

Brevi riflessioni sul test di Turing

di Astro Calisi

Ricorre quest'anno il centenario della nascita di Alan Turing, il matematico inglese noto al grande pubblico per essere riuscito a decifrare il codice *Enigma*, con cui, durante la seconda guerra mondiale, venivano crittografate le comunicazioni dei nazisti. In realtà, l'importanza di Turing è legata soprattutto ai suoi contributi fondamentali alla teoria della computazione, sulla quale si basa la quasi totalità degli elaboratori d'informazione oggi esistenti, dai comuni computer commerciali, fino ai più sofisticati supercomputer impiegati in campo militare e nella ricerca scientifica.

Per Turing c'erano pochi dubbi che entro poche decine di anni i sistemi computazionali avrebbero raggiunto un livello d'intelligenza paragonabile a quello degli esseri umani. Proprio in vista di questo obiettivo, egli si pose il problema di come valutare il livello di intelligenza di una macchina. Si rese conto che il modo più diretto per affrontare la questione richiedeva in qualche modo di definire con precisione il significato di termini come "macchina", "intelligenza" e "pensiero". Termini assai difficili da racchiudere in una definizione esaustiva che trovi tutti d'accordo.

Invece che tentare una strada del genere, secondo Turing, è preferibile sostituire la domanda "le macchine possono pensare?" con una situazione concreta in cui vengano messe a confronto le prestazioni intelligenti di una macchina con quelle di un uomo. Si procede così.

Collochiamo una macchina computazionale e un essere umano in due stanze separate. Da una postazione esterna, un esaminatore, senza conoscere l'identità degli interlocutori, rivolge loro delle domande tramite un terminale, ricevendone le relative risposte. Se dopo un tempo ragionevolmente lungo, l'esaminatore non riesce a riconoscere la macchina dall'uomo, si conclude che la prima si è comportata esattamente come un essere umano, tanto da non essere distinguibile da questo. Significa che non esiste una sostanziale differenza tra le modalità di elaborazione di una macchina e il pensiero dell'uomo e che quindi le macchine sono intelligenti. (1)

Turing parte dall'assunto che l'intelligenza umana, e anzi l'intera attività del nostro pensiero sia riducibile a procedimenti di tipo logico-matematico, accettando sostanzialmente la ben nota concezione di Hobbes, secondo la quale pensare equivale a calcolare. Questa concezione sopravvive largamente ancora oggi nelle tesi dei cosiddetti *funzionalisti*, i quali si mostrano convinti che non solo l'intelligenza dell'uomo, ma anche ogni altra facoltà della mente – compresa la coscienza – possano venir riprodotte da un computer sufficientemente potente. In tale ottica, qualsiasi limite mostrato da una macchina computazionale rispetto alle capacità cognitive di un essere umano va imputato

unicamente a una insufficiente implementazione della macchina a livello hardware e/o software. Accrescendo la potenza di elaborazione della macchina e dotandola dei programmi adatti, essa non potrà che arrivare a eguagliare (e magari anche a superare) le prestazioni di un essere umano.

Prima di esaminare criticamente questa idea, vorrei approfondire un altro aspetto del test di Turing. Molto si è scritto su questo test ma, a quel che mi risulta, nessuno si è soffermato a sufficienza sul rilievo che i risultati ottenuti sono inevitabilmente legati alla *soggettività* dell'esaminatore, vale a dire alla sua capacità di cogliere differenze significative tra le risposte fornite dai due interlocutori, ma anche alla sua abilità nell'inventare domande in grado di mettere in difficoltà il sistema computazionale. Questo vuol dire che il test di Turing non può in alcun caso venir considerato una prova *oggettiva* per dimostrare in maniera assolutamente affidabile quando le macchine raggiungeranno un livello di intelligenza in tutto simile a quello degli esseri umani, ