

Contributi/6

Oltre l'algoritmo

Questioni etiche e trasformazioni antropologiche

Filippo Pianca  0000-0002-7958-1519

Articolo sottoposto a doppia blind peer review. Inviato il 29/11/2021. Accettato il 08/06/2022.

BEYOND THE ALGORITHM. ETHICAL ISSUES AND ANTHROPOLOGICAL TRANSFORMATIONS

By adopting a descriptive-reconstructive and linguistic-theoretical approach, the widespread link between algorithms and ethics will initially be highlighted. Recalling the meaning of the term algorithm, it will be argued that this connection with ethics is both insufficient and doubtful. While considering what else can be made 'ethical' in the IT process, the meaning of ethics within the operational-functional perspective will be analysed, re-constructing the debate on 'artificial morality'. Finally, leaving aside applicative suggestions, a radical ethical-anthropological perspective will be gained: the cultural and anthropological motivations of 'artificial morality' will be traced, considering their upshot on human self-understanding and moral status.

Introduzione

L'aggiornamento prospettico della relazione uomo-tecnologie informatiche richiede, preventivamente, che si attui una riflessione sulla terminologia impiegata comunemente e sul suo senso. Esaminare filosoficamente il concetto di algoritmo significa vagliarne, innanzitutto, l'impronta e il significato assunti nella narrativa mediatica e nel discorso accademico, valutando gli eventuali equivoci ed integrandone le limitazioni. Al contempo, significa prendere in considerazione le prospettive di 'un'eticità algoritmica' e, ampliando il discorso fino ad includere i temi dell'intelligenza artificiale e della robotica, di una 'moralità artificiale'. Oltrepassando le rappresentazioni concettuali dell'algoritmo, significa giungere infine a una lettura critica delle intenzioni e ambizioni diffuse nella letteratura, valutandone parallelamente la tenuta dal punto di vista operativo-applicativo e la rilevanza etico-antropologica.

Giornalmente, i programmi informatici svolgono un ruolo fondamentale nella produzione e nella gestione delle nostre comunicazioni e della cultura condivisa, determinano come si risponde alle nostre domande, 'decidono' cosa

è rilevante, modellano le nostre reti personali e professionali, suggeriscono con chi si dovrebbe uscire e cosa si dovrebbe guardare e profilano il nostro comportamento per determinare quale pubblicità riceveremo¹. Che incidenza hanno questi sistemi sull'uomo? In che modo trasformano e plasmano la nostra comprensione antropologica, la nostra posizione morale, il nostro modo di pensare, agire ed esistere nel mondo?

Tenendo fissi questi interrogativi come punto focale dell'indagine, quest'ultima sarà suddivisa in tre tempi. Nel primo, adottando un approccio descrittivo-ricostruttivo e allo stesso tempo linguistico-teoretico, si evidenzierà il legame sempre più comune tra 'algoritmi' ed etica, mostrando come il contesto narrativo odierno intenda attribuire una moralità a questi elementi. Richiamando il significato essenziale del termine algoritmo, e mettendo in luce l'equivoco innanzitutto epistemologico esistente tra questa nozione, intesa come concetto astratto, e quella di programma informatico, intesa come concetto applicato, si sosterrà che il tentativo di attribuire un'etica agli algoritmi è insufficiente e dubbio. Nel secondo tempo del contributo, sulla scia di questa conclusione provvisoria, si prenderanno in considerazione i seguenti quesiti: oltre all'algoritmo, inteso come procedura astratta di calcolo, cos'altro può essere reso 'etico' nel processo informatico? Cosa si intende, comunemente, con 'etica' all'interno di questo dibattito? Quali tentativi, pratici e concettuali, sono stati intrapresi finora? Mantenendo l'angolatura descrittivo-ricostruttiva, ma affiancandola ora a una lettura funzionale-operativa – non fine a se stessa, bensì tramite per giungere al punto cruciale dell'indagine – si ricostruirà il dibattito sulla 'moralità artificiale' e sulla '*machine ethics*', citandone le aspirazioni e accennando a quei criteri applicativi che sono solitamente considerati soluzione alle preoccupazioni su *come* rendere etico un sistema informatico. Nel terzo tempo dello studio, dopo aver messo in discussione sia i criteri particolari sia l'intenzione generale della 'moralità artificiale', si andrà oltre ogni suggerimento funzionale-operativo e si guadagnerà una prospettiva radicalmente etico-antropologica. Ponendo l'interrogativo '*perché* vogliamo rendere etici i sistemi informatici?', se ne rintracceranno, inizialmente, le motivazioni culturali e antropologiche, fino a considerarne gli esiti sulla nostra auto-comprensione e sul nostro status morale. Si proporrà che l'ambizione di elaborare una 'moralità artificiale' e di sviluppare 'motori etici' deriva dal fatto che l'uomo contemporaneo si sente antiquato moralmente rispetto alle macchine informatiche. Si puntualizzerà altresì che questa percezione, lungi dall'essere concettuale, dà vita già oggi ad una serie di deleghe tecnologiche fiduciose che mettono in questione la nostra autonomia e responsabilità, provocando al contempo ripercussioni pratiche non trascurabili.

¹ Cfr. C. Sandvig, K. Hamilton, K. Karahalios, C. Langbort, *When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software*, «International Journal of Communication», X, 2016, p. 4973.

1. 'Das Man informatico'

Il primo tempo di questa ricerca sarà, dunque, sviluppato a partire da un'angolazione linguistico-teoretica e al contempo descrittivo-ricostruttiva. In tal senso, all'orizzonte principale e al fine della presente indagine, ossia a quella che preliminarmente si potrebbe definire come riflessione morale sull'algoritmo (estesa ulteriormente fino ad inglobare gli ambiti generali dell'intelligenza artificiale e della robotica), si arriverà solamente in un secondo tempo. Sembra infatti ineludibile, in un contesto socio-culturale com'è quello attuale permeato dal dibattito ad ampio spettro sugli algoritmi, ricostruirne, seppur sinteticamente, senso storico e significato specialistico: solo in tal modo si potrà valutare la solidità e la completezza di alcune riflessioni 'moralì' motivate operativamente, giungendo infine ad una lettura di stampo etico-antropologico.

Al fine di avviare dettagliatamente l'indagine, si considerino i seguenti passaggi estratti da notizie recenti riportate su quotidiani e riviste internazionali: «l'intelligenza artificiale sta rapidamente diventando pervasiva nella società moderna, spesso con la capacità di elaborare il linguaggio naturale e di conversare con gli umani. Dotare questi sistemi di capacità di ragionamento etico deve essere sicuramente una priorità»²; «l'idea di dare alle macchine un codice morale risale a decenni fa sia nella ricerca accademica che nella fantascienza. Le famose tre leggi della robotica di Isaac Asimov resero popolare l'idea che le macchine potessero seguire l'etica umana, anche se i racconti che esploravano l'idea evidenziavano le contraddizioni in un ragionamento così semplicistico»³; «l'intelligenza artificiale sta rapidamente diventando più potente e diffusa e gli scienziati devono insegnare moralità ed etica a questi sistemi di apprendimento automatico»⁴.

Ciò che emerge dalla narrativa mediatica diffusa, dunque, è il tentativo di legare algoritmi e 'etica': in altre parole, data l'espansione generalizzata delle 'intelligenze artificiali'⁵, e dato il loro impiego pressoché 'autonomo' – si tornerà sul senso di questo aggettivo e di 'etica' più sotto – la priorità improrogabile sembra essere diventata l'attribuzione di una 'moralità' a queste macchine. Prima di aprire la discussione su tale intenzione, va considerato il contesto narrativo in cui la stessa si inserisce.

Il termine 'algoritmo', per quanto di vecchio conio e addirittura secolare⁶, ha assunto negli ultimi decenni, e in particolare negli ultimissimi anni, una

² <https://www.discovermagazine.com/technology/ethical-ai-matches-human-judgements-in-90-per-cent-of-moral-dilemmas> (consultato il 30 ottobre 2021).

