

SETTIMO TERMINI

L'IA TRA KATZUO ISHIGURO E FEDERICO FAGGIN

1. Introduzione
2. Ricezione pubblica e impatto sulla società e la vita quotidiana
3. Legami storici e rapporti con la fisica, con la cibernetica (e con Napoli)
4. Utopia di un mondo liberato dalla fatica
5. Scenari
6. Qualche ultima riflessione prima di concludere
7. A mo' di conclusione

ABSTRACT: AI BETWEEN KATZUO ISHIGURO AND FEDERICO FAGGIN

The themes proposed by the Journal for a critical discussion are crucial for understanding the impact that AI can have in Society (and also in everyday life), as is the case with all "science-based" cutting-edge technologies, i.e. those founded on and closely intertwined with basic frontier research. The questions posed can, in my view, be more satisfactorily addressed if seen from a long-term perspective with a historical reference to the origins and development of AI, its progress over the last few decades, the fluctuating consideration of it by both public opinion and the scientific community, its relations with other fields of research, in particular Cybernetics, whose heir it can in many respects be considered to be from the point of view of the sociology of science. This is the perspective from which useful paths can be indicated to clarify and deepen the themes posed even if no definitive answers can, presently, be given. The unusual reference to books such as those by Ishiguro and Faggin is intended to indicate the profound complexity and intricacies of the questions posed by AI.



1. Introduzione

Le tematiche proposte nel "Call for papers" sono cruciali per capire l'impatto che l'IA può avere nella società (e anche nella vita quotidiana), come peraltro tutte le tecnologie di punta "science-based", cioè che non sono applicazione di risultati ben noti e consolidati ma sono basate su nuovi risultati scientifici di frontiera. Le domande poste possono, a mio avviso, essere più agevolmente affrontate se viste in una prospettiva di lungo periodo con un richiamo storico alle origini e allo sviluppo dell'IA, al suo andamento nel corso degli ultimi decenni,

all'oscillazione della sua considerazione da parte sia dell'opinione pubblica sia della comunità scientifica, ai suoi rapporti con altri settori di ricerca, in particolare con la cibernetica di cui, per molti versi, può considerarsi l'erede. Per molti di questi temi mi limiterò essenzialmente a fornire una sorta di mappa sintetica con "puntatori" ad alcuni riferimenti bibliografici per chi volesse approfondire quanto solo accennato. Lo spazio dedicato a presentare e discutere suggerimenti provenienti da testi "insoliti" deriva dalla convinzione che le domande poste dalla Rivista possano meglio essere affrontate da una prospettiva che includa punti di vista di natura varia quando significativi. Una ricezione pubblica insolita e un programma di ricerca in una direzione non ortodossa sono indicate già nel titolo. Rinviano a quanto sarà scritto successivamente, desidero anticipare che Ishiguro narrando¹ del robot Klara e del suo rapporto con gli umani (e con la natura) ci fa riflettere - nella maniera più coinvolgente possibile - di come le sfide dell'IA possano contribuire a farci crescere interiormente e a conoscere meglio anche noi stessi. Faggin, pioniere tecnologico e inventore, tra l'altro, del microprocessore, pone in un modo assolutamente radicale la tesi che la coscienza non sia "riducibile" alla materia². L'IA, quindi, pone molte sfide concettuali, che sono cruciali per l'oggi pur essendo fortemente radicate nel passato.

2. Ricezione pubblica e impatto sulla società e La vita quotidiana

Partiamo da alcune domande. Che impatto hanno l'IA e le nuove tecnologie nella società e nella vita quotidiana? Le enormi possibilità che si aprono e i pericoli possibili sono qualcosa di cui il cittadino comune è cosciente? Sono questi dei temi su cui la stampa svolge una funzione di sensibilizzazione? La risposta

¹ K. Ishiguro, *Klara e il sole* (2021), tr. it. Einaudi, Torino 2021.

² F. Faggin, *Silicio. Dall'invenzione del microprocessore alla nuova scienza della consapevolezza*, Mondadori, Milano 2019. Vedi anche il sito della Fondazione Faggin <http://www.fagginfoundation.org/it/>.

non può che essere affermativa se possiamo leggere su un giornale come il New York Times frasi come quelle che seguono scritte da uno scienziato di altissimo livello:

By this time the public is well aware that a new age of machines is upon us ... The tendency of these new machines is to replace human judgment on all levels ... It is already clear that this new replacement will have a profound influence on our lives, but it is not clear to the man of the street what this influence will be. ... These New machine have a great capacity for upsetting the present basis of Industry, and of reducing the economic value of the routine factory employee to a point at which he is not worth hiring at any price. ... Moreover, if we move in the direction of making machines which learn and whose behaviour is modified by experience, we must face the fact that every degree of independence we give the machine is a degree of possible defiance of our wishes. ... It is only a humanity which is capable of awe, which will also be capable of controlling the new potentials which we are opening for ourselves³.

Prima di continuare, chiedo al lettore di indovinare l'anno in cui i brani precedenti sono stati scritti e chi ne sia l'autore. Ho cercato di indurre i lettori in una trappola temporale e non so con quanti questo gioco sia riuscito. I brani riportati sono stati scritti da Norbert Wiener e sono stati presi da un articolo che *avrebbe dovuto* essere pubblicato sul NYT nel 1949. Non lo è stato, in realtà ma solo per motivi casuali, come riportato dallo stesso NYT⁴ nel 2013, a causa di successive richieste di chiarimenti e revisioni da parte del direttore del giornale a cui l'autore ebbe difficoltà a rispondere trovandosi fuori degli Stati Uniti. Ma il punto che a noi interessa è che l'articolo è stato scritto nel 1949. Ci siamo accorti, leggendo quegli estratti, che l'articolo era vecchio di settantatré anni? Forse da sfumature linguistiche ma non certo per i contenuti. Questo ci permette di fare immediatamente due commenti. Il primo è che le domande che oggi ci stiamo giustamente ponendo si ponevano già prima che il nome IA fosse introdotto: solo un anno dopo che lo stesso Wiener aveva

³Cfr. N. Wiener, *The Machine Age*, http://monoskop.org/images/3/31/Wiener_Norbert_The_Machine_Age_v3_1949.pdf.

