

Intelligenza artificiale e algoritmi: datificazione, politica, epistemologia

Teresa Numerico

Abstract: The aim of the present paper is to show the evolution of the concept of Artificial Intelligence (AI) and of the different technical methods that progressively informed and organized this concept. The article presents a view of the evolution of such a notion: 1) with special regards to the social, political and epistemological consequences of the chosen technical solutions, 2) with special attention to the parallel transformation of the human intelligence concept. The recent major successes of AI are based on datification and availability of huge quantity of information relative to traces left behind by people online behaviours. Big Data methods together with machine learning algorithms have the purposes to interpret data and create pattern recognition methods that discover correlations between data series. Algorithms exploit such correlations, which are not precisely causation categories, in order to produce anticipations of future behaviours, inferring regularities and measuring probabilities grounded on past actions. Moreover algorithms work on the clusterization of people according to their activities and other personal characteristics, such as where they live, who their friends are, etc. The implicit foundation of data science is the induction principle, which 'guarantees' that the past will be similar to the future and that people that share some peculiarities tend to behave similarly in corresponding situations. It is an interpretative data organization, which is obtained via the datification of online traces and the implementation of adequate machine learning algorithms. The datification itself implies that data is cleaned and arranged in a form that the program can understand. The pretence of neutrality of such complex procedures blurs the activity of interpretation that is implicitly embedded in the system, by giving the allure of neutrality of measuring methods. The radical success of Big Data and machine learning algorithms invites to assign decisions responsibility to machines because they are the unique agents capable of managing the huge quantity of available data. It is more and more difficult to control the output of complex technical systems even when the results of the procedures impact human beings' lives. As Norbert Wiener already suggested, technical systems could exclude humans from feedback loops because they are too slow to catch up with the rhythm of the technical decision process. This is the first issue under discussion in the present paper. The second issue is that the machine, as Turing underlined, must only pretend to be intelligent enough to be able to take in inexperienced judges. If it is not possible to control the actions of the devices because they are too fast and complex to be explicitly understood – and the system is programmed to take in humans – how can we trust machines? The third issue regards technology as a socio-technical system that, differently from

Università degli Studi Roma Tre (teresa.numerico@uniroma3.it)

science does not aim at understanding the external world, but it is a medium, a representation and an intervention that orientates the world according to social and political criteria. It is necessary to ask who is in charge of the governance of such a system and which are the objectives of such a transformation. It is crucial then to delineate rules, powers and intentions that underline the design of the sociotechnical systems in order to choose democratically which of the methods are more favourable to the whole society.

Keywords: Artificial Intelligence; Algorithms; Data science; Big Data; Technology as politics; Turing Test; Cybernetics.

1. Introduzione

L'obiettivo del presente articolo è mostrare l'evoluzione del concetto di Intelligenza Artificiale (IA) e i diversi contenuti tecnici che progressivamente hanno informato di sé questo mutevole concetto. L'articolo presenta una panoramica dell'evoluzione di tale concetto: 1) in relazione alle conseguenze sociali, politiche ed epistemologiche delle scelte tecniche, 2) in relazione alla parallela trasformazione del concetto di intelligenza umana.

La prima versione dell'idea di intelligenza da attribuire alla macchina si proponeva di costruire uno strumento in grado di fingere di saper maneggiare il linguaggio, obiettivo questo esplicitamente assunto da Alan Turing nella realizzazione del suo famoso Test, o gioco dell'imitazione. Dall'attenzione di Turing per il linguaggio e la sua difficile maneggiabilità da parte della macchina si passa a una visione riduzionista dell'IA secondo cui le macchine sarebbero in grado di sostituire la mente attraverso un sistema fisico-simbolico di impostazione logicista. Contemporaneamente allo sviluppo del paradigma dell'IA come simulazione della mente si sviluppa la cibernetica, una disciplina transdisciplinare il cui obiettivo è abbattere la differenza tra organico e inorganico attraverso l'analisi dell'attività degli agenti, a partire dalle loro prerogative di comunicazione e controllo. Il paradigma di integrazione tra vivente e macchina conduce a immaginare l'intelligenza meccanica come l'esito di un processo di integrazione con le capacità umane, attraverso strategie di comunicazione e controllo.

La prospettiva dell'interazione uomo-macchina si è evoluta negli ultimi anni in una progressiva datificazione dell'azione umana, possibile grazie alla diffusione della rete Internet che ha consentito una registrazione di tutte le azioni compiute nell'ambito della comunicazione digitale. L'Intelligenza Artificiale si è trasformata nella costruzione di algoritmi in grado di prevedere, anticipare e orientare il comportamento umano, dopo averlo discretizzato in un flusso di dati misurabili. Non si tratta più di una si-

mulazione dell'attività umana, ma in una sua riproduzione dopo averla datificata.

Il computer, infatti, a queste condizioni, svolge il ruolo per il quale è più adatto: quello di una macchina per eseguire calcoli. Dopo aver cercato, quindi, invano di esorcizzare il corpo chiudendolo tra parentesi nell'ipotesi di riuscire a costruire sistemi simbolici del tutto privi di riferimento fisico, attraverso la mediazione della cibernetica, l'Intelligenza Artificiale ha riorientato la sua organizzazione nel senso di un'interazione comunicativa tra esseri umani e macchine. Tale interazione in un primo tempo si è manifestata come integrazione in cui la macchina era sotto il controllo dell'operatore umano e successivamente, con la datificazione dell'agire umano, ha sussunto sotto di sé la capacità di prendere decisioni, sia pure dipendendo dall'operatore umano per la costruzione degli algoritmi usati per interpretare i dati. Tale evoluzione propone una serie di questioni di natura epistemologica, etica, sociale e politica che devono essere analizzate accuratamente.

La prima questione riguarda la presa di decisione. Non è facile, infatti, sottoporre al controllo di operatori umani le decisioni algoritmiche. Come suggeriva preoccupato Wiener, Gli esseri umani rischiano di essere lasciati fuori dal *feedback* della presa di decisione perché troppo lenti.

La seconda questione è che la macchina, come segnalava Turing, deve solo fingere di essere intelligente ed essere capace di ingannare giudici non esperti di tecnologia. Ma se non si può controllare quello che il dispositivo fa – e il meccanismo è programmato per ingannare gli umani – come possiamo fidarci delle sue decisioni?

La terza questione riguarda la tecnologia come sistema sociotecnico che diversamente dalla scienza non ha come obiettivo la conoscenza di come funzioni il mondo esterno, ma è piuttosto un metodo di organizzazione del mondo esterno, una mediazione e un intervento sulla realtà per orientarla secondo precisi criteri politici e sociali. È necessario perciò chiedersi chi governa questo sistema e quali siano i suoi obiettivi. Si tratta cioè di definire le regole di una condivisione di spazi e obiettivi tra esseri viventi, in particolare esseri umani e dispositivi tecnici.

2. Intelligenza meccanica, intelligenza artificiale

Alan Turing – l'inventore del modello teorico della macchina – viene considerato anche il padre del progetto di Intelligenza Artificiale, cioè di quella disciplina che si occupa di costruire macchine in grado di svolgere compiti

considerati intelligenti nel caso fossero svolti da esseri umani. Il progetto di Turing era fin dalla seconda metà degli anni Quaranta, costruire un computer per sperimentare la possibilità di ottenere macchine intelligenti. In una lettera scritta tra il 1946 e il 1947 a Werner Ross Ashby, un cibernetico inglese con il quale era in corrispondenza, Turing dichiarava esplicitamente il suo interesse per l'intelligenza meccanica: «I am more interested in the possibility of producing models of the action of the brain than in the practical applications to computing»¹. Questa lettera insieme con altre evidenze testuali², mostrano che Turing fosse interessato non tanto a una macchina per eseguire calcoli, quanto a una macchina per svolgere compiti intelligenti.

La sua attenzione anticipava anche la nascita ufficiale della disciplina che viene fatta risalire convenzionalmente a una *Summer School* organizzata al Dartmouth college nel 1956³. Nel progetto della *Summer School* – firmato nel 1955 da scienziati che divennero alcuni tra i più influenti scienziati dell'intelligenza artificiale quali John McCarthy, Claude Shannon, Marvin Minsky, Nathaniel Rochester – si introduce nel titolo l'espressione 'Intelligenza Artificiale' che diventerà il termine ombrello sotto il quale riunire le ricerche sulla possibilità per le macchine di svolgere compiti intelligenti, se adeguatamente programmate.

Furono diverse le opzioni affrontate in quell'ambito di ricerca: dalla riproduzione delle capacità euristiche umane attraverso sistemi esperti, alla programmazione con strumenti presi a prestito dalla logica, all'emulazione del funzionamento neurale cerebrale nella macchina, fino ad arrivare a programmi come SHRDLU scritto da Terry Winograd, che tra il 1968 e il 1970 si ripromettevano di insegnare alla macchina a comprendere il linguaggio naturale. In quegli stessi anni vi fu l'avvento dei *chatbot*, dei piccoli software che permettevano alla macchina di intrattenere una pseudo-conversazione con gli esseri umani e quindi di passare per agenti intelligenti. Uno di questi dispositivi, di nome Eliza, inventato nel 1966 quasi per gioco da Roger Weizenbaum che riproduceva il comportamento di uno psicoterapeuta di impostazione Rogeriana ebbe un grandissimo suc-

1 Lettera di Turing a W. Ross Ashby, non datata, scritta probabilmente nel 1946 e sicuramente non dopo ottobre 1947, citata in Copeland (2004, 374-375).