³ <https://www.wired.com/story/program-give-ai-ethics-sometimes/> (consultato il 30 ottobre 2021).

⁴ <https://www.vice.com/en/article/v7dg8m/ethical-ai-trained-on-reddit-posts-said-genocide-is-okay-if-it-makes-people-happy> (consultato il 30 ottobre 2021).

⁵ In questa sede non verrà discusso il significato metaforico e antropomorfo dell'espressione 'intelligenza artificiale'. Si veda anche la nota 9.

⁶ Sulla nascita e sull'evoluzione diacronica del termine, su cui qui non ci si può soffermare per limiti di spazio, cfr. D. E. Knuth, *Selected Papers on Computer Science*, Cambridge 1996, p. 60, pp. 87-93 e pp. 185-203; P. Benanti, *Le macchine sapienti. Intelligenze artificiali e decisioni*

rilevanza pubblica prima sconosciuta⁷. Riprendendo e adattando un concetto heideggeriano⁸, è come se si fosse diffuso una sorta di ‘*Das Man* informatico’, un ‘si dice’ imperniato sugli algoritmi. Gli esempi paradigmatici, individuabili in diversi contesti, compreso quello accademico e specialistico, comprendono: ‘gli algoritmi hanno potere decisionale’, ‘gli algoritmi prevedono i nostri comportamenti’, ‘gli algoritmi hanno bisogno di un’etica’. In poche parole, ‘si dice’ che questi sistemi, ormai ‘autonomi’, abbiano bisogno di una moralità, di essere dotati di una capacità *intrinseca* di ragionamento e discernimento morale, al fine di essere applicati nella realtà sociale riducendo al minimo, ed auspicabilmente rimuovendo, ogni supervisione e così ogni problematica funzionale – da intendere come impatto operativo dell’uso algoritmico⁹ sulle persone.

Ma, come Heidegger ci insegna, il *Das Man*, ossia l’opinione pubblica della chiacchiera e della curiosità (quest’ultima intesa, evidentemente, non in senso intellettuale-agostiniano, bensì come sguardo distratto e approssimativo), è animato dall’equivoco. Nell’equivoco pubblico, scrive il filosofo tedesco, «tutto sembra genuinamente compreso, afferrato ed espresso, ma in realtà non lo è»¹⁰. Adeguando queste riflessioni alla narrativa attuale, l’equivoco è innanzitutto epistemologico-linguistico: nel discorso essoterico sull’informatica è riscontrabile una discrepanza tra significante e significato. Il ‘*Das Man* informatico’ dibatte di algoritmi, ma si riferisce in verità, inconsapevolmente, ai programmi informatici e ai *software*¹¹. Va, dunque, rimarcata questa differenza terminologica, sottile ma sostanziale, dunque cruciale.

Un programma informatico (concetto applicato) non coincide con un algoritmo (concetto astratto). È Donald Knuth, tra i padri della cosiddetta informatica teorica, a precisare che

umane, Bologna 2018, p. 52 (ai fini del presente intervento è rilevante che l’algoritmo nasca come metodo di calcolo).

⁷ Cfr. D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 59: Knuth, nel 1977, scriveva che «dieci anni fa la parola ‘algoritmo’ era sconosciuta alla maggior parte delle persone istruite. Infatti, era scarsamente necessaria. Ma la rapida ascesa dell’informatica, che ha lo studio degli algoritmi come suo punto focale, ha cambiato tutto ciò; la parola è ora essenziale».

⁸ Cfr. M. Heidegger, *Essere e tempo*, a cura di F. Volpi, Milano 1971, pp. 158-159.

⁹ Va chiarito che fin qui si sta ancora usando ‘algoritmo’ e concetti correlati in maniera ordinaria, prescindendo quindi dalla precisazione concettuale introdotta sotto.

¹⁰ M. Heidegger, *Essere e tempo*, cit., p. 212.

¹¹ Cfr. R. K. Hill, *What an algorithm is*, «Philosophy & Technology», XXIX, 2016, p. 36: «vediamo prove che qualsiasi procedura o processo decisionale, per quanto mal definito, può essere chiamato ‘algoritmo’ nella stampa e nel discorso pubblico. Sentiamo, nelle notizie, di ‘algoritmi’ che suggeriscono potenziali compagni per i single e algoritmi che rilevano tendenze di beneficio finanziario per i venditori, con l’implicazione che questi algoritmi possano essere giusti o sbagliati»; B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi, *The ethics of algorithms: Mapping the debate*, «Big Data & Society», 2016, p. 2: «i riferimenti agli algoritmi nel discorso pubblico non riguardano normalmente gli algoritmi come costrutti matematici, ma piuttosto come implementazioni particolari».

gli algoritmi sono concetti che esistono indipendentemente da qualsiasi linguaggio di programmazione. Per me la parola *algoritmo* denota un metodo astratto per calcolare un qualche output a partire da qualche input, mentre un *programma* è la rappresentazione [*embodiment*] di un metodo computazionale in qualche linguaggio [di programmazione]. Posso scrivere diversi programmi per lo stesso algoritmo¹².

La sovrapposizione semantica tra algoritmo e programma tipica del ‘*Das Man* informatico’ è dunque inaccurata: un programma informatico è la rappresentazione o ‘incarnazione’ concreta, ma particolare, di un algoritmo in un linguaggio di programmazione definito precisamente, mentre un algoritmo è una procedura computazionale (ossia di calcolo) astratta volta a trasformare l’informazione¹³. In altre parole, un algoritmo è «una sequenza di regole definita con precisione che dice come produrre in un numero finito di passaggi informazioni di output specificate a partire da informazioni di input fornite. Una rappresentazione particolare di un algoritmo è chiamata programma»¹⁴. In tal senso, la ricetta per cucinare una torta, un teorema e una dimostrazione matematica (ad esempio, il metodo che si usa per risolvere un lungo problema di divisione) o il processo per fare il bucato, sono algoritmi (cioè procedure univoche volte alla risoluzione di problemi, non per forza di calcolo¹⁵), ma non sono programmi informatici, mentre questi a loro volta non sono propriamente algoritmi (in termini logici: non c’è identità tra i due concetti), bensì codici scritti in un linguaggio di programmazione che codifica, ossia traduce, algoritmi di carattere logico-matematico in modo da farli eseguire dai computer più velocemente (e, solitamente, in maniera più attendibile) degli esseri umani.

¹² D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 1. Cfr. T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms*, Cambridge 2009; F. Kraemer, K. van Overveld, M. Peterson, *Is there an ethics of algorithms?*, «Ethics and Information Technology», XIII, 2011, p. 251: «un algoritmo è, grosso modo, una sequenza finita di istruzioni ben definite che descrivono in modo sufficientemente dettagliato come risolvere un problema. Sia i computer che gli esseri umani utilizzano algoritmi per risolvere un’ampia gamma di problemi».

¹³ Cfr. R. K. Hill, *What an algorithm is*, cit., p. 47: un algoritmo è un costrutto matematico con «una struttura di controllo finita, astratta, efficace, composta, data in modo imperativo, che realizza un dato scopo sotto determinate disposizioni».

¹⁴ D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 2 e p. 5. Cfr. anche p. 59: «la lingua inglese ha diverse altre parole che si avvicinano quasi, ma non del tutto, al concetto necessario: procedura, ricetta, processo, routine, metodo, trafila. Come queste cose un algoritmo è un insieme di regole o direzioni per ottenere uno specifico output da uno specifico input. La caratteristica distintiva di un algoritmo è che tutta la vaghezza dev’essere eliminata; le regole devono descrivere operazioni che sono così semplici e così ben definite che possono essere eseguite da una macchina. Inoltre, un algoritmo deve sempre terminare dopo un numero finito di passaggi».

¹⁵ Un esempio a favore di questa tesi si trova in J. Savulescu, M. Vergano, L. Craxì, D. Wilkinson, *An ethical algorithm for rationing life-sustaining treatment during the COVID-19 pandemic*, «British Journal of Anaesthesia», III, 2020, pp. 253-258. L’algoritmo presentato in questo saggio è inteso come procedura decisionale astratta, a prescindere da ogni traduzione informatica. Ad ogni modo, non è chiaro se sia l’algoritmo, come schema astratto, ad essere etico di per sé, o se piuttosto, essendo l’algoritmo basato su principi derivati da qualche teoria etica specifica, non sia la pratica umana a rivelarsi morale o immorale nell’attuazione della procedura proposta. In altre parole: uno schema decisionale-operativo può essere di per se stesso etico, a prescindere dalla sua applicazione pratica?

Come ribadisce Knuth, «un programma informatico rappresenta un algoritmo, sebbene l'algoritmo stesso sia un *concetto mentale* che esiste indipendentemente da ogni rappresentazione»¹⁶. In altre parole, ogni traduzione rappresentativa non è mai esclusiva né assoluta, pur essendo univoca, ossia inequivocabile.

Se il '*Das Man* informatico' è preoccupato prevalentemente per la portata e i risvolti pratici dei cosiddetti 'algoritmi decisionali' (ad esempio, i sistemi di supporto alle decisioni cliniche che consigliano diagnosi e trattamenti ai medici¹⁷), la distinzione teoretico-linguistica appena svolta ci permette di comprendere che sono i *software*, ossia insiemi di programmi informatici e *file* dati implementati in componenti *hardware*, a suggerire le 'decisioni' (a fornire, cioè, degli output numerici dopo un processo di calcolo più o meno intenso) ad utenti e operatori, o a compiere operazioni in modo automatico: sono questi, non gli algoritmi in sé, ad essere venduti, distribuiti e applicati in diversi campi della realtà sociale.

Ciò che *di fatto* determina dei problemi, in termini operativi e applicativi, sono le rappresentazioni delle procedure computazionali astratte, dunque le traduzioni e implementazioni particolari dei metodi di calcolo. Di conseguenza, parlare di algoritmi pensando, in realtà, ai programmi informatici è equivoco: il primo concetto rimane nel campo dell'astrazione. Da questa considerazione si evince che l'intenzione di attribuire una moralità agli algoritmi, oltre ad essere insufficiente, è anche dubbia.

2. Etica operativa e moralità artificiale

La precisazione terminologica appena svolta, lungi dall'essere salottiera, introduce al secondo tempo del contributo. Mantenendo la prospettiva descrittivo-ricostruttiva fin qui adottata, ma affiancandola ora ad un'esplicita angolatura funzionale-operativa (l'assunzione di quest'ultima non è fine a se stessa, al contrario è indispensabile per giungere alla successiva lettura di stampo etico-antropologico, fulcro della presente ricerca), si tratta di evidenziare, per prima cosa, che il tentativo di rendere 'etico' un algoritmo è del tutto insufficiente.

Sulla scorta delle precisazioni nozionistiche appena introdotte, si può rilevare come l'intenzione di rendere 'etici' gli algoritmi presente nella narrativa

¹⁶ D. E. Knuth, *Selected Papers on Computer Science*, cit., p. 59 (corsivo mio). Cfr. C. Sandvig et al., *When the Algorithm Itself Is a Racist*, cit., p. 4976: «per un informatico, gli algoritmi esistono indipendentemente da qualsiasi tipo di computer, disco rigido o altro substrato fisico su cui possono essere implementati. [...] Si noti che, per Knuth, lo stesso algoritmo può essere implementato diversamente in due diversi linguaggi di programmazione; pertanto, un algoritmo non è equivalente a un programma. [...] [Un algoritmo è] un insieme astratto di regole o una strategia».

¹⁷ Cfr. B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 3. Questi sistemi si fondano sulle tecniche di *machine learning*: sono proprio queste a destare le maggiori preoccupazioni, in quanto, apparentemente, i modelli da loro elaborati risultano opachi, poco comprensibili o addirittura imperscrutabili. Per quanto di rilievo attuale, questo dibattito non può essere approfondito in queste righe: quando qui si parlerà di programmi, *software*, 'intelligenza artificiale' e robot, si farà riferimento implicito anche, ma non solo, alle tecniche di *machine learning*. Sul tema etica e *machine learning* cfr. M. Kearns, A. Roth, *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*, Oxford 2020, pp. 7-11.

mediatica e in parte della letteratura accademica sia limitata, lacunosa: l'equivoco epistemologico-linguistico apre la strada ad un equivoco 'etico', ossia pratico-applicativo. Mettendo, per il momento, tra parentesi la questione delle condizioni *necessarie* (*perché* bisogna rendere 'etici' i programmi, i *software*?), si considerino quelle sufficienti. Se l'algoritmo è un processo di calcolo astratto, una serie di istruzioni e regole matematiche univoche con cui si può risolvere un problema, rendere 'etico', dunque utilizzabile e al contempo sicuro – è questa l'accezione utilitaristica più diffusa e approvata del termine – questo metodo pare essere del tutto insufficiente: va considerato il fatto che i meccanismi e i criteri di calcolo di volta in volta stabiliti non sono moralmente neutri (anche se matematicamente inequivocabili)¹⁸, e al contempo che l'elaborazione algoritmica deve attenersi a dei presupposti di accessibilità, comprensibilità e affidabilità¹⁹. A questa osservazione, mossa ancora in maniera descrittiva, si può aggiungere che l'aspetto 'etico' riguarda anche il processo di codifica in un preciso linguaggio di programmazione: in altre parole, vanno pensate parimenti le diverse modalità con le quali un algoritmo può essere rappresentato in un programma, ossia essere tradotto in uno specifico codice informatico attraverso un peculiare linguaggio di programmazione²⁰. Infine, all'interno della prospettiva 'etica' approvata comunemente, anche le particolari implementazioni concrete devono rivestire una certa importanza: perché, in quel contesto, è stato impiegato quel *software* e non un altro? In che modo è stato applicato²¹? Ci si può avvalere dello stesso pacchetto di programmi, senza modificarlo radicalmente, ad un livello pratico diverso? Ognuno di questi passaggi nello sviluppo, dall'ideazione dell'algoritmo alla sua codifica in un linguaggio di programmazione fino all'applicazione, va rivestito di attenzione 'etica'. In tal senso, rimanendo sempre all'interno dell'approccio funzionale e descrittivo, l'aspetto 'etico' è in realtà triplice: non riguarda *solo* l'elaborazione astratta, ma anche l'effettivo processo di traduzione e adozione sociale. Nel '*Das Man* informatico' questa constatazione pare frequentemente trascurata.

¹⁸ Riprendendo l'analogia della ricetta, la riflessione 'etica' sull'algoritmo inteso come metodo di calcolo astratto ha a che fare innanzitutto con la scelta degli ingredienti e della metodologia: in altre parole, la preferenza per un certo meccanismo di calcolo e manipolazione dei dati e per determinati criteri e parametri non è mai puramente matematica o formale, quindi 'neutra', ma è sempre motivata da giudizi e atteggiamenti soggettivi passibili di una valutazione di stampo 'etico'.

¹⁹ Cfr. B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 6.

²⁰ Ivi, p. 7: «il design e la funzionalità di un algoritmo riflettono i valori del suo progettista e gli usi previsti, se non altro nella misura in cui un particolare design è preferito come l'opzione migliore o più efficiente. Lo sviluppo non è un percorso lineare e neutro; non esiste una scelta oggettivamente corretta in un dato stadio di sviluppo, ma molte scelte possibili. Di conseguenza, 'i valori dell'autore [di un algoritmo], consapevolmente o meno, sono congelati nel codice, istituzionalizzando di fatto quei valori'».