⁴ J. Markoff, in 1949, *He Imagined an Age of Robots - Norbert Wiener, the visionary mathematician whose essay "The Machine Age" languished for six decades in the M.I.T. archives*, New York Times, May 20, 2013. <https://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html>

pubblicato il suo libro sulla cibernetica. E un anno prima di un suo successivo libro nel quale riportava sue considerazioni sull'impatto di queste nuove tecnologie che già dal titolo indicava la direzione nella quale avremmo dovuto muoverci⁵. Certo dopo settant'anni di sviluppi possiamo esemplificare in modo più sofisticato e ricco ogni tema ma i problemi generali, di fondo, sono gli stessi. E se volessimo evidenziare continuità e differenze potremmo leggere un articolo appena uscito⁶ in parallelo a quello fortemente voluto dal NYT, pur senza successo, nel 1949. Gli argomenti specifici trattati sono, in questo articolo recente, molto più specifici ma le domande di fondo sono essenzialmente le stesse.

Anche per quanto riguarda i possibili pericoli (dell'IA), questi non sono qualcosa di nuovo legato agli ultimissimi sviluppi ma si possono utilmente rintracciare in quanto ipotizzato già a suo tempo da Wiener (considerati così cruciali già allora da avere avuto una richiesta di articolo dal New York Times, come abbiamo visto). Come scrive Katherine Hayles: «the problem that Wiener encountered was how to restrain this revolutionary potential of Cybernetics so that it would not threaten the liberal humanism that so deeply informed his thinking»⁷. Il problema, dunque, non è totalmente nuovo, tutt'altro. I problemi sono da aggiornare, ovviamente, anche alla luce di quello che nel frattempo è successo. Katherine Hayles, proprio nella stessa pagina, poche righe più sotto - riportando alcune analisi fatte sulla "militarizzazione dell'apprendimento" nei percorsi educativi degli Stati Uniti - riferendosi a una di queste, scrive che essa «indicates that Wiener's antimilitary stance was not sufficient to

⁵ N. Wiener, *The Human Use of Human Beings: Cybernetics and Society*, Houghton Mifflin, Boston 1950.

⁶ S. Johnson, *A.I. Is Mastering Language. Should We Trust What It Says?*, New York Times, April 15, 2022 (ultimo accesso 17 aprile 2022).

⁷ K. Hayles, *How we became post human*, The University of Chicago Press, Chicago 1999, p. 119.

prevent the marriage of war and cybernetics, a union that he both feared and helped to initiate».

Vorrei invitare il lettore a confrontare le sagge considerazioni di Wiener (nonostante la contraddizione appena riportata) con le “preoccupazioni” di Henry Kissinger⁸, recentemente espresse in un libro⁹. Libro e preoccupazioni da esaminare e maneggiare con estrema cura e direi anche un po’ di diffidenza, perché - a differenza delle preoccupazioni profondamente umanistiche di Wiener - Kissinger, presentando le paure dell’uomo comune per aspetti inquietanti di queste nuove tecnologie, non ci induce ad andare in direzione di un rafforzamento della trasparenza democratica - ad esempio, un maggiore controllo decentrato da associare a uno sviluppo della “cittadinanza scientifica” - ma spinge in direzione di un controllo verticistico della politica sulla ricerca. Lo stesso, a mio avviso, può dirsi delle perplessità avanzate contro l’IA da Bill Gates, così come quelle di Elon Musk e Jeff Bezos. Giuste, ovviamente, le critiche a un uso non controllato e non etico delle nuove tecnologie soprattutto in campo militare ma, come abbiamo notato, tutto questo dibattito era già presente da molti decenni e il modo con cui questi nuovi “convertiti” al problema ne parlano non presenta nessuna argomentazione innovativa. Inoltre, questi temi sono inseriti in un discorso generale su cosa sia l’IA di una vaghezza estrema e senza mai parlare dei meccanismi di potere nei quali queste considerazioni devono essere inserite se non si vuole che restino inutilmente retoriche (per chi le ascolta) o consolatorie (per chi le fa). Parzialmente incomprensibili mi appaiono, poi, le affermazioni di Stephen Hawking nelle quali la paura si alterna

⁸H. Kissinger, How the Enlightenment Ends, *Atlantic Monthly*, June 2018 (<https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/>), apparso in italiano il 21 luglio 2018 su Robinson, La Repubblica (https://www.repubblica.it/robinson/2018/07/21/news/un_intelligenza_da_paura_kissinger_mette_in_guardia_dalle_ai-300769791/).

⁹ H. Kissinger, E. Schmidt, D. Huttenlocher, *The Age of AI and our Human Future*, Little, Brown and Company, New York 2021.

all'ottimismo, in un miscuglio intelligente e banale allo stesso tempo. La saggezza di Wiener la ritroviamo, invece, in molti recenti interventi pubblici di Giorgio Parisi. Tra le analisi puntuali si devono ricordare quelle di Karen Hao come accennerò - brevemente - in conclusione. Le analisi devono essere fatte con serietà e tenendo conto dell'estrema complessità del problema. Wiener rimane un riferimento basilare anche oggi, grazie alla ricchezza del suo pensiero. E per apprezzarlo pienamente ci è di guida la profonda analisi effettuata da Leone Montagnini¹⁰ che ce lo presenta - correttamente - non solo come un raro matematico-filosofo ma anche come un grande interprete del nostro tempo. Parlando di Wiener e raccontando l'episodio del 1949 che lo riguarda, ho introdotto praticamente tutti i temi che tratterò e che formano il mio abbozzo di risposta a (quasi) tutte le domande poste nel Dossier.

3. Legami storici e rapporti con La fisica, con La cibernetica (e con Napoli)

In quest'ambito la domanda relativa alle implicazioni di ciò che è definito "incontro tra IA e meccanica quantistica" - pionieristicamente avanzata da Roger Penrose - può essere maggiormente capita e approfondita se vista nell'ottica di un esame più vasto dei problemi cruciali dell'IA, come ad esempio delineato recentemente da Giorgio Parisi che ha dato indicazioni su aspetti da approfondire ribadendo che la maggior parte delle *applicazioni* dell'IA si possono tutte ricondurre a un settore *molto specifico* di indagini. Si fa riferimento a idee (reti neurali) che hanno visto la comunità scientifica italiana (e Napoli in modo specifico) in una posizione originale nella quale è stata la comunità dei fisici a introdurre quelle che negli anni

¹⁰ L. Montagnini, *Harmonies of Disorder: Norbert Wiener: A Mathematician-Philosopher of Our Time*, Springer Biographies, Springer International Publishing, 2017.

'60 erano tecnologie di punta associate a una disciplina scientifica allora emergente, la cibernetica, appunto.

È importante ricordare che Wiener è venuto varie volte in Italia e in particolare proprio a Napoli¹¹ nel contesto di quello che è stato chiamato il *risveglio scientifico* di questa città¹². Per quanto riguarda l'aspetto che stiamo considerando l'autore del risveglio è stato Eduardo Caianiello¹³. Il suo interesse per la cibernetica si focalizzò sulla possibilità di costruire modelli del cervello adeguati allo studio dei processi mentali, sulla scia, in particolare, di un libro di Ashby¹⁴. Il suo contributo scientifico fondamentale in questo nuovo campo¹⁵ è apparso nel 1961 sul primo numero di una nuova rivista dal nome - per allora - insolito (biologia teorica) che lascia intravedere contenuti fortemente innovativi per l'epoca. Questo lavoro non solo è alla base di tutti i suoi lavori successivi sulle reti neurali, ma contiene anche tutti gli elementi significativi del suo modo di affrontare i problemi della cibernetica. Lo scritto, oltre a contenere risultati originali, è un vero e proprio manifesto programmatico ancora oggi significativo¹⁶. Questo lavoro, un modello matematico di rete neurale, è scientificamente innovativo rispetto a modelli precedenti perché basato su intuizioni di tipo fisico (piuttosto che precipuamente logico) sul suo funzionamento. Le reti neurali hanno svolto e svolgono un ruolo centrale anche su alcuni degli sviluppi più importanti dell'IA,

¹¹ L. Montagnini Quando Wiener era di casa a Napoli, *La rivista del Centro studi di Città della scienza*, 21 marzo 2016, <http://www.cittadellascienza.it/centrostudi/2016/03/quando-heisenberg-e-wiener-erano-di-casa-a-napoli/>

¹² G. Barone, P. Greco, L. Mazzarella, A.M. Liquori, *Il risveglio scientifico negli anni '60 a Napoli*, *Bibliopolis*, Napoli 2013.

¹³ Cfr. S. Termini, *Caianiello, Eduardo Renato*, *Dizionario Biografico degli Italiani*, 2017.

https://www.treccani.it/enciclopedia/eduardo-renato-caianiello_%28Dizionario-Biografico%29/ e anche S. Termini (ed), *Imagination and Rigor, Essays on Eduardo Caianiello's Scientific Heritage*, Springer, 2006.

¹⁴ W. Ross Ashby, *Design for a brain*, London, 1952.

¹⁵ E.R. Caianiello, *Outline of a theory of thought-processes and thinking machines* in «*Journal of theoretical biology*», I (1961), 2, pp. 204-235.

¹⁶ Cfr. <http://www.treccani.it/enciclopedia/la-cibernetica> (*IL-Contributo-italiano-alla-storia-del-Pensiero:-Scienze*).

come richiamato anche da parte di Giorgio Parisi nel suo libro rivolto al grande pubblico¹⁷, parlando dell'introduzione del disordine. Il rapporto di Parisi con l'IA è testimoniato oltre che da molte sue dichiarazioni pubbliche anche dal conferimento del titolo di Socio onorario che l'Associazione Italiana per l'Intelligenza Artificiale gli ha conferito. Se ricordiamo questo aspetto è per mettere in evidenza una sorta di positiva anomalia italiana che ha visto la comunità dei fisici interessati a introdurre e finanziare nuove ricerche in campi di frontiera di nuove tecnologie, come ha fatto negli anni '60 del secolo scorso con la cibernetica. Sia pure in forma minore, quanto riportato sopra va nella stessa linea e con grandi potenzialità, se perseguito efficacemente perché è stata proprio l'interazione interdisciplinare di metodologie scientifiche diverse che dialogavano tra loro a produrre gli intensi e proficui dibattiti avvenuti a metà del secolo scorso¹⁸.

