

La robotica tra fantascienza e filosofia

Alberto Giovanni BIUSO *

La fantascienza affronta tematiche metafisiche che affondano nelle correnti più profonde del pensiero occidentale

M*atrix*, 1999. Il film dei fratelli Wachowsky ha ottenuto un successo dal significato complesso, perché si tratta di un'opera che si può leggere – come di fatto è stata letta – su una molteplicità di livelli, con strumenti ermeneutici assai differenziati, in riferimento a obiettivi e tesi fra di loro lontani, un coacervo che soddisfa tutti i gusti, dai videogames alla filosofia. Lo scenario di *Matrix* è certamente poco plausibile. Non solo non si comprende come le macchine siano arrivate a prendere il potere e a sottomettere gli umani, ma non si spiega neppure perché mai dei corpi in una bottiglia possano fornire così tanta energia da sostituire quella mancante del Sole,

oscurata dagli stessi umani nel disperato e suicida tentativo di privare le macchine della loro fonte energetica.

E poi, se le Intelligenze Artificiali arrivate al potere hanno bisogno solo di energia, perché non privano quei corpi umani della loro funzione cerebrale, in modo da evitare ogni rischio di ribellione? E ancora: l'energia necessaria a creare e a mantenere la Matrice sarebbe di gran lunga superiore a quella fornita da dei corpi sotto vetro. Sono queste delle contraddizioni logiche e insensatezze empiriche che indirizzano il film verso qualcosa d'altro da una fantascienza ipertecnologica che si ribella alla tecnologia, qualcosa d'altro da un'opera che utilizza tutti i più raffinati strumenti messi a disposizione dalla scienza del digitale. Da che cosa muove e verso che cosa va, quindi, *Matrix*? Il

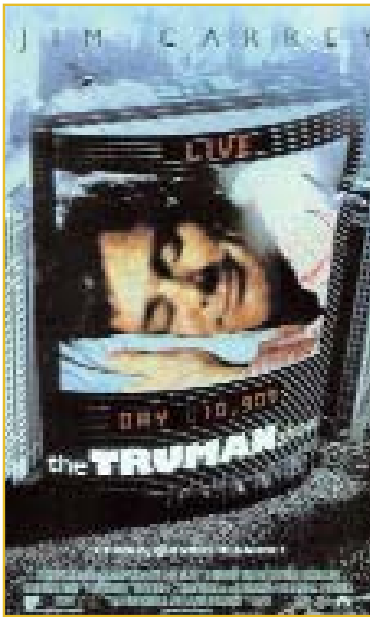
punto di partenza più remoto può essere, come spesso accade alla nostra cultura, ... Platone, che teorizzò in forme complete, profonde, plausibili, l'ipotesi che il mondo in cui ci muoviamo sia soltanto l'ombra e il riflesso di una realtà vera che ci sfugge. Dopo Platone, l'ipotesi è stata ripresa innumerevoli volte, da Descartes che immaginava un genio maligno (uguale, in pratica, alle I.A. di *Matrix*) volto a ingannarci sempre – persino sulla verità delle matematiche –, sino alla finta cittadina nella quale è rinchiuso l'ignaro protagonista del *Truman Show*.

Analogo orientale dell'ipotesi platonica è il celebre sogno di Chuang Tzu, che al monaco taoista fece sospettare di essere una farfalla che stesse sognando d'essere un uomo, piuttosto che un uomo che stesse sognando d'essere una farfalla. Dietro *Matrix* c'è il mondo dei fumetti giapponesi e americani, un puro esercizio di stile digitale che fa della Matrice un avanzatissimo videogioco, come diventa evidente fino alla nausea nei due successivi episodi dell'opera.

Anche tutto questo è *Matrix*; e tuttavia il film affronta, in modo certo scaltro e spettacolare, autentiche tematiche metafisiche e affonda in una delle correnti sotterranee più profonde del pensiero occidentale: la Gnosì. Al di là dello gnosticismo di Philip Dick che lo ispira; al di là degli evidenti simbolismi che fanno di *Morpheus* un Giovanni Battista, di *Trinity* una Maddalena, di *Cypher* un Giuda e di *Neo* un novello Messia; al di là dell'essere il Signor *Smith* e i suoi cloni delle malvagie divinità intermedie (gli Arconti della tradizione gnostica); al di là dell'itinerario nella Matrice presentato come un cammino dalle Tenebre dell'ignoranza alla Luce dell'autoconsapevolezza; al di là di tutto questo – che pure sostanzia il film – “il balzo concettuale gnostico dall'illusione alla comprensione è infatti fondamentale per l'intera struttura di *Matrix*”⁽¹⁾. Un balzo che è possibile compiere solo per umani non più soltanto umani, per dei Cyborg ormai ibridati con componenti digitali e avviati verso il post-umano mediante le nanotecnologie. Non può darsi, quindi, che le vere I.A. siamo noi?



* Università di Catania, agbiuso@unict.it
www.biuso.it



Matrix e la letteratura

Dietro *Matrix* c'è soprattutto la grande narrativa di fantascienza filosofica che ha in Dick il suo maggiore rappresentante; sta il cyberspazio inventato da William Gibson e da lui definito "un'allucinazione vissuta consensualmente ogni giorno"⁽²⁾. Quella di Gibson è una narrativa totalmente visionaria eppure a suo modo razionale, nella quale il passato della cibernetica, il presente delle Reti, il futuro delle Intelligenze Artificiali si mescolano in una danza macabra di sinapsi deliranti, morti congelate, esistenze in bilico fra immaginazione e realtà. Ma che cos'è effettivamente *reale*? Che cosa significa realtà, in un mondo fatto di dati, di bit, di guizzanti luci elettroniche dentro la grande Matrice del calcolo, del

tempo, dei corpi? L'antico sogno del patto coi demoni ritorna nelle forme di impalpabili ma potentissime Intelligenze Artificiali, i cui nomi – *Invernomuto* e *Negromante* – costituiscono da soli l'ammissione che, anche nel più astratto mondo della computazione, sono sempre le paure, i sogni, le memorie degli umani a formare, distruggere, ricreare gli strumenti del pensiero. Lo spessore filosofico e metafisico di Gibson si genera dal suo incrociare i temi della mente, del corpo, dell'Intelligenza Artificiale e della morte.

L'I.A. che si fa chiamare Negromante dice "Io sono i morti, e la loro terra"⁽³⁾ e quando *Case*, il protagonista di *Neuromancer*, porta a compimento la sua missione, che consiste nel fondere e distruggere le due supreme Intelligenze Artificiali facendole tornare alla Matrice originaria, le eleganti parole che scandiscono questa sofferta vittoria sono ancora una volta parole di morte: "Si muoveva al di là dell'ego, al di là della personalità, al di là della coscienza, e il Kuang si muoveva con lui, evitando i suoi aggressori con un'antica danza, la danza di Hideo, la grazia concessa dall'interfaccia mente-corpo, in quel secondo, dalla chiarezza e dalla precisa unicità del suo desiderio di morte"⁽⁴⁾.

Solo letteratura? In realtà la rivoluzione della fisica quantistica ha aperto ipotesi sulla materia e sul tempo rispetto alle quali le tradizionali – e troppo spesso disprezzate – speculazioni metafisiche rappresentano delle tesi persino moderate. Non soltanto, infatti, una metafisica *implicita* è presente in ogni ipotesi scientifica – teoria dei quanti compresa – ma sulla base delle equazioni quantistiche matematici e fisici discutono seriamente di possibilità quali "un'intelligenza cosmica così potente da essere in grado di resuscitare tutte le persone vissute in ogni epoca"⁽⁵⁾; o come

i viaggi nel tempo, sia verso il passato, sia verso il futuro (visto che "al livello atomico tutti i processi fisici sono completamente simmetrici rispetto al tempo" e quindi sono reversibili⁽⁶⁾); non escludendo neppure che *Matrix* esista davvero. La fisica dei quanti rende plausibile anche l'ipotesi che quello in cui siamo immersi sia uno dei tanti universi, che esista il *multiverso*, un insieme interminabile di universi paralleli. Uno dei modi per confermare queste e altre ipotesi è la costruzione di computer quantistici, che non solo avrebbero una potenza non paragonabile a quella di qualsiasi altro strumento mai ideato prima, ma che "potrebbero diventare più capaci degli esseri umani e a quel punto potrebbero iniziare a riprogettarsi da soli"⁽⁷⁾.

I Robot di Asimov

È quello che sembra accadere anche alle macchine pensanti immaginate da Isaac Asimov. In un mondo nel quale le Macchine e gli Umani convivono e lavorano in stretta collaborazione e *dipendenza*, i Robot di Asimov possiedono due importanti qualità di fondo: hanno la capacità di imparare da ciò che fanno, dai lavori per i quali sono impiegati, dal tempo che trascorrono, e sono onesti nei loro comportamenti. Nonostante il suo evidente entusiasmo nei confronti della tecnologia, dei Robot e dell'avvenire, Asimov sembra però condividere la critica all'identificazione dell'intelligenza con la semplice capacità di calcolo e persino con la struttura logica della mente.

La perfezione algoritmica delle macchine rappresenta, infatti, anche la loro debolezza. Una volta posto di fronte a un dilemma insolubile, a un circolo vizioso logico, il robot Herbie impazzisce e muore. Ciò significa che la *contraddizione* è invece una delle caratteristiche peculiari della specie umana, una delle vie d'uscita dalle situazioni apparentemente senza scampo, uno degli strumenti più potenti della nostra esistenza, una delle strade aperte al transito verso una maggiore perfezione. L'invenzione fondamentale di Asimov, la più interessante, quella che costituisce il vero nucleo generatore non solo dei suoi racconti ma anche di gran parte della fantascienza successiva, è costituita dalle *Tre Leggi della Robotica*. Diamole qui nella loro formulazione completa poiché esse delineano una vera e propria *etica delle Macchine*⁽⁸⁾:

1. nn robot non può recar danno a un essere umano, né permettere che, a causa della propria negligenza, un essere umano patisca danno;
2. nn robot deve sempre obbedire agli ordini degli esseri umani, a meno che contrastino con la Prima Legge;
3. nn robot deve proteggere la propria esistenza, purché questo non contrasti con la Prima o la Seconda Legge.

In queste Leggi sembra tutto molto chiaro, semplice, lineare... eppure da regole così coerenti

Note 1

⁽¹⁾ P. De Filippo, in AA. VV., *Visioni da Matrix. Tracce di un presente Cyber*, Sperling & Kupfer, Milano, 2003, pag. 42

⁽²⁾ W. Gibson, *Neuromante*, Mondadori, Milano, 2003, pag. 73

⁽³⁾ Ivi, pag. 304

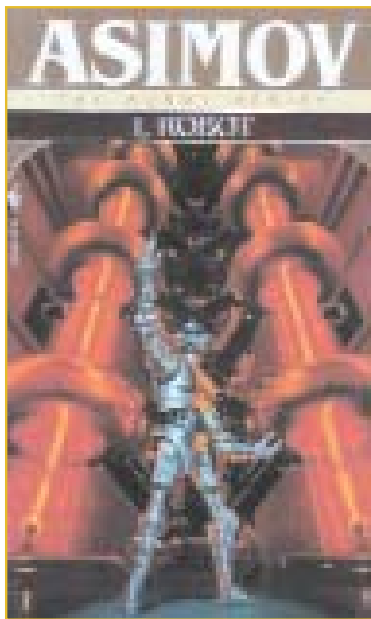
⁽⁴⁾ Ivi, pag. 324

⁽⁵⁾ J. Brown, *Menti, macchine e multiverso. Alla ricerca del computer quantistico*, Einaudi, Torino, 2003, pag. 371