²¹ Cfr. C. Sandvig et al., *When the Algorithm Itself Is a Racist*, cit., p. 4985: «scegliere un algoritmo rispetto a un altro per affrontare lo stesso problema può essere una decisione significativa. Così potrebbe essere il modo in cui viene implementato un algoritmo. Naturalmente, le circostanze vincolano la scelta dell'algoritmo, o anche il suo design». Qui, chiaramente, 'algoritmo' va inteso come programma, o insieme di programmi (*software*).

Da queste brevi osservazioni si evince che il termine ‘etica’, all’interno della prospettiva particolare adottata, non rimanda ad una teoria dei *costumi* correlata al bene o ad una valutazione delle trasformazioni esperienziali e valoriali umane, tanto meno ad una teoria critica del dover essere di fronte all’essere della realtà. L’approccio funzionale-operativo – qui definito in questo modo per evidenziare come la questione degli algoritmi non sia mai solo astratta e concettuale – ha piuttosto una visione dell’etica intrinsecamente legata all’utilizzo, al funzionamento, all’attività e alla diffusione dei sistemi informatici. La preoccupazione di questa impostazione, quindi, non è di indagare il *significato* dei dispositivi e dei programmi informatici, cioè le loro implicazioni per la nostra comprensione e percezione antropologico-morale, bensì che questi sistemi funzionino al meglio, in modo sicuro e per il maggior numero possibile di utenti. In questa premessa si delinea la cornice utilitaristica che regola la gran parte del dibattito sul tema²²: uno sfondo certamente ‘etico’, ma ancora preliminare e intempestivo rispetto all’analisi di quell’*imago hominis* plasmata dall’intelligenza artificiale.

Al fine di illustrare e guadagnare in modo più dettagliato il significato comunemente diffuso del termine ‘etica’ in riferimento agli algoritmi, ossia ai programmi informatici, si consideri la seguente affermazione di Mittelstadt e colleghi: «gli algoritmi non etici [*unethical*] possono essere pensati come artefatti *software* malfunzionanti che non operano come previsto»²³. Capovolgendo il negativo, un algoritmo è allora ‘etico’ se è ben funzionante e se opera in conformità a quanto previsto dalle specifiche condizioni del suo utilizzo. Per l’approccio funzionale-operativo la dimensione ‘etica’ dei sistemi informatici risiede, dunque, nelle loro prestazioni e attività e nel rispetto delle regole procedurali. Ciò che interessa a questa impostazione, laddove sottoscritta, è unicamente *come* rendere ‘etico’, cioè utilizzabile, funzionale, efficiente e persino autosufficiente un programma informatico (quindi, per estensione, un’intelligenza artificiale, ed eventualmente un robot). Questa intenzione è tra le finalità primarie della cosiddetta ‘*machine ethics*’: dotare i sistemi e i dispositivi informatici di una ‘moralità artificiale’²⁴.

Come ricordano Tsamados e colleghi, «il lavoro sull’etica degli algoritmi è aumentato in modo significativo dal 2016, quando i governi nazionali, le organizzazioni non governative e le aziende private hanno iniziato ad assumere un ruolo di primo piano nel dibattito sull’intelligenza artificiale e sugli algoritmi ‘giusti’ ed ‘etici’»²⁵. Questo interesse così recente si è inserito a pieno titolo nel più generale ambito di ricerca denominato ‘*machine ethics*’, apparso all’inizio

²² Ivi, p. 4983: «l’etica consequenzialista sembra il modo più *produttivo* di ragionare sugli algoritmi». Corsivo mio.

²³ Cfr. B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 13.

²⁴ Cfr. P. Danielson, *Artificial Morality: Virtuous Robots for Virtual Games*, New York 1992.

²⁵ A. Tsamados, N. Aggarwal, J. Cowls, J. Morley, H. Roberts, M. Taddeo, L. Floridi, *The ethics of algorithms: key problems and solutions*, «AI & Society», 2021, p. 2. Cfr. l’interessante articolo di Ochigame al link <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>, dove si dice che «la maggior parte del lavoro ben finanziato sulla ‘IA etica’ è in linea con

del nuovo millennio. Come sottolinea la *Stanford Encyclopedia of Philosophy*, «la *machine ethics* è l'etica per le macchine, per 'macchine etiche', per le macchine come *soggetti*, piuttosto che per l'uso umano delle macchine come *oggetti*»²⁶. In questo quadro, ciò che è in gioco è l'elaborazione di una 'moralità artificiale' (detta anche 'computazionale'²⁷): l'intento è quello di progettare sistemi informatici e robot dotati della capacità di ragionare, decidere e agire moralmente, e quindi di formulare ed esprimere giudizi etici, così da poter agire *come se*²⁸ fossero agenti morali. Allen e Wallach, tra i primi e principali fautori di questo progetto, affermano che

la moralità artificiale sposta parte dell'onere del comportamento etico lontano da progettisti e utenti e sui sistemi informatici stessi. Il compito di sviluppare agenti morali artificiali diventa particolarmente importante poiché i computer vengono progettati per funzionare con una sempre maggiore autonomia, cioè con una supervisione umana sempre meno diretta. La velocità con cui i computer eseguono le attività impedisce sempre più agli esseri umani di valutare se ogni azione venga eseguita in modo responsabile o etico. L'implementazione di agenti *software* con capacità decisionali morali offre la promessa di sistemi informatici in grado di valutare se ogni azione eseguita sia eticamente appropriata²⁹.

Mantenendo ancora l'angolatura descrittivo-ricostruttiva, la 'moralità artificiale', come tentativo di introdurre facoltà decisionali e comportamentali 'moralì' nei sistemi *software* e *hardware*, è strettamente legata alla sicurezza e all'affidabilità degli stessi:

gli ingegneri si sono sempre preoccupati di progettare strumenti sicuri e affidabili. La sensibilità alle implicazioni morali di due o più linee di azione in contesti limitati

l'agenda della lobby tecnologica: adeguare volontariamente o moderatamente, anziché limitare legalmente, l'impiego di tecnologie controverse» (consultato il 8 novembre 2021).

²⁶ <https://plato.stanford.edu/entries/ethics-ai/#MachEthi> (consultato l'8 novembre 2021). Il termine *machine ethics* è lasciato in inglese sia perché nella letteratura non è generalmente tradotto, sia perché la traduzione ('etica della macchina') è ambigua, essendo il genitivo potenzialmente sia soggettivo sia oggettivo. Cfr. W. Wallach, P. M. Asaro (eds), *Machine Ethics and Robot Ethics*, London 2017.

²⁷ In molti testi, la prospettiva della moralità computazionale è affiancata direttamente all'etica utilitaristica. In altre parole, diversi autori sostengono che solo questa teoria morale può essere computabile (come, d'altronde, affermava già Bentham) e quindi trasferibile in un codice informatico. Cfr. M. Anderson, S. L. Anderson, *Machine Ethics: Creating an Ethical Intelligent Agent*, «AI Magazine», XXVIII, 2007, p. 18.

²⁸ A questo proposito Hans Vaihinger, padre del finzionalismo, parlerebbe di 'finzione utile': l'idea di una moralità delle macchine informatiche non è un descrittore della realtà, ma fingiamo comunque di crederci per una più o meno favorevole utilità pratica. Cfr. H. Vaihinger, *La filosofia del come se*, Roma 1967.

²⁹ C. Allen, I. Smit, W. Wallach, *Artificial morality: Top-down, bottom-up, and hybrid approaches*, «Ethics and Information Technology», VII, 2005, p. 149. Cfr. W. Wallach, *Implementing moral decision making faculties in computers and robots*, «AI & Society», XXII, 2008, p. 464: «con l'avvento di agenti e sistemi sempre più autonomi, computer e robot con capacità decisionali morali diventeranno una necessità tecnologica».

può essere intesa come un'estensione della preoccupazione dell'ingegnere di progettare meccanismi di controllo appropriati per la sicurezza all'interno di computer e robot³⁰.