4. Utopia di un mondo liberato dalla fatica

Per quanto riguarda "L'utopia di un mondo liberato dalla fatica" questa non è un'utopia e non ha bisogno di attendere "un'industria (5.0)". Profetizzata, o, in modo più accurato, semplicemente "prevista" da Keynes negli anni '30 del secolo scorso¹⁹, è realtà già oggi - *in linea di principio*. Cerco di essere più chiaro. Le sue previsioni riguardo allo sviluppo della tecnologia si sono puntualmente avverate. Quindi non avremmo bisogno di attendere "un'industria (5.0)". Certo ulteriori passi nella stessa direzione sono auspicabili. Ma il problema da affrontare è, come detto

¹⁷ G. Parisi, *In un volo di storni. Le meraviglie dei sistemi complessi*, Rizzoli, Milano 2021,

¹⁸ Cfr. L. Montagnini, M.E. Tabacchi, S. Termini, *Out of a creative jumble of ideas in the middle of last Century: Wiener, interdisciplinarity, and all that*, in «Biophysical Chemistry», 208, 2016, pp. 84-91; M.E. Tabacchi, S. Termini, *The human use of human beings": Interdisciplinarity, transdisciplinarity and all that in Biophysics and beyond*, in «Biophysical Chemistry», 229, 2017, pp. 165-172

¹⁹ J.M. Keynes, *Possibilità economiche per i nostri nipoti* (1930), Adelphi, Milano 2009.

prima, quello della reazione della società di fronte a un cambiamento di questo tipo. E il Keynes così preveggenente riguardo allo sviluppo tecnico si è invece completamente sbagliato riguardo al secondo aspetto. La società non è riuscita a modificare un modello di sviluppo che impedisce il pieno dispiegarsi degli effetti dello sviluppo tecnologico. Ma, a mio avviso, se oggi è aumentata la disoccupazione e la precarietà - nonostante le previsioni di Keynes si siano verificate per quanto riguarda lo sviluppo della tecnologia -, volendo affrontare il contributo che lo sviluppo di “un’industria (5.0)” può dare per la realizzazione di “un mondo liberato dalla fatica”, dobbiamo pur porci la domanda di quanto le condizioni al contorno - sociali, economiche, politiche - possano favorire o soffocare nella culla tale possibilità, indipendentemente da quanto siamo riusciti a sviluppare “un’industria (5.0)”. Per questi motivi, questa domanda la derubricherei dalla lista di ciò che riguarda l’IA da sola e creerei una nuova lista che, con un titolo alla Lina Wertmuller, chiamerei: «Ciò che l’IA può aiutarci a fare se riusciamo a risolvere contemporaneamente qualche altro problema». Ma torniamo a Keynes. Le sue idee - mi riferisco proprio e solo alle *idee*, in questo momento, anche se un discorso più ampio sarebbe da fare proprio in relazione ai temi della tecnologia e dell’IA²⁰ - a proposito del tema del liberarsi dalla fatica, sono state considerate ancora degne di essere ridiscusse se nel 2009 è uscito un volume²¹ nel quale vari economisti rileggono questo piccolo breve scritto degli anni '30 del Novecento. Ricordiamo che anche nell’edizione italiana il suo breve scritto non compare da solo ma è seguito da una nota di Guido Rossi, con lo stesso titolo ma col punto interrogativo. È quindi il caso che chi si interessa di

²⁰ Vedi, ad esempio, Daniela Palma, *L’Italia, il declino e il falso mito dello “sviluppo senza ricerca”*, in «Nuova Lettera Matematica», 4, 2021, pp. 9-20; e A. Bellocchi, G. Travaglini, *Attualità di Keynes*, in «Nuova Lettera Matematica», 4, 2021, pp. 37-44, nonché tutto il dossier su *Ricerca e Innovazione* contenuto nello stesso fascicolo della Rivista.

²¹ L. Pecchi, G. Piga (a cura di), *Revisiting Keynes: Economic Possibilities for our Grandchildren*, MIT Press, 2008.

industria 5.0 lo faccia rileggendo queste considerazioni. Da sviluppare, da adattare alla situazione odierna ma fondamentalmente corrette per quanto riguarda il tema centrale.

5. *Scenari*

Avanzando un tentativo di risposta all'ultima domanda, direi che gli "scenari narrativi ... [che] si squadernano di fronte a *sapiens*" sono fondamentalmente gli stessi di cinquanta/settanta anni fa quando l'IA (e, un po' prima, la cibernetica) fecero la loro comparsa. Naturalmente con una maggiore consapevolezza delle vie che possono essere più agevolmente percorse. Un esempio cruciale è fornito dalla traduzione automatica. È interessante confrontare il modo di impostare il problema negli anni '50 (e i magri risultati conseguiti) con le impostazioni più recenti che invece hanno prodotto risultati notevoli. La metodologia che sta dietro è profondamente diversa. Il tema è da collegare anche a domande sulla natura epistemologica di tecniche come quelle del *deep Learning*.

Un esempio significativo che può contribuire a chiarire il problema è dato dai tempi intercorsi tra la formulazione di un problema e la sua soluzione. Nel caso del gioco degli scacchi i tempi sono stati lunghi, in quello del gioco del GO, ridottissimi, pur essendo quest'ultimo molto più complesso del primo. Ciò è dovuto al fatto che di recente abbiamo cominciato a usare tecniche e tecnologie diverse, modificando l'impostazione metodologica. Purtroppo queste nuove tecnologie non sono frutto di nuove acquisizioni teoriche.