⁽⁶⁾ Ivi, pag. 48

⁽⁷⁾ Ivi, pag. 331

⁽⁸⁾ I. Asimov, *Io, robot*, Mondadori, Milano, 2003, pag. 13



possono derivare conseguenze imprevedibili, atteggiamenti ambigui, pratiche contraddittorie. Chi, infatti, stabilisce che cosa sia un *danno*? La consapevolezza di che cosa siano il bene e il male è già qualcosa di profondamente umano e come possono, quindi, le Macchine farla propria? Il danno è riferito a ogni singolo essere umano o a un gruppo, a una comunità, alla specie? E se un singolo uomo recasse danno a tanti altri, il Robot che cosa dovrebbe fare, sarebbe autorizzato ad attaccarlo oppure dovrebbe lasciare che i tanti patiscano danno? Che cosa succede se un umano impartisce ordini contraddittori o se due umani prescrivono atteggiamenti opposti? Ingannare un uomo affinché non patisca danno è un comportamento coerente o no con la Prima Legge? Come escludere che le conoscenze cui i robot perverranno non faranno loro comprendere quale sia il bene degli esseri umani meglio di quanto lo comprendiamo noi stessi? E a quel punto dovranno ancora obbedirci – rischiando di infrangere la Prima Legge – o saranno autorizzati a prendere iniziative in forma autonoma – infrangendo la Seconda Legge?

E se, come accade a Hal 9000 – la vera *Mente* del viaggio descritto in *2001 Odissea nello spazio* – il Robot intuisse la propria centralità per portare a termine l'incarico affidatogli sulla Terra dai suoi costruttori, e quindi per compiere la propria missione si convincesse di dover eliminare gli umani? In altri termini, se il rispetto della Prima Legge comportasse la necessità di porre per un certo tempo in primo piano la Terza?

Non si pensi che tutto questo sia pura teoria o semplice fantascienza. Lo sviluppo delle *macchine per pensare* procede regolare e continuo e il salto qualitativo potrà avvenire in qualsiasi momento attraverso le prime forme di ibridazione fra uomo e computer. A quel punto, quando la dimensione post-umana sarà diventata realtà, ci sarà ancora differenza gerarchica fra la Prima e la Terza Legge? E quali saranno le possibili conseguenze di un simile cortocircuito etico? La robotica, allora, sarà strettamente legata all'antropologia e l'antropologia alla robotica. Conviene cominciare a pensarci... anche se all'interno del quadro filosofico e fantascientifico che stiamo cercando di delineare la questione dell'Intelligenza Artificiale mostra una significativa ambiguità.

Matrix e Blade Runner

Dick, ad esempio, ha creato alcuni dei romanzi e dei personaggi più efficaci nel descrivere le possibilità, le forme e i limiti di una intelligenza non biologica. I protagonisti di *Do androids dream of Electric Sheep?* – da cui Ridley Scott ha tratto il suo *Blade Runner* – sono dei replicanti del tutto indistinguibili dagli umani. La loro capacità di muoversi nello spazio, parlare, interagire e reagire è del tutto al di fuori di qualunque realizzabilità pratica. E ciò non solo per le carenze delle tecnologie attualmente disponibili ma per una ragione sostanziale: la mente non è una semplice facoltà sintattica, il pensare non è solo un calcolo, parte decisiva di ogni intelligenza è l'immersione nel continuum spazio-temporale che solo il corpo sa dare perché intriso di temporalità: memoria, attenzione, attesa.

E proprio perché ha intuito questa dimensione qualitativa dell'intelligenza Dick scrive che “forse siamo noi umani – teneri e buoni d'aspetto – con i nostri occhi penserosi, le vere macchine⁽⁹⁾”. Una reale I.A. deve, infatti, consistere nella percezione anche fisica del mondo attraverso il corpo, nella simbiosi fra la capacità di calcolo dei computer e quella emotiva dell'essere umano, in un incontro “a metà strada”, come si esprime Dick, fra “noi e i nostri computer in complicatissima evoluzione”. Molto per tempo e con grande acutezza, lo scrittore ha delineato la realtà già in qualche modo in atto dell'ibridazione: “una graduale fusione della natura generale delle attività e delle funzioni umane con le attività e le funzioni di ciò che noi umani abbiamo costruito e di cui ci siamo circondati⁽¹⁰⁾”.

