TROVA L'ERRORE: ALAN TURING, PHILIP K. DICK E L'INTELLIGENZA ARTIFICIALE (I)

Data: 27 Febbraio 2023 - Di Stefano Suozzi

Rubrica: Pensare il pluriverso

Nel 1966 Philip Dick pubblica nel numero di maggio di «Worlds of Tomorrow» il racconto *Holy Quarrel*. Sono passati quasi sessant'anni e i computer immaginati da Dick erano ancora elettromeccanici e gli operatori introducevano dati e istruzioni e ricevevano risposte attraverso schede perforate. Eppure, un po' nascosto tra i tratti tipici della fantascienza dell'epoca e gli immancabili colpi di scena, il racconto contiene alcune riflessioni sull'intelligenza dei computer e di chi se ne serve che possono ancora essere utili per gettare uno sguardo disincantato, equidistante da critiche ed entusiasmi incondizionati, su alcuni aspetti dell'intelligenza artificiale (AI) e del suo, e nostro, futuro. Ma prima di discutere di *Holy Quarrel* cerchiamo di ricostruire alcuni tratti caratteristici dello sviluppo attuale dell'AI, necessari per meglio cogliere la specificità delle riflessioni dickiane.

1. Sin dai suoi albori, la fantascienza si è divertita ad affascinare o a terrorizzare i suoi lettori con le conseguenze della cosiddetta *singolarità*, ovvero la possibilità che macchine e computer così potenti e "intelligenti" da diventare coscienti di sé iniziassero ad agire di propria volontà, rifiutando i compiti per cui erano stati programmati e scegliendo nuovi obiettivi di propria iniziativa. Tutti abbiamo invidiato Lazarus Long per il suo rapporto con Minerva e Dora, supercomputer "femminili", la prima anche capace di

trasferirsi in un corpo biologico (cfr. R. Heinlein, *Lazarus Long l'immortale* [1973], Nord, Milano 2002); così come più o meno tutti, quando sentiamo parlare di intelligenza artificiale, robot, droni e automobili a guida autonoma, pensiamo al futuro di *Terminator*, in cui le macchine dominano la Terra e vogliono estinguere quell'inutile parassita che è l'uomo.

Tuttavia, la possibile manifestazione di una *singolarità* non è certamente il problema più urgente dell'AI. Innanzitutto, perché è davvero difficile immaginare che ciò che, nel corso di milioni di anni, si è prodotto sulla Terra solo nel caso dell'essere umano possa ripetersi in poche decine di anni per i supercomputer (per quanto riguarda il resto dell'universo resta ancora valido l'assunto che l'esistenza di altre forme di vita intelligenti è dimostrata dal fatto che *non* hanno cercato di mettersi in contatto con noi). Inoltre, e soprattutto, perché l'intelligenza, di qualunque forma essa sia, non basta a produrre la *singolarità*. Maurizio Ferraris, per esempio, nel suo recente *Documanità* (Laterza, Roma-Bari 2021), insiste sul fatto che l'unica differenza tra intelligenza artificiale e intelligenza naturale è nella quantità delle prestazioni:

> *La differenza tra intelligenza naturale e intelligenza artificiale non consiste nella qualità delle operazioni garantite, che sono comunque una manipolazione di segni, bensì nel fatto che nel caso dell'intelligenza naturale queste operazioni hanno luogo in un organismo, come tale destinato alla interruzione [morte], mentre nel caso dell'intelligenza artificiale hanno luogo in un meccanismo, come tale destinato alla iterazione. Ciò che di diverso troviamo nell'intelligenza umana – e sarebbe meglio parlare di «ragione» [...] – come il fatto di manifestare degli scopi, di essere condizionata da emozioni, di avere un'apertura sociale, dipende non tanto dal fatto che gli umani posseggano un cervello, ma dal fatto che quel cervello sia un organo vivente che è a sua volta parte di un organismo, che si inserisce in un contesto tecnico e sociale. [...] La differenza tra intelligenza naturale e intelligenza artificiale non si gioca anzitutto sul piano dell'intendere, bensì su quello della finalità (Ferraris,*

> *215-216).*

Ferraris insiste perciò sul ruolo della mano (Ferraris, 126ss) e sul concetto di *incarnazione* (Ferraris, 212ss) per sottolineare come sia la presenza di un corpo – della sua capacità di movimento, di provare sensazioni di piacere e dolore, di dover soddisfare bisogni e necessità, e soprattutto della sua finitezza e mortalità – che comporta la costituzione di fini e conseguentemente di una qualche forma di autocoscienza.

A tutto questo aveva in qualche modo già pensato Alan Turing nel celebre articolo *Computing Machinery and Intelligence* («Mind», LIX, 236, October 1950, 433-460). Innanzitutto, riflettendo sulla differenza tra intelligenza umana e intelligenza artificiale, Turing sosteneva che la domanda dalla quale l'articolo prende le mosse, «Can machines think?», è «così priva di senso che non merita di essere discussa» e confidava nel fatto che «per la fine del secolo l'uso delle parole e l'opinione comune più colta saranno cambiati in modo così significativo che sarà possibile parlare del pensiero delle macchine senza aspettarsi di essere contraddetti» (Turing, 442). In secondo luogo, consapevole del fatto che il lavoro di programmazione necessario per ottenere un computer in grado di simulare una mente adulta sarebbe stato troppo lungo e gravoso, ipotizzò la possibilità di progettare un computer più elementare, «bambino», ma in grado di apprendere. Il fatto è che un computer può apprendere solo attraverso la programmazione (di cui si sono già detti i problemi) o attraverso un sistema di «premi e punizioni», capace di procedere in autonomia, ma comunque lungo. È a questo punto che Turing introduce la possibilità di velocizzare il processo di apprendimento del computer attraverso l'educazione e l'esperienza, ma per fare questo sarebbe stato necessario dotarlo di un corpo. Purtroppo, osserva Turing, se anche l'ingegneria fosse progredita al punto di costruire un computer dotato di gambe, mano e cinque sensi, «non sarà possibile mandare a scuola una simile creatura senza che gli altri bambini se ne prendano continuamente gioco» (Turing, 456). Infine, Turing supera ogni perplessità e ogni difficoltà relativa alla natura del pensiero delle macchine

sostituendo la domanda iniziale – *Can machines think?* – con il celebre *imitation game* (oggi *test di Turing*), nel quale la macchina non deve dimostrare di saper pensare o di essere intelligente, ma deve limitarsi a rispondere alle domande di un osservatore ignaro (l'esperimento immagina che la comunicazione avvenga tra stanze separate e con domande e risposte dattiloscritte) senza che questi sia in grado di capire se sta comunicando con un uomo o con una macchina.

2. La misura del tempo passato dall'articolo di Turing è data dal fatto che oggi sono i computer ad avere la necessità di verificare se stanno davvero comunicando con un essere umano: è questo infatti lo scopo dei CAPTCHA, *Completely Automated Public Turing test to tell Computers and Humans Apart*. Ciononostante, l'aspetto rivoluzionario del suo approccio resta tale ed è quello di aver immediatamente rinunciato a porsi la questione della natura dell'intelligenza delle macchine e di essersi posto il problema, apparentemente più semplice, di ottenere la riproduzione degli effetti della comunicazione, ovvero di fare in modo che le risposte fornite dalla macchina *sembrino* quelle di un essere umano. Ed è questo ciò che ha consentito lo straordinario sviluppo della AI degli ultimi anni. Lo sottolinea Elena Esposito in *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale* (Bocconi University Press, Milano 2022)[1]:

> *Come gli esseri umani sono riusciti a volare quando hanno abbandonato l'idea di costruire delle macchine che sbattessero le ali come gli uccelli, così l'elaborazione digitale delle informazioni è riuscita a raggiungere i risultati che vediamo oggi quando ha abbandonato l'ambizione di riprodurre in forma digitale i processi della mente umana (Esposito, 22-23).*

In questo senso, il problema della natura dell'intelligenza delle macchine passa in secondo piano: non importa comprendere se si tratta di una forma di intelligenza diversa da quella degli umani o se, come sostiene Ferraris, la

distinzione debba essere fatta tra intelligenza (umana e artificiale) da un lato, e ragione dall'altro. Che tra di esse esista una differenza di grado o una differenza di natura non influisce sul fatto che per realizzare i processi di calcolo dell'intelligenza artificiale non è necessario conoscere e riprodurre i processi biologici attraverso i quali lavora l'intelligenza naturale (senza dimenticare che ciò che avviene nel cervello, per non parlare della mente, è ancor meno noto di quanto accade nell'AI). Ciò che occorre sottolineare, insiste Esposito, è innanzitutto la differenza tra intelligenza artificiale e comunicazione artificiale:

> *Ciò che possiamo osservare nelle interazioni con gli algoritmi non è necessariamente una forma artificiale di intelligenza, ma piuttosto una forma artificiale di comunicazione. Intelligenza e capacità comunicativa non sono la stessa cosa. Gli algoritmi sono in grado di agire come partner di comunicazione – che siano intelligenti o meno è un'altra questione. I recenti algoritmi di machine learning sono così efficienti non perché hanno imparato a imitare l'intelligenza umana e a comprendere le informazioni, ma piuttosto perché i loro programmatori hanno abbandonato il tentativo e l'ambizione di farlo, e si sono orientati verso un modello differente. Gli algoritmi di* ***machine learning*** *che utilizzano i big data, a mio parere, stanno riproducendo artificialmente non l'intelligenza ma la capacità di comunicare, e lo fanno sfruttando in modo parassitario la partecipazione degli utenti sul web (Esposito, 21).*

Gli assistenti virtuali come Alexa o i chatbot delle banche online o di qualsiasi altro servizio commerciale conducono e riproducono la comunicazione senza pensare e senza comprendere ciò che fanno o ciò di cui si sta parlando in generale, così come di ogni singola parola in particolare. Nella mole straordinaria di informazioni resa disponibile dalle continue interazioni sul web (big data), l'AI non fa che identificare *pattern*, ovvero modelli e strutture che si ripetono, e riprodurle nel contesto appropriato. Per esempio,

confrontando il numero incalcolabile di conversazioni condotte sul web, alla parola «Grazie» l'AI reagisce con «Prego» e alla parola «Merci» con «De rien», pur non sapendo cosa siano la lingua italiana e francese, che cosa significhino *grazie*, *prego*, *merci*, *de rien*, e nemmeno cosa siano la gratitudine, le buone maniere, la cortesia, ecc. Se infatti un numero sufficiente di persone sul web rispondesse al «Grazie» con «Albicocca», l'AI inizierebbe a rispondere nello stesso modo, senza nemmeno fare un *plissé*. In altri termini:

> *gli algoritmi non impar[ano] a pensare ma a partecipare alla comunicazione, cioè a sviluppare (artificialmente) una prospettiva autonoma che consente loro di reagire in modo appropriato e di generare informazioni nell'interazione con gli altri partecipanti. (...) Attraverso l'addestramento, gli algoritmi non diventano più intelligenti; imparano solo a giocare meglio (Esposito, p. 38).*

I computer hanno dunque imparato a giocare l'*imitation game* e a superare egregiamente il test di Turing. Ciò che potrebbe destare qualche perplessità è che, proprio come ipotizzava Turing, hanno imparato a farlo *da soli*. È vero che i processi di calcolo dell'AI sono determinati dalla programmazione inziale, ma lo sviluppo è stato tale che, ormai operando su una mole di dati che sarebbe ingestibile da qualunque uomo o gruppo di uomini, i processi attraverso i quali riconosce in questa massa informe i *pattern* di dati che possono essere trasformati in «nuove» informazioni, sono divenuti ormai imperscrutabili. Al punto che è nata una nuova corrente di studi espressamente dedicata all'AI spiegabile, ovvero che si propone di descrivere come procede l'elaborazione dati da parte degli algoritmi o, addirittura, di chiedere agli algoritmi stessi di spiegare i propri procedimenti di calcolo. Ma i risultati non sono particolarmente confortanti. Emerge, pertanto, un problema del controllo. Sottolinea Esposito:

Il risultato, che può essere osservato oggi, è una condizione in cui disponiamo di informazioni di cui spesso nessuno può ricostruire né

comprendere la genesi, ma che ciononostante non sono arbitrarie. Le informazioni generate autonomamente dagli algoritmi non sono affatto casuali e sono del tutto controllate, ma non dai processi della mente umana. Come possiamo controllare questo controllo, che per noi può essere anche incomprensibile? Questa è, a mio parere, la vera sfida che ci pongono oggi le tecniche di machine learning e l'uso dei big data (Esposito, 14).

Si tratta dunque di una questione attuale e urgente, ma, come spesso accade, era già stata anticipata dalla fantascienza e, in particolare, dal racconto di Philip K. Dick, *Holy Quarrel*.

(*fine Prima Parte*)

NOTE

[1] Poiché cito dall'edizione Kindle, possono sussistere leggere imprecisioni sull'indicazione del numero di pagina. Per quanto attiene ai più recenti sviluppi dei diversi settori dell'AI (big data, deep learning, profilazione, algoritmi predittiva, ecc.) e per la relativa bibliografia aggiornata rimandiamo a questo testo estremamente informato.