2 Per altri dettagli su questo argomento cfr. Numerico (2005, capp. 4 e 5)

3 McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (31 agosto, 1955), A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, ristampato in AI Magazine Volume 27 Number 4 (2006), pp. 12-14, URL: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>; trad it «Sistemi intelligenti», XVIII, Dic. 2006, 413-428. Si segnala che tutto il numero di dicembre della rivista è dedicato al dibattito sull'intelligenza artificiale.

cesso tanto da spaventare lo stesso inventore e spingerlo a pubblicare nel 1976 un famoso testo contro l'intelligenza artificiale dal titolo *Computer power and human reason*, che costituiva un atto d'accusa contro le scelte considerate dei tecnologi. Al di là di quello che sarebbe stato possibile realizzare con la tecnologia, bisognava chiedersi cosa sarebbe stato giusto fare. Era giusto sostituire dottori e psicoterapeuti con i computer?

Fin dal principio, infatti, la questione della macchina intelligente non era soltanto un tema di carattere scientifico ma anche un tema dai profondi risvolti etici politici e sociali. Il tema etico del resto non ha mai abbandonato la discussione tecnologica. Ne è prova il grande dibattito sollevato dalla proposta fatta da Google a fine marzo 2019 per la costituzione di un comitato etico indipendente per discutere delle tante iniziative all'avanguardia nel campo dell'intelligenza artificiale, progetto che è stato sommerso di critiche per la scelta di membri che avrebbero potuto dimostrare interessi contrastanti con l'obiettivo etico. L'Advanced Technology External Advisory Council (ATEAC) ha provocato tante polemiche che un consistente gruppo di dipendenti della piattaforma hanno firmato una lettera di protesta. Il movimento è stato così repentino e impressionante da indurre in pochi giorni la *governance* di Google a abbandonare l'iniziativa per il momento⁴.

Il rapporto tra intelligenza umana e meccanica era già affrontato in un famoso articolo di Turing⁵, nel quale veniva proposto un metodo per capire se una macchina potesse essere scambiata per un essere umano, chiamato gioco dell'imitazione, divenuto poi famoso come Test di Turing. La dinamica del gioco prevedeva due agenti chiusi in stanze separate capaci di comunicare con l'esterno attraverso telescriventi. I giudici, inesperti di tecnologia, avrebbero potuto fare domande su ogni argomento per un tempo breve e poi decidere se uno dei due interlocutori fosse una macchina. L'idea era che se una parte dei giudici si ingannava in un certo numero di occasioni era possibile prendere in considerazione una risposta positiva alla più complessa domanda «se le macchine possano pensare».

La proposta di Turing ebbe molto successo e fu oggetto di un grande dibattito, come mostrano le tante pubblicazioni dedicate a questo tema nel

⁴ La scelta di rinunciare al comitato etico esterno è stata comunicata sul blog di Google pochi giorni dopo la sua istituzione, il 4 Aprile 2019 con un post a firma di Kent Walker, dal titolo: An external advisory council to help advance the responsible development of AI, reperibile alla URL: <https://www.blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai/>. La proposta è durata quindi solo dal 26 marzo al 4 aprile. Un resoconto della vicenda si può trovare in Knight (2019).

⁵ Turing (1950).

tempo⁶. Al Test di Turing è anche dedicato un premio, il Loebner Prize⁷ che viene consegnato ogni anno al software più efficiente nel superare il gioco dell'imitazione.

Turing disse espressamente, in una conversazione radiofonica del 1952, pubblicata successivamente, che la macchina deve fare finta di comportarsi come un essere umano fino a ingannare il giudice.

The idea of the test is that the machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing. A considerable proportion of a jury, who should not be expert about machines, must be taken in by the pretence.⁸

L'idea che la macchina faccia finta e che debba avere capacità di ingannare una giuria di non esperti è interessante e ci riporta alla questione più generale di cosa intendiamo quando ci riferiamo a una macchina intelligente. Ancora Turing ci fornisce degli spunti di riflessione in un articolo del 1948 sull'*Intelligent machinery*, nel quale sostenne che l'attribuzione dell'intelligenza a un comportamento riguarda sia le caratteristiche dell'agente che viene valutato sia la prospettiva dell'osservatore «the extent to which we regard something as behaving in an intelligent manner is determined much by our state of mind and training as by the properties of the object under consideration»⁹.

Se riconosciamo che l'intelligenza è attribuita socialmente, e conta molto la condizione di chi deve valutare, tale prospettiva offre un primo punto di vista rivoluzionario sull'uso dei calcolatori nella nostra vita quotidiana. Se chi programma le macchine è consapevole che sono in grado di ingannare gli inesperti e far loro ritenere che i loro comportamenti siano intelligenti è possibile che il nostro concetto di intelligenza debba essere considerato flessibile. L'inesperienza delle regole e dei metodi di costruzione e sviluppo dei dispositivi software e hardware può spingere le persone a ritenere che la macchina svolga compiti intelligenti e soprattutto trasforma la nozione di intelligenza includendo i dispositivi meccanici come candi-

6 Cfr. tra l'altro Shieber (2004), Moor (2006), Epstein, Roberts, Beber (2009) per una rassegna degli articoli dedicati al Test di Turing.

7 Si tratta di un premio che prende spunto dal test di Turing per premiare il software che meglio di altri mostra di saper rispondere alle domande di un giudice. Per maggiori dettagli sul premio si può vedere il sito che riporta la storia fin dalle sue origini <http://www.loebner.net/Prizef/loebner-prize.html> oppure quello dove sono descritte le edizioni più recenti della competizione <http://www.aisb.org.uk/events/loebner-prize>.

8 Turing (1952/2004, 495)

9 Turing (1948/2004, 431)

dati a esibirla, proprio come altri agenti sociali. Tale cambiamento ha effetto retroattivamente sul giudizio che abbiamo dei comportamenti umani.

Come suggeriva Norbert Wiener, il tema dell'intelligenza delle macchine è strettamente connesso a quello della presa di decisione e spesso l'attribuzione dell'intelligenza alle macchine si basa sulle superiori capacità computazionali dei computer, in grado di gestire grandi quantità di dati strutturati. Tali caratteristiche spingono ad attribuire alla macchina la responsabilità della presa di decisione anche in campi delicati per la vita delle persone, senza sapere esattamente come saranno gestiti.

For the man who is not aware of this, to throw the problem of his responsibility on the machine, whether it can learn or not, to cast his responsibility to the winds, and to find it coming back seated on the whirlwind.¹⁰

La questione della responsabilità nella presa di decisione della macchina è cruciale nel dibattito sull'intelligenza artificiale. Il progetto di costruzione di macchine sempre più autonome in grado di agire su diversi terreni implica che questi dispositivi siano dotati di una autonomia decisionale e presuppongono che la scelta intorno alla decisione da intraprendere sia inscritta già dal programma che ne governi le prestazioni. Si possono fare molti esempi calzanti dal soldato robot sul campo militare alla guida autonoma, passando per le macchine capaci di compiere diagnosi mediche o di supportare il giudice nella presa di decisione intorno alla pena da infliggere a un colpevole¹¹. L'intelligenza meccanica, quindi, si accompagna all'autonomia decisionale senza che sia sempre possibile rinvenire i criteri adottati, è soprattutto senza che la macchina possa esibire alcuna autoriflessione sulle proprie scelte. La *agency* intelligente è quindi sganciata ormai inevitabilmente dalla coscienza e dall'agire umano in senso stretto, sebbene comunque per ora siano gli esseri umani a progettare i dispositivi autonomi e quindi a definire le regole sulla base delle quali l'intelligenza meccanica esercita la sua azione.

Ma di quale nozione di intelligenza stiamo parlando? Non possiamo fare a meno di notare che nel corso della storia dell'intelligenza artificiale (poco meno di 70 anni) questa abbia più volte cambiato rotta, spostando spesso i metodi, ma anche i risultati sulla base dei quali definire una macchina come intelligente. Nei prossimi paragrafi proveremo a individuare le diverse definizioni di intelligenza in particolare distinguendo le diverse soluzioni rispetto alla rilevanza o meno del corpo.

¹⁰ Wiener (1950/1954, 185).

¹¹ Fry (2019, cap. 1).

3. La rappresentazione dell'intelligenza meccanica e il rapporto col corpo

La tecnologia ha un carattere diverso dalla scienza dal punto di vista epistemologico. Almeno in teoria, infatti, la scienza si occupa di comprendere come stanno le cose, mentre la tecnologia si interessa di come devono essere le cose per funzionare; per esempio, quanto cemento deve stare nel ponte perché possa reggere i carichi ai quali è sottoposto?

Gli strumenti, cioè, sono, a certe condizioni, dei modi per guardare al funzionamento della nostra interazione con loro e costituiscono una mediazione tra noi e il mondo esterno. In tal modo consentono di riprogettare il mondo che ci circonda per renderlo più alla nostra portata. Secondo Weizenbaum la tecnologia, però, non ha impatto solo sulla realtà circostante come processo di trasformazione delle sue caratteristiche, ma anche sul modo in cui riflettiamo la nostra immagine nel mondo modificato dalla nostra volontà di intervento.

They [tools] are then part of the stuff out of which man fashions his imaginative reconstruction of the world. It is within the intellectual and social world he himself creates that the individual prehearses and rehearses countless dramatic enactments of how the world might have been and what it might become. That world is the repository of his subjectivity. Therefore it is the stimulator of his consciousness and finally the constructor of the material world itself. It is this self-constructed world that the individual encounters as an apparently external force. But he contains it within himself; what confronts him is his own model of a universe, and, since he is part of it, his model of himself.¹²

La tesi di Weizenbaum, quindi, è che quando si inventano gli strumenti, contestualmente si attiva una trasformazione del mondo che implica un cambiamento anche di ciò che percepiamo come umano, in quanto l'essere umano è parte del mondo che con lo strumento si sta riorganizzando. La definizione delle tecniche non si limita a modificare la relazione con l'esterno nel senso di una facilitazione e di una riprogettazione di alcune pratiche, ma spinge a una nuova definizione della soggettività, che per Weizenbaum riguarda il riflettersi della soggettività nell'oggetto tecnico, in quanto parte integrante della rappresentazione del mondo esterno.