Ma nel romanzo Dick sembra oltrepassare la grande tradizione gnostica rendendola ancora una volta uno sguardo della mente teso a comprendere qualcosa che c'è già nel nostro presente: l'ibridazione. L'ironia estrema del romanzo di Dick sta nel fatto che non c'è alcun criterio che consenta davvero di distinguere *chi* fra i vari personaggi sia un umano e *chi* un androide: “E così, la distinzione tra esseri umani autentici, vivi, e strutture umanoidi andava a farsi benedire⁽¹¹⁾”, come ammette il protagonista in un momento cruciale del romanzo.

E i *veri* androidi? Che cosa sentono, provano, vivono queste macchine quasi perfette, i robot così avanzati della serie Nexus-6? Essi sembrano rivelare la propria natura artificiale nella facilità con cui si fanno rintracciare e uccidere, nella *rassegnazione* con la quale cedono ancor prima di combattere, nella *consapevole tristezza* con la quale intuiscono che il loro ciclo vitale non va oltre i quattro anni, nella freddezza inquietante e rigida del loro essere, nonostante la forza del loro intelletto: “era come se una particolare e malevola *astrattezza* pervadesse tutti i loro processi mentali”, privandoli della “consapevolezza emotiva”, di quella “percezione sensi-



bile del vero significato. [...] Solo la vuota definizione formale e intellettuale dei singoli termini⁽¹²⁾. Intelligenze puramente sintattiche, prive della profondità semantica e incapaci, pertanto, di stare davvero al mondo. E, tuttavia, la splendida replicante che è Rachel afferma che “noi androidi non riusciamo a controllare le nostre passioni fisiche e sensuali⁽¹³⁾” e per definire i pensieri di queste entità abbiamo utilizzato parole come coscienza, tristezza, rassegnazione che sono termini, appunto, umani.

Il fatto è che questi androidi sono creature hegeliane, sono “il servo [...] divenuto più abile e sagace del padrone⁽¹⁴⁾”, sono l’altro dell’umanità in cui l’umanità si è trasformata, in una ibridazione estrema proprio perché indistinguibile, in

una contaminazione che per il solo fatto di essere in Dick tutta negativa non per questo risulta meno significativa del destino di complementarietà verso cui siamo – come specie e come singoli – avviati. La capacità di intuire ciò che sarà, di comprendere a fondo il proprio tempo e i suoi sviluppi, di pervenire al cuore delle cose, è la dimensione propria della filosofia. La *Science Fiction* offre, per Dick, la stessa forma di felicità.

Le macchine e l’identità umana

Il limite di fondo della cybercultura, e in generale della fantascienza letteraria e cinematografica, sta invece nella convinzione che l’*homo sapiens* possa trasformarsi in *homo cyber* abbandonando la corporeità. Ma noi non soltanto *abbiamo* un corpo, noi *siamo* corpo. Solo sulla base di una antropologia che riconosca la corporeità integrale dell’umano sarà poi possibile – e anche auspicabile – potenziare questo corpo attraverso l’implementazione in esso di elementi – *chip* o analoghi sistemi – in grado di incrementare le facoltà percettive e logiche che già possediamo.

Facoltà che non sono riproducibili, trasportabili o *uploadabili* su diversi supporti materiali, neppure se fosse tecnicamente possibile effettuare delle copie perfette dello sterminato numero di neuroni di cui ogni cervello è costituito. Non è, infatti, solo il cervello che produce la mente a pensare, è – probabilmente – l’intero corpo. Già nel 1991 il neurologo Richard Restak si chiedeva: “è possibile che la nostra definizione di cervello sia troppo limitata? Che i processi di regolazione che oggi consideriamo localizzati nelle nostre teste siano in realtà distribuiti in modo molto più ampio?⁽¹⁵⁾”. La ricerca – da Varela a Clark, da von Glasersfeld a Damasio, da Dreyfus alla *Teoria della Mente Allargata* di

Manzotti e Tagliasco – va rispondendo positivamente a questi interrogativi.

In realtà, la cybercultura è un fenomeno complesso, che non riguarda soltanto i rapporti tra mente e corpo, toccando anche aspetti religiosi, politici, letterari, di costume. La nuova narrativa fantascientifica ha svolto una funzione centrale nel coagulare intorno ad alcuni miti l’arcipelago di tendenze artistiche, religiose e filosofiche che intesono le prospettive più visionarie e radicali della cultura contemporanea. Incubo per alcuni o speranza mistica per altri, non saranno comunque le macchine a diventare intelligenti ed è invece assai più probabile che saranno gli umani a trasformarsi o, se si preferisce, a evolversi in intelligenze ancora più raffinate, potenti, sintetiche.

Poiché è vero quello che affermano McLuhan e Freud, quando il primo osserva che “sul piano fisiologico, l’uomo è perpetuamente modificato dall’uso normale della tecnologia (o del proprio corpo variamente esteso) e trova a sua volta modi sempre nuovi per modificarla⁽¹⁶⁾” e il secondo sa (e lo dice nel 1930) che “l’uomo è per così dire divenuto una specie di dio-protesi, veramente magnifico quando è equipaggiato di tutti i suoi organi accessori; questi, però, non formano un tutt’uno con lui e ogni tanto gli danno ancora del filo da torcere. [...] Le età future riservano nuovi e forse inimmaginabili passi avanti in questo campo che appartiene alla civiltà, e accresceranno ancora la somiglianza dell’uomo con Dio⁽¹⁷⁾”.

Autocostruzione, autoreplicazione, autoriferimento, sono i rischi più temuti. Ciò di cui si ha paura, in fondo, è che le macchine possano dare a se stesse, od ottenerla da noi, la dimensione specifica e caratterizzante l’umano: l’autocoscienza. Ma si tratta di timori e speranze forse entrambe illusorie, se è vero che la ragione più profonda e concreta del fallimento delle ottimistiche previsioni scaturite dalla conferenza di Dartmouth del 1956 che fondò l’*Artificial Intelligence* consiste nel fatto che manca ancora una teoria adeguata e plausibile dell’intelligenza e della coscienza.

Difficilmente l’Intelligenza Artificiale emergerà dall’accumulo di dati o dall’assemblaggio di circuiti o dalla miniaturizzazione spinta, ma – se accadrà – sarà assai più probabilmente il frutto della ibridazione fra la specie umana e le macchine, in cui saremo *noi* a rimanere la coscienza che nessun algoritmo è in grado di dare a se stesso. E questo termine – coscienza – va inteso proprio nella duplice accezione consentita, ad esempio, dalla lingua inglese: sia come *consciousness*, consapevolezza di sé e della soggettività che si è, sia come *conscience*, dimensione morale dell’agire. L’orizzonte della contaminazione, dell’alterità, dello scambio continuo con il diverso, è in fondo parte essenziale dell’orizzonte umano da sempre. Non c’è ragione perché con le macchine – anche se digitali – l’esito debba essere diverso. •

Note 2

⁽⁹⁾ P. K. Dick, *Mutazioni. Scritti inediti, filosofici, autobiografici e letterari*, Feltrinelli, Milano, 1997, pag. 268

⁽¹⁰⁾ Ivi, pagg. 227 e 224

⁽¹¹⁾ Id., *Ma gli androidi sognano pecore elettriche?*, Fanucci, Roma, 2003, pag. 164

⁽¹²⁾ Ivi, pagg. 179 e 213

⁽¹³⁾ Ivi, pag. 219

⁽¹⁴⁾ Ivi, pag. 54

⁽¹⁵⁾ M. Dery, *Velocità di fuga. Cyberculture a fine millennio*, Feltrinelli, Milano, 1997, pag. 346

⁽¹⁶⁾ M. McLuhan, *Gli strumenti del comunicare*, Il Saggiatore, Milano, 1997, pag. 56

⁽¹⁷⁾ S. Freud, *Il disagio della civiltà*, in “Opere”, vol. 10, Bollati Boringhieri, 1992, pag. 582