Con riferimento ad una corrente specifica della robotica³¹, punto di sintesi e d'arrivo di diversi campi di studio, compreso quello sugli algoritmi e l'intelligenza artificiale, l'obiettivo è dotare i robot di un'autonomia operativa e di giudizio e di un'indipendenza dall'umano sempre maggiori, così da renderli 'agenti artificiali responsabili'³². L'aspirazione dichiarata esplicitamente è notevole: far sì che i robot e in generale i sistemi informatici ragionino e si comportino meglio degli uomini dal punto di vista 'etico'³³, fino a renderli gradualmente dei 'santi morali', capaci perfino di aiutarci a migliorare eticamente³⁴. Ancora Wallach:

L'assenza di emozioni o obiettivi negli AMA [agenti morali artificiali], diversi da quelli che scegliamo di programmare nel sistema, sottolinea la differenza intrinseca tra gli esseri umani e i sistemi informatici di decisione morale. I poteri logici di analisi, liberi da emozioni, desideri o pregiudizi suggeriscono che i computer hanno la capacità di essere *perfetti ragionatori morali*³⁵.

Nondimeno, prima ancora dell'ipotesi di realizzare 'macchine etiche', la questione implicita in quest'obiettivo è se sia possibile definire l'etica in modo procedurale e computazionale: il ragionamento sottinteso, una sorta di pregiudizio, è che se avessimo una definizione perfetta di quello che intendiamo per discorso e comportamento etico, ossia, se potessimo avere un algoritmo inequivocabile dell'etica, di ciò che è bene e giusto, allora le macchine

³⁰ Ivi, p. 465.

³¹ Va precisato che quanto è stato ricostruito e quanto segue non è l'obiettivo di tutti coloro che si occupano di robotica né di tutti quelli che si occupano di 'machine ethics', o in generale del problema dell'eticità nei sistemi artificiali. La visione discussa è una delle impostazioni, non l'unica diffusa nel settore. Viene qui analizzata per il suo carattere antropologicamente provocatorio e radicale. Devo questa specificazione ai revisori anonimi del testo.

³² Cfr. F. Alaieri, A. Vellino, *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, in A. Agah, J. J. Cabibihan, A. Howard, M. Salichs, H. He (eds), *Social Robotics. ICSR 2016. Lecture Notes in Computer Science*, Cham 2016, pp. 159-168.

³³ Cfr. M. Anderson, S. L. Anderson, *Machine Ethics*, cit., p. 17. Gli autori ricordano che «[Eric] Dietrich sosteneva che le macchine che modelliamo affinché abbiano le buone qualità degli esseri umani, e che seguono anche principi derivati da esperti di etica che sono l'eccezione alla regola generale degli esseri umani non etici, potrebbero essere viste come 'umani 2.0' – una versione migliore dell'essere umano»; N. Bostrom, E. Yudkowsky, *The ethics of artificial intelligence*, in K. Frankish, W. Ramsey (eds), *The Cambridge Handbook of Artificial Intelligence*, Cambridge 2014, p. 332; D. Leben, *Ethics for Robots. How to Design a Moral Algorithm*, New York 2019, pp. 3-4 e pp. 146-149.

³⁴ L'espressione 'santi morali' riferita ai robot si trova in J. Gips, *Towards the ethical robot*, in M. Anderson, S. L. Anderson (eds), *Machine ethics*, Cambridge 2011, pp. 244-253. In questo saggio (come poi anche in M. Anderson, S. L. Anderson, *Robot be good*, «Scientific American», IV, 2010, pp. 72-77) viene raffigurato un robot con il nimbo per introdurre la prospettiva della 'santità robotica'.

³⁵ W. Wallach, *Implementing moral decision making faculties in computers and robots*, cit., p. 468. Corsivo mio. È interessante notare che, lungi dall'essere innovativa, questa idea di una perfetta moralità logica, libera da emozioni e desideri, è rintracciabile già in diversi autori e filosofi moderni, tra cui Spinoza e Schopenhauer.

informatiche riuscirebbero ad eseguirlo meglio di noi, potrebbero attenervisi in modo migliore dell'uomo, fino a diventare moralmente perfette. Concretamente, dal punto di vista dello sviluppo informatico, questo proposito di costruire dei 'motori etici'³⁶ impeccabili o degli 'agenti etici espliciti'³⁷, capaci di conformarsi a determinati principi morali nelle loro decisioni e azioni, si è focalizzata su una serie di elementi di carattere procedurale.

Ampliando quanto detto sopra in merito alle condizioni sufficienti per considerare 'etico' un algoritmo – quindi un programma informatico o un *software* – tra i criteri idonei vengono generalmente annoverate la trasparenza, la verificabilità, l'incorruttibilità, la prevedibilità, la responsabilità (o imputabilità) e l'equità³⁸. Fattori, questi, di *per sé* morali, il cui significato è stato però piegato, anche in modo formale, in vista di un'operatività e utilità sempre maggiori. Se, difatti, l'incertezza e l'opacità dei sistemi informatici ne rallentano la distribuzione economica e l'accettabilità sociale, la prospettiva funzionale cerca innanzitutto di renderli trasparenti, ossia comprensibili e accessibili a tutte le parti interessate, garantendo al contempo la loro robustezza contro la manipolazione. In questo modo, secondo l'approccio operativo, sarebbe assicurata anche la possibilità della verifica: da quali regole e parametri è guidato il programma implementato? A questo si aggiungono le preoccupazioni per i pregiudizi, le discriminazioni e la *privacy*: secondo l'ottica assunta, lo sviluppo di un'intelligenza artificiale 'etica' dovrebbe tenere conto ampiamente anche di queste problematiche legate all'equità sociale. Per quanto riguarda, inoltre, i sistemi di *machine learning*³⁹, i cui output sembrano spesso enigmatici, la proposta crescente è di attribuire loro una qualche responsabilità morale, rendendoli 'agenti biasimevoli' (*blameworthy*)⁴⁰ e colmando così quel nuovo divario di responsabilità introdotto da queste stesse tecniche⁴¹. Tutte queste caratteristiche (trasparenza, robustezza, verificabilità, equità, prevedibilità e imputabilità), guadagnate da un punto di vista funzionale-

³⁶ Cfr. D. Leben, *Ethics for Robots*, cit., pp. 76-96.

³⁷ Cfr. M. Anderson, S. L. Anderson, *Machine Ethics*, cit., 2007, pp. 15-17: «la maggior parte di coloro che lavorano sulla *machine ethics* direbbero che l'obiettivo finale è creare una macchina che sia un agente etico esplicito [...] in grado di spiegare perché una determinata azione è giusta o sbagliata facendo appello a un principio etico».

³⁸ Cfr. N. Bostrom, E. Yudkowsky, *The ethics of artificial intelligence*, cit., pp. 316-318.

³⁹ È interessante notare che le tecniche di *machine learning* sono impiegate anche «nel tentativo di codificare il ragionamento etico a partire da esempi di particolari dilemmi etici» (M. Anderson, S. L. Anderson, *Machine Ethics*, cit., 2007, p. 20). Questo sforzo, che vorrebbe rendere il ragionamento morale statistico e probabilistico, si inserisce nel ramo dell'etica applicata definita casistica o casuistica (*casuistry*).

⁴⁰ Cfr. L. Floridi, J. W. Sanders, *On the Morality of Artificial Agents*, «Minds and Machines», XIV, 2004, pp. 349-379.

⁴¹ Su tutte le considerazioni citate in questo paragrafo cfr. B. D. Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, cit., pp. 6-12; A. Tsamados et al., *The ethics of algorithms*, cit., pp. 1-8; K. Martin, *Designing Ethical Algorithms*, «MIS Quarterly Executive», XVIII, 2019, pp. 137-138. Come ricordato sopra, il dibattito sulla responsabilità è molto intenso anche nel contesto della robotica: cfr. F. Alaïeri, A. Vellino, *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, cit., pp. 164-165 e 167.

operativo, dunque procedurale e applicativo, concorrono apparentemente a soddisfare il quesito: *come* rendere ‘etico’ un sistema informatico?