Quindi, questi sono temi e settori nei quali è cambiato moltissimo ma, apparentemente, gli scenari narrativi non sono cambiati. Oggi siamo riusciti a fare cose che 50 anni fa non si riuscivano a fare. Basta. Sembrerebbe che non ci sia molto altro da dire. Se non che è cambiata la metodologia usata e l'epistemologia che questo cambio di metodologia ha giustificato. Ma questo riguarda

un non detto, soprattutto perché realizzato in forma autocelebrativa e poco critica. Occorrerebbe, cioè, una profonda analisi critica delle metodologie usate.

Per quanto riguarda gli “scenari ... auto-narrativi“, rimanendo alla superficie, potremmo dire che essi risentono sia del diverso contesto socio-economico sia dell'illusione che il “nuovo” sia veramente (e, radicalmente) nuovo e non un (sia pur profondissimo) aggiornamento di tematiche novecentesche. E mi riferisco qui alla profonda connessione esistente tra quanto si fa oggi e i progetti di ricerca di settant'anni fa.

Parlando di “scenari”, desidero adesso introdurre i progetti di ricerca (secondo me è il modo corretto di chiamarli) di Ishiguro e Faggin. Progetti visionari ed entrambi di fantasia sia pure di fantasie diverse (ma, poi, quanto realmente diverse?). Quello di Faggin è un classico progetto di ricerca scientifico, almeno così appare. In realtà non è un progetto qualsiasi. Oltre a essere estremamente ambizioso si muove in una zona scivolosa e spigolosa insieme, come direbbero molti colleghi scienziati. Quello di Ishiguro è un romanzo, non un progetto di ricerca, mi sembra che stiano obiettando altri, colleghi e non colleghi all'unisono. Desidero controbattere a mia volta: sembra un romanzo ma in realtà è il *report* di un progetto di ricerca condotto a termine e già attuato. Esempio di quei momenti felici nei quali la ricerca di punta, di frontiera, riesce a trasformarsi in applicazioni tecnologiche di punta che poi vanno pure in produzione. Diventano “commodities”. Nella *realtà dell'immaginazione*, ovviamente. Ma nel mondo di *Klara e il sole* i risultati tecnologici raggiunti presuppongono che un qualche progetto di ricerca teorico “alla Faggin” sia stato portato a termine con successo. Un progetto, cioè, nel quale sia stato risolto il problema di cosa sia la coscienza, analizzandola scientificamente.

Ma di che parla Faggin e cosa si propone di fare? Riporterò qui di seguito alcuni passi tratti - in modo non sistematico - dal suo

libro, lasciando che il lettore si faccia un'idea, senza miei commenti.

... avevo cominciato a prendere sul serio l'idea che la coscienza potesse essere un aspetto fondamentale della natura, presente in qualche modo già negli atomi e nelle molecole di cui tutto è fatto. Quest'idea affiorò un po' alla volta, vista l'impossibilità di spiegare l'emergere della coscienza dalla complessità del nostro cervello. ... Pensavo: Come può una struttura fisica fatta solo di aspetti esteriori produrre aspetti interiori semantici? Il concetto di complessità non ha nulla a che fare con le sensazioni e i sentimenti del mondo interiore. Di fatto, i computer d'oggi, con tutta la loro complessità, non hanno neanche una briciola di consapevolezza. Per me non rimaneva alternativa: il mondo interiore dev'essere fin dall'inizio una proprietà irriducibile di tutto ciò che esiste²².

e, ancora:

Molti scienziati e filosofi affermano che non può essere così perché le implicazioni di quest'idea porterebbero a contraddire le interpretazioni riguardanti la natura della realtà, basate sulle teorie matematiche che hanno già ottenuto verifiche sperimentali rigorose. Altri sostengono che quest'idea sa di religione. E scienza e religione non si devono mischiare.

Questo argomento mi affascinava perché aveva il potenziale di spiegare e di unificare l'esistenza delle realtà esteriore e interiore che stavo esplorando da quasi vent'anni. Scienza e spiritualità, finora irconciliabili, avrebbero potuto trovare un'unione profonda anziché una semplice giustapposizione di convenienza.

Decisi quindi di ritirarmi completamente da tutte le mie altre attività per concentrarmi sullo sviluppo di un modello della realtà basato sul presupposto che la consapevolezza sia una proprietà fondamentale e irriducibile della natura... Non sono d'accordo con quegli scienziati che sostengono che l'universo non ha uno scopo e un significato e che noi siamo macchine destinate a essere superate dai nostri computer. Prima di prendere sul serio quest'interpretazione della realtà, dobbiamo riconoscere che essa si basa sul postulato che solo la materia esiste, una premessa non provata che non può spiegare l'esistenza della coscienza che ciascuno di noi sperimenta nella sua interiorità²³.

Rinvio il lettore a una lettura attenta dell'ultimo capitolo del libro di Faggin, riportando qui due passaggi da un'appendice conclusiva:

L'intelligenza artificiale, che oggi rappresenta la punta di diamante del progresso tecnologico umano e grande motivo di orgoglio, potrebbe fornirci l'esperienza necessaria - ma non sufficiente - per riportare l'umanità a riscoprire il vero mistero della sua natura, perché finirebbe per dimostrare clamorosamente l'impossibilità di creare macchine coscienti²⁴.

²² F. Faggin, *op. cit.*, p. 225

²³ *Ibid.*, pp. 226-227.

²⁴ *Ibid.*, pp. 265-266.

A mio avviso, il vero pericolo dei progressi della robotica e dell'intelligenza artificiale non è quello di creare macchine che prenderanno il sopravvento sull'umanità perché saranno più perfette di noi. Il vero pericolo è che uomini malvagi possano causare danni seri all'umanità e all'ecosistema controllando computer e robot sempre più potenti soltanto per il loro interesse. Ma allora sarà l'uomo, non la macchina, a causare il problema. Questa è una grande sfida che la società dovrà affrontare il più presto possibile²⁵.

