente indicatore prognostico: «Il 90% delle suore che avevano sviluppato il morbo di Alzheimer aveva dimostrato, in gioventù, una “capacità linguistica limitata”, mentre solo il 13% delle suore che avevano conservato le proprie facoltà cognitive anche in tarda età aveva scritto un saggio “a bassa densità concettuale”»¹¹.

I ricercatori si sono anche chiesti quale dei due fattori condizioni l'altro: se, in altre parole, un basso livello di introspezione e comunicazione predisponga alla malattia mentale, o se, al contrario, non sia l'effetto di un deficit in atto (in tal caso un disturbo come il morbo di Alzheimer sarebbe presente fin dalla più tenera età). Ciò che rimane degno di rilievo è la correlazione tra i due aspetti.

33

Comunicare e comprendere

Il rapporto tra linguaggio e salute mentale dice della dimensione essenzialmente biologica, corporea, vivente del linguaggio umano, la cui semantica presenta un insieme di regole estremamente complesse e non codificabili, ma nello stesso tempo note a tutti. In questa prospettiva, una macchina o un robot non potrebbero mai «parlare» come un essere umano, perché non comprendono una lingua, ma utilizzano strisce di informazioni: la possibilità di una coscienza in

11. H. FRY, *Hello World...*, cit., 92; cfr D. SNOWDON, «Healthy Aging and Dementia: Findings from the Nun Study», in *Annals of Internal Medicine* 139 (2003) 450-454. «Densità intellettuale» è un parametro utilizzato per indicare il grado di complessità linguistica, come, ad esempio, quanti concetti sono presenti in una frase, e se sono impiegati in maniera appropriata (cfr L. FRAZIER, «Syntactic complexity», in D. DOWTY - L. KARTTUNEN - A. ZWICKY [eds], *Natural Language Parsing, Psychological, Computational, and Theoretical Perspectives*, Cambridge, CUP, 1985, 129-189; O. DAHL, *The Growth and Maintenance of Linguistic Complexity*, Amsterdam - Philadelphia, Benjamins, 2004).

ARTICOLI

una macchina rimanda all'annoso e inestricabile problema del rapporto tra mente e cervello, e all'altro problema, altrettanto complesso, del rapporto tra mente e corpo nell'essere umano. Tematiche che, oltre a interpellare discipline molto differenti (filosofia, linguistica, psicologia, psicanalisi, neurologia, genetica, fisica, chimica, tutto l'ambito delle neuroscienze), mostrano al loro interno un proliferare di ipotesi e teorie variegata e tra loro contraddittorie, a conferma della complessità e difficoltà di giungere a conclusioni definitive e universalmente condivisibili. Anche il tentativo di eliminare uno dei due termini – la mente –, non ha avuto successo. A differenza del cervello, la mente presenta una molteplicità di significati difficili da precisare (coscienza, anima) e che non sono riconducibili al cervello, ma hanno a che fare con l'intero organismo¹². In riferimento all'intelligenza artificiale, ci si chiede se la macchina, oltre a un «cervello», possa avere una «mente», una coscienza, che nell'uomo ha una stretta relazione con il corpo¹³.

Per comprendere questa differenza basilare il filosofo della mente John Searle ideò un esperimento mentale divenuto celebre, chiamato «stanza cinese». Un volontario viene posto in una stanza dove si trovano delle strisce di testo in lingua cinese, che egli non conosce. Successivamente riceve un manuale di istruzioni che indica quali simboli dovrà usare per rispondere a ciascuna striscia. Alla fine le risposte risulteranno corrette, anche se il volontario continua a ignorare la lingua cinese. L'esperimento della stanza mostra la differenza tra il linguaggio umano e un programma informatico dove si ricevono e inviano strisce di informazioni senza poterle capire.

Il programma utilizza una procedura, mentre il linguaggio fa soprattutto riferimento al significato delle parole e utilizza simboli, ciò

12. Per una panoramica della questione, cfr S. MORAVIA, «Mente», in *Enciclopedia filosofica*, Milano, Bompiani, 2006, vol. 8, 7268-7286; M. KÄUSBAUER, «Mente, Filosofia della», ivi, 7286-7299; A. ANTONIETTI, «La mente tra cervello e anima», in *Rivista di Filosofia Neo-scolastica* 97 (2005/2) 211-242.

13. Siegel riprende in qualche modo la verità del detto «mens sana in corpore sano» nella sua definizione della mente: «La mente è un processo incarnato e relazionale che regola i flussi di energia e informazioni» (*La mente relazionale...*, cit., 1). Per questo essa non si limita alle attività cerebrali e non cessa di espandersi una volta che il cervello ha terminato il suo sviluppo, ma ha a che fare con la globalità della persona, con la sua coscienza, quello che gli antichi e i medievali chiamavano «anima razionale» (cfr ARISTOTELE, *De anima* II, 2, 413 b 26; *Sum. Theol.* I, q. 75).

PER UN UMANESIMO DIGITALE

che in linguistica è ricompreso sotto il termine «semantica» e che è assente nei programmi dei computer, proprio perché non è «programmabile», e soprattutto perché presenta connotazioni biologiche e affettive. Come abbiamo fatto notare, questo è uno dei grandi enigmi della linguistica, e insieme un aspetto peculiare dell'essere umano.

Searle riassume così la sua posizione al riguardo: «1) I programmi sono completamente sintattici; 2) La mente ha una semantica; 3) La sintassi non è la stessa cosa e non è di per sé sufficiente per la semantica»¹⁴. In altre parole, c'è un salto di qualità tra programma e significato. Gli esseri umani, a differenza delle macchine, non sono programmati da algoritmi. Per questo si utilizza per lo più il termine «emergenza» (nel senso di qualcosa che sopravviene, una eccedenza) per indicare una dimensione della mente umana irriducibile non solo a un algoritmo o a una macchina, ma anche a una dimensione meramente materiale¹⁵.

Questo salto di qualità è stato reso in maniera visivamente esemplare dal celebre film di Stanley Kubrick *2001, Odissea nello spazio*. In esso si presenta il dialogo tra un astronauta e il megacomputer di bordo, Hal 9000, che ha il controllo assoluto di tutte le operazioni. Quando il comandante decide di disinstallarlo, in seguito a un errore di calcolo, Hal inizia a uccidere i membri dell'equipaggio. Inutilmente l'astronauta sopravvissuto cerca di mostrargli il significato della missione e il valore della vita degli astronauti; Hal non è in grado di comprendere questa problematica, e continua a ripetere ciò per cui è stato programmato, fino a interrompere il collegamento.

Le macchine hanno una morale?

Come il linguaggio, anche l'esperienza morale presenta complessità irriducibili a qualunque tentativo di ricondurla a un sistema

14. J. SEARLE, *Il mistero della coscienza*, Milano, Raffaello Cortina, 1998, 9; cfr 8.

15. Cfr A. VENTURA, «Inconsistenza fondativa dell'ereditismo naturalistico», in *Rivista di Filosofia Neo-scolastica* 111 (2019/1) 45-49; S. GALVAN, «Godel e il modello computazionale della mente», ivi 96 (2004/1) 145-174; A. CORRADINI - N. GAY - G. LO DICO, «Emergenza: le origini di un concetto», ivi, 263-279. Sono grato al prof. Antonino Ventura, docente presso l'Università Cattolica di Brescia, per queste indicazioni.

ARTICOLI

o a una teoria esaustiva. Questo è un problema che ha assillato da lungo tempo la filosofia. La riflessione di Kant a tale riguardo è emblematica. Egli ha cercato di formulare le leggi supreme della conoscenza e dell'agire, ma alla fine si è trovato costretto a riconoscere che la caratteristica peculiare della moralità è di presentarsi come un ideale *a priori* che si impone da sé: essa non può essere appresa, e nello stesso tempo è presente in tutti gli uomini.

Nella *Fondazione della metafisica dei costumi* Kant riassume la complessità dell'esigenza etica con il termine «imperativo categorico», che regola l'esperienza e nello stesso tempo è sempre al di là di essa. E conclude la sua analisi notando che la moralità è un mistero, dal momento che è impossibile indicarne con precisione le condizioni di possibilità. È qualcosa che orienta le scelte senza mai essere pienamente realizzata; ed è anche ciò che attesta la dignità dell'uomo¹⁶.

Un enigma simile è dato dall'esperienza del male morale. Nel *Saggio sul male radicale* il filosofo di Königsberg riconosce che la sua radice non è l'ignoranza sul da farsi, perché la problematica morale non è legata all'entità delle informazioni a disposizione. Al contrario, la ragione ha spesso elementi per intervenire, eppure, stranamente, rinuncia a farlo. Perché? Kant non è in grado di dirlo, ma è certo che non si tratta di ignoranza, bensì piuttosto di cattiveria¹⁷. Per questo l'esperienza morale – in altre parole, la presenza del bene e del male – smentisce un approccio meramente razionale all'azione umana; c'è un «fondo» enigmatico, irriducibile alla norma: «Questo male è *radicale*, perché corrompe il principio di tutte le massime e nello stesso tempo, d'altra parte, in quanto tendenza naturale, non può essere *distrutto* dalle forze umane, perché ciò non potrebbe av-

16. «Se non conosciamo certamente la necessità pratica incondizionata dell'imperativo morale, ne conosciamo però la inconoscibilità, che è quanto di ragionevole si può chiedere a una filosofia che si spinge nei principi fino al limite della ragione umana» (I. KANT, *Fondazione della metafisica dei costumi*, Bari, Laterza, 1979, 108; cfr M. KUEHN, *Kant. Una biografia*, Bologna, il Mulino, 2011, 421 s).

17. «La frase: l'uomo è *cattivo*, non può [...] voler dire altra cosa che questo: l'uomo è consapevole della legge morale, ed ha tuttavia adottato per massima di allontanarsi (occasionalmente) da questa legge» (I. KANT, «Della coesistenza del principio cattivo accanto a quello buono o del male radicale nella natura umana», in *Id.*, *La religione entro i limiti della sola ragione*, Bari - Roma, Laterza, 1985, 32).

PER UN UMANESIMO DIGITALE

venire se non per mezzo di buone massime, il che è impossibile, se il principio soggettivo supremo di tutte le massime è presupposto corrotto [...]. Per noi, non c'è alcuna causa comprensibile dalla quale il male morale possa per la prima volta essere venuto in noi»¹⁸.

In maniera analoga all'apprendimento di una lingua, anche nella morale c'è una sorta di «grammatica universale», alla quale la coscienza non può accedere, ma che semplicemente scopre in sé, e che è legata, più che alla competenza, ai sentimenti. E difatti le persone anaffettive, psicopatiche, presentano dei deficit proprio in questo ambito. Le ricerche svolte in ambito neurocerebrale da Antonio Damasio hanno mostrato come la lobotomizzazione, cioè l'asportazione dei lobi frontali della neocorteccia cerebrale (sede delle emozioni), comporti gravi deficit nella valutazione, smentendo un pregiudizio piuttosto comune, che cioè una mente fredda e priva di emozioni si trovi nelle condizioni ottimali per prendere decisioni sagge. In realtà, si nota esattamente il contrario. Damasio riferisce il caso di un paziente, Elliot, che in seguito a una lobotomizzazione (effettuata per rimuovere un tumore) mantenne intatte le sue capacità intellettive, linguistiche, di comunicazione e di comprensione, ma divenne completamente privo di emozioni. Tale privazione lo aveva reso «l'essere umano intelligente più freddo e meno emotivo che si potesse immaginare, la cui ragione pratica, però, era talmente menomata da produrre, nelle vicende della vita quotidiana, un seguito di errori, una perpetua violazione di quel che voi e io riterremmo socialmente appropriato e vantaggioso dal punto di vista personale»¹⁹.