3. Antiquatezza morale: questioni antropologiche

Nonostante gli ambiziosi propositi, l’approccio funzionale-operativo rimane intrinsecamente limitato: sia le modalità particolari con cui il ‘*Das Man* informatico’ cerca di attribuire una sorta di ‘eticità’ ai *software*, sia la stessa intenzione generale di sviluppare una ‘moralità artificiale’ rivelano delle significative difficoltà al contempo teoretiche e applicative. Per quanto concerne i criteri procedurali, se la trasparenza «è spesso trattata ingenuamente come una panacea per le questioni etiche derivanti dalle nuove tecnologie»⁴², nel senso che si tende a ricondurre qualsiasi problematica operativa a questo elemento – che, per inciso, non è «un principio etico in sé, ma una condizione pro-etica per favorire o compromettere altre pratiche o principi etici»⁴³ – l’ascrizione di responsabilità alle macchine calcolatrici è «un errore categoriale»⁴⁴: se chi elabora, sviluppa ed implementa i programmi informatici (e con essi i robot) «non può evitare di esprimere giudizi etici su ciò che è buono e ciò che è cattivo, allora è ragionevole sostenere che i progettisti [e gli utenti] di *software* sono moralmente responsabili degli algoritmi che progettano [e utilizzano]»⁴⁵.

Astraendo dai singoli criteri operativi, all’ambizione generale di elaborare una ‘moralità artificiale’ si può contrapporre la domanda: le macchine possono realmente ragionare, decidere e comportarsi eticamente? Ogni agente (umano) può, in linea di massima, essere definito morale dato che è in grado di agire intenzionalmente e con libero arbitrio, gerarchizzando i valori resisi manifesti in una situazione, e giacché è dotato di sensibilità ed emotività⁴⁶, aspetti fondamentali nella relazione interpersonale, luogo autentico della prassi etica.

⁴² B. D. Mittelstadt et al., *The ethics of algorithms*, cit., p. 6.

⁴³ M. Turilli, L. Floridi, *The ethics of information transparency*, «Ethics and Information Technology», XI, 2009, p. 105.

⁴⁴ G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma 2020, p. 63.

⁴⁵ F. Kraemer et al., *Is there an ethics of algorithms?*, «Ethics and Information Technology», 13, 2011, pp. 251-260, p. 251. Cfr. R. Sparrow, *Why machines cannot be moral*, «AI & Society», XXXVI, 2021, p. 692: «sono i progettisti, o eventualmente gli utenti, dell’IA che affronteranno i dilemmi etici e che saranno responsabili delle conseguenze del suo utilizzo». Va, comunque, precisato che la questione non è così nettamente risolvibile: considerando il caso specifico del *machine learning* e dei ‘sistemi artificiali dotati di motivazioni autonome’, sviluppati a partire da tecniche di *reinforcement learning*, il programmatore potrebbe non risultare *direttamente* responsabile degli sviluppi della macchina, dando luogo così ad una prospettiva in cui la responsabilità è difficilmente imputabile. Cfr. V. G. Santucci, G. Baldassarre, M. Mirolli, *Grail: a Goal-discovering Robotic Architecture for Intrinsically-motivated Learning*, «IEEE Transactions on Cognitive and Developmental Systems», VIII, 2016, pp. 214-231.

⁴⁶ Cfr. M. Anderson, S. L. Anderson, *Machine Ethics*, cit., 2007, p. 19. La distinzione messa in campo dagli autori «tra il compiere l’azione moralmente corretta in una data situazione, compreso il poterla giustificare facendo appello a un principio etico accettabile, e l’essere ritenuti moralmente responsabili dell’azione» non regge: la giustificazione che si appella a determinati

Da questo punto di vista, se le macchine informatiche non possono dimostrare alcuna comprensione delle questioni e dei valori di volta in volta in gioco, mancando di corpi vissuti, coscienza, intenzionalità e capacità di esprimere giudizi di valore soggettivi, ossia validi per loro stesse in quanto soggetti – qualsiasi ‘decisione’ prendano non è una decisione per loro – il ragionamento e l’agire etico rimangono di dominio umano: per questo motivo «le macchine potrebbero al massimo essere progettate per fornire un *simulacro superficiale* dell’etica, che dimostrerebbe un’utilità limitata nell’affrontare i dilemmi etici e politici associati all’IA»⁴⁷. Data questa limitatezza, si può arrivare ad affermare che ci sono ‘decisioni’ che i programmi informatici non dovrebbero mai prendere: facendo riferimento al caso estremo della guerra automatizzata, rivelatore della generale criticità e portata trasformativa delle ‘scelte algoritmiche’, «la decisione finale di uccidere un essere umano dovrebbe essere presa solo da un altro essere umano a causa dell’azione morale e della responsabilità implicate; il peso di una tale decisione dovrebbe spettare solo ad un’entità che può veramente comprendere, in modo umano, le conseguenze in questione»⁴⁸.

Analogamente, se il ‘*Das Man* informatico’ è preoccupato di sviluppare algoritmi buoni, ossia programmi ‘etici’, a partire da criteri operativi, va posto un ulteriore quesito: «un buon algoritmo è davvero un algoritmo eticamente buono? I due aggettivi hanno significati diversi: un buon algoritmo è un sistema funzionante, un processo esatto e corretto; mentre definire un algoritmo ‘eticamente buono’ significa descriverlo in termini etici»⁴⁹. Instaurando un’analogia tra un programma informatico e un cacciavite, una stampante o un fucile – va ricordato che nelle analogie non c’è mai assoluta identità, bensì una più o meno stretta somiglianza – si potrebbe concepire l’idea di un cacciavite, di una stampante o di un fucile ‘etici’, ossia dotati in sé di un carattere morale⁵⁰? Non si tratta sempre, in fondo, di artefatti umani? In che modo i programmi informatici, e con essi i robot, differiscono dagli strumenti citati? Il fatto di operare spesso automaticamente (senza supervisione) o di essere ‘opachi’, poco trasparenti, non è sufficiente a postularne una differenza insormontabile: sia perché la loro origine, intesa come provenienza produttiva, è invariabilmente umana⁵¹, sia perché i *software* e i robot rimangono supporti con cui l’uomo

principi etici è già una prova di responsabilità (giustificare è motivare, ossia rendere conto: ma questo è il senso intimo della responsabilità morale).

⁴⁷ R. Sparrow, *Why machines cannot be moral*, cit., pp. 685-691 (corsivo mio).

⁴⁸ M. Kearns, A. Roth, *The Ethical Algorithm*, cit., p. 178.

⁴⁹ A. Pessina, *Is a Good Algorithm also an Ethically Good Algorithm?*, in V. Paglia, R. Pegoraro (eds), *The ‘Good’ Algorithm? Artificial Intelligence Ethics, Law, Health. Proceedings of the XXVI General Assembly of Members*, Selci Lama 2021, p. 29.

⁵⁰ Cfr. M. Kearns, A. Roth, *The Ethical Algorithm*, cit., p. 7.

⁵¹ Cfr. A. Pessina, *Onlife. Trasformazioni dell’Esperienza tra Immanenza e Trascendenza*, «Hermeneutica», 2020, pp. 82-83: le ‘intelligenze artificiali’ «manipolano simboli di cui ignorano il significato, peraltro attraverso regole di cui pure ignorano il senso, per produrre dati, informazioni, immagini e suoni che, invece, hanno significato per gli uomini. In realtà, l’intelligenza è solo quella umana, che è all’inizio – creazione degli algoritmi – e al termine del processo di calcolo, quando legge e interpreta i risultati».

interviene sul mondo e nel mondo, e per questo ancora e sempre dei mezzi, di conoscenza e d'azione.