La discussione sulla possibilità di una simulazione o emulazione da parte della macchina dell'intelligenza prevede una definizione delle componenti della capacità umana replicabili meccanicamente, o simulabili rispetto all'esito dell'esecuzione del compito.

¹² Weizenbaum (1976, 18).

Quando Turing discuteva su quali fossero gli ambiti sui quali far esercitare l'attività della macchina intelligente, descriveva una serie di compiti scelti tra quelli nei quali il ruolo del corpo è marginale. Quella che segue è la lista degli ambiti a suo avviso più promettenti.

We propose to try and see what can be done with a 'brain' which is more or less without a body, providing at most organs of sight, speech and hearing. We are then faced with the problem of finding suitable branches of thought for the machine to exercise its powers in. The following fields appear to me to have advantages:

- (i) Various games e.g. chess, noughts and crosses, bridge, poker
- (ii) The learning of languages
- (iii) Translation of languages
- (iv) Cryptography
- (v) Mathematics.¹³

L'assenza del corpo caratterizza la maggior parte dei compiti che verranno a poco a poco attribuiti alle macchine intelligenti. Le attività sono preferite perché hanno relativamente poco a che fare con il corpo, perché le altre sono ritenute troppo complesse per la macchina. L'apprendimento del linguaggio sarebbe stato, secondo Turing, «the most impressive, since it is the most human of these activities»¹⁴. Tale posizione quindi associa la possibilità dell'intelligenza a una macchina, suggerendo implicitamente che per svolgere attività intelligenti il corpo non sia necessario. L'apprendimento del linguaggio costituisce un'attività limite per lui, perché il funzionamento del linguaggio si potrebbe descrivere come manipolazioni di simboli. Il progetto di educazione della macchina potrebbe avvenire attraverso l'intervento di un'interferenza proveniente dall'esterno capace di orientare il comportamento della macchina nella direzione desiderata.

Nell'articolo scritto nel 1950, invece, Turing si preoccupa di più dell'assenza di corpo nel caso dell'evoluzione del processo educativo della macchina, non tanto per la mancanza di organi di senso o arti preposti al movimento, quanto per la relativa presunta assenza della capacità emotiva. In un processo educativo di un dispositivo sarà necessario predisporre «some other unemotional channels of communication»¹⁵, a meno di utilizzare metodi troppo poco ortodossi e, in un certo senso, rischiare di torturare la macchina.

L'idea di Turing sembra essere basata su una completa scissione tra attività del corpo e intelligenza, pur mantenendo l'eccezione del linguaggio,

¹³ Turing (1948/2004, 420).

¹⁴ Turing (1948/2004, 421).

¹⁵ Turing (1950/2004, 461).

che sarebbe una specie di testa di ponte di capacità collegate al corpo nella macchina, e sulla cui possibilità di Turing mantiene qualche perplessità. La visione di Turing sul ruolo del linguaggio nell'educazione della macchina, lascerebbe supporre che questo, pur potendo essere compreso fino in fondo solo nel suo carattere di intenzionalità inaccessibile alla macchina, potrebbe tuttavia essere usato dal computer con l'inganno. Si tratterebbe di insegnare alla macchina a fare finta di maneggiare il linguaggio abbastanza da ingannare giudici inesperti. Questo era il tipo di intelligenza alla quale ritengo pensasse Turing, considerando il carattere sociale della sua definizione.

I *caveat* che, nonostante il grande ottimismo, permeavano il lavoro pionieristico di Turing, furono del tutto dimenticati dai successori. Il ruolo attribuito al linguaggio da Turing, che conservava un certo carattere di ambivalenza proprio della comunicazione interpersonale, non venne raccolto dai successori pionieri nell'ambito dell'intelligenza artificiale come John McCarthy, Herbert Simon, Marvin Minsky e altri. La letteratura sull'IA mantiene la centralità del linguaggio, ma considerandolo solo come strumento per rappresentare la conoscenza nel software di macchine intelligenti, pensandolo quindi come una riproduzione delle caratteristiche della logica.

La logica, infatti, pur usando un sistema formale che prevedeva la presenza di simboli, formule e espressioni composte secondo regole ben formate, che poteva essere scambiato per un linguaggio, non consentiva la possibilità di riprodurre la caratteristica ambivalenza, polisemicità, e variabilità proprie delle lingue naturali. Nella già citata Conferenza di Dartmouth¹⁶, per esempio, tutti i riferimenti al linguaggio non riguardano il contesto nel quale Turing aveva usato il concetto, e cioè il carattere comunicativo e di interferenza da lui indicato. Il linguaggio non era che un metodo, uno strumento per rappresentare informazioni e ottenere conclusioni applicando precise regole di inferenza, proprie della programmazione logica.

It therefore seems to be desirable to attempt to construct an artificial language which a computer can be programmed to use on problems requiring conjectures and self-reference. It should correspond to English in the sense that short English statements about the given subject matter should have short correspondents in the language [...]. I hope to try to formulate a language having these properties and in addition to contain the notions of physical object, event, etc. with the hope of using this language it will be possible to program a machine to learn to play games well and do other tasks.¹⁷

¹⁶ McCarthy *et al.* (1955).

¹⁷ McCarthy *et al.* (1955).

Da questo passaggio è chiaro che il modello epistemico che presidiava il progetto dell'intelligenza artificiale dopo Turing non ripropose l'importanza del ruolo ambivalente del linguaggio come strumento di comunicazione, capace di produrre effetti trasformativi attraverso l'interazione con la macchina. In *Programs with common sense* McCarthy¹⁸ stabiliva gli obiettivi del suo lavoro nell'IA: 1) un programma per manipolare affermazioni strumentali comuni usando un linguaggio formale adeguato (una parte del calcolo dei predicati); 2) il programma base sarebbe stato in grado di trarre conclusioni immediate a partire da una lista di premesse; 3) nel suo programma le procedure dovevano essere descritte usando un linguaggio preciso, e così doveva essere per le attività euristiche. Anche dopo quel periodo McCarthy era completamente convinto che il problema della generalità in IA consistesse nella definizione di un linguaggio capace di rappresentare la conoscenza di senso comune in modo che fosse possibile includerla in un database strutturato¹⁹.

Il gruppo di ricerca di Herbert Simon, un altro dei gruppi più importanti per l'IA, riguardo il linguaggio e la comunicazione con la macchina sosteneva un approccio simile a quello di McCarthy, sebbene non condivideva con lui la stessa fiducia in una rappresentazione logica della conoscenza. In un articolo dell'epoca si legge:

It seems that the rise of effective communication between man and computer will coincide with the rise in the intelligence in the computer [...]. But at this point in history, the only way we can obtain more intelligent machines is to design them – we cannot yet grow them, or breed them, or train them by the blind procedures that work with humans.²⁰

L'idea quindi di addestrare la macchina, come avrebbe desiderato Turing, era stata accantonata dal gruppo di studiosi all'epoca *mainstream* dell'IA. La loro proposta era piuttosto costruire un sistema di regole, anche di carattere euristico, capaci di spingere le macchine a prendere decisioni che sarebbero considerate intelligenti, qualora fossero il frutto dell'attività umana.

Marvin Minsky, un altro dei pionieri dell'intelligenza artificiale, sosteneva la tesi secondo la quale l'intelligenza poteva essere definita come quello che ancora non siamo capaci di comprendere, sebbene siamo pronti ad ammirare:

¹⁸ McCarthy (1959).

¹⁹ McCarthy (1971/1987).

²⁰ Newell, Shaw, Simon (1958/1963, 41)

To me “intelligence” seems to denote little more than the complex of performances which we happen to respect, but do not understand. [...] But we should not let our inability to discern a locus of intelligence lead us to conclude that programmed computer cannot think. For it may be so with man, as with machine, that when we understand finally the structure and program, the feeling of mystery (and self-approbation) will weaken.²¹

Anche nel caso di Minsky, l'intelligenza era caratterizzata da un comportamento che non si riusciva in prima battuta a spiegare indipendentemente da quali fossero gli agenti a esibirlo. I primi sostenitori dell'intelligenza artificiale abbracciavano senza riserve il funzionalismo²², secondo il quale l'attribuzione dell'intelligenza e soprattutto degli stati mentali non aveva a che fare con delle caratteristiche di tipo fisico, ma soltanto con la capacità dell'agente di svolgere un compito considerato intelligente, come manipolare simboli in input alla ricerca di una soluzione ottenuta come output della manipolazione.

Il primo progetto dell'intelligenza artificiale, al di là di una differenza intorno al ruolo del linguaggio, si appoggiava quindi su una rivisitazione della vulgata del dualismo cartesiano. Secondo la visione di Cartesio dei pionieri di questo campo, infatti, la mente sarebbe la *res cogitans* priva di estensione che allo stesso tempo sarebbe stata in grado di partecipare e interagire causalmente con il corpo, senza però essere in se stessa un fenomeno causale. Sequestrata nella sua posizione, con un rapporto incerto con il mondo fisico, per Cartesio, la mente ha un oggetto tipico e privilegiato del suo interesse nelle idee matematiche.

Secondo Philip Agre²³, gli scienziati della prima ondata dell'IA furono ispirati dalle idee di Cartesio, facendo però un salto epistemologico rispetto alla sua posizione: «the mind does not simply contemplate mathematics, they asserted; the mind is itself mathematical, and the mathematics of mind is precisely a technical specification for the causally explicable operation of the brain»²⁴.

Agre suggerisce un'interessante interpretazione della prima ondata dell'intelligenza artificiale. Questa sarebbe basata sulla metafora cognitiva della mente come intessuta di matematica, capace di dar conto del funzionamento del cervello: «Computational studies [...] were studies of

21 Minsky (1961/1963, 447)

22 Una possibile definizione del funzionalismo è che quello che rende qualcosa uno stato mentale è piuttosto quello che produce, che quello di cui è fatto. Per maggiori dettagli su questo argomento rimando alla voce Functionalism della Stanford Encyclopedia of philosophy <https://plato.stanford.edu/entries/functionalist/>

23 Agre (1997).

24 Agre (1997, 3).

faculties of *intelligence* and processes of *thought*, as part of a kind of cult of cognition whose icons were the rocket scientist, the symbolism of mathematics, and the computer itself»²⁵.