Le visioni di progetti scientifici molto futuribili (estremamente stimolanti nonostante abbiano una probabilità alta di fallimento) e la visionarietà dell'alta letteratura (o delle arti figurative) si intrecciano e potenziano a vicenda ma non si tratta di fantasie da quattro soldi, di sogni a buon mercato. L'alta letteratura, dal suo lato, prende sul serio l'ipotesi scientifica e la usa per approfondire la nostra conoscenza di noi stessi, svelandoci di noi molto di più di quanto molti articoli scientifici di psicologia riescono a fare. Per almeno due motivi; perché ci analizza, analizza i nostri comportamenti possibili di fronte a una profonda novità - non all'ignoto in sé ma a ciò che non si conosce di qualcosa che è di fronte a noi per la prima volta (come di fronte allo scoppio della prima bomba atomica e alla fine del progetto "Human genome"). Ma qui l'impatto è ancora più coinvolgente e Ishiguro lo descrive in modo delicatissimo e profondo nello stesso tempo, come se fosse già avvenuto. Ci viene da pensare che se dovesse avvenire non potrebbe che svolgersi con le implicazioni e le modalità da lui descritte.

Non possiamo non notare (e osservare con rispetto e attenzione) le paure che riporta Ishiguro: «Allora, Klara, il fatto è che l'ansia riguardo agli AA», cioè gli amici artificiali, «sta dilagando. La gente sostiene che siete diventati troppo intelligenti»²⁶.

E fin qui è ciò che noi ci potremmo aspettare - magari con un semplice cambiamento del tempo verbale: "diventerete" al posto di "siete diventati". È, invece, assolutamente imprevisto (almeno per me) quello che segue, ciò che Ishiguro scrive immediatamente dopo:

²⁵ *Ibid.*, p. 305.

²⁶ K. Ishiguro, *op. cit.*, p. 260.

Le persone si spaventano perché non riescono a seguire quello che vi succede dentro. Vedono quello che fate. Riconoscono che le vostre decisioni e i vostri suggerimenti sono sensati e affidabili, quasi sempre corretti. Ma non sono contente di non sapere come ci arrivate. Da questo nasce la loro reazione, il pregiudizio. E noi dobbiamo combatterlo. Dobbiamo dir loro, d'accordo, siete preoccupati perché non capite come pensa un AA. Perfetto, allora diamo un'occhiata al motore. Decodifichiamo. Non vi piacciono le scatole nere sigillate? Benissimo, apriamole, allora. Quando ci avremo guardato dentro, non soltanto le cose faranno meno paura, ma impareremo. Impareremo meraviglie nuove²⁷.

Qui invece il nostro raffinato (intellettualmente raffinato) giapponese, naturalizzato britannico, va oltre toccando punti estremamente importanti dal punto di vista scientifico ed epistemologico. Intanto vorrei osservare che non sono queste le paure che, oggi almeno, toccano un cittadino medio (italiano almeno); non so per uno britannico o giapponese. Il nostro tocca invece con sottile intelligenza un punto chiave che riguarda la paura, la paura in generale, la paura vera, non un timore generico per il nuovo. E, poi, anche un altro punto che riguarda esplicitamente l'aspetto scientifico dell'IA, ma su questo tornerò dopo. Abbiamo paura dell'ignoto, di ciò che non conosciamo quando questo ci contorna e ci avvolge. Il bambino ha paura del buio perché il buio avvolge e vela, nasconde quello che c'è. Ciò che afferma Ishiguro è più profondo. Afferma - nel brano riportato sopra - che: «*Le persone si spaventano perché non riescono a seguire quello che vi succede dentro ... non sono contente di non sapere come ci arrivate*» a quei suggerimenti e decisioni che pur si apprezzano per la loro affidabilità e sensatezza.

Qui si toccano due punti cruciali: uno comportamentale. Se sappiamo come si arriva a un certo risultato non dovremmo averne paura anche se si tratta di qualcosa di nuovo e di insolito. Questo è un punto cruciale per il cittadino comune di fronte a situazioni inattese. Queste derivano normalmente dalla scienza. Oppure, lasciatemelo dire, dalla magia. Entrambe presentano al cittadino comune risultati imprevisti. Ma mentre in quest'ultimo

²⁷ *Ibid.*, p. 260.

caso (la magia) non viene detto come si è arrivati al risultato (e qui importa poco se quello a cui ci troviamo esposti è vero, al di fuori delle leggi naturali, inganno, suggestione o cos'altro ancora). Il punto chiave, cruciale è che non viene (e, in un certo senso, non può venire) detto "come ci si è arrivati". Nel caso della scienza, invece, tutti possono sapere com'è andata. In linea di principio, perché poi, in pratica, bisogna avere studiato e, per esperimenti come quelli del CERN, non possiamo controllare di persona. Ma la trasparenza c'è ed è esplicitamente esibita. Sono chiaramente enunciati, per chiunque voglia e possa eseguirli, i protocolli da seguire per capire, verificare, controllare (magari in ordine inverso).

Passiamo adesso all'altro punto che, incredibilmente, è introdotto da Ishiguro. Potremmo dire, ma in senso buono, che lo fa surrettiziamente. Comunica al lettore una caratteristica dell'IA odierna quasi senza fargliene accorgere. Gli algoritmi migliori, come, ad esempio quelli legati alla traduzione automatica funzionano oggi molto bene (se non benissimo) soprattutto se paragonati a quelli di cinquant'anni fa. Ma questo non avviene perché oggi abbiamo una teoria linguistica più sviluppata di quelle di allora per quanto riguarda questo aspetto.

La maggior parte dei risultati recenti più interessanti dell'IA sono stati ottenuti proprio mediante tecniche e percorsi che *non permettono di capire*, almeno oggi, *come* ai risultati si sia arrivati. Proprio quello che lui scrive. Coincidenza, caso, intuizione creativa di uno scrittore di genio oppure - fatto salvo quanto scritto prima - anche frutto di un accurato studio della letteratura (scientifica, in questo caso) rilevante per questo problema? Sarei indotto a credere a quest'ultima ipotesi. D'altronde il libro è del 2021. Il suo libro precedente è del 2015. Gli anni in cui si è concentrato su questo tema andrebbero quindi dal 2015 al 2021. E la vittoria dei computer sui campioni di GO è del 2016. Un risultato da citare perché è proprio quello

più eclatante (e inaspettato) e che ha avuto una risonanza giornalistica notevole che ha coinvolto anche il cittadino comune. E poi ci sono le profonde osservazioni di von Neumann sulla complessità degli automi su cui, adesso, non ci possiamo soffermare ma che devono far parte di una riflessione complessiva sul percorso seguita dall'IA. Ed è proprio quello di cui Ishiguro, indirettamente e in modo lieve, ci ha parlato.

6. Qualche ultima riflessione prima di concludere

Non sappiamo cosa veramente ci attende nel futuro, grazie all'IA o per sua colpa. Come non lo abbiamo mai saputo per nessuna di tutte le altre innovazioni, concettuali e tecnologiche, che l'essere umano ha ideato e realizzato. Possiamo solo avere indicazioni. Da usare come guida. Da interpretare. Vorrei solo notare che sembra che gli storici affermino che ogni grande svolta - economica, tecnologica, culturale - ogni passaggio che viene definito *epocale*, in realtà raramente sono sentiti come tali da coloro che li stanno vivendo mentre, a volte, viene così definito - da chi lo vive - qualcosa che i futuri storici considereranno di interesse men che mediocre.

Inoltre, cosa forse più importante, anche nelle trasformazioni più radicali vi è sempre una forte continuità con ciò che è avvenuto prima. Anzi, potremmo dire che i passaggi più innovativi sono proprio quelli che sono profondamente intrecciati con ciò che li ha preceduti. Cioè, un cambiamento è profondo e non effimero se esprime qualcosa di significativo che, trasformandola, si *radica* nella storia precedente.

Per questo ritengo che un'analisi dell'IA debba tener conto di molte delle cose che l'hanno preceduta e accompagnata e di altri aspetti dell'agire umano che - a prima vista - possono sembrare non particolarmente o strettamente connessi. A questo riguardo particolarmente significative appaiono analisi come quelle che si possono trovare sul *MIT Technological review* - almeno fino a

quando era diretta da Karen Hao²⁸. E di cui dobbiamo tener conto quando andiamo ad analizzare l'influenza a cui sistemi apparentemente oggettivi sono soggetti. Risentono – in modo devastante e inaspettato – di pregiudizi, trasmessi nel modo in cui sono stati progettati, pur senza volerlo. Oppure di come la potenza di queste nuove tecnologie possa amplificare tendenze negative proprie del sistema economico sociale. Ma dobbiamo essere in grado di distinguere e sceverare. I segnali di pericolo avanzati da Karen Hao sono in continuità e vanno nella direzione indicata da Wiener nel 1949. Le paure di Kissinger non riguardano questi aspetti ma il timore che il “potere”, qualunque definizione vogliamo darne, possa perdere il controllo su ciò che avviene. E la risposta che dovremmo dare a lui è che dobbiamo andare verso un forte decentramento di tutti gli apparati, restituendo, in questo modo, alla “rivoluzione informatica” tutte le possibilità umanistiche che al suo apparire erano state, correttamente, intraviste recuperando anche lo spirito umanistico dell'analisi wieneriana.

7. A mo' di conclusione

L'aver inserito l'IA tra le futuristiche utopie letterarie di Ishiguro e l'ambizioso programma di ricerca di Faggin, che non si colloca tra le idee pacificamente accettate è stata un po' provocatoria ma, contemporaneamente ha voluto – costruttivamente – indicare lo spettro enorme entro il quale considerazioni che a essa si riferiscono possono localizzarsi in modo significativo. È vero, infatti, che, in forme diverse, l'IA si può trovare “quasi ovunque”. Una conclusione che, se da un lato rafforza la nostra convinzione di avere a che fare con qualcosa che può profondamente

²⁸ Cfr. K. Hao, A.P. Hernández, *How the AI industry profits from catastrophe*, April 20, 2022 <https://www.technologyreview.com/2022/04/20/1050392/ai-industry-appen-scale-data-labels/>; e K. Hao, *Artificial intelligence is creating a new colonial world order*, April 19, 2022 <https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/>.

incidere nello sviluppo della società e anche nella quotidianità delle nostre vite, da un altro lato ci lascia un po' insoddisfatti pensando al vecchio detto che "SE tutto è X ALLORA niente è X". Quindi è bene cominciare a fare, proprio al momento di concludere, alcune distinzioni. L'IA per la sua presentazione che tocca l'immaginario tende a svolgere una funzione mitopoietica, come ha fatto a suo tempo la cibernetica. Forse uno dei pochi settori che sono stati capaci di generare nuovi miti dopo l'aurea temperie del mondo greco. Mi chiedo cosa avrebbe detto Carl Gustav Jung di questo fenomeno e auspico che qualche suo brillante nipotino ce lo dica al suo posto.