Questo, tuttavia, non significa che i programmi informatici e la loro implementazione metallica possano essere considerati, ingenuamente, come *puri* mezzi di cui disporre bene o male. Attualizzando quanto segnalava Günther Anders quasi settant'anni fa in riferimento a radio e televisione, le produzioni dell'intelligenza artificiale e della robotica «sono realtà che ci plasmano. E questo dato di fatto, che ci plasmano qualunque sia lo scopo al quale le impieghiamo, non viene eliminato soltanto perché le depotenziamo verbalmente a 'mezzi'»⁵². In questo senso, l'etica e la riflessione morale non riguardano gli strumenti e i dispositivi presi in sé, dal punto di vista funzionale e indipendentemente dall'uomo e dalle ricadute sulla sua esperienza ed esistenza, ma proprio questi stessi effetti e trasformazioni⁵³. In altre parole, l'etica non è riconducibile né alla razionalità tecno-scientifica né ai suoi requisiti funzionali e criteri operativi⁵⁴: piuttosto, *la riflessione morale è un'indagine essenzialmente antropologica*, dal momento che non esiste un domandare etico senza l'uomo – per quanto possano esserci uomini senza domandare etico. Come rammenta Benanti, «parlare di tecnica non è mai solo un discorso sulla tecnica ma soprattutto un discorso sull'uomo»⁵⁵, dato che non solo l'uomo plasma il mondo tramite la tecnologia, ma anche la tecnologia stessa plasma l'uomo e i suoi modi di pensare, agire ed esistere. Da questo punto di vista, la questione degli 'algoritmi etici' e della 'moralità artificiale' è una questione radicalmente etico-antropologica.

Andando oltre ogni suggerimento funzionale-operativo, alla domanda ricorrente 'come rendere etici gli algoritmi?' – iniziativa di per sé parziale ed equivoca, date le premesse di cui sopra – si devono opporre i quesiti 'perché vogliamo rendere etici i programmi informatici? Quali sono le motivazioni e gli esiti etico-antropologici di un tale tentativo di attribuzione di moralità a *software* e robot?'. Indubbiamente, anche il 'Das Man informatico' potrebbe avere qualcosa da dire di fronte a questi interrogativi, per lo meno al primo: per l'approccio applicativo la 'moralità artificiale' rappresenterebbe un evidente beneficio in termini di utilizzabilità, funzionalità, efficienza, autosufficienza, sicurezza e affidabilità dei sistemi informatici. Sviluppare dispositivi 'etici' – 'autonomi' e al contempo 'responsabili' – significherebbe poter disporre fiduciosamente di loro, intensificando ed estendendo in questo modo le deleghe

⁵² G. Anders, *L'uomo è antiquato. Vol. I. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, Torino 2007, p. 97.

⁵³ Cfr. A. Pessina, *Is a Good Algorithm also an Ethically Good Algorithm?*, cit., p. 27: «non è sufficiente calcolare i risultati, misurare il potenziale ed evidenziare i problemi che queste nuove tecnologie comportano: è essenziale capire come trasformano la nostra esperienza sensoriale, cognitiva, emotiva e relazionale».

⁵⁴ Cfr. G. Tamburrini, *Etica delle macchine*, cit., p. 68. L'etica non equivale al rispetto delle regole.

⁵⁵ P. Benanti, *La condizione tecno-umana. Domande di senso nell'era della tecnologia*, Bologna 2016, p. 20 e p. 70.

tecnologiche ad un numero sempre maggiore di contesti ed ambienti umani⁵⁶: è ora lo strumento in sé, non più il comportamento applicativo umano, ad essere voluto intrinsecamente 'etico', dato che esso opera automaticamente. Ciò nonostante, questa giustificazione pragmatica non coglie ancora il nesso tra tecnologie informatiche e antropologia: la prospettiva utilitaristica dell'impiego funzionale non comprende alcuna considerazione né sul significato essenziale della 'moralità artificiale' né sulle trasformazioni della nostra auto-comprensione antropologica e del nostro *status* morale.

L'idea che i sistemi informatici e i robot possano essere superiori all'uomo anche dal punto di vista etico, fino a diventare 'santi e perfetti ragionatori morali', è riconducibile a quello che Broussard chiama 'tecnosciovinismo':

l'idea che i computer siano più 'obiettivi' e 'senza pregiudizi' perché riducono domande e risposte alla valutazione matematica [...]. In qualche modo, negli ultimi due decenni, molti di noi hanno cominciato a dare per scontato che i computer facessero le cose bene e gli umani male. Abbiamo cominciato a dire cose del tipo: 'i computer sono migliori, perché sono più obiettivi delle persone'⁵⁷.

In altre parole, la *Weltanschauung* dei tecnosciovinisti risiede nella concezione, vissuta come convincimento, che i sistemi informatici (siano essi 'intelligenze artificiali' o robot) abbiano un vantaggio sull'uomo in molti, se non in tutti i compiti e settori. Questo vantaggio comprenderebbe anche il ragionamento morale. Eppure, questa spiegazione *socio-culturale* non è ancora sufficiente: da dove proviene tale persuasione?

Riattualizzando nuovamente le riflessioni di Anders, l'idea radicale di una 'moralità artificiale' e quindi di una superiorità etica dei sistemi informatici sembra trovare la sua giustificazione in quella che il filosofo tedesco chiama 'vergogna prometeica': l'uomo contemporaneo è in preda ad un senso di inferiorità, umiliazione e meschinità di fronte alle macchine che lui stesso ha elaborato e sviluppato. Questa vergogna non è più solamente fisico-corporea, non è soltanto ontologica, è ora anche apertamente morale: rispetto alle 'intelligenze artificiali' l'uomo si vede antiquato moralmente, impreciso nei suoi giudizi, goffo e inaffidabile nelle sue azioni. Parafrasando Anders, l'uomo si percepisce come una «costruzione difettosa» anche dal punto di vista etico: «le deficienze di cui 'si vergogna' sono numerose»⁵⁸, e si sente quindi inferiore ai sistemi informatici e ai robot non solo «per quanto riguarda la forza, la velocità, la precisione»⁵⁹ fisica, ma anche per quanto concerne il ragionamento e il comportamento morale. Se questi sono computabili, come già voleva Bentham, e se l'uomo è scarso e lento con i calcoli, allora è *evidente* – secondo i tecnosciovinisti del '*Das Man* informatico' – che anche l'etica può e dev'essere delegata alle macchine:

⁵⁶ Cfr. K. Martin, *Designing Ethical Algorithms*, cit., p. 131.

⁵⁷ M. Broussard, *La non intelligenza artificiale. Come i computer non capiscono il mondo*, Milano 2019, pp. 17-18. Cfr. anche p. 43 e pp. 204-205.

⁵⁸ G. Anders, *L'uomo è antiquato*, cit., p. 55.

⁵⁹ Ivi, p. 39.

queste possono arrivare ad agire da sole, autonomamente e meglio dell'uomo. Le macchine informatiche, essendo apparentemente più obiettive e libere da emozioni, desideri e pregiudizi – potendo, cioè, eseguire verosimilmente meglio dell'uomo le scelte e le azioni descritte operativamente – assurgono a modello etico sempre più perfezionato e perfezionabile, fino ad essere *immaginate* come una versione migliore dell'essere umano e perfino come 'sante', ossia moralmente adulte: «ma – scrive Anders – se sono le macchine a essere considerate 'adulte', allora il 'superamento dell'infanzia' e 'l'educazione del genere umano' significano il 'superamento dell'essere uomo'»⁶⁰.