Damasio aggiunge che se le lesioni avvengono nell'età dello sviluppo, le persone risultano incapaci persino di apprendere le regole basilari dell'etica. La conclusione del medico portoghese è che l'agire morale è essenzialmente legato agli affetti e, in loro mancanza, non può essere compensato da alcun tipo di istruzione: «È quindi ancor più sorprendente e nuovo che l'assenza di emozione e sentimento sia non meno dannosa, non meno capace di compromettere la razionalità che ci rende peculiarmente umani e ci permette di decidere in

18. Ivi, 38.45-46; corsivi nel testo.

19. A. DAMASIO, *L'errore di Cartesio*, Milano, Adelphi, 1995, 17 s.

ARTICOLI

armonia con un senso di futuro personale, di convenzione sociale e di moralità. Né ciò equivale a dire che sono i sentimenti, quando esercitano un'azione, a decidere per noi, o che noi non siamo esseri razionali. Io suggerisco soltanto che certi aspetti del processo dell'emozione e del sentimento sono indispensabili per la razionalità»²⁰.

Le analisi di Kant e le ricerche di Damasio possono aiutarci a individuare un'altra significativa differenza rispetto al modo di procedere dell'intelligenza artificiale: è il caso del cosiddetto «dilemma etico».

Il dilemma etico

Il dilemma etico è legato alla riflessione sul duplice effetto²¹, divenuto celebre soprattutto ad opera della filosofa inglese Philippa Ruth Foot. Ella ricorre all'esempio di un carrello ferroviario fuori controllo, che rischia di finire su un gruppo di operai che stanno lavorando; non c'è alcuna possibilità di avvisarli, ma è possibile azionare una leva di scambio che porterà il carrello su un binario morto dove però si trova un solo operaio. È la scelta, terribile ma inevitabile, del male minore. La differenza non è solo quantitativa (un gruppo di persone contro una), ma si pone soprattutto in termini di responsabilità, perché in quest'ultimo caso la persona indirizza *volutamente* il carrello verso il binario morto, uccidendo l'operaio²². Una versione ancora più tragica del dilemma precisa che si tratta di una persona con cui l'operatore ha un legame affettivo (un amico o un parente).

Le macchine come reagiscono a un dilemma etico? Un'automobile senza autista opterebbe anch'essa per il male minore, ma non si sentirebbe in colpa per aver ucciso. Per l'essere umano non

20. Ivi, 19.

21. Il duplice effetto è una teoria etica che giustifica la liceità di un'azione che ha due diverse conseguenze: una positiva (voluta direttamente) e una negativa (indiretta).

22. Cfr PH. FOOT, «The Problem of Abortion and the Doctrine of Double Effect», in *Oxford Review* 5 (1967) 5-15; ID., *Moral Dilemmas: and Other Topics in Moral Philosophy*, Oxford, Clarendon, 2002; D. EDMONDS, *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Milano, Raffaello Cortina, 2014; J. D. GREENE ET AL., «An fMRI investigation of emotional engagement in moral judgment», in *Science* 293 (2001) 2105-2108.

PER UN UMANESIMO DIGITALE

è così: la mancanza di alternative non attenua il dispiacere e il rimorso per aver attuato una scelta terribile. Ciò è ben mostrato dalla vicenda del romanzo di William Clark Styron *La scelta di Sophie*, divenuto un celebre film ad opera del regista Alan Pakula. La protagonista, Sophie, quando viene deportata ad Auschwitz si trova di fronte a una tremenda scelta: una sadica guardia le impone di decidere quale dei due figli sarebbe finito nella camera a gas. Se si rifiuterà, moriranno entrambi. Sophie, disperata, sceglie di tenere il figlio maschio, sperando che il suo aguzzino cambi idea, cosa che purtroppo non accade. E Sophie si porterà dentro il peso dell'accaduto per tutti gli anni a venire, fino al suicidio.