Dobbiamo riconoscere questo piano e collegarlo - mantenendolo strettamente distinto - a quello molto più "concreto" dei risultati specifici. Molto disparati questi ultimi e frammentari. Che non delineano in alcun modo la forza e la potenza di una teoria unitaria, come sono state - ai loro tempi - la teoria della gravitazione di Newton (non a caso definita sistema del mondo) o la precedente *unificazione* dei moti terrestri e celesti compiuta da Galileo o, infine, la teoria di Maxwell che, a un tempo, unificava elettricità e magnetismo e forniva - con le sue equazioni - uno strumento tecnico per la progettazione di sistemi concreti. Nel nostro caso abbiamo quasi sempre a che fare con tecniche che producono risultati estremamente efficienti - come, ad esempio, quelli elencati nel NYT citato nella nota 6 - MA, allo stesso tempo, frammentari e incomprensibili. *Incomprensibili* in un senso molto specifico: non è del tutto chiaro *come* certi risultati siano stati ottenuti. Rimane il disappunto che - nonostante l'efficienza tecnologica di alcuni risultati - nessun vero passo in avanti sia stato compiuto nella direzione di "capire" veramente la natura dei problemi affrontati. Nel caso della "traduzione", ad esempio, i risultati non hanno alcuna relazione con progressi compiuti nella comprensione di *cosa sia* un linguaggio naturale. È questo che può generare, assieme alla soddisfazione per

l'avanzamento tecnico, allo stesso tempo frustrazione e paura. Frustrazione per l'incapacità, di fronte a una tecnica che produce risultati inaspettati, di riuscire a capire le ragioni teoriche di tale successo. La paura nasce, a sua volta, dalla sensazione che i risultati tecnici possano prendere il sopravvento sugli intendimenti di chi a tali potenti tecnologie si è affidato o ha preso decisioni in base a esse. Quante volte, anche a sproposito, abbiamo sentito accusare "l'algoritmo" di esiti incongrui.

Una situazione inedita rispetto ad altre nelle quali le scienze coinvolte erano la meccanica (crollo di ponti o dighe), l'elettromagnetismo (i due famosi Black-out di New York) o la chimica (Bhopal). Tutti casi nei quali in nessun modo ci si è nascosti dietro la poca trasparenza delle tecnologie utilizzate e delle scienze che a esse stavano dietro.

Possiamo pensare che il tempo trascorso sia troppo breve per avere permesso l'emergere e la costruzione di una teoria unitaria dei fenomeni legati all'emergere della nuova nozione di informazione, in tutti i suoi risvolti, e quindi si tratti solo di continuare a lavorare lungo la direzione già percorsa e aspettare. Ma come abbiamo già notato vi sono alcune anomalie: c'è tutta una famiglia di discipline che negli ultimi settanta anni sono nate, divenute centrali, poi declinate o scomparse, sovrapponendosi senza coincidere, differenziandosi per metodologie e ambizioni. Ma non si tratta solo di questo perché poi ci si scontra sempre con problemi concettuali non risolti, come già osservato.

Le scienze dell'informazione, di cui, oggi, l'IA è uno dei protagonisti principali (come ieri era stata la cibernetica) hanno bisogno di una profonda analisi che chiarisca il loro *status* epistemologico, evidenzi le loro caratteristiche specifiche e puntualizzi le differenze con altre discipline. Non c'è nulla di strano in questa richiesta. Rimanendo all'interno della sola fisica, la meccanica ha alcune caratteristiche che la

differenziano dall'elettromagnetismo. Ed entrambe le teorie hanno forti differenze con la termodinamica.

La situazione attuale potrebbe anche nuovamente portare al ritiro, da parte della società, di quella concessione di fiducia incondizionata che è stata periodicamente concessa (e ritirata) all'IA (e, in generale, alla scienza). Forse può essere utile descrivere con maggiore dettaglio questo aspetto. C'è stato uno sviluppo incredibile e inimmaginabile di risultati (tecnici) nuovi; questo ha corrisposto a una concessione ampia di fiducia sia pure associata sempre, ovviamente, a riserve e paure. Fiducia connessa, in qualche modo, all'assunzione che la via seguita e intrapresa fosse quella giusta e, come tale, degna di essere perseguita. Una volta che quel percorso mostrava i suoi limiti subentrava un calo di fiducia. Quest'ultimo fenomeno in sé non è negativo perché indice di una reattività della società che mostra di saper distinguere - attraverso cammini suoi propri - la sottile differenza epistemologica che intercorre tra le tecnologie frutto e sviluppo di teorie scientifiche pienamente sviluppate e tecnologie molto potenti che non hanno dietro di loro uno scudo teorico forte.

L'IA è, al momento, protagonista indiscussa della società del nostro tempo. Perché lo rimanga sono necessarie almeno due cose. Che la comunità scientifica comunichi i suoi risultati in modo corretto; cioè, critico, modesto e problematico. E che abbia una visione generale delle questioni con un progetto ambizioso ma non velleitario e fumoso. Abbia conoscenza della storia precedente e sia cosciente degli scacchi, degli errori e dei vicoli ciechi del passato. Sia in grado di proporre assieme ai grandi obiettivi strumenti scientifici e tecnici potenzialmente adeguati a poterli raggiungere.

L'importanza dell'intersecarsi di questi diversi piani, nel rispetto dell'autonomia di ciascuno di essi, è quello che si è cercato di mettere in evidenza in queste pagine, sia pure in modo

frammentario ed episodico. In modo leggermente provocatorio sono stati presi come estremi simbolici due diversi *visionari* che sono entrambi, a rigore, al di fuori del campo proprio dell'IA ma che - proprio per questo - possono indicarci l'orizzonte verso cui muoverci.