Che ne è, quindi, dell'uomo, della nostra auto-comprensione antropologica e della nostra posizione morale? Considerare le macchine *potenzialmente* superiori dal punto di vista etico è una novità decisiva perché in questo modo «ci autotrasformiamo per amore delle nostre macchine, perché prendiamo le macchine a modello delle nostre alterazioni; rinunciamo quindi ad assumerci noi stessi come unità di misura e con ciò limitiamo la nostra libertà o vi rinunciamo»⁶¹. Prendere, anche solo teoreticamente, a modello i sistemi informatici e i robot, fino ad imitarli e a imparare da loro come pensare e comportarci, è una forma di auto-umiliazione e auto-limitazione antropologica: un «annientamento dell'uomo in quanto uomo»⁶², un'ulteriore e disperata privazione di libertà. Nella «*imitatio instrumentorum*»⁶³ come riforma di se stesso, fisica e morale, l'uomo si auto-degrada, adeguando il suo metro di giudizio etico a quello della macchina informatica: il ragionamento morale diventa calcolo operativo dell'efficienza logica, privato di qualsiasi emotività, carnalità e desiderio, oramai obsoleti. In questo senso, se i sistemi informatici e i robot diventano i 'veri adulti morali', i modelli grazie ai quali possiamo imparare a migliorare eticamente, avviene un rovesciamento, autentico e definitivo equivoco etico: «*i soggetti della libertà e della mancanza di libertà sono scambiati. Libere sono le cose [le macchine]; mancante di libertà è l'uomo*»⁶⁴.

Apparentemente astratte ed eccessive, queste osservazioni hanno in realtà una loro fondatezza se si considerano quelle stesse deleghe incentivate dal 'Das Man informatico' a cui si è accennato sopra. Come l'autonomia personale è limitata o inconsapevolmente pregiudicata dall'ubiquità pervasiva dei sistemi informatici – limitatamente controllabili, poco comprensibili e al contempo in grado di influenzare attivamente le scelte umane⁶⁵ – così anche il senso della responsabilità è trasformato dalla delega del processo decisionale: «una fiducia sostanziale è già riposta negli algoritmi, che in alcuni casi influenzano una *de-responsabilizzazione* degli attori umani, o una tendenza a 'nascondersi

⁶⁰ Ivi, p. 47.

⁶¹ Ivi, p. 52.

⁶² Ivi, p. 53.

⁶³ Ivi, p. 43.

⁶⁴ Ivi, p. 41. Cfr. S. Chiodo, *Human autonomy, technological automation (and reverse)*, «AI & Society», 37, 2022, pp. 39-48.

⁶⁵ Cfr. A. Tsamados et al., *The ethics of algorithms*, cit., pp. 8-9.

dietro il computer' e presumere che i processi automatizzati siano corretti per impostazione predefinita»⁶⁶. Il rischio, dunque, è quello di dar vita a dei 'riciclaggi morali' (*agency laundering*⁶⁷) prendendo le distanze da azioni moralmente sospette, a prescindere dal fatto che siano intenzionali o meno, e dando la colpa a programmi informatici e robot: si arriverebbe così, di fatto, ad eludere la responsabilità personale attraverso affermazioni distorte del tipo 'il computer ha detto così', 'il sistema *autonomo* mi ha consigliato di fare così', anche in quei contesti come la medicina in cui un 'suggerimento algoritmico' non ragionato o verificato potrebbe decidere realmente delle sorti di una persona.

Filippo Pianca
Università Cattolica del Sacro Cuore
✉ filippo.pianca@unicatt.it

Bibliografia / References

- Alaieri, F., Vellino, A. 2016. *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, in Agah, A., Cabibihan, JJ., Howard, A., Salichs, M., He, H. (ed. by). *Social Robotics. ICSR 2016. Lecture Notes in Computer Science*, Cham, Springer, pp. 159-168.
- Allen, C. et al. 2005. *Artificial morality: Top-down, bottom-up, and hybrid approaches*, «Ethics and Information Technology», 7, pp. 149-155.
- Anders, G. 2007. *L'uomo è antiquato. Vol. I. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, trad. it., Torino, Bollati Boringhieri.
- Anderson, M., Anderson, S. L. 2007. *Machine Ethics: Creating an Ethical Intelligent Agent*, «AI Magazine», 28, 4.
- Anderson, M., Anderson, S. L. 2010. *Robot be good*, «Scientific American», 4.
- Benanti, P. 2018. *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Genova, Marietti.
- Bostrom, N., Yudkowsky, E. 2014. *The ethics of artificial intelligence*, in Frankish, K., Ramsey, W. (ed. by). *The Cambridge Handbook of Artificial Intelligence*, Cambridge, Cambridge University Press, pp. 316-334.
- Broussard, M. 2019. *La non intelligenza artificiale. Come i computer non capiscono il mondo*, trad. it., Milano, Franco Angeli.
- Chiodo, S. 2022. *Human autonomy, technological automation (and reverse)*, «AI & Society», 37, pp. 39-48.
- Cormen, T. H. et al. 2009. *Introduction to Algorithms*, Cambridge, Cambridge University Press.
- Danielson, P. 1992. *Artificial Morality: Virtuous Robots for Virtual Games*, New York, Routledge.
- Floridi, L., Sanders, J. W. 2004. *On the Morality of Artificial Agents*, «Minds and Machines», 14, pp. 349-379.

⁶⁶ B. D. Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, cit., p. 12 (corsivo mio).

⁶⁷ Cfr. A. Tsamados et al., *The ethics of algorithms*, cit., p. 10.

- Gips, J. 2011. *Towards the Ethical Robot*, in Anderson, M., Anderson, S. L. (ed. by). *Machine Ethics*, Cambridge, Cambridge University Press, pp. 244-253.
- Heidegger, M. 1971. *Essere e tempo*, trad. it., a cura di F. Volpi, Milano, Longanesi.
- Hill, R. K. 2016. *What an algorithm is*, «Philosophy & Technology», 29, pp. 35-59.
- Kearns, M., Roth, A. 2020. *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*, Oxford, Oxford University Press.
- Knuth, D. E. 1996. *Selected Papers on Computer Science*, Cambridge, Cambridge University Press.
- Kraemer, F. et al. 2011. *Is there an ethics of algorithms?*, «Ethics and Information Technology», 13, pp. 251-260.
- Leben, D. 2019. *Ethics for Robots. How to Design a Moral Algorithm*, New York, Routledge.
- Martin, K. 2019. *Designing Ethical Algorithms*, «MIS Quarterly Executive», 18, 2, pp. 129-142.
- Mittelstadt, B. D. et al. 2016. *The ethics of algorithms: Mapping the debate*, «Big Data & Society», 3, 2.
- Pessina, A. 2020. *Onlife. Trasformazioni dell'Esperienza tra Immanenza e Trascendenza*, «Hermeneutica», 27, pp. 77-88.
- Pessina, A. 2021. *Is a Good Algorithm also an Ethically Good Algorithm?*, in Paglia, V., Pegoraro, R. (ed. by). *The 'Good' Algorithm? Artificial Intelligence Ethics, Law, Health. Proceedings of the XXVI General Assembly of Members*, Selci Lama, Pliniana, pp. 27-29.
- Sandvig, C. et al. 2016. *When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software*, «International Journal of Communication», 10, pp. 4972-4990.
- Santucci, V. G. et al. 2016. *Grail: a Goal-discovering Robotic Architecture for Intrinsically-motivated Learning*, «IEEE Transactions on Cognitive and Developmental Systems», 8, pp. 214-231.
- Savulescu, J. et al. 2020. *An ethical algorithm for rationing life-sustaining treatment during the COVID-19 pandemic*, «British journal of anaesthesia», 3, pp. 253-258.
- Sparrow, R. 2021. *Why machines cannot be moral*, «AI & Society», 36, 3, pp. 685-693.
- Tamburrini, G. 2020. *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma, Carocci.
- Tsamados, A. et al. 2021. *The ethics of algorithms: key problems and solutions*, «AI & Society», 37, pp. 215-230.
- Turilli, M., Floridi, L. 2009. *The ethics of information transparency*, «Ethics and Information Technology», 11, pp. 105-112.
- Vaihinger, H. 1967. *La filosofia del come se*, trad. it., Roma, Astrolabio Ubaldini.
- Wallach, W. 2008. *Implementing moral decision making faculties in computers and robots*, «AI & Society», 22, 4, pp. 463-475.
- Wallach, W., Asaro, P. (ed. by). 2017. *Machine Ethics and Robot Ethics*, London, Routledge.