Il dilemma etico non può essere risolto sulla base di un algoritmo: calcolare il massimo risultato con il minimo costo. Per gli esseri umani ciò non è sufficiente; sentirsi responsabili rimanda a qualcosa di qualitativamente differente, come il rimorso, il pentimento, la tristezza, la colpa, la redenzione: «L'etica utilitaristica (conseguenzialistica) nega l'esistenza di genuini dilemmi morali. La ragione di questa negazione è evidente: se l'agire viene valutato sulla base di un criterio di ottimizzazione, non può esistere alcun conflitto [...]. I computer digitali sono definiti macchine di Turing e forniscono risultati univoci. Già solo per questo non possono essere un modello di ragion pratica»²³. Di fronte al dilemma una macchina resterà nell'incertezza, ma non si sentirà in colpa. E, al contrario di Sophie, non giungerà a suicidarsi.

39

Per un umanesimo digitale

Il rapporto con l'innovazione e le possibili ricadute che le nuove scoperte possono comportare per la vita umana sono un tema lungamente esplorato dalla filosofia e dalla letteratura di tutti i tempi, puntualizzando come ogni innovazione, mentre dispiega nuove possibilità, ripropone interrogativi antichi. In particolare, ne sono stati menzionati alcuni. Il tema della coscienza, della responsabilità,

23. J. NIDA-RÜMELIN - N. WEIDENFELD, *Umanesimo digitale. Un'etica per l'epoca dell'Intelligenza Artificiale*, Milano, FrancoAngeli, 2019, 98; cfr J. J. SMART - B. WILLIAMS, *Utilitarismo: un confronto*, Napoli, Bibliopolis, 1985.

ARTICOLI

della comunicazione, del potere, dell'etica rimangono questioni che interpellano ogni epoca e società, e che non possono essere disattesi.

Martin Heidegger, nella sua celebre analisi del predominio della tecnica nell'era moderna, notava come il problema centrale non risiedesse nella rilevazione di un tale dominio, ma piuttosto nel fatto che l'uomo non fosse preparato a viverlo in maniera critica e consapevole, soppesando possibili vantaggi e perdite. Ma, precludendosi un confronto con la dimensione sapienziale della vita, la mentalità tecnologica rischia di avvicinarsi pericolosamente alla follia e alla perdita del senso²⁴.

Julian Nida-Rümelin e Nathalic Weidenfeld, commentando il dialogo tra l'astronauta e Hal 9000 nel film *2001 Odissea nello spazio*, notano che non è un caso che il computer abbia le fattezze di un occhio di vetro rosso e nero, i colori che l'immaginario cristiano attribuisce all'inferno: «L'inferno è un luogo nel quale l'uomo ha dato il potere di decidere sulla vita e sulla morte a dei computer programmati in modo consequenzialistico, che sono incapaci di pensare veramente»²⁵.

La mole accresciuta di dati e risorse richiede alla fine una valutazione che l'intelligenza artificiale può al massimo suggerire, ma che non può mai diventare l'istanza ultima. La posta in gioco è stata efficacemente riassunta da Naief Yehya: «Con un computer possiamo trasformare quasi tutti i problemi umani in statistiche, grafici, equazioni. La cosa davvero inquietante, però, è che così facendo creiamo l'illusione che questi problemi siano risolvibili con i computer»²⁶.

Da qui la necessità di un dialogo sempre più attento tra le innovazioni tecnologiche e le scienze umane, in vista di un «umanesimo digitale» (per riprendere il titolo del libro di Nida-Rümelin e Weidenfeld). Il dibattito sulla posta in gioco e la decisione da porre in atto devono rimanere, in ultima analisi, sempre nelle mani di un uomo che è, da sempre, *homo sapiens*.

24. Cfr M. HEIDEGGER, «La questione della tecnica», in ID., *Saggi e discorsi*, Milano, Mursia, 1985, 5-27; ID., *L'abbandono*, Genova, il melangolo, 1986, 36.

25. J. NIDA-RÜMELIN - N. WEIDENFELD, *Umanesimo digitale...*, cit., 109.

26. N. YEHYA, *Homo cyborg. Il corpo postumano tra realtà e fantascienza*, Milano, Eleuthera, 2005, 15.