

*Ivan Rotella*

## **Etica e fantascienza. Io, robot e i dilemmi etici**

*ABSTRACT: Isaac Asimov's collection of short stories I, Robot is best known for his formulation of the "Three Laws of Robotics". It is a science-fiction novel, thus a work of literary art, but also a true ethical-philosophical reflection. The three laws, in fact, apart from being principles of computer programming through which the robots and machines imagined by Asimov are constructed, are also actual rules of behaviour for them. Asimov himself admits that he was inspired by human ethics in formulating the laws. Asimov, however, is by no means an apologist for the three laws; on the contrary, it could be said that his reflection is to be a questioning of any normative approach to ethics, be it human or artificial. The aim of this essay will therefore be to investigate the extent to which Asimov's imagery can serve as a litmus test of real, human problems.*

*KEYWORDS: Ethics, Science Fiction, Dilemma, Anthropology, Uncanny.*

Formulare giudizi etici. A oggi, questa è la funzionalità robotica più complessa che sia stata sviluppata.<sup>1</sup>

Hannah Arendt, nel suo *Vita activa*, riporta la notizia del primo satellite lanciato nello spazio, lo *Sputnik 1*, sostenendo che al successo sovietico dell'impresa "uno dei giornali americani più rispettabili riportò in prima pagina ciò che era confinato fino allora in una letteratura non precisamente rispettabile, la fantascienza"<sup>2</sup>. In realtà, è la stessa Arendt a lanciarsi immediatamente in un commento più conciliativo nei confronti di questo genere letterario, sostenendo che alla fantascienza: "purtroppo, nessuno ancora ha dedicato l'attenzione che merita come veicolo di sentimenti e di desideri di massa"<sup>3</sup>.

Un grande scrittore di fantascienza, Philip José Farmer, nel suo *Il grande disegno*, terzo volume del *Ciclo del mondo del fiume*, mette in bocca a uno dei suoi personaggi – Frigate, una sorta di pseudonimo per mascherare se stesso – una interessante definizione di fantascienza, quella di "letteratura speculativa"<sup>4</sup>.

1 Asimov 2021.

2 Arendt 2017, 38.

3 Arendt 2017, 38.

4 Farmer 2012, 36.

In effetti la fantascienza e la speculazione filosofica hanno molto in comune. Basti pensare all'esperimento del *Gran Navilio*<sup>5</sup> di Galilei che, oggi, appare come un esperimento mentale di grande valore scientifico ed esplicativo del movimento terrestre, ma, all'epoca, poteva tranquillamente essere interpretato come "letteratura speculativa". Un ragionamento analogo si potrebbe fare per gli zombie di Chalmers<sup>6</sup>, la teiera di Russell<sup>7</sup>, la stanza di Mary di Jackson<sup>8</sup>, il violinista di Thompson<sup>9</sup> e così via. Ma se la filosofia ha da sempre speculato attraverso il ricorso ad esperimenti mentali offrendo materiale alla fantascienza – l'esempio forse più clamoroso sono i cervelli in una vasca di Putnam<sup>10</sup> poi ripresi in numerosi film, serie tv, cartoni animati, su tutti basti ricordare *Matrix* o *Bliss* – è vero anche il contrario, che la fantascienza, intesa come letteratura speculativa, ha offerto e può offrire materiale alla riflessione filosofica. In questo senso l'obiettivo del presente saggio sarà quello di indagare uno dei possibili contributi che l'etica robotica asimoviana può offrire all'etica umana e, in particolare, quanto il confronto con le macchine immaginate da Asimov possa aiutare a delineare un'antropologia filosofica centrata sulla saggezza fronetica più che sulla razionalità calcolante.

I romanzi e i racconti di Asimov vantano numerose traduzioni e diverse trasposizioni cinematografiche che hanno indubbiamente contribuito ad aumentarne la fama. La raccolta di racconti *Io, robot* è forse l'opera asimoviana più celebre. Il motivo di questa notorietà è individuabile con ogni probabilità nelle famose *Tre leggi della robotica* che proprio in questa raccolta vengono formulate per la prima volta<sup>11</sup>. All'interno di *Io, robot* Asimov distingue tra i robot, protagonisti di quasi tutti i racconti, e le macchine, personaggi maggiormente presenti nella parte finale della raccolta, in particolare nell'ultimo racconto, *Il conflitto evitabile* [*The Evitable Conflict*, 1950]. La distinzione offerta da Asimov tra robot e macchine potrebbe ricalcare l'attuale distinzione tra robot e computer, più specificamente quella tra robot e AI, in quanto i robot di Asimov hanno un aspetto umanoide, mentre le macchine sono dei grandi cubi, collegati tra loro in una rete globale, dotati di altissime capacità di calcolo e di risoluzione dei problemi. La conformazione fisica umanoide dei robot e l'inferiore capacità di calcolo rispetto alle macchine non sono le uniche differenze, ma una distinzione forse ancora più importante è individuabile nel fatto che i robot sono a servizio degli umani e questi ultimi – coerentemente all'etimologia stessa della parola "robot" – li utilizzano come dei veri e propri schiavi<sup>12</sup>. Al contrario, le macchine governano l'umanità

5 Galilei 2014.

6 Chalmers 1996.

7 Russell 1952.

8 Jackson 1986.

9 Thomson 1971.

10 Putnam 1981.

11 In realtà le *Tre Leggi* compaiono soltanto nella seconda edizione del racconto di apertura della raccolta, *Robbie*. Inoltre, in seguito Asimov aggiungerà la *Legge zero* nel romanzo *I robot e l'impero*.