Secondo l'autore, quindi; la visione dell'intelligenza dell'IA prevedeva un completo distacco dalle concrete situazioni vissute dagli esseri umani dotati di corpo immersi nel mondo. Il modo stesso di vedere i problemi a cui l'IA si accingeva a fornire delle soluzioni era tarato sul modo di percepire i problemi come astrattamente matematici, cioè come problemi strumentali per i quali le tecnologie potevano determinare effetti che avrebbero riguardato il mondo.

Tuttavia questa intelligenza computazionale di impostazione logicista, o di carattere simbolico-cognitivista, non ebbe sufficienti successi. I problemi per i quali forniva soluzioni erano solo i cosiddetti problemi giocattolo e non riuscì mai a uscire dal laboratorio con le sue pratiche.

Le questioni sollevate nel fallimento di questo progetto di intelligenza artificiale avevano a che fare con le stesse domande inevase che riguardavano l'impostazione filosofica dualista che implicitamente guidava le tracce degli studiosi²⁶. La mancanza di attenzione per il corpo e per l'intelligenza come *embodiment* furono le probabili cause dell'insuccesso che sostanzialmente viene attribuito ai progetti di IA di quegli anni²⁷. La data che consacra una specie di atto di morte di quel modello di ricerca può essere considerata il 1993 quando il governo americano decise di chiudere Strategic Computing Program finanziato da DARPA (Defence Advanced Research Project Agency), nato dieci anni prima per unificare tutte le ricerche finanziate dal governo americano nell'ambito dell'IA²⁸. Tale chiusura fu accompagnata dalla nascita di un nuovo programma di ricerca High-Performance Computing (HPC) che invece puntava sull'amplificazione delle capacità computazionali piuttosto che su progetti di IA. Attualmente Darpa finanzia di nuovo ricerche di IA, ma queste riguardano tutte metodi di

25 Agre (1997, 3).

26 Agre (2005).

27 Segnalo brevemente che al tema dell'incorporazione o *embodiment* dell'intelligenza nel corpo si sono dedicati alcuni studiosi dell'ambito dell'IA, soprattutto nel versante filosofico. Per un approfondimento sul tema dell'*embodiment* Varela, Thompson, Rosch (1993). In questo testo si apre anche a un altro aspetto importante circa lo studio dell'intelligenza, il concetto di *extended mind* (mente estesa). Non solo la mente è configurata dal suo stare nel corpo, ma è anche orientata dall'ambiente nel quale si trova e dall'organizzazione culturale della società in cui vive. Per questa prospettiva si veda Clark, Chalmers (1998).

28 Franchi e Güzeldere (2005, 94-96).

*machine learning*²⁹, che come vedremo (Cfr. § 5) costituiscono l'orizzonte più promettente di questo settore.

Come suggeriscono gli autori dell'articolo sulla *Revanche des neurones*, Cardon, Cointet e Mazières³⁰, l'intelligenza artificiale non è più quella che era stata al centro delle ricerche dei suoi primi pionieri, cioè l'intelligenza computazionale di carattere simbolico, di cui ho parlato in questo paragrafo. Si tratta di un cambio di paradigma avvenuto silenziosamente, in cui è stata mantenuta in vita l'etichetta ma sostituita con un paradigma metodologico e epistemico completamente nuovo, nel quale, come vedremo, è meno marcata la distinzione tra intelligenza simbolica (che dovrebbe rappresentare il funzionamento del mentale) e corpo (che costituirebbe il contesto pratico dell'esperienza del mondo esterno).

Nel prossimo paragrafo affronteremo le origini cibernetiche della svolta del *machine learning* e le sue conseguenze epistemologiche.

4. La cibernetica e l'interazione con la macchina

Il progetto dell'intelligenza artificiale era affiancato fin dal principio da un altro modello di sviluppo delle macchine, quello della cibernetica, un campo di ricerche transdisciplinare, inventato da Norbert Wiener. La cibernetica si proponeva di guardare alle macchine come a dispositivi indistinguibili da altri agenti organici. Il titolo del testo del 1948 che diede vita a questa transdisciplina era: *Cybernetics or Control and Communication in the Animal and the Machine*. L'obiettivo della cibernetica era analizzare i dispositivi sotto il profilo della comunicazione e del controllo. Si considerava il controllo come un tipo particolare di comunicazione in cui era necessario non solo trasmettere un messaggio informativo, ma anche essere sicuri che l'altro agente avesse ricevuto ed eseguito l'ordine. Secondo tale rappresentazione esseri viventi e dispositivi meccanici potevano essere analizzati attraverso lo stesso modello scientifico di comportamento.

Il modello si basava sul funzionamento del *feedback* che permetteva all'agente che riceveva il messaggio di rispondere adattando il proprio comportamento al messaggio ricevuto dall'altro agente o dall'ambiente stesso. Vista da questa prospettiva, la capacità degli esseri viventi di ade-

29 Per una descrizione dei progetti finanziati da Darpa sotto la voce AI next campaign. Accelerating the third Wave è possibile consultare il sito della Darpa con la descrizione dettagliata dei finanziamenti attivati a partire da settembre 2018, accessibile nel sito dell'agenzia americana del Department of Defence: <https://www.darpa.mil/work-with-us/ai-next-campaign>

30 Mazières (2018).

guarsi all'ambiente per sopravvivere poteva essere valutata in modo simile alla capacità delle macchine di obbedire agli ordini impartiti o meglio di comportarsi in maniera teleologica, essendo in grado di portare a termine il proprio compito.

Il principale modello di riferimento della cibernetica era la recente esperienza bellica nella quale radar, missili balistici e altri dispositivi meccanici funzionavano per realizzare un obiettivo, mostrando così di possedere una loro finalità nel raggiungere lo scopo che era loro prefissato, qualunque fosse: dal riconoscere un velivolo nemico sul radar, a inseguire un obiettivo in movimento per colpirlo:

Wiener came to see the predictor as a prototype not only of the mind of an inaccessible Axis opponent but of the Allied anti-aircraft gunner as well, and then even more widely to include the vast array of human proprioceptive and electro-physiological feedback systems. The model then expanded to become a new science known after the war as "cybernetics," a science that would embrace intentionality, learning, and much else within the human mind.³¹

La tesi di Peter Galison – uno studioso di cibernetica membro della *Stanford School*³² – è che la guerra aveva trasformato ogni nemico in un nemico meccanico:

On the mechanized battlefield, the enemy was neither invisible nor irrational; this was an enemy at home in the world of strategy, tactics, and maneuver, all the while thoroughly inaccessible to us, separated by a gulf of distance, speed, and metal. It was a vision in which the enemy pilot was so merged with machinery that (his) human-nonhuman status was blurred. In fighting this cybernetic enemy, Wiener and his team began to conceive of the Allied anti-aircraft operators as resembling the foe, and it was a short step from this elision of the human and the nonhuman in the ally to a blurring of the human-machine boundary in general.³³

L'idea quindi che il nemico non fosse un altro essere umano, ma un agente strettamente connesso con una macchina spingeva a ritenere che anche le forze aeree alleate fossero simili al nemico³⁴. Sul campo di battaglia

31 Galison (1994, 229)

32 La Stanford School è una scuola di filosofia della scienza di impostazione pluralista, cioè contro l'idea della scienza come un tutto unitario, i cui membri originari erano passati per la Stanford University tra gli anni Ottanta e Novanta. Le tesi di questo gruppo di studiosi si concentravano sulla natura plurale dei modelli scientifici e su un approccio post-positivistico a basato sulle pratiche della conoscenza scientifica.

33 Galison (1994, 233)

34 Cfr. Galison 1994 "I will argue that the system of weaponry and people that Wiener had in mind was predicated on a picture of a particular kind of enemy. On the mechanized battlefield, the enemy was neither invisible nor irrational; this was an enemy

avveniva questa elisione, questa confusione del confine tra essere umano e macchina in generale.

Tale sconfinamento della macchina nell'ambito del vivente era stata già suggerita da un articolo scritto da Wiener con altri due suoi colleghi, Arturo Rosenblueth e Julian Bigelow, dal titolo *Behavior, purpose, teleology*, proprio durante la seconda guerra mondiale.

A further comparison of living organisms and machines leads to the following inferences. The methods of study for the two groups are at present similar. Whether they should always be the same may depend on whether or not there are one or more qualitatively distinct, unique characteristics present in one group and absent in the other. Such qualitative differences have not appeared so far. The broad classes of behavior are the same in machines and in living organisms.³⁵

In questo articolo si mettevano le basi per affermare che non c'erano differenze di rilievo nei metodi di studio relativi a esseri viventi e alle macchine dal punto di vista della comunicazione e del controllo. Tale conclusione, quindi, suggeriva un abbattimento sostanziale della barriera che separava organico e inorganico. La prospettiva di un possibile studio congiunto di esseri viventi e macchine rispetto alla comunicazione – di cui il controllo era solo un caso particolare – consentiva di analizzare i dispositivi meccanici come casi analoghi a creature organiche, ma prospettava soprattutto la possibilità di un incontro, di una relazione, di un'integrazione tra viventi, in particolare umani, e macchine.

Se quindi le macchine potevano essere integrate negli esseri umani come una loro parte, come suggerivano le ricerche della cibernetica, il problema della radicale separazione tra loro sarebbe stato superato. In particolare sarebbe stata superata la principale mancanza dell'intelligenza artificiale, quella relativa alle capacità del corpo, a cui si sarebbe potuto sopperire attraverso una integrazione con il corpo dell'utente dell'interazione con la macchina.

L'idea di una integrazione tra esseri umani e macchine e di una loro collaborazione basata sul principio della comunicazione e del controllo fu estremamente proficua, molto più di quella dell'intelligenza artificiale almeno fino agli anni Duemila. Dall'interazione uomo-macchina derivarono alcuni dei maggiori successi dell'informatica, dal sistema operativo che si propone come *user-friendly*, basato sulla programmazione orientata agli oggetti (*object-oriented programming*), che introduce icone, finestre,

at home in the world of strategy, tactics, and maneuver, all the while thoroughly inaccessible to us, separated by a gulf of distance, speed, and metal" (p. 233).

³⁵ Rosenblueth, Wiener, Bigelow (1943, 21).

groupware, e la capacità intuitiva di cliccare con il mouse o col dito sulle icone, all'idea di una vera e propria simbiosi umano meccanica suggerita da Licklider³⁶. L'integrazione con la macchina era alternativa alla sostituzione dell'attività umana da parte della macchina nello svolgere compiti intelligenti, prospettata dall'IA. L'integrazione, infatti, prevedeva una collaborazione nella quale ciascun agente avrebbe mantenuto le proprie caratteristiche più qualificanti integrandosi con l'altro: «To think in interaction with a computer in the same way that you think with a colleague whose competence supplements your own will require much tighter coupling between man and machine»³⁷.

L'idea stessa di interazione con la macchina derivava dalla cibernetica come disciplina che si occupava della comunicazione e del controllo. La mia tesi è che il vero contributo rivoluzionario all'IA da parte della cibernetica consista nella riflessione sulla centralità della comunicazione tra esseri umani e macchine, e che dalle conseguenze di questa impostazione derivino i maggiori successi della disciplina in generale, e l'entrata in una nuova agenda di ricerca. Comunemente invece si ritiene che mentre l'IA era concentrata sulla riproduzione della mente, la cibernetica fosse interessata alla riproduzione del cervello. Sebbene fossero in atto questi due diversi orientamenti, ritengo che si trattasse comunque di modelli astratti da riprodurre nella macchina, nessuno dei quali aveva davvero a che fare né con la mente è col cervello umano. Si trattava di diverse impostazioni su come costruire un modello efficiente dell'attività umana. L'aspetto della riproduzione dell'attività neurale era oggetto del progetto di ricerca del connessionismo³⁸, che spesso viene associato allo sviluppo delle macchine intelligenti della cibernetica³⁹. L'identificazione tra cibernetica e interesse per una replica meccanica del funzionamento del cervello, poi incarnato dal connessionismo, rimanda alla tradizionale descrizione della storia dell'intelligenza meccanica di impostazione cibernetica⁴⁰. Eppure l'articolazione dell'interazione uomo-macchina è stata l'eredità della cibernetica che ha avuto gli esiti di maggior impatto sull'IA, perché fu in grado di

36 Licklider (1960).

37 Licklider (1960, 4).

38 Cfr. McClelland, Rumelhart, Hinton, (1988)

39 A mio avviso l'esito del connessionismo come influenza della cibernetica sull'IA rappresenta un aspetto e nemmeno quello più rilevante, citato per esempio in Cardon *et al.* (2018), della cibernetica e del suo impatto sull'IA.

40 Vedi Cordeschi (2002) per una descrizione precisa dei diversi modelli di intelligenza in gioco all'epoca: dall'emulazione della mente attraverso un sistema fisico simbolico al connessionismo che invece si proponeva di emulare una rete neurale in cui i singoli nodi erano unità elementari che agivano in parallelo con tutti gli altri nodi collegati da una rete che si attivava a seconda della quantità di stimoli ricevuti.

costruire un metodo per la cattura dei modelli cognitivi umani necessari per l'evoluzione delle macchine e la loro progressiva autonomizzazione.

Fu Licklider, inoltre, che pure aveva partecipato alle fasi progettuali della cibernetica⁴¹ insieme a Robert Taylor, a progettare l'idea del computer come strumento di comunicazione⁴². I due suggerivano l'esigenza di un dispositivo in grado di connettere macchine differenti, capaci sia di interagire tra loro, sia di mediare con gli esseri umani. L'obiettivo delle macchine interconnesse era aperto a ogni possibile uso, anche a usi imprevisti o non ancora ipotizzati:

These new computer systems we are describing differ from other computer systems advertised with the same labels: interactive, time-sharing, multiaccess. They differ by having a greater degree of open-endedness, by rendering more services, and above all by providing facilities that foster a working sense of community among their users.⁴³

Tale articolo fu l'anticipazione della nascita di Arpanet, l'antesignana della rete Internet che si realizzò solo un anno dopo, sotto la guida di Taylor nell'ambito di un ufficio di ARPA (successivamente divenuta DARPA) chiamato IPTO (Information Processing Technique Office).

A very important part of each man's interaction with his on-line community will be mediated by his OLIVER.[...] At your command, your OLIVER will take notes (or refrain from taking notes) on what you do, what you read, what you buy and where you buy it. It will know who your friends are, your mere acquaintances. It will know your value structure, who is prestigious in your eyes, for whom you will do what with what priority, and who can have access to which of your personal files. It will know your organization's rules pertaining to proprietary information and the government's rules relating to security classification.⁴⁴

La lungimiranza di Licklider e Taylor non solo anticipò e indirizzò la nascita della rete Arpanet, ma si spinse anche a introdurre il tema della datificazione e della possibilità di ottenere dati e connessioni sui comportamenti delle persone ai fini di anticiparne e orientarne i comportamenti⁴⁵. OLIVER, che rappresenta una specie di assistente personale onnisciente che i due studiosi prevedono, è l'antesignano di tutta la nuova serie di

41 Numerico (2010).

42 Licklider, Taylor (1968).

43 Licklider, Taylor (1968, 31)

44 Licklider, Taylor (1968,38-39).

45 Cfr. Numerico (2017) per una trattazione di come Licklider abbia anticipato il tema della datificazione nel suo libro sul futuro delle biblioteche (Licklider 1965).

servizi e progetti legati all'analisi dei Big Data nell'ambito della ricerca sociale.

È a questo progetto di datificazione che si riferiscono i nuovi e più promettenti sviluppi della *data science* e degli algoritmi di *machine learning* che attualmente rappresentano la nuova ondata di intelligenza artificiale.

5. Data science e algoritmi

La diffusione della rete Internet e dei massicci processi di datificazione ha messo a disposizione dell'analisi tecnologica una grande quantità di dati prima impensabile. Fu proprio ARPA a finanziare il progetto della rete Arpanet e a sostenere il processo di integrazione, che al principio riguardò solo centri di ricerca e successivamente, con la diffusione del World Wide Web⁴⁶, incluse anche tutti i computer che potevano essere connessi in rete e i loro utilizzatori. La grande disponibilità di dati ha riaperto le speranze per un nuovo modello di intelligenza artificiale che non si basi sul metodo top-down della definizione di regole per risolvere problemi, come avveniva in passato, ma si concentri piuttosto su una logica induttiva: sulla capacità di trovare delle correlazioni utili per diversi obiettivi di carattere predittivo, tra le ingenti quantità di dati rilasciate in rete.

Questo è il mantra dei Big Data e degli algoritmi che vengono messi in azione per analizzare e organizzare le grandi quantità di dati. La statistica è la scienza principale utilizzata in questo contesto, ma gli esperti di *data science* sostengono che la quantità ingente di dati implichi una trasformazione parziale delle metodologie e delle pratiche per la loro gestione, richiedendo soluzioni completamente originali. L'obiettivo della *data science* che rappresenta la frontiera più promettente dell'attuale orientamento dell'IA, non è tanto spiegare i dati, fornendo un modello che ne renda conto, quanto costruire algoritmi in grado di fare previsioni sulla serie futura dei dati, indipendentemente dalle loro capacità di spiegazione. Cambia quindi la prospettiva generale attraverso la quale affrontare i dati rispetto alla più tradizionale ricerca statistica, che è orientata alla spiegazione⁴⁷.

I dati sono analizzati con l'aiuto di algoritmi di *machine learning*. Tra questi quelli di maggior successo sono attualmente quelli di *deep learning*. Si tratta di metodi per analizzare i dati ricercando modelli che possano

⁴⁶ Il progetto promosso da Tim Berners-Lee nel 1989 decollò davvero solo nel 1993 con il lancio di Mosaic, il primo browser che aveva una diffusione globale indipendente dalla piattaforma informatica utilizzata.

⁴⁷ Kelleher, Tierney (2018, cap.1).

essere generalizzati e usati per fare delle previsioni per il futuro comportamento delle serie analizzate. Per esempio, nel caso di algoritmi per la diagnosi medica, si analizzano i dati di pazienti che hanno avuto un infarto, alla ricerca di caratteristiche ricorrenti che possano consentire a chi fa la diagnosi di prevedere l'avvento dell'infarto in anticipo e mettere in atto delle cure preventive⁴⁸. Ma affinché un tale algoritmo possa funzionare, c'è bisogno di costruire i dati in un certo modo, costruire cioè dei cluster di pazienti a rischio di infarto che abbiano tutti le stesse caratteristiche.

Il metodo della *data science* è alla ricerca di correlazioni. Le correlazioni tra le serie di dati non possono essere considerate delle vere e proprie relazioni di causa effetto, ma rispondono piuttosto all'instaurazione di una reciprocità tra due serie di dati, per usare una categoria kantiana. La forza della correlazione deve essere misurata ed è rilevante solo a partire da una certa soglia⁴⁹. Tuttavia, proprio in presenza di ingenti quantità di dati è possibile che le correlazioni, quando pure fossero rilevanti dal punto di vista quantitativo, potrebbero essere spurie, come mostra un brillante articolo di Calude e Longo⁵⁰.

La datificazione e la costruzione di algoritmi che si propongono di fare previsioni dalle serie di dati, inoltre, ha bisogno di un presupposto imprescindibile, quello del controverso principio di induzione, tanto discusso in filosofia: il fatto che quello che è accaduto in passato continuerà a verificarsi in futuro. Inoltre, presuppongono la possibilità implicita di costruire cluster di individui che sono simili e condividono preferenze e comportamenti. La previsione appoggiata sulla serie di dati pregressi assume che il futuro non potrà che corrispondere al passato, principio che da Hume⁵¹ in poi deve essere usato con cautela.

48 Kelleher, Tierney (2018, p. 98).

49 Kelleher, Tierney (2018, 105 e segg.).

50 Calude, Longo (2017).

51 La questione storicamente nota come il problema dell'induzione introdotto da David Hume (*A treatise of human nature*, 1739) è: come si fa partendo dalle passate esperienze a dimostrare che quelle stesse esperienze si ripresenteranno anche in futuro? «Our foregoing method of reasoning will easily convince us, that there can be no demonstrative arguments to prove, *that those instances, of which we have had no experience, resemble those, of which we have had experience*» (Hume 1888, Sect. VI, p.89). Questo problema classico dell'epistemologia dell'inferenza induttiva aleggia irrisolto anche nella ricerca di generalizzazione degli algoritmi. Gli algoritmi ad apprendimento supervisionato usano una base dati di training a partire dalla quale estrarre categorie e modelli che poi applicheranno sui dati oggetto delle inferenze. La replica nel futuro di quello che è già accaduto è tutta da dimostrare e non può essere assunta senza condizioni, soprattutto quando i dati sono normalizzati e la quantificazione è costruita ad arte per il reperimento dei meccanismi di riconoscimento delle regolarità (pattern recognition).

Il metodo generale che utilizza la nuova ondata di IA, in particolare quella applicata ai comportamenti umani, si basa sulla datificazione dell'agire umano e sull'identificazione di metodi per il *pattern recognition*, il riconoscimento di regolarità e probabilità di ripetizione di alcuni schemi di comportamento. Tuttavia tali regolarità, che si misurano sulla probabilità di una ripetizione, di un'abitudine – potremmo dire – hanno delle caratteristiche molto particolari. Da un lato si propongono come meccanismi di cattura dei comportamenti umani⁵², laddove la capacità di replicarli attraverso da definizione di regole top-down non è stata sufficientemente efficace, e dall'altro lato hanno una forte componente prescrittiva più che descrittiva⁵³, perché tendono a orientare il comportamento delle persone a partire dagli schemi che si fanno delle loro preferenze o di quelle di altre persone che sono considerate appartenenti al medesimo cluster comportamentale. Inoltre, dal momento che la capacità di fare previsioni a partire dai dati è di carattere probabilistico, i giudizi che vengono espressi sulla pericolosità di alcuni individui, così come sulla tendenza a un certo stile di comportamento, mantengono comunque un carattere di interpretazione, che però viene occultato nel segreto sui meccanismi algoritmici prescelti⁵⁴.

Sarebbe necessario, quindi, fare attenzione a quali siano gli ambiti nei quali adottare senza controindicazioni la capacità previsionale degli algoritmi di *machine intelligence* che sono tanto promettenti e tanto potenti. Il rischio di performatività e prescrittività degli algoritmi di *machine e deep learning* è segnalato anche da un report della Rand Corporation, dal titolo esplicativo: *An intelligence in our image*⁵⁵. La Rand è un'agenzia di ricerca attiva nell'ambito della presa di decisione pubblica, che ha svolto un ruolo importante nello sviluppo dell'informatica negli US. Il risultato di questo report di monitoraggio e analisi dell'uso degli strumenti di Intelligenza Artificiale per prendere decisioni pubbliche e politiche è piuttosto problematico. Il report, dedicato all'analisi delle applicazioni sociali degli algoritmi di *machine learning* nell'ambito dell'IA, segnala che:

The opacity of algorithms makes it harder to judge correctness, evaluate risk, and assess fairness in social applications. It can also obscure the causal understanding behind decisions. These issues might be harmless if algorithms were (near) infallible. But most algorithms have only probabilistic guarantees of accuracy. And this is in the best possible scenarios, in which the right models and algorithms are applied appropriately, with the best intention to “perfect” data. Algorithm designers and users

52 Rouvroy, Berns (2013).

53 Chun (2019).

54 O'Neil 2016, Cramer, (2019).

55 Osoba e Welser (2017).

rarely have the luxury of such perfect scenarios. They must rely on assumptions that can fail and lead to unexpected results.⁵⁶

La mancanza di una conoscenza totale delle situazioni nelle quali si fanno previsioni di tipo probabilistico, quindi, rischia di favorire – nell’ambito dell’uso sociale di questi metodi per la presa di decisione guidata dagli algoritmi intelligenti – pratiche non abbastanza affidabili e garantite, anzi opache e incomprensibili che si prestano a esiti iniqui e discriminatori nei confronti di gruppi di popolazione più deboli e incapaci di difendere i propri diritti.

Il rapporto tra presa di decisione, intelligenza e macchine è uno dei più controversi della rivoluzione informatica. Sebbene i metodi usati per l’intelligenza meccanica siano cambiati nel tempo, resta il problema che l’attribuzione alla macchina di sempre maggiori responsabilità decisionali e governamentali non permette di controllare le ragioni delle scelte che si compiono, proprio come sosteneva Wiener nel 1950. Esiste al momento un ampio dibattito nella letteratura sull’etica degli agenti artificiali a proposito di quanto e se si possa attribuire a tali macchine l’autonomia necessaria per essere un agente etico e quindi per essere responsabile delle proprie azioni⁵⁷.

Sul problema è difficile dire una parola definitiva e non è questa la sede per gli approfondimenti, possiamo però segnalare la grande difficoltà di comprendere anche solo le ragioni di certe decisioni perché le macchine sono governate da algoritmi spesso opachi, come suggerisce Frank Pasquale⁵⁸ nel suo libro *Black box society*. Nel libro si discute dell’estrema segretezza di molti algoritmi governamentali che rendono impossibile controllare le pratiche di presa di decisione in settori come la finanza e l’informazione, tenendo il sistema e gli utenti in una situazione di grave asimmetria informativa. Tale stato di cose rischia di spostare gli *arcana imperii* della presa di decisione politica fuori dalle stanze dei governanti e dentro quelle degli algoritmisti che producono e diffondono “algoritmi intelligenti” per pren-

56 Osoba, Welser (2017, p. 3).

57 Per una ricognizione del problema della responsabilità legale degli agenti artificiali segnalo l’articolo di Peter Asaro (2016); inoltre la centralità del rapporto tra etica e tecnologia è testimoniata dall’aumento di riviste scientifiche dedicate a Etica e Tecnologia (per es. «Science and Engineering Ethics», «Science, Technology, & Human Values», «Ethics and Information Technology», oltre a «Philosophy and Technology» ecc.) a dimostrazione che esiste un problema sul rapporto tra tecnologia e questioni etiche, politiche e sociali. Nulla è più improbabile che ritenere che tra tecnologia e questioni politiche non ci siano collegamenti e che le decisioni sull’innovazione debbano essere indipendentemente dall’ingaggio politico e sociale che impongono, cfr. Latour (2010/2013).

58 Pasquale (2015).

dere decisioni in vari ambiti dello spazio pubblico, come segnala anche Rodotà. La sua posizione è chiarita nel passo che segue:

Ma quando l'algoritmo diviene il fondamento stesso del potere esercitato da un soggetto, com'è nel caso enfatizzato di Google, e tutto ciò che lo riguarda è avvolto dalla massima segretezza, allora siamo davvero di fronte alla nuova versione degli arcana imperii, che non tutelano solo la libertà di impresa ma si impadroniscono, direttamente o indirettamente, della vita stessa delle persone.⁵⁹

E non si tratta solo di Google. Siamo di fronte a un processo di progressiva delegittimazione della decisione umana in favore di dispositivi commerciali, oscuri e non trasparenti, di proprietà di aziende for-profit che agiscono in modi impossibili da spiegare razionalmente che riguardano i cittadini e che impattano persino sulla libertà personale dei soggetti. Un esempio è l'algoritmo COMPAS dell'azienda Northpointe usato a supporto di alcuni tribunali americani per misurare il rischio di recidività dei colpevoli di un reato. Gli algoritmi di questo tipo misurano il *risk assessment score*, la valutazione del rischio del potenziale di recidività a commettere di nuovo il reato per il quale si viene condannati. Secondo un'indagine di Propublica, un network indipendente di giornalismo investigativo, poi supportato dai risultati di alcuni scienziati, è impossibile creare un algoritmo predittivo sulla recidività che sia corretto e insieme equo rispetto al pregiudizio razziale⁶⁰.

6. Osservazioni finali

Abbiamo visto (Cfr. § 4) che l'ultima ondata dell'intelligenza artificiale basata sulla datificazione e sulla struttura algoritmica della ricerca quantitativa della capacità predittiva, di anticipazione e di influenza dell'azione umana e sociale ha la sua origine in una metafora del nemico come un agente artificiale parzialmente meccanico, metafora che si riverbera nell'umanità intera come suggerito da Galison⁶¹ e Edwards⁶². Tale metafora ha permesso da un lato di incorporare la macchina nel corpo e quindi di accedere alla struttura della sensibilità e dell'esperienza umana trasformandola in dati da elaborare, e dall'altro ha costruito il meccanismo della datificazione

⁵⁹ Rodotà (2014, 38).

⁶⁰ Julia Ankwin, Jeff Larson (30 dic. 2016) "Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say", Propublica: URL: <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>.

⁶¹ Galison (1994).

⁶² Edwards (1997).

come una discretizzazione, una riduzione alla quantità misurabile valutata attraverso il metodo dell'induzione. L'agire del corpo, le sue abitudini, i suoi comportamenti sono ridotti a dati con tutto quello che ne consegue in fatto di normalizzazione, rettificazione e cosiddetta pulitura delle informazioni. Il flusso così prodotto è analizzato attraverso algoritmi basati sulla teoria della probabilità che hanno il compito di prevedere, anticipare e talvolta persino provocare alcuni comportamenti.

Tale incorporazione della macchina nell'interazione umana porta con sé una riorganizzazione della struttura psichica e mentale dell'umano in una direzione difficile da valutare. La mediazione della macchina provoca una nuova individuazione psichica e collettiva della soggettività umana, suggerisce Bernard Stiegler⁶³, introducendo il concetto di *pharmakon* per definire i dispositivi digitali: di possibile rimedio, ma anche potenziale veleno. Ogni tecnologia che costituisce un nuovo supporto per la memorizzazione, infatti, modifica il rapporto con l'interiorità e con la capacità cognitiva umana. Dürrenmatt, in un capitolo del racconto *L'incarico*⁶⁴, descrive la trasformazione violenta di uno dei piloti bombardieri e l'incapacità del pilota di conservare la capacità di comprendere quello che sta facendo:

Lui pilota di un bombardiere, all'occorrenza poteva soltanto indicare quante città e quanti villaggi aveva bombardato, ma non quante persone aveva ucciso, e neppure come le aveva uccise[...] non lo sapeva si limitava a osservare i suoi strumenti e seguiva le indicazioni del suo marconista per guidare l'aereo [...] ma tra lui, Achille, e le sue vittime non c'era confronto le vittime non erano più nemmeno una razza inferiore, bensì qualcosa di vago, era come sterminare degli insetti [...] era una cosa astratta, puramente tecnica da considerare ormai soltanto per sommi capi, o meglio dal punto di vista finanziario [...] la morale era estirpata come un tumore maligno, l'odio era iniettato come uno stimolante contro un nemico che era un fantasma, se lui avesse visto un nemico reale prigioniero non avrebbe potuto odiarlo, certo combatteva contro un sistema contrario alla sua opinione politica, ma qualsiasi sistema, anche il più criminale, era un misto di innocenti e di colpevoli, [...] gli sembrava di non essere più una persona, ma soltanto un osservatore di lancette di orologi [...] il loro aereo era un computer volante, partiva volava fino alla meta, sganciava le bombe, tutto automatico, loro due avrebbero avuto soltanto una funzione di osservatori [...] l'uomo era un'illusione, o diventava una macchina senz'anima un apparecchio fotografico, un computer oppure una bestia".⁶⁵

Dürrenmatt rimanda al carattere strutturalmente militare dei dispositivi intelligenti. Tale suggestione riguarda anche tecniche che non erano ancora in funzione quando venne scritto il testo. Si va dalle tavole da tiro

63 Stiegler (2015).

64 Dürrenmatt (1986).

65 Dürrenmatt (1986/1993, 1025-1026 §XXI).

dei missili durante la seconda guerra mondiale, alle bombe intelligenti e ai droni che uccidono manovrati da esseri umani che si trovano a migliaia di chilometri dal teatro di guerra, fino al soldato robot. L'intelligenza artificiale nasce e si applica in primo luogo allo scenario di guerra⁶⁶ e ha come proprio obiettivo la perdita di vista del nemico concreto e la costruzione di una rappresentazione parziale e distante dell'oggetto dell'azione. La questione al centro della follia di Achille è la completa automatizzazione del processo e la sua velocità di esecuzione che impediva all'essere umano di essere tenuto nel processo decisionale, mentre il dispositivo non smetteva mai di sorvegliare l'obiettivo e il pilota stesso.

L'uomo non è più padrone della decisione su come agire, perché questa dipende dai dispositivi di sorveglianza. Come in ogni processo di automazione l'azione è parcellizzata e deindividualizzata in modo che nessuno sappia esattamente cosa stia facendo. La macchina è quindi in capo a un processo di progressiva deresponsabilizzazione della presa di decisione. Questo meccanismo vale sia quando si tratti di un nemico da battere, sia quando si tratti di un progetto di marketing in cui si debba ricercare il target giusto per essere sollecitato dal prodotto da comunicare, come nel caso dello scandalo Cambridge Analytica⁶⁷, sia quando si prende una decisione relativa alla valutazione del rischio di recidività di un condannato.

I dispositivi digitali interattivi sono, come ha suggerito Stiegler⁶⁸, delle ritenzioni terziarie in quanto costituiscono un'esternalizzazione della memoria, e conseguentemente della capacità cognitiva umana. Si frappongono tra gli esseri umani in capo al processo di interpretazione e altri i cui comportamenti vengono analizzati attraverso la traccia digitale dei loro agiti raccolta conservata, processata, manipolata per fare supposizioni su come interpretare quei comportamenti. Tale esternalizzazione della memoria umana finisce per privare l'umano della sua capacità cognitiva che viene delegata alla macchina, lasciando all'essere umano nulla se non il processo di proletarianizzazione e un forte sentimento di risentimento e di odio, che è, secondo Stiegler, causato da questa assenza di capacità:

l'assenza dell'epoca essendo da parte a parte dominata dalla tecnologia reticolare costitutiva del "capitalismo cognitivo" che è innanzi tutto lo *sfruttamento totalizzante*

66 Edwards (1997).

67 Per i dettagli su questo scandalo si veda il recente intervento a Ted Talk di Carole Cadwalladr che aveva lavorato a scoprire come funzionava il sistema di acquisizione dei dati degli utenti Facebook per essere il target di messaggi pubblicitari politici prima a favore di Brexit e poi nelle elezioni americane a favore di Donald Trump https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy

68 Stiegler (2018).

del farmaco digitale a favore del capitalismo puramente e semplicemente computazionale, che raggiunge *in ciò* lo stadio della *proletarizzazione totale* nel non-sapere assoluto, che è la denoetizzazione compiuta: come *nichilismo funzionale* – questa funzione essendo quella dell'*informazione* di cui la nozione rimante gravemente impensata.⁶⁹

Il farmaco rappresentato dal dispositivo digitale è completamente asseruito al capitalismo e consegna l'essere umano al non sapere dal momento che lo priva delle sue capacità cognitive necessarie alla presa di decisione. Abbiamo visto che Wiener, per esempio, suggeriva che, anche nell'integrazione con la macchina, le decisioni avrebbero dovuto restare nelle mani dell'essere umano⁷⁰ in quanto unico agente in grado di anticipare con l'immaginazione l'esito finale dell'azione.

Il processo di integrazione con la macchina promosso dalla nuova ondata di Intelligenza Artificiale, tuttavia non consente di lasciare la presa di decisione agli esseri umani. Come abbiamo visto gli strati di funzionamento dell'algoritmo di *machine learning* – in particolare le soluzioni attualmente più efficienti dette *deep learning* – non consentono di seguire il funzionamento della procedura in tutto il suo progresso. È inevitabile come aveva già compreso Wiener⁷¹ che: «It may be seen that the result of a programming technique of automatization is to remove from the mind of the designer and operator an effective understanding of many of the stages by which the machine comes to its conclusions and of what the real tactical intentions of many of its operations may be»⁷².

Tale asimmetria esperienziale della macchina è tattica perché i Big Data non possono essere controllati direttamente dall'operatore umano. Da una asimmetria gestionale si producono in conseguenza sia l'asimmetria cognitiva rispetto all'operatore umano sia la difficoltà oggettiva di contenere e sostenere la responsabilità dell'agire che viene artificialmente attribuita alla macchina, senza che la macchina abbia una piena *agency* morale. Se, come suggeriva Turing⁷³, l'individuazione dell'intelligenza è costruita socialmen-

69 Stiegler (2018, 63, trad. mia).

70 Cfr. § 2 per la discussione sulla responsabilità decisionali che avrebbero dovuto rimanere saldamente in mano all'essere umano, come dimostra anche discutendo il problema sulla cosiddetta, *push button war* Wiener (1960). «If we use, to achieve our purposes, a mechanical agency with whose operation we cannot efficiently interfere once we have started it, because the action is so fast and irrevocable that we have not the data to intervene before the action is complete, then we had better be quite sure that the purpose put into the machine is the purpose which we really desire and not merely a colourful imitation of it» (p.1358).

71 Wiener (1960).

72 Wiener (1960, 1357).

73 Turing (1948).

te⁷⁴ ed è anche il frutto dell'ignoranza di chi la attribuisce, e se la struttura della digitalizzazione come esternalizzazione della memoria produce il non-sapere assoluto e la proletarizzazione dell'umano⁷⁵, allora nessuno è in grado di dubitare dell'intelligenza della macchina, che determina il suo primato a partire dall'esternalizzazione delle capacità di memorizzazione dell'umano.

Da questa prospettiva, quindi, i metodi dell'IA, in particolare i metodi quantitativi usati dagli algoritmi per interpretare i Big Data di natura sociale non si porrebbero in una logica della competizione rispetto agli esseri umani, ma proporrebbero piuttosto un cambiamento del paradigma epistemologico e conseguentemente anche politico della presa di decisione e dell'analisi nell'ambito delle scienze sociali. Questo cambio di paradigma, però, come suggerisce ancora Stiegler⁷⁶, ha un carattere politico prima che epistemologico, e avrebbe come conseguenza, se non viene adottato con molta cautela, la perdita della capacità cognitiva umana di valutarne l'efficacia e la correttezza. Le decisioni prese dalle macchine non possono essere controllate, se non con sistemi *ad hoc*. Costruire dispositivi psicosociali di controllo degli algoritmi è una scelta che attiene al sistema di regole sociali e politiche per l'uso di questi strumenti, a meno di sposare l'interpretazione del determinismo tecnologico. Del resto abbiamo visto che la tecnologia si propone come possibile organizzazione della società, suggerisce un dispositivo complesso che introietta e proietta un certo modello di convivenza progettato da forze geo-politicamente e storicamente collocate, come nel caso dei Big Data⁷⁷. Tale disegno politico di impostazione tecnologica può essere controllato sia dal sistema del diritto, sia dalle regole di convivenza di una collettività. Come implementare gli algoritmi di intelligenza artificiale e a che cosa applicarli è una decisione sociale e politica che non può esserci imposta come un obbligo inevitabile, ma deve essere oggetto di una costante negoziazione.

Bibliografia

Agre, P. E. (2005), *The soul gained and lost: artificial intelligence as a philosophical project*, in Franchi S. Güzeldere G. (2005) (Eds.) *Mechanical*

74 Cfr. § 2 per maggiori dettagli sulla posizione di Turing al riguardo.

75 Stiegler (2015a).

76 Stiegler (2015b).

77 Come è ben noto la culla della maggior parte delle aziende Internet capaci di trattare i Big è nella Silicon Valley, in California (Morozov 2017).

- bodies, computational minds: artificial intelligence from automata to cyborgs*, Cambridge (Mass.): MIT Press: 153-173.
- Agre, P.E. (1997), *Computation and human Experience*, Cambridge: Cambridge Univ. Press.
- Ankwin, J., Larson, J. (2016) *Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*, in «Propublica» (30 dec. 2016), URL: <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>.
- Asaro, P. (2016) *The liability problem for autonomous artificial agents*, «AAAI Spring Symposium Series» Ethical and Moral Considerations in Non-Human Agents, URL: <https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12699/11949>
- Calude, C.S. & Longo, G. (2017) “The Deluge of Spurious Correlations in Big Data”. *Found Sci* (2017) 22: 595. <https://doi.org/10.1007/s10699-016-9489-4>
- Cardon, D., Cointet, J.P., Mazières, A. (2018) *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle*, «Réseaux» 2018/5 (n° 211): 173-220. DOI 10.3917/res.211.0173.
- Chun, W. H. (2019), *Queerying Homophily*, In Apprich, C., Chun, W. H., Cramer, F., Steyerl, H. *Pattern Discrimination*, Minneapolis: University of Minnesota Press and Meson Press, 59-97.
- Clark A., Chalmers D. (1998) *The extended mind* in «Analysis» 58:10-23.
- Cordeschi R. (2002) *The discovery of the artificial*. Dordrecht: Kluwer.
- Cramer, F. (2019), *Crapularity hermeneutics: interpretation as the blind spot of analytics, artificial intelligence, and other algorithmic producers of the postapocalyptic present*. In Apprich, C., Chun, W. H., Cramer, F., Steyerl H. *Pattern Discrimination*, Minneapolis: University of Minnesota Press and Meson Press, 23- 58.
- Dürrenmatt F. (1986/1993), *Der Auftrag*, Zurich: Diogenes, 1986; trad. It. *L'incarico*, in *Romanzi e Racconti* (a cura di Eugenio Bernardi), Torino: Einaudi- Gallimard, pp. 974-1032.
- Edwards P. (1997), *The Closed World*, Cambridge (Mass.): MIT Press.
- Epstein, R., Roberts, G., Beber, G. (Eds.) (2009), *Parsing the Turing Test*, Berlin: Springer Science.
- Franchi S. Güzeldere G. (2005), *Machinations of the mind: Cybernetics and Artificial intelligence from automat to cyborgs*, in Franchi S. Güzeldere G. (Eds.) *Mechanical bodies, computational minds: artificial intelligence from automata to cyborgs*, Cambridge (Mass.): MIT Press: 15-149.

- Fry H. (2019), *Hello World. Being Human in the Age of Algorithms*, New York: W. W. Norton & Company.
- Galison P. (1994) *The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision*, «Critical Inquiry», Vol. 21, No. 1 (Autumn, 1994), pp. 228-266.
- Hume, D. (1888) *A Treatise of human nature* [1739]. Ed. by Selby.Bigge L.A. Oxford: Clarendon Press.
- Kelleher, J. D., Tierney, B. (2018), *Data Science*, Cambridge (Mass.): Mit Press.
- Knight W. (2019) Google employees are lining up to trash Google's AI ethics council. Mit Technology Review, 1 apr. 2019. <https://www.technologyreview.com/s/613253/googles-ai-council-faces-blowback-over-a-conservative-member/>
- Latour B. (2010), *Cogitamus. Six lettres sur les humanités scientifiques*, Paris: La Découverte, (trad. it. *Cogitamus. Sei lettere sull'umanesimo scientifico*, Bologna: Il Mulino 2013).
- Licklider, J.C.R. (1960), *Man computer symbiosis*, «IRE Transactions on Human Factors in Electronics», Volume: HFE-1 , Issue: 1 , March 1960: 4 – 11; ristampato In Memoriam: J. C. R. Licklider 1915-1990 Research Report 61 Digital Equipment Corporation Systems Research Center August 1990 <http://memex.org/licklider.pdf>.
- Licklider, J. C. R. (1965), *Libraries of the Future*, Cambridge (Mass.): Mit Press.
- Licklider J. C. R., Taylor R. W. (1968), The Computer as a Communication Device. *Science and Technology*, no.76, Apr. 1968, pp. 21 31; ristampato In Memoriam: J. C. R. Licklider 1915-1990 Research Report 61 Digital Equipment Corporation Systems Research Center August 1990 <http://memex.org/licklider.pdf>
- McCarthy, J. (1959), *Programs with common sense*, in *Mechanization of thought processes*, in *Proceedings of a symposium held at the National Physical Laboratory on 24th, 25th, 26th and 27th November 1958*, Her Majesty's Stationary Office, London, URL: <http://www-formal.stanford.edu/jmc/mcc59.pdf>.
- McCarthy, J. (1971/1987), *Generality in artificial intelligence*, «Communication of the ACM», 30: 1030-1035. URL: <http://www-formal.stanford.edu/jmc/generality.pdf>.
- McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (31 agosto, 1955), A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, ristampato in AI Magazine Volume 27 Number 4 (2006), pp. 12-14, URL: <http://www-formal.stanford.edu>

- edu/jmc/history/dartmouth/dartmouth.html; (trad it «Sistemi intelligenti», XVIII, Dic. 2006, pp. 413-428.
- McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1988), *The appeal of parallel distributed processing*, New York: Morgan Kaufmann.
- Moor, J. H. (Ed.), (2006), *The Turing Test. The elusive standard of Artificial Intelligence*, Dordrecht: Kluwer Academic Publishers.
- Morozov E. (2017), *Silicon Valley: i signori del silicio*. Codice, Milano.
- Newell, A., Shaw, J.C., Simon, H. (1958), *Chess Playing programs and the problem of complexity*, «IBM Journal of research and development», October 1958, 2: 320-335; Reprinted in Feigenbaum E.A., Feldman J. (Eds) *Computers and thought*, Cambridge (Mass.): MIT Press, 1963: 39-70.
- Numerico, T. (2005), *Alan Turing e l'intelligenza delle macchine*, Milano: FrancoAngeli.
- Numerico, T. (2010), *Alle origini di Arpanet. Il contributo cognitivo di Norbert Wiener*. «Sistemi Intelligenti», vol. XXI: 533-542.
- Numerico, T. (2017) *La memoria e la rete*, in Bertollini A. Finelli R. (a cura di) *Soglie del linguaggio. Corpo, mondi, Società*, Roma: Roma Tre University Press, 81-102.
- O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown; trad. it. *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*, Milano: Bompiani, 2017.
- Osoba, O. A., & Welser IV, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation.
- Pasquale, F. (2015), *The black box society*, Cambridge (Mass): Harvard University Press.
- Rodotà, S. (2014), *Il mondo nella rete. Quali i diritti quali i vincoli*, Roma: Laterza, Repubblica.
- Rosenblueth, A. Wiener, N. Bigelow J. (1943) *Behavior, Purpose and Teleology*, «Philosophy of Science»,10,1:18-24.
- Rouvroy, A. & Berns, T. (2013), *Gouvernementalité algorithmique et perspectives d'émancipation: Le disparate comme condition d'individuation par la relation ?* «Réseaux», no 177,(1): 163-196. doi:10.3917/res.177.0163.
- Shieber, S. (Ed.) (2004), *The Turing Test. Verbal behavior as the hallmark of intelligence*, Cambridge (Mass.): Mit Press.
- Stiegler, B. (2015a), *La société automatique : Tome 1, L'avenir du travail*, Paris: Fayard.
- Stiegler, B. (2015b), *Platone Digitale*, Milano: Mimesis.
- Stiegler, B. (2018), *Qu'appelle-t-on panser?*, Paris: Les Liens qui Libèrent.

- Turing, A.M. (1948/2004), *Intelligent machinery*, in Copeland J. (Ed.) (2004) *The essential Turing*, Oxford: Clarendon Press, 410-432.
- Turing, A. (1950), *Computing machinery and intelligence*, «Mind», 59: 433-460; reprinted in Copeland (2004), *The essential Turing*, Oxford: Clarendon Press, 441-464.
- Turing, A.M. (1952/2004), *Can Automatic calculating machines be said to think?* discussione registrata alla BBC il 10 gennaio 1952 tra Turing, Max Newman, Richard Braithwaite e Geoffrey Jefferson, pubblicata in Copeland J. (Ed.) (2004), *The essential Turing*, Oxford: Clarendon Press 494-506.
- Varela, F. J., Thompson, E. Rosch, E.(1993) *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge (Mass.): Mit Press.
- Weizenbaum, J. (1966), *ELIZA-A Computer Program For the Study of Natural Language Communication Between Man and Machine*, «Communication of the ACM», 9, Number 1 (January 1966): 33-35.
- Weizenbaum, J. (1976), *Computer power and human reason. From judgement to calculation*, New York: W.H. Freeman and Company.
- Wiener, N. (1948/1965), *Cybernetics or Control and Communication in the Animal and the Machine*, Cambridge (Mass.): MIT Press.
- Wiener, N. (1950/1954), *The human use of human being*, Boston: Houghton Mifflin.
- Wiener, N. (1960) Some Moral and Technical Consequences of Automation. «Science». 06 May 1960: 131, Issue 3410:1355-1358.

