

ALCUNE OSSERVAZIONI SUI FONDAMENTI DELL'INTELLIGENZA ARTIFICIALE*

Settimo Termini

Abstract

In the present paper we shall discuss some foundational issues of Artificial Intelligence. In particular we shall relate the scientific aims of A.I. with those of Cognitive Science and of the original programme of Cybernetic. The role played, in this context, both by the theory of computability, seen as the general theoretical background, and of some new innovative notions will be also briefly discussed.

1. Introduzione

Il presente lavoro si prefigge lo scopo di individuare e discutere alcuni temi centrali dell'Intelligenza Artificiale ponendoli in relazione da un lato con il programma originario della Cibernetica e dall'altro lato con quello delle Scienze Cognitive. Le considerazioni che seguono sono del tutto preliminari e non presumono di essere in alcun modo né complete né, tantomeno, definitive limitandosi a svolgere il ruolo di semplici riflessioni epistemologiche. In quanto tali, tuttavia, esse ambirebbero a stimolare sia ulteriori indagini epistemologiche sia -soprattutto- la consapevolezza che in settori di indagine relativamente nuovi le chiarificazioni di tipo concettuale sono assolutamente necessarie per mettere a fuoco i problemi tecnici e formali al giusto livello di generalità.

Come ribadiremo anche successivamente, un possibile abbaglio che tuttavia è facile prendere nel seguire questa via è dato dal fatto di credere, o di illudersi, che i problemi studiati siano stati risolti una volta chiarite le «coordinate concettuali» dei problemi stessi.

Un'obiezione che viene comunemente fatta contro le ambizioni e le pretese dell'Intelligenza Artificiale è la seguente: in base a quali principi, fatti o evidenze

* Questo lavoro è stato svolto nell'ambito dell'accordo di collaborazioni scientifiche tra Italia e Spagna, all'interno del progetto di ricerca congiunto tra l'Istituto di Cibernetica del CNR ed il Centro de Estudios Avanzados de Blanes del CSIC.

Il lavoro è stato parzialmente sostenuto anche dal Progetto Finalizzato Robotica del CNR. L'autore desidera infine ringraziare l'Università di Santiago di Compostela ed in particolare il Dipartimento di Logica e Filosofia della Scienza per aver permesso di discutere alcuni dei temi qui trattati nell'Incontro tenutosi nello scorso anno.

sperimentali, possiamo o vogliamo sostenere che sia possibile ottenere un'intelligenza «non naturale»? Cioé, detto in altro modo, in base a cosa presumiamo che degli artefatti costruiti dall'uomo possano presentare realmente caratteristiche o sprazzi di ciò che noi riteniamo essere intelligenza?

A mio avviso è molto difficile controbattere questa obiezione se si rimane all'interno di una concezione dell'I.A. fondamentalmente centrata sui suoi risvolti tecnologici in quanto anche i suoi successi più spettacolari non sono tali -da un punto di vista qualitativo- se presi singolarmente da corroborare una tale pretesa.

La situazione, però, cambia completamente, anzi si rivolta, se guardiamo all'I.A. da una prospettiva scientifica. Se l'I.A. ambisce a costituirsi come Scienza dell'Intelligenza, cioè se si pone lo scopo di comprendere i caratteri essenziali dei fenomeni intelligenti e di costruire una loro teoria, allora passa ad avversari ed obiettori l'onere della prova che esista un'impossibilità teorica, di principio, alla realizzazione di un tale programma. Perché infatti i fenomeni descritti dalla nostra teoria non dovrebbero potere realizzarsi in nessun'altra struttura, in nessun altro supporto che quello biologico ben noto. In base a quale principio, cioè possiamo oggi supporre che i principi della nostra teoria non possano concretizzarsi, essere implementati, in nessun altro tipo di struttura, di materiale ma solo possano «incarnarsi» su un unico tipo di struttura biologica?

Non abbiamo nessun altro esempio di teoria scientifica che sia teoria di un *unicum* tranne forse la cosmologia, ma, in questo caso, forse più per mancanza di «altri universi», per la specificità cioè dell'oggetto studiato che per caratteristiche proprie della teoria.

Si potrebbe, è vero, obiettare che è proprio una teoria come quella cercata che non è possibile ottenere data la specificità del suo oggetto, ma per essere efficace quest'obiezione dovrebbe essere argomentata a partire da considerazioni non riguardanti, appunto, semplicemente l'oggetto della teoria, pena il rischio della circolarità.

Inizialmente avevo pensato di inserire nel titolo di questo paragrafo la frase «un punto di vista eretico», in quanto ho la sensazione che le argomentazioni portate possano essere considerate non ortodosse dai punti di vista standard dell'I.A. (nonché variante solo minimale di essa dai critici di tale punto di vista). Poiché però sarei molto felice se le considerazioni generali accennate in questo lavoro fossero considerate pienamente ortodosse e generalmente condivise ho deciso di eliminare tale aggiunta nel titolo del paragrafo. Vorrei, però, qui brevemente accennare a due punti di vista dai quali forse esse possono non esserlo. Entrambi hanno a che fare col ruolo centrale che ritengo debba giocare la teoria generale della calcolabilità nell'ambito di una Intelligenza Artificiale vista come Scienza (delle attività intelligenti). Una prima obiezione è quindi quella fatta dagli oppositori ad una impostazione algoritmica ai problemi della mente. Tuttavia nelle considerazioni da me proposte ho aggiunto che tale teoria deve tenere conto in modo essenziale, centrale, dall'interno, di nozioni di limitazione, finetezza, approssimazione. Che cosa sarà una tale teoria così arricchita? Solo una piccola variante della teoria generale che oggi conosciamo? (Tale previsione è tra l'altro corroborata dall'estrema robustezza ed invarianza della teoria classica così come oggi la conosciamo). O, viceversa, l'immissione dall'interno di nozioni innovative come quelle elencate prima può completamente trasformare la teoria così come oggi la conosciamo?

La posizione che desidero sostenere può essere così brevemente riassunta: l'I.A. può presentarsi come programma di ricerca scientifico e non solo come pura e semplice tecnologia, ingegnosa attività artigianale o soluzione/dissoluzione dei problemi della mente ad un livello puramente linguistico. La base per tale pretesa è fornita proprio dai grandi risultati di ordine generale delle teorie della calcolabilità e della complessità. I risultati di carattere limitativo non essendo altro che manifestazioni, indici, sintomi (o qual altro sinonimo voglia usarsi) del fatto che le teorie suddette hanno sì raggiunto un livello di notevole «profondità» e quindi colto alcuni processi essenziali dei processi studiati ma tuttavia sono ancora a grana troppo grossa per mettere a fuoco gli aspetti specifici dei sistemi intelligenti.

2. Una panoramica sui rapporti esistenti tra A.I., Cibernetica e Scienze Cognitive

Se conveniamo di vedere, come abbiamo fatto, l'I.A. come *scienza* dei sistemi intelligenti, allora per una valutazione complessiva del programma dobbiamo da un lato necessariamente rifarci al programma originario della Cibernetica e dall'altro esaminare i progetti e programmi delle Scienze Cognitive.

a. Rapporti con la Cibernetica

Nel volere stabilire una relazione tra il programma originario della Cibernetica anni cinquanta ed i programmi coevi o immediatamente successivi dell'Intelligenza Artificiale bisogna fare un'opera, diciamo così, di «ricostruzione razionale» (anche se il termine è un po' infelice dato che è stato usato a volte nel senso di operare forzature sullo svolgimento reale di alcuni «fatti storici»). Dichiaro subito che non ho alcuna intenzione di operare forzature e che il termine precedente è usato solo in mancanza di uno migliore.

Ciò che è necessario sottolineare invece è il fatto che la stretta relazione (e forse «filiazione») di cui ho parlato prima è percettibile solo se si guarda al fenomeno da una prospettiva abbastanza generale.

Il nucleo del programma originario della Cibernetica può individuarsi nel tentativo di costruire una teoria unificata in grado di incorporare nozioni nuove per l'epoca in oggetto, come Informazione, Complessità, Feedback, interdisciplinarietà, ed abbastanza generale da essere applicabile a tutti i sistemi in cui entravano tali concetti, indipendentemente dalla loro natura. In maniera equivalente si potrebbe identificare la Cibernetica con un programma che cerca di individuare un livello di descrizione «informazionale», in grado di rendere conto in maniera soddisfacente sia delle nozioni descritte prima sia di una trattazione unificata di sistemi di tipo diverso.

Possiamo dire che la Cibernetica nacque realmente come programma unificante di ricerche differenti e si può caratterizzare proprio come un tentativo di trattazione interdisciplinare di alcuni aspetti comuni a varie classi di sistemi complessi.

Alla base di tale tentativo stava la convinzione che un certo numero di meccanismi di autoregolazione che sono presenti in sistemi ed organismi di natura diversa devono essere guidati dalle medesime leggi indipendentemente dalla funzione specifica che essi svolgono nei differenti sistemi. Tale programma era destinato ad andare incontro al fallimento fondamentalmente per due motivi. In primo luogo, andando a studiare a fondo varie proprietà di tali fenomeni apparentemente simili si vide che il contesto nel quale si verificano (cioè il particolare tipo di sistema) condiziona le modalità di studio e di risoluzione dei problemi stessi. In secondo luogo, le tecniche formali proposte per l'unificazione cercata, in primo luogo la teoria del feedback e la teoria dell'Informazione di Shannon, non erano sufficientemente potenti dal punto di vista tecnico, da potere esse stesse mettere in evidenza le variazioni del comportamento specifico di tali nozioni al variare del contesto. Per dissipare possibili dubbi, dico subito che non voglio sostenere che i risultati ottenuti in un campo siano inessenziali per l'altro ma solo che la loro interazione può non essere maggiore di quella esistente tra altre discipline considerate come differenti. Quello che storicamente è accaduto è che dal tessuto comune si sono via separati settori autonomi di studio non appena essi sono riusciti a darsi una loro formalizzazione autonoma. Ad esempio, la teoria degli automi si è sviluppata a partire dal teorema di Kleene che a sua volta aveva cercato di caratterizzare/delimitare quello che potevano effettivamente fare le reti di McCulloch e Pitts, presentate dagli autori come modello logico matematico del cervello, e, quindi, nella linea di un risultato prettamente cibernetico.

Ciò che la Cibernetica non riuscì a produrre è una unificazione profonda come era nelle sue ambizioni e come è avvenuto altre volte nella storia della scienza. Ciò è un fallimento tanto più profondo in quanto tra gli scopi della Cibernetica vi era quello di presentare se stessa non solo come disciplina specifica ma anche come paradigma unificante. Per essere più precisi anzi come nuovo paradigma della razionalità scientifica che era anche schema unificante di altre discipline già esistenti. Si osservi però che ciò che è fallita è solo la pretesa di riuscire a realizzare il suo programma con gli strumenti tecnici proposti e nella prospettiva limitata inizialmente proposta. Ma l'enorme lavoro di dissodamento concettuale operato non poteva non continuare a dare i suoi frutti, cosa che continua ancora oggi anch'esse tali frutti li vediamo apparire spesso su rami molto lontani dal tronco originale.

La tesi che vogliamo sostenere in questo lavoro è che vi è una sorta di continuità ideale tra il programma originario della Cibernetica e l'Intelligenza Artificiale. La differenza essenziale tra i due programmi di ricerca consistendo a mio avviso nel fatto che mentre la prima non pone delle ipotesi esplicite sul tipo di tecniche utilizzabili per realizzare il programma, la seconda ne fa di fortissime e cioè che la soluzione al problema posto non possa che essere -se mi è permesso usare un linguaggio impreciso- la scoperta (potremmo dire l'«invenzione») di programmi intelligenti.

L'affermazione precedente potrebbe sembrare discutibile dal momento che proprio il programma originario della Cibernetica (e come ho io stesso sostenuto altrove) è fallito, e proprio per la non corrispondenza della potenzialità delle tecniche formali usate alle ambizioni del progetto stesso. Questo mostrerebbe che delle ipotesi sulle tecniche da usare erano state dopotutto poste. Di contro a ciò l'IA, nonostante nuove e vecchie critiche, può dopotutto vantare vari risultati ottenuti se non altro a livello tecnologico.

La differenza tra le due impostazioni consiste, a mio avviso, nel fatto che mentre la prima (la Cibernetica) fece implicitamente delle scelte sulle tecniche formali utilizzabili, ma non le fece mai assurgere ad opzioni ontologiche, la seconda (IA) -almeno nella sua vulgata ortodossa- ha considerato l'opzione «programmativa», come una sua caratteristica irrinunciabile (e in un certo senso definitoria).

Paradossalmente, in pratica, si è dimostrata molto più rigida la cibernetica dell'intelligenza artificiale, e infatti, di fronte al fallimento del programma originario, si è assistito o ad un ripiegamento-scomparsa del campo o ad una fuga in avanti, come quella rappresentata ad esempio da alcuni aspetti della teoria generale dei sistemi. Ma, e questo è ciò che mi preme sottolineare, si è sempre trattato di reazioni più o meno giustificate e razionali, ma dirette a salvare il programma di ricerca nonostante risultati di tipo limitativo. Cioè detto altrimenti, la tecniche utilizzate sono state ritenute le uniche possibili, se si vuole per limitazione culturale ma mai per scelta (ideo)-ontologica.

Nel caso dell'IA, invece, la scelta ontologica è chiarissima e fermissima, tanto da far pensare ad una «ipotesi algoritmica sulla mente umana-ed-artificiale»; ma nello stesso tempo si è assistito ad una estrema spregiudicatezza, nell'introdurre nuovi concetti e categorie (si pensi ai *frames* o anche ai *micro worlds*) tutte le volte che questo è servito a superare ostacoli notevoli nella realizzazione, sia pure parziale, degli scopi che ci si era prefissi.

In breve potremmo dire che mentre nel caso della Cibernetica, c'è stato un atteggiamento fortemente non ideologico, in linea di principio, che si è scontrato, in pratica, con scelte ideologiche o, quantomeno, metodologiche rigidissime, nel caso dell'intelligenza artificiale c'è stata un'opzione ontologica rigidissima che è stata temperata in pratica da scelte che di volta in volta apparivano le più convenienti per la realizzazione del programma.

Vorrei sottolineare che ritengo questo tipo di riflessioni rilevanti da un punto di vista scientifico, e non solo da quello della storia della scienza in quanto le opzioni epistemologiche giocano sempre un ruolo nell'indirizzare l'attività scientifica *tout court* in una direzione o in un'altra ed inoltre questo ruolo è tanto maggiore quanto minore è la coscienza esplicita della loro esistenza; e questo tanto più oggi in quanto si assiste da un lato ad una tecnicizzazione completa e teorizzata dell'IA, via Sistemi Esperti, e dall'altro ad un rilancio di versioni dell'IA, rimodellati sulla Cibernetica.

b. Rapporti con le Scienze Cognitive

Per quanto riguarda il rapporto con le Scienze Cognitive ci limitiamo a fare riferimento ai volumi di Johnson Laird [2] e di Longuet-Higgins [1] e Gardner [4] per notare quanto le tematiche tra Intelligenza Artificiale, Cibernetica e Scienze Cognitive siano tra loro strettamente collegate. Sarebbe interessante un'indagine accurata che metta in evidenza la sostanziale comune unità di intenti e contemporaneamente le notevoli differenze programmatiche delle impostazioni di questi ed altri autori, fra cui sono certamente da ricordare quella delineata da M. Minsky nel suo recente volume e quella che è in un certo senso la ripresentazione aggiornata del programma originario

della Cibernetica, cioè il programma connessionista. Questo tuttavia esula dagli scopi che si riprometteva il presente lavoro e non è consentito dallo spazio a disposizione. Ci limiteremo quindi a presentare solo dei brevi commenti. In primo luogo notiamo che il termine «Scienze Cognitive» è stato usato per la prima volta da Longuet-Higgins in una sua breve nota pubblicata come commento al famoso (e famigerato) rapporto Lighthill e che ora è contenuto nel volume del medesimo autore recentemente pubblicato [1].

In questo breve commento che si proponeva assieme ad altri di raddrizzare le conclusioni di Lighthill sull'Intelligenza Artificiale, Longuet-Higgins enuncia e sostiene alcuni punti che qui brevemente richiameremo.

Longuet-Higgins sostiene che il lavoro più rilevante dell'Intelligenza Artificiale, almeno per un certo tempo sarà proprio quello di cercare di esprimere sotto forma di programmi per calcolatore teorie astratte di varie facoltà cognitive degli esseri umani più che cercare modelli matematici del cervello.

Questa sua prima affermazione mostra che Longuet-Higgins condivide il contributo dato dall'Intelligenza Artificiale al superamento di una visione complessiva della vecchia cibernetica diventata ormai paralizzante, una volta mostrata l'incapacità del programma originario a risolvere ad un livello unitario i problemi che esso stesso si era posto, a causa dell'inadeguatezza degli strumenti tecnici che aveva a disposizione. Valore scientifico dunque e riproposizione scientifica degli obiettivi del programma tanto più utile in quanto essa veniva a soppiantare una visione puramente behavioristica in psicologia ed anch'essa insoddisfacente.

Longuet-Higgins continua dicendo che: «in short, whatever the technological prospects of Artificial Intelligence, its principal scientific value, in my view, is that *it sets new standards of precision and detail in the formulation of models of cognitive processes, these models being open to direct and immediate test*» (corsivo suo) ed infine conclude affermando che l'Intelligenza Artificiale può contribuire ad arricchire tutte quelle scienze che sono direttamente rilevanti per il pensiero e la percezione umani.

Dopo avere infine individuato quattro raggruppamenti principali nei quali possono essere suddivise le attività rilevanti per scienze cognitive e cioè: matematico, linguistico, psicologico e fisiologico, Longuet-Higgins conclude dicendo che forse l'uso del termine scienza cognitiva (al singolare) è preferibile a quello al plurale perchè i settori descritti sono in realtà strettamente correlati e che proprio gli studi di I.A. stanno mostrando interessanti correlazioni logiche emergenti tra vari modelli di esperienza.

Dunque proprio in quello che segna l'atto di nascita del termine «scienza cognitiva» se non proprio dell'attività scientifica stessa, vi è il riconoscimento dell'importanza dell'I.A. come generatrice di interessanti ipotesi scientifiche -e non solo di interessanti soluzioni e sviluppi tecnologici. Nello stesso tempo queste vengono individuate e nella direzione del completo recupero delle ambizioni del programma originario della Cibernetica -anche se l'autore non ne parla mai- e della consapevolezza che queste erano state sussunte dall'I.A., vista come generatrice di ipotesi scientifiche e di meccanismi di verifica empirico-sperimentali. Rimandando ad altra sede una più accurata indagine in questo senso completiamo questa breve carrellata sulle scienze cognitive con un breve riferimento ad una tesi di Johnson-Laird. Secondo questo autore, le nozioni di autoriferimento sarebbero sì molto importanti ma non risolutive per la

definizione e costruzione di macchine e sistemi intelligenti, in quanto al più avremmo dei sistemi capaci di autodescriversi ma non introspettivi in senso stretto. Altri autori, come ad esempio Hofstadter, vedono proprio nella capacità di autoriferimento il passo formale essenziale per costituire la nozione di un Sé. La coscienza, invece secondo Johnson Laird, sarebbe da attribuire ad una particolare classe di algoritmi, e sarebbe caratterizzata più dal come fanno certe cose che dalle cose che essi calcolano. Una posizione che se verificata metterebbe in scacco la possibilità di determinare, a partire dal comportamento, il possesso o meno di concetti significativi per l'Intelligenza Artificiale¹. Rimandiamo al libro *Mental Models* ed in particolare al capitolo conclusivo intitolato «Coscienza e Computazione» per una conoscenza delle posizioni di questo autore. Qui ci si limita ad osservare che le considerazioni conclusive di Johnson Laird non si discostano poi molto dall'impostazione che si vuole sostenere ed in ogni caso sono compatibili con essa. Esse sono sì molto più generali e tendono a fornire un quadro di riferimento generale. In questo senso egli è portato ed obbligato ad usare il termine «coscienza». Ma quando egli scrive che «ovviamente alcuni eventuali aspetti della mente, come spiritualità, moralità e capacità immaginativa, sono tali da sfuggire alla possibilità stessa di una loro implementazione in programmi per computer» dice qualcosa che è molto vicino ad una impostazione prudentiale, secondo la quale non si vogliono esplicitamente affrontare problemi relativi al Sé o alla coscienza a questo stadio di sviluppo della teoria. A mio avviso si può difendere la tesi che è possibile costruire una teoria dei processi mentali e che i «fenomeni intelligenti» posseggono una loro autonomia indipendente dal sé e dalla coscienza, intesi nell'accezione più generale.

Johnson-Laird giustamente sostiene che «qualsiasi teoria scientifica della mente deve per forza considerarla come un automa», ma è proprio per questo che sono, allo stadio attuale di sviluppo, personalmente riluttante a presentare come punto centrale le nozioni di riflessione ed autoriferimento in relazione ad una possibile teoria che getti luce anche su questioni relative alla coscienza, per le implicazioni antropomorfe che esse inevitabilmente inducono. Sono profondamente convinto che esse siano nozioni centrali e rilevanti ma soprattutto per il fatto che forniscono uno (e forse lo) strumento tecnico fondamentale per affrontare il problema del finitismo, dell'approssimazione e delle strategie «tecniche» che si devono mettere in atto per superare tale limitazione delle risorse.

Una teoria della mente e dei processi mentali è molto probabile che sarà una teoria algoritmica, computazionale o non sarà, ma poichè i fenomeni di finitezza ed appros-

¹ Vorrei ancora osservare che l'attributo che viene immediatamente individuato come significativo per la discussione da Johnson-Laird è quello di «coscienza». Probabilmente anche Hofstadter sarebbe favorevole ad usare tale termine in relazione a questioni di autoriferimento, anche se assieme ad altri termini. Dal punto di vista più asettico mantenuto nelle osservazioni fatte nel presente lavoro si è preferito evitare di assumere posizione su tali concetti. In realtà la posizione che personalmente preferisco sostenere è di distinguere nettamente fra intelligenza e coscienza. Le attività intelligenti hanno una loro autonomia e possiamo percorrere un lungo tratto di strada esaminando problemi riguardanti la costruzione di una buona teoria delle attività (e dei sistemi) intelligenti senza dovere affrontare problemi connessi al Sé, o alla coscienza.

Lascio anche questo problema, come altri, solo accennato come pura indicazione del tipo di problemi che un'analisi comparativa delle impostazioni presenti in IA e nelle Scienze Cognitive potrebbe, e dovrebbe affrontare.

simulazione giocano all'interno di ed insieme ai meccanismi computazionali, un ruolo centrale, a livello di un'analisi dei fondamenti dell'I.A., dobbiamo in primo luogo individuare ed analizzare strategie per rappresentare e descrivere tali processi all'interno stesso della Teoria della Calcolabilità che rimane lo schema teorico e lo sfondo di riferimento di ogni tentativo teorico generale di comprensione dei processi mentali.

3. Computabilità e Complessità (ed oltre)

Se vogliamo che l'Intelligenza Artificiale vista come Scienza ed intesa in particolare come Scienza dei sistemi intelligenti, non corra il rischio di incappare nello stesso pericolo corso dal programma originario della Cibernetica, dobbiamo allora stabilire qual è lo sfondo teorico -solido- di riferimento, individuando nello stesso tempo quali sono le sue proposte e richieste innovative a livello concettuale in modo da individuare quali possano essere le tecniche formali-matematiche necessarie a tradurre in «teoria», ad incarnare scientificamente le richieste concettuali/epistemologiche e, nel caso in cui queste tecniche, come è possibile², non siano già disponibili sul mercato, a forgiare nuovi strumenti formali adatti allo scopo. E' accettato da una parte rilevante della comunità scientifica dell'IA, ma non da tutta, che un utile punto di partenza è l'assunzione dell'algorithmicità dei processi mentali o, detto in altromodo, l'ipotesi computazionale sulla mente.

Tale ipotesi può venir fatta a vari livelli di compromissione; quella di livello più basso si limita a riconoscere l'utilità di poter fare delle simulazioni algoritmiche mediante programmi concreti che girano su macchine reali di alcuni processi mentali senza chiedersi se vi sia una identità tra la simulazione e ciò che viene simulato. Lo scopo è semplicemente quello di avere un utile strumento per sperimentare e vagliare i risultati di alcune ipotesi fatte su alcuni processi.

All'altro estremo troviamo la posizione di chi sostiene che esiste una reale indentità tra i processi mentali ed una rappresentazione algoritmica. Spingendo avanti questa posizione possiamo dunque dire che una teoria dei processi mentali non sarà altro che un capitolo -sia pure estremamente sofisticato- di una teoria generale della calcolabilità, quindi, se accettiamo la tesi di Church-Turing, della Teoria delle Funzioni Ricorsive.

Oltre ad obiezioni di tipo filosofico concettuale l'obiezione teorica, avanzata tra gli altri da Michael Rabin, che a questa pretesa viene fatta riguarda essenzialmente problemi di complessità. Gli algoritmi propri dell'I.A. sono solitamente algoritmi non «trattabili» e quindi -si presume- le vie forgiate dal cervello saranno diverse da quelle proposte da noi.

Ricordiamo che tra le caratteristiche essenziali della Teoria della Calcolabilità emergono da un lato l'esistenza di algoritmi universali e dall'altro risultati di indicibi-

² Se guardiamo alla storia della scienza questo è già avvenuto altre volte. Euclide ha forgiato la teoria delle proporzioni per trattare e risolvere lo «scandalo pitagoreo» dell'esistenza degli irrazionali. Newton inventò la teoria delle flussioni per dare corpo matematico alle sue intuizioni sulla gravitazione. Einstein aspettò di essere a conoscenza del calcolo di Ricci per poter tradurre in teoria le sue intuizioni sull'equivalenza tra inerzia e gravitazione.

lità come, ad esempio, il teorema della fermata. Ricordiamo ancora che mentre la nozione di calcolabilità ha mostrato di essere una nozione molto robusta, invariante al variare di tutte le formalizzazioni adeguate che sono state proposte, la nozione intuitiva di complessità ha trovato finora varie, utilissime e profonde formalizzazioni, ciascuna delle quali, però carpisce solo alcuni aspetti della nozione intuitiva.

Cerchiamo adesso di esaminare alcuni aspetti del significato epistemologico dell'esistenza di un programma universale in relazione alle nozioni *informali* di calcolabilità e complessità.

L'esistenza di un tale programma ci dice che esiste un livello critico di complessità (informale: cioè quantità di materiale che possiamo elaborare) tale che al di qua di questa soglia non riusciamo a raggiungere un livello generale mentre al di là di questa soglia non esistono ulteriori salti qualitativi: non si può fare più di prima se non in senso puramente quantitativo.

L'universalità sembra quindi nascere in modo spontaneo, è quasi una proprietà emergente al crescere della complessità (opportuna strutturata).

D'altro canto anche le proprietà di tipo limitativo sono emergenti al crescere della complessità. Sembrerebbe, quindi che al crescere della complessità emergano spontaneamente sia proprietà positive che negative.

Ma anche l'intelligenza e la mente sembrerebbero essere proprietà emergenti al crescere della complessità. Vi è un rapporto sia pure puramente epistemologico tra queste cose? Si tenga inoltre ancora conto dell'ipotesi di von Neumann sui sistemi complessi.

In questo quadro quindi ritengo che la teoria della calcolabilità non possa non essere lo sfondo teorico di riferimento (i prolegomena scientifici) di una IA che voglia presentarsi come scienza. Nello stesso tempo non si può non rilevare che l'attuale teoria della calcolabilità è sì uno sfondo, ma in realtà esso è troppo remoto ed a trama troppo poco fine per riuscire a mettere in evidenza quegli aspetti che sono i più rilevanti per i fondamenti dell'IA stessa. In questa sede non possiamo fare altro che indicare, molto rapidamente, quali a nostro avviso possano essere le direzioni nelle quali cercare un utile sviluppo.

Tutta una serie di nozioni innovative come approssimazione e vaghezza (viste in senso intrinseco, alla von Neumann), complessità, autoriferimento non sono nè immerse fin dall'inizio del quadro generale della teoria della calcolabilità nè emergono come conseguenza di opportune caratterizzazioni e specificazioni della teoria stessa ma in generale³ sono, per così dire, sovrapposte o aggiunte alla teoria base. Tutte queste nozioni, tra l'altro, sono fondamentali per una fondazione scientifica dell'I.A. perchè, a mio avviso, sono vari modi in cui emerge, si manifesta, il «finitismo» basilico della mente umana e le varie strategie che -evoluzionisticamente- essa ha messo in atto per superare tale finitezza, in modo analogo a quanto facciamo noi quando nel programmare un calcolatore conoscendo la limitatezza delle sue risorse, operiamo in modo che

³ La complessità assiomatica di Manuel Blum e la definizione di complessità di Kolmogorov-Chaitin soddisfano alle nostre richieste, ma come abbiamo già osservato, la nozione intuitiva di complessità non viene completamente esaurita da nessuna delle teorie proposte, quindi a parte l'analisi di alcuni problemi posti anche dalle teorie citate prima, la nostra obiezione rimane valida anche nel caso della complessità, se vista nella sua generalità.

si comporti come una macchina di Turing e non come un automa finito. Se uno sviluppo nella direzione proposta rafforzerà il ruolo centrale della nozione di algoritmo nell'I.A. o, al contrario, metterà in evidenza nuove e più profonde basi teoriche di riferimento lo potremo stabilire solo a posteriori⁴.

6. Conclusioni

Le osservazioni fatte nel corso di questa rapida carrellata mostrano che nel campo di indagine velocemente delineato c'è ancora bisogno di un enorme lavoro di dissodamento concettuale e filosofico. E' inutile ed a volte può risultare anche dannoso, procedere ad un semplice «rivestimento» formale di un ossatura ancora concettualmente immatura e non consolidata; tanto più se questa -come è avvenuto per alcuni aspetti e settori della Teoria Generale dei Sistemi- si limita ad essere un linguaggio descrittivo è non *teoria* che forza alcuni risultati che saranno poi confrontabili con «l'evidenza empirica». La posta in gioco dunque è concettualmente troppo sofisticata per essere lasciata in mano ai soli tecnici senza un'opportuna guida concettuale ed unitaria. Ma nello stesso tempo si deve prendere coscienza che tale posta non può essere vinta se non viene data una risposta ad una serie di problemi che non possono essere risolti ad un livello puramente concettuale. Potremmo anzi prendere come misura della *non* soluzione reale quelle proposte che si limitano ad un uso normale o banale di tecniche tradizionali (Vedi ad esempio i commenti di Berlinski [5] su alcuni aspetti della Teoria Generale dei Sistemi e quello di Haugeland [8] sulla nozione di feedback in Cibernetica). In realtà come minimo dovremmo avere o la presentazione di tecniche realmente innovative (anche se usate ad un livello non estremamente sofisticato) o un uso estremamente sofisticato e profondo di tecniche tradizionali. Personalmente sarei indotto a credere che entrambi gli aspetti dovranno essere presenti in una qualsiasi nuova teoria emergente di tali fenomeni nella loro complessità se dovranno essere presenti in una soluzione reale. La posta in gioco è, quindi, tecnicamente troppo intricata e complessa per essere lasciata in mano solo agli epistemologi.

Ringraziamenti

Desidero ringraziare Juan Vázquez ed Alejandro Sobrino per avermi costretto a chiarirmi le idee sugli aspetti concettuali dell'Intelligenza Artificiale nel corso di varie ed interessanti discussioni.

⁴ Oltre a non avere approfondito i problemi posti da nozioni innovative come approssimazione, complessità, vaghezza, finitismo, non abbiamo neanche accennato alle critiche all'impostazione ortodossa dell'I.A. avanzate da Winograd e Flores, ai risultati di M. Pour-El su calcolabilità/non calcolabilità, alle considerazioni di G. Kreisel sulla tesi di Church, alla rilevanza del costruttivismo e delle impostazioni ultrafinitiste in matematica per i nostri temi. Ci si riserva di illustrare questi problemi, nel contesto qui delineato, in un successivo lavoro.

Bibliografia

- [1] H.C. Longuet-Higgins, *Mental Processes. Studies in Cognitive Science*, M.I.T. Press (1987).
- [2] P.N. Johnson-Laird, *Modelli Mentali*, Il Mulino (1989), ed. originale: *Mental Models*, Cambridge University Press (1983).
- [3] T. Winograd, F. Flores, *Understanding Computers and Cognition*, Ablex Publ. Co. (1986).
- [4] H. Gardner, *The Mind's New Science. A History of the Cognitive Revolution*, Basic Books (1987).
- [5] D. Berlinsky, *On System Analysis*, MIT Press, 1976.
- [6] M. Minsky, *La Società della Mente*, Adelphi, 1989.
- [7] G. Tamburrini, *Reflections on Mechanism*, Ph.D. Thesis, Columbia University, New York, 1987.
- [8] J. Haugeland, *Artificial Intelligence: the very idea*. MIT Press (1987).

Settimo TERMINI
Dipartimento di Matematica
Università di Perugia, Italia