12 La parola "robot" deriva dal ceco "robota", che significa lavoro servile. Generalmente, nel romanzo di Karel Čapek, *I robot universali di Rossum*, è riconosciuta la prima comparsa

all'insaputa di quest'ultima, sebbene il loro governo si configuri nei fatti come un governo illuminato e filantropico. Infine, i robot prestano servizio al di fuori del pianeta Terra, poiché le pressioni politiche della *Coalizione per l'Umanità* – una organizzazione tecnofobica e conservatrice – sono riuscite a ottenerne la messa al bando, mentre le macchine operano liberamente sul pianeta Terra, sebbene quasi nessuno sia davvero consapevole della intera portata del loro operato.

Tuttavia, nonostante queste numerose differenze, Asimov sostiene che “benché le Macchine, in fondo, non siano altro che i più vasti conglomerati di circuiti di calcolo mai inventati, ai fini della Prima Legge restano pur sempre dei robot [*they are still robots within the meaning of First Law*]”<sup>13</sup>. Adombrando quasi la possibilità che le *Tre Leggi della Robotica* valgano tanto per i robot in senso stretto, quanto per le Macchine che non possiedono forma umanoide, sebbene sia forse più corretto limitare l'identità di programmazione alla sola Prima Legge.

## 1. Eteronomia e le Tre Leggi della Robotica

Prima di proseguire è forse opportuno richiamare le Tre Leggi della Robotica così come formulate da Asimov:

1. Un robot non può arrecare danno a un essere umano, o, mancando di agire, lasciare che un essere umano subisca un danno [*A robot may not injure a human being or, through inaction, allow a human being to come to harm*];
2. Un robot deve obbedire agli ordini che riceve da un essere umano, se ciò non è in conflitto con la Prima Legge [*A robot must obey the orders given it by human beings except where such orders would conflict with the First Law*];
3. Un robot deve proteggere la propria esistenza, se ciò non è in conflitto con la Prima e la Seconda Legge [*A robot must protect its own existence as long as such protection does not conflict with the First or Second Law*]<sup>14</sup>.

Le Tre Leggi della Robotica sono state accostate a una moralità di stampo kantiano a causa del loro carattere imperativo e della loro natura deontologica e universalistica. Tuttavia, non si è mancato di far notare come la legge morale kantiana sia autonoma, impartita dalla ragione per sé stessa, mentre le leggi robotiche siano eteronome, cioè impartite da una ragione umana nei confronti di un'intelligenza robotica deputata a riceverle dall'esterno<sup>15</sup>. In realtà, è lo stesso

del termine “robot”. Sul tema robot e AI interpretati in termini di schiavitù, con particolare riferimento alla dialettica servo-padrone hegeliana, si veda Bodei 2019.

13 Asimov 2021, 225.

14 Asimov 2021, 49.

15 Gisondi 2013, 212: “L'imperativo categorico di Kant e le leggi robotiche di Asimov mostrano due chiare similitudini e una profondissima differenza: in tutti e due i casi alla centralità dell'umano si associa la necessità di rendere universale la norma che lo tutela, ma se per Kant

Asimov a mettere in luce l'ispirazione etica dalla quale ha evinto le Tre Leggi della Robotica e questa ispirazione non sembra affatto coincidente con la riflessione di Kant:

se ci riflette un attimo, le Leggi della Robotica sono i principi cardine [*essential guiding principles*] di gran parte dei sistemi etici dell'umanità [*of a good many of the world's ethical systems*]. Ovviamente ogni essere umano dovrebbe essere dotato di un istinto di autoconservazione [*instinct of self-preservation*]. Per un robot questa è la Terza Legge. Ogni essere umano "buono", inoltre, se con ciò si intende un essere umano provvisto di un minimo di coscienza sociale e senso di responsabilità [*a social conscience and a sense of responsibility*], accetta di rispettare una qualche misura di autorità: ascolta i medici, i superiori, il governo, lo psichiatra, i propri simili; non infrange le leggi, segue le regole, si adatta alle usanze [*to conform to custom*] – anche quando ciò risulta scomodo o pericoloso [*even when they interfere with his comfort or his safety*]. Per un robot questa è la Seconda Legge. Non solo: ogni essere umano "buono" dovrebbe amare il prossimo suo come se stesso [*to love others as himself*], proteggerlo, e rischiare la propria vita per salvarne un'altra. Per un robot questa è la Prima Legge. In parole povere: se Byerley rispetta le Leggi della Robotica, potrebbe essere un robot [*he may be a robot*], oppure semplicemente un ottimo essere umano [*may simply be a very good man*].<sup>16</sup>

La parte conclusiva della citazione si riferisce alle difficoltà che la robopsicologa Susan Calvin riscontra nel distinguere un essere umano, rispettoso delle norme etiche più comuni, da un robot programmato per rispettare le Tre Leggi. In particolare, il Byerley di cui parla si scoprirà essere effettivamente un robot, il quale riuscirà, non senza incontrare ostacoli, a farsi eleggere governatore mondiale, ma, data la messa al bando dei robot sulla Terra, per riuscire nel suo scopo dovrà fingersi un essere umano.

Ciò che Asimov, attraverso la voce di Susan Calvin, dimentica di evidenziare è che, se pure si possano paragonare l'amore per il prossimo alla Prima Legge, il rispetto delle regole sociali alla Seconda Legge e l'istinto di sopravvivenza alla Terza, tuttavia, per gli umani, non esiste una gerarchia chiara e precisa di questi tre elementi, come invece avviene per le Tre Leggi della Robotica, in cui l'istinto di sopravvivenza è chiaramente subordinato alle prime due leggi, analogamente a come il rispetto delle regole sociali è subordinato alla Prima e, dunque, all'amore del prossimo. Per gli esseri umani, a parità di principi, non è data alcuna gerarchia<sup>17</sup> e, forse, è plausibile pensare che nella maggior parte dei casi sia data una gerarchia esattamente opposta a quella disegnata da Asimov per i suoi robot. Inoltre, se per gli umani il libero arbitrio si pone come problema, per i robot rimane un aspetto del loro comportamento completamente assente. Tuttavia, l'intervento di Susan Calvin rimane di grande interesse filosofico, poiché evidenzia non soltanto la difficoltà con la quale si confronta da sempre l'antropologia filosofica, cioè il

il senziante tutelante è un soggetto che dà a sé stesso la norma, per Asimov esso è vincolato dall'esterno".

<sup>16</sup> Asimov 2021, 204.

<sup>17</sup> Fonnesu 1998.

rinvenimento di una definizione di “essere umano”<sup>18</sup>, ma anche e soprattutto il fatto che, a differenza di quanto portato avanti dalla riflessione filosofico-antropologica classica, il confronto forse più fecondo al fine di una maggiore consapevolezza su che cosa sia un essere umano non è il confronto con gli animali non-umani, sul modello dei tre grandi padri dell’antropologia filosofica, Scheler, Plessner, Gehlen<sup>19</sup>, bensì con le macchine, avvicinandosi in questo all’approccio de *L’uomo è antiquato* di Anders<sup>20</sup>, quasi coevo all’uscita di *Io, robot*. Inoltre, le difficoltà di riuscire a distinguere un essere umano da un replicante robotico chiamano in causa quei sentimenti e desideri di massa di cui parlava Arendt, sentimenti e desideri che la fantascienza è in grado di veicolare e scandagliare. In questo caso, il sentimento *sui generis* evocato dalla difficoltà di distinguere un umano da un robot richiama chiaramente il perturbante freudiano<sup>21</sup> e la conseguente riflessione sul doppio<sup>22</sup> inteso come altro nel quale rispecchiare se stesso.

Al di là della narrativa espressa in *Io, robot*, dei personaggi e delle situazioni raccontate, sembra quasi che a legare tutti i racconti sia esattamente la domanda antropologica di fondo: cosa fa di un essere umano un essere umano? E in che cosa continua a distinguersi un essere umano da un robot capace di replicare qualsiasi aspetto fisico e qualsiasi comportamento degli esseri umani?

Il vero protagonista di *Io, robot* è la robopsicologa Susan Calvin: tutti i racconti sono risposte che la dottoressa Calvin offre a un intervistatore che ha il compito di scrivere un articolo biografico su questa importante figura dell’azienda leader nella produzione di robot e macchine nel mondo immaginato da Asimov. Nelle primissime battute di *Io, robot*, Susan Calvin afferma: “io stessa sono stata definita un robot più di una volta [*I’ve been called a robot myself*]. Glielo avranno detto che non sono umana [*Surely, they’ve told you I’m not human*]”<sup>23</sup>. Byerly è un robot che però tutti credono un umano, Calvin è una umana che però alcuni credono sia un robot: i confini tra umano e robotico, in questi due personaggi, sono a tal punto sfumati da generare volutamente confusione. Un indizio importante di questa commistione tra umano e robotico viene offerto da Asimov nelle primissime battute della raccolta, esplicitando il nome dall’azienda produttrice di robot per cui lavora Susan Calvin: *US Robots and Mechanical Men Corp*. Il nome è indicativo del fatto che l’azienda produce sì robot, ma anche *Mechanical Men*. Se per “mechanical Men” si intendessero i robot, non ci sarebbe bisogno di specificare due volte l’oggetto di produzione. Sembra quindi adombrato, all’interno del nome dell’azienda, l’assunto implicito che la *US Robots* produca – in aggiunta rispetto a

18 Scheler 1988, 257-258: “noi siamo la prima epoca in cui l’uomo è divenuto completamente e interamente ‘problematico’ per se stesso; in cui egli non sa più che cosa è, ma nello stesso tempo sa anche che non lo sa”.

19 Per una visuale completa sui padri fondatori dell’antropologia filosofica, si veda Pansera 2019.

20 Anders 2003.

21 Freud 1989.

22 Per una visuale sul rapporto tra perturbante e rispecchiamento tra umani e macchine, si veda Pisano 2021a.

23 Asimov 2021, 11.

robot, macchine e cervelli positronici – anche esseri umani non più completamente organici, ma in parte meccanizzati. Sebbene Asimov non usi la parola “cyborg”, è come se con “mechanical Men” intendesse esattamente dei cyborg e, Susan Calvin fosse un esemplare di anello mancante tra umano e robot: un cyborg per l'appunto<sup>24</sup>.

## **2. Conflitto e collasso delle Tre Leggi della Robotica**

Le Tre Leggi della Robotica, sebbene costituiscano la programmazione di base dei robot e delle macchine immaginate da Asimov, in realtà entrano in crisi quasi immediatamente. Già nel secondo racconto, *Girotondo* [*Runaround*] del 1942, il robot Speedy, programmato per recuperare selenio sul pianeta Mercurio, si trova a dover affrontare una situazione dilemmatica. Il selenio rappresenta per il robot un pericolo, quindi, obbedendo alla Terza Legge, il robot è portato ad allontanarsi da questo elemento pericoloso per la sua sopravvivenza. Tuttavia, gli umani presenti sulla stazione di Mercurio avevano ordinato al robot di svolgere questa operazione di recupero, quindi il robot avrebbe dovuto obbedire agli umani in conformità alla Seconda Legge. I robot della serie Speedy, però, proprio a causa del loro compito altamente pericoloso e del loro elevato costo di produzione, avevano ricevuto una programmazione tale che la Terza Legge ne risultava rafforzata, sebbene rimanesse vincolata all'ordine gerarchico delle prime due. Nel momento in cui il robot riceve un ordine di recupero non particolarmente esplicito, la Seconda Legge funziona, ma non al massimo delle sue potenzialità, mentre la Terza Legge che era stata rafforzata in sede di programmazione, risulta abbastanza forte da controbilanciare l'ordine poco convinto che gli era stato impartito. Si viene così a formare un conflitto tra Seconda e Terza Legge che porta il robot a eseguire l'ordine, abbandonando la stazione e avvicinandosi alla pozza di selenio, ma, una volta giunto nei pressi della pozza, scatta nel robot l'impulso a evitare di svolgere l'operazione di recupero rischiosa per la sua incolumità, costringendolo di fatto a un girotondo continuo tra l'impulso a seguire la Seconda Legge obbedendo agli umani e l'impulso a seguire la Terza preservando se stesso dai pericoli.

Se le Tre Leggi della robotica non sono altro che metafore utilizzate da Asimov per descrivere la moralità umana, allora è chiaro l'intento asimoviano di sostenere una posizione etica che, sebbene si presenti come una posizione deontologica, sul modello kantiano, non sia tuttavia cieca e acritica nei confronti dei problemi che possono nascere in caso di conflitti tra doveri. Inoltre, se un conflitto di doveri può presentarsi come irrisolvibile in un robot, il quale è sprovvisto di libero arbitrio ed è fornito di una chiara gerarchia di doveri, allora, un conflitto di doveri analogo può presentarsi, per un umano, come possibilità di scelta tragica e foriera di sensi di colpa.

24 Sul tema dei cyborg come possibile salto evolutivo all'interno di una visuale olistica delle forme di vita sul pianeta, si veda l'interessante libro-intervista Lovelock 2020.

Un problema simile si ripresenta in *Caccia alla lepre* [*Catch that Rabbit*, 1944]: il robot DV-5, soprannominato Dave, è un robot che deve recuperare materiale minerario ed è chiamato a coordinare il lavoro di altri sei robot che lo aiutano nelle operazioni di recupero. Il lavoro in miniera è ovviamente pericoloso e potrebbero verificarsi dei crolli i quali metterebbero a rischio l'incolumità dei robot. La programmazione di Dave è stata modificata in modo da poter agire in autonomia, finanche in totale assenza di esseri umani durante le operazioni, ma quando gli esseri umani sono presenti, Dave riesce a svolgere le operazioni e portare a termine il lavoro, quando invece si trova in compagnia di soli altri robot, fallisce nello svolgere l'ordine ricevuto. Anche in questo caso si tratta di un conflitto tra la Seconda e la Terza Legge, esacerbato, in questo caso, dal fatto che Dave, oltre a eseguire gli ordini, avrebbe dovuto anche coordinare il lavoro di altri sei robot e impartire loro a sua volta degli altri ordini. Gregory Powell e Mike Donovan, due dipendenti della *US Robots*, vengono inviati a risolvere il malfunzionamento. Per prima cosa sottopongono Dave a una serie di controlli volti a verificare il corretto funzionamento di ogni componente del suo cervello positronico e, a questo punto, Asimov dimostra ancora una volta il suo interesse per l'etica:

All'inizio andò tutto liscio. Il modello DV-5 eseguì alcune moltiplicazioni a cinque cifre al ritmo spietato di un cronometro. Elencò i numeri primi fra mille e diecimila. Calcolò alcune radici cubiche e funzioni integrali via via più complesse. Eseguì le reazioni meccaniche in ordine crescente di difficoltà. E infine approcciò il suo raffinato cervello positronico alle funzioni *più elevate del mondo robotico, ovvero risolvere problemi di etica decisionale* [*the highest function of the robot world – the solutions of problems in judgment and ethics*].<sup>25</sup>

L'interesse per le Tre Leggi della Robotica, se si interpreta il confronto umani-robot in termini di antropologia filosofica, si dimostra dunque in un senso ancora più profondo un interesse per l'etica. Immaginare, come fa Asimov, un'etica robotica, significa *ipso facto* immaginare un'etica umana<sup>26</sup>. Ma se le Tre Leggi della Robotica espresse da Asimov in *Io, robot* entrano immediatamente in conflitto tra loro e mettono fuori gioco un approccio puramente deontologico, è allora necessario cercare altrove un modello etico più conforme all'approccio asimoviano.

### 3. Oltre le Tre Leggi della Robotica

L'etica robotica descritta da Asimov, in realtà, è comprensiva di una quarta legge, antecedente alle famose Tre Leggi e, proprio per questo, chiamata Legge Zero. Canonicamente viene fatta risalire all'opera *I robot e l'impero*, in cui il robot Daneel Olivaw enuncia le seguenti parole: "C'è una legge superiore alla Prima Legge [...]: *Un robot non può fare del male all'umanità o, tramite l'inazione,*

<sup>25</sup> Asimov 2021, 85. Corsivo mio.

<sup>26</sup> Anderson 2008; Murphy 2009.

*permettere che l'umanità riceva danno [A robot may not injure humanity or, through inaction, allow humanity to come to harm]. Io la considero la Legge Zero della Robotica*<sup>27</sup>. Sebbene la prima enunciazione di questa legge si trovi in *I robot e l'impero*, tuttavia è possibile rinvenire anche in *Io, robot* qualcosa di simile, pur se non espressamente enunciata nei termini formali di una legge. Si tratta, ancora una volta, dei personaggi di Byerley, il robot che si finge umano per governare il mondo meglio di quanto farebbe qualsiasi umano, e Susan Calvin, la robopsicologa algida e misantropa. Susan Calvin e Alfred Lanning, un altro dipendente della *US Robots*, stanno discutendo della possibilità che Byerley possa effettivamente essere un robot sotto mentite spoglie e, Lanning, propone di escludere una tale eventualità in ragione del fatto che Byerley aveva lavorato come procuratore distrettuale, trovandosi quindi nella situazione di dover accusare degli esseri umani di aver commesso dei reati e, così facendo, correre il rischio di procurare loro degli eventuali danni. Il lavoro di procuratore distrettuale sarebbe dunque stato, secondo Lanning, incompatibile con l'essere un robot, poiché accusare di un reato un essere umano, nel caso in cui l'accusa avesse condotto effettivamente a una condanna, avrebbe potuto essere considerato *ipso facto* un modo per infrangere la Prima Legge, in particolare se la condanna eventualmente espressa dal giudice si fosse poi rivelata una condanna a morte. Susan Calvin risponde a Lanning che il robot Byerley, in qualità di procuratore, non ha mai chiesto condanne a morte, né mai ne sono state eseguite a partire dal suo lavoro di indagine. Anzi, Byerley era stato anche un membro attivo di movimenti che si ponevano l'obiettivo di abolire la pena di morte e di riformare il sistema penale in una direzione maggiormente rieducativa e niente affatto punitiva. La robopsicologa Susan Calvin, a questo punto, può smentire l'interpretazione di Lanning secondo cui un robot non potrebbe lavorare come procuratore distrettuale e, in questo senso, afferma: "non ha ucciso nessuno. Ha esposto delle circostanze che potrebbero convincere una giuria che un dato essere umano rappresenta un pericolo per la [...] società. Protegge la maggioranza, quindi rispetta la Prima Legge nel suo potenziale massimo [*He protects the greater number and thus adheres to Rule One at maximum potential*]. Non si spinge oltre"<sup>28</sup>. In altri termini, Calvin sta sostenendo che proteggere l'umanità è la massima espressione della Prima Legge, una sorta di Legge Zero *ante litteram*.

L'etica robotica asimoviana, dunque, è un'etica che si fonda su quattro leggi, le quali potrebbero essere anche interpretate come quattro principi e, a questo punto, l'accostamento più funzionale non sarebbe più con il modello kantiano, dovuto semplicemente al carattere deontologico delle Tre/Quattro Leggi, bensì quello con il principialismo etico di Beauchamp e Childress<sup>29</sup>, sebbene quest'ultimo sia in realtà una trentina d'anni più tardo rispetto alla formulazione asimoviana.

La Prima Legge della Robotica – un robot non può arrecare danno a un essere umano, o, mancando di agire, lasciare che un essere umano subisca un danno – potrebbe essere accostata al principio della *non-maleficence*, nel senso del *primum*

27 Asimov 1986, 268.

28 Asimov 2021, 205.

29 Beauchamp and Childress 1979.



*non nocere*, sebbene nella classificazione di Beauchamp e Childress questo *primum* sia in realtà il terzo principio. La Seconda Legge della robotica – un robot deve obbedire agli ordini che riceve da un essere umano, se ciò non è in conflitto con la Prima Legge – sembra adattarsi con più difficoltà al principio della *beneficence*. Tuttavia, non è forse completamente impossibile l'accostamento. Il brocardo *primum non nocere* è immediatamente seguito da *secundum cavere, tertium sanare*, poiché si presume, almeno nel paradigma ippocratico antico, che la richiesta del paziente sia sempre una richiesta di guarigione o comunque di alleviamento del dolore. Si viene così a costituire un legame molto stretto tra il principio della *non-maleficence* e della *beneficence*. Il *sanare* è vincolato a un preventivo *non nocere*. Analogamente, l'etica robotica asimoviana presuppone, in linea generale, che gli ordini provenienti dagli umani siano sempre ordini volti a migliorare il loro benessere e, quindi, degni di essere soddisfatti, ma anche in questo caso l'obbedienza che presume un comando benevolo è vincolata a un preventivo discrimine sull'effettiva bontà del comando. In altri termini, il contenuto della Seconda Legge, sebbene radicalmente differente rispetto al principio della *beneficence*, tuttavia istituisce con la Prima Legge un rapporto analogo a quello che il principio della *beneficence* istituisce con quello della *non-maleficence*. La Terza Legge – un robot deve proteggere la propria esistenza, se ciò non è in conflitto con la Prima e la Seconda Legge – è quella più difficilmente avvicinabile al principio di autonomia. Nel principialismo si tratta dell'autonomia del paziente sottoposto a cure mediche che può rifiutare il trattamento suggeritogli; nel caso della Terza Legge, l'autonomia andrebbe interpretata come una sorta di libertà d'azione e di autodeterminazione del robot che può esprimersi pienamente e in qualsiasi forma, ma soltanto a patto di aver precedentemente adempiuto e rispettato le prime due leggi. La Terza Legge potrebbe dunque essere interpretata come un principio di salvaguardia e rispetto delle libertà individuali, a patto che queste non infrangano i due principi precedenti di *non-maleficence* e *beneficence*. Infine, l'aspetto probabilmente più interessante, è l'accostamento tra la Legge Zero e il principio di giustizia. Nel caso in cui si diano due pazienti, entrambi bisognosi di un trattamento di terapia intensiva, ma un solo posto letto disponibile, i principi di *beneficence*, autonomia e, in parte, di *non-maleficence*, vengono meno. Nel momento in cui bisogna necessariamente scegliere uno dei due pazienti, di fatto condannando l'altro, la scelta deve essere guidata da criteri di giustizia. Analogamente, la Legge Zero, precedendo le altre tre, permette di nuocere a un essere umano, nel caso in cui quell'essere umano metta a rischio l'umanità intera, di non obbedire a un essere umano, nel caso in cui l'ordine ricevuto metta a rischio l'umanità intera e di proteggere l'esistenza del robot, sebbene questa protezione implichi il mancato rispetto di Prima e Seconda Legge.

Floridi, nel suo *Etica dell'intelligenza artificiale*<sup>30</sup>, pur senza fare riferimento ad Asimov, arriva a sostenere una tesi simile a quella del paragone tra le quattro leggi della robotica e il principialismo biomedico, cioè che “di tutte le aree dell'etica

applicata, la bioetica è quella che più assomiglia all'etica digitale<sup>31</sup> e, proprio in ragione di questa somiglianza, lo stesso Floridi riduce tutti i principi guida che dovrebbero guidare l'etica dell'intelligenza artificiale ai quattro principi del principlismo biomedico, aggiungendone però un quinto, l'esplicabilità<sup>32</sup>, intesa sia in termini di trasparenza, sia in termini di *accountability*.

Ciò che continua a mancare, tanto nel principlismo biomedico quanto nell'etica dell'intelligenza artificiale, è una chiara gerarchia tra i principi. In questo senso le leggi della robotica di Asimov potrebbero addirittura andare al di là di principlismo ed etica dell'intelligenza artificiale, offrendo sì un'impostazione simile a entrambi i modelli, ma aggiungendovi anche un rigido ordine gerarchico tra i principi.

Se si associa a ognuna delle leggi della robotica un paradigma etico di riferimento, si può ottenere un'ulteriore corrispondenza tra la Legge Zero e l'etica della virtù, una corrispondenza tra Prima e Seconda Legge con il deontologismo e, infine, tra la Terza Legge e il consequenzialismo. Una volta istituite queste corrispondenze, la numerazione da zero a tre potrebbe essere interpretata in termini cronologici, nel senso che l'etica della virtù ha preceduto storicamente quella deontologica, la quale a sua volta ha preceduto quella consequenzialista, ma, se ci si avvicina alla questione cercando di tener conto dei possibili conflitti tra principi e/o leggi, l'ordine potrebbe rovesciarsi: ognuno è libero di agire autonomamente, godendo al massimo delle proprie libertà, se e solo se questo godimento non infrange i principi di *beneficence* e *non-maleficence*, cioè il consequenzialismo, finanche interpretato in direzione dell'egoismo etico, può configurarsi come legittimo principio dell'agire, un principio *prima facie* valido a patto di non infrangere i doveri di un sistema etico deontologico, il quale, a sua volta, impone dei doveri di *beneficence* come doveri a loro volta *prima facie*, vincolati a un principio preliminare di *non-maleficence*; infine, anche il sistema etico deontologico non può essere interpretato in versione *strong*, ma deve strutturarsi anch'esso come un deontologismo a sua volta *prima facie* rispetto al più fondante principio di giustizia, inteso come fondamento etico ed *extrema ratio* in caso di conflitto tra principi. La giustizia si configura quindi come l'alfa della moralità, in quanto fondamento, ma, allo stesso tempo, omega in quanto *extrema ratio* e ultimo appello quando gli altri principi hanno già evidenziato i loro limiti, quando non si tratta più dell'applicazione di un principio, ma della risoluzione di un dilemma.

A questo punto può forse essere reso evidente il contenuto implicito del ragionamento asimoviano: ciò che rende umano un essere umano è la capacità, tragica, di risoluzione di un dilemma etico. I robot, pur dotati delle tre/quattro leggi della robotica, non riescono mai a divincolarsi nel momento in cui queste leggi entrano in conflitto tra loro, presentando guasti e malfunzionamenti. Un essere umano, al contrario, è proprio nel momento della scelta tragica e dilemmatica che mostra tutta la sua umanità.

31 Floridi 2022, 96.

32 Floridi 2022, 96.

La trasposizione cinematografica di *Io, robot* (Proyas 2004) si apre proprio con una scena, diegeticamente assente in Asimov, che porta il segno di tale tragicità, cogliendo lo spirito del testo asimoviano: un incidente stradale tra due auto che genera una situazione dilemmatica<sup>33</sup>. Due automobili, a causa dell'incidente, sono uscite dalla carreggiata e si trovano catapultate nel fiume che costeggia la strada. All'interno della prima auto c'è un uomo adulto – l'agente Spooner interpretato da Will Smith – e, nell'altra auto, ci sono un padre morto sul colpo e una bambina, Sarah. Un robot che si trova nei paraggi assiste alla scena e si tuffa in acqua nel tentativo di salvare le persone coinvolte. Grazie ai suoi sensi robotici riesce a vedere, meglio di quanto potrebbe fare un qualsiasi essere umano, attraverso le acque torbide del fiume. Individua l'agente Spooner in un'auto e la bambina nell'altra. Grazie alla sua intelligenza robotica, in pochissimi decimi di secondo, calcola in percentuale le *chances* di sopravvivenza dei due soggetti. Spooner si trova nell'auto più vicina al pelo dell'acqua; la bambina è più in profondità. Spooner ha una maggiore capacità polmonare e quindi maggiore autonomia di ossigeno, mentre la bambina va in iperventilazione, presa dal panico per aver assistito alla morte sul colpo del padre. Il robot, programmato per salvare in una situazione del genere la persona con maggiori probabilità di sopravvivenza, salva l'agente Spooner. Raccontando l'incidente e la dinamica di salvataggio alla robopsicologa Calvin, Spooner afferma che il robot ha fatto la scelta più logica e razionale, cioè salvare colui che presentava il 45% di *chances* di sopravvivenza, quindi lo stesso Spooner, lasciando andare colei che ne aveva appena l'11%, quindi Sarah. Ma poi aggiunge: “per chi amava quella bambina l'11% era più che sufficiente. Un umano l'avrebbe capito”.

Il regista Proyas, dunque, pur discostandosi dagli eventi narrati da Asimov, coglie però lo spirito del testo: il riconoscimento del dilemma etico come cifra stessa dell'umano. A questo punto, però, potrebbe ritornare utile il modello kantiano, ma non nel senso di un confronto tra la legge morale e le Tre Leggi della Robotica, bensì per una prospettiva antropologica strutturale che potrebbe aver ispirato il lavoro di Asimov. Nell'*Antropologia pragmatica*, Kant sostiene che un uomo “di giudizio”<sup>34</sup> è un uomo in grado di sapere quando applicare una regola e quando non applicarla. Se i robot di Asimov possono confondersi con gli esseri umani, perché possono imitare perfettamente il linguaggio umano, possono riprodurne perfettamente l'aspetto e le capacità sensomotorie, allora è forse nel giudizio (*Urteilskraft*) inteso come saggezza pratica (*phronesis*) da esercitare in situazioni dilemmatiche, che bisogna ricercare la specificità umana<sup>35</sup>. I robot asimoviani possiedono vista, udito, olfatto, gusto e tatto, in misura anche maggiore degli esseri umani, possiedono anche intelletto e intelligenza in misura maggiore

33 Sulla programmazione necessaria, in particolare per le auto a guida autonoma, in caso di dilemmi etici, si veda Tamburrini 2020.

34 Kant 2009, 85: “l'ufficiale a cui viene prescritta soltanto la regola generale per l'ufficio che gli è affidato, e a cui vien lasciato di stabilire da sé quel che all'occorrenza sia da fare, ha bisogno di Giudizio”.

35 Pisano 2021b.

rispetto agli esseri umani, ma non possiedono giudizio, non possiedono la capacità di scegliere quando rispettare una regola e quando non farlo, quando rispettare la programmazione ricevuta e quando non farlo. Al di là di Asimov negli ultimi anni si stanno aprendo le porte all'intelligenza artificiale, la quale ha fatto enormi progressi, ma non esiste ancora alcuna capacità di giudizio artificiale, né paiono esserci in tal senso sviluppi all'orizzonte. I processi di *machine learning* sembrano anch'essi, kantianamente, addestramenti dell'intelligenza, ma non del giudizio:

L'intelletto naturale [o artificiale] può arricchirsi di molte idee e munirsi di regole; ma la seconda facoltà intellettuale, quella per cui si decide se sia il caso della regola o no, cioè il Giudizio (*iudicium*), non può essere istruito, ma soltanto esercitato; quindi il suo sviluppo si chiama maturità, ed è tale che non viene prima del tempo. Ed è facile vedere che non può essere diversamente; poiché l'istruzione avviene soltanto con la comunicazione delle regole. Se dunque ci dovesse essere una istruzione per il Giudizio, allora ci dovrebbero essere delle regole generali, secondo cui si possa distinguere se qualche cosa rientra o non nelle regole: il che rimanda la questione all'infinito<sup>36</sup>.

Le macchine, i robot, le AI possono invecchiare, diventare obsolete, ma non possono maturare; possono accrescere le capacità di calcolo, ma non possono crescere in saggezza; si possono istruire con regole e con regole per l'applicazione delle regole, ma ci sarà sempre un livello di astrazione ulteriore non raggiungibile dalla programmazione. Una sorta di deficit fonetico<sup>37</sup> inteso come facoltà dell'infinito, oppure come capacità di “darci un taglio” quando ci si trova in un *loop* analogo a quelli in cui si ritrovano i robot di Asimov che, invece, rimangono incastrati in un girotondo perpetuo.

In una retroazione riflessiva che riconosce nella capacità di giudizio la differenza specifica dell'essere umano rispetto alla macchina, si potrebbe addirittura sostenere che l'individuo Adolf Eichmann – appartenente per costituzione biologica alla specie *Homo sapiens* –, nel momento in cui tenta di giustificarsi affermando di aver semplicemente “eseguito gli ordini”, dimostra una natura robotica priva di capacità di giudizio<sup>38</sup>. Ironia della sorte, Eichmann che sostiene di aderire a un'impostazione kantiana in quanto rispettoso della legge, è assolutamente antikantiano proprio perché incapace di comprendere quando rispettare la legge e quando non farlo, soprattutto considerando che la legge – razzista, ingiusta, ingiustificabile e non razionale – è per giunta proveniente dall'esterno, in un movimento eteronomo analogo a quello dei robot asimoviani. E ancor più ironicamente, il nazista Eichmann che considerava gli ebrei *Untermenschen*, si trova, limitandosi ad obbedire agli ordini senza alcun giudizio critico, egli stesso in una situazione robotica effettivamente *untermenschlich*: un automa privo di giudizio, mero esecutore materiale di ordini.

36 Kant 2009, 85-86.

37 Pisano 2021b.

38 Arendt 2013.

## Bibliografia

- Anders, Günther. 2003. *L'uomo è antiquato I. Considerazioni sull'anima nell'era della seconda rivoluzione industriale*. Tr. it. Laura Dallapiccola, Torino: Bollati Boringhieri.
- Anderson, Susan Leigh. 2008. "Three Laws of Robotics and Machine Metaethics", *AI&Soc*, n. 22, 477-493.
- Arendt, Hannah. 2013. *La banalità del male*. Tr. it. Piero Bernardini, Milano: Feltrinelli.
- . 2017. *Vita activa. La condizione umana*. Tr. it. Sergio Finzi, Milano: Bompiani.
- Asimov, Isaac. 1986. *I Robot e l'impero*. Tr. it. Piero Anselmi, Milano: Mondadori.
- . 2021. *Io, robot*. Tr. it. Vincenzo Latronico, Milano: Mondadori.
- Beauchamp, Tom Lamar, and Childress, James Franklin. 1979. *Principles of Biomedical Ethics*. Oxford: Oxford University Press.
- Bodei, Remo. 2019. *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*. Bologna: il Mulino.
- Chalmers, David. 1996. *The Conscious Mind. In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Farmer, Philip José. 2012. *Il grande disegno*. Tr. it. Roberta Rambelli, Roma: Fanucci.
- Floridi, Luciano. 2022. *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*. Milano: Raffaello Cortina.
- Fonnesu, Luca. 1998. *Dovere*. Venezia: La Nuova Italia.
- Freud, Sigmund. 1989. *Il perturbante*. Tr. it. Silvano Daniele, Torino: Bollati Boringhieri.
- Galilei, Galileo. 2014. *Dialogo sopra i due massimi sistemi del mondo*. Milano: Rizzoli.
- Gisondi, Marcello. 2013. "Ma gli androidi leggono Kant? Le leggi della robotica: un possibile percorso epistemologico dalla letteratura al diritto". *ISLL Papers. The Online Collection of the Italian Society for Law and Literature*, n. 4: 209-216. DOI – 10.6092/unibo/amsacta/5558.
- Jackson, Frank. 1986. "What Mary Didn't Know", *The Journal of Philosophy*, vol. 83, n. 5: 291-295. <https://doi.org/10.2307/2026143>.
- Kant, Immanuel. 2009. *Antropologia pragmatica*. Tr. it. Giovanni Vidari, riveduta da Augusto Guerra, Roma-Bari: Laterza.
- Lovelock, James. 2020. *Novacene. L'era dell'iperintelligenza*. Tr. it. Allegra Panini, Torino: Bollati Boringhieri.
- Murphy, Robin. 2009. "Beyond Asimov: The Three Laws of Responsible Robotics", *IEEE Intelligent Systems*, n. 24, 14-20.
- Pansera, Maria Teresa. 2019. *La specificità dell'umano. Percorsi di antropologia filosofica*. Roma: InSchibboleth.
- Pisano, Aldo. 2021<sup>a</sup>. "Il perturbante artificiale. Livelli di complessità e interfacce possibili fra umano e AI", *Kaiak. A Philosophical Journey*, n. 8. <https://philpapers.org/erc/PISIPA-2>.
- . 2021<sup>b</sup>. "La macchina e le forme dell'azione: deficit fonetico e autonomia artificiale", *Mechané*, n. 1, 37-53. <https://www.iris.unina.it/retrieve/e268a733-832a-4c8f-e053-1705fe0a812c/Mechane%201%20DEFINITIVO.pdf>.
- Putnam, Hilary. 1981. *Reason, Truth, and History*, Cambridge: Cambridge University Press, Cambridge.
- Russell, Bertrand. 1952. "Is There a God?", in *The Collected Papers of Bertrand Russell, vol. XI: Last Philosophical Testament 1943-1968*. ed. John G. Slater and Peter Köllner. London: Routledge 1997, pp. 543-548. [articolo commissionato, ma mai pubblicato per «Illustrated Magazine» e oggi disponibile su [https://www.cfpf.org.uk/articles/religion/br/br\\_god.html](https://www.cfpf.org.uk/articles/religion/br/br_god.html)].
- Scheler, Max. 1988. *Uomo e storia*, in *Lo spirito del capitalismo e altri saggi*. Tr. it. Roberto Racinaro, Napoli: Guida.
- Tamburrini, Guglielmo. 2020. *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*. Roma: Carocci.
- Thomson, Judith Jarvis. 1971. "A Defense of Abortion", *Philosophy & Public Affairs*, vol. 1, n. 1, pp. 47-66. Stable URL: <http://www.jstor.org/stable/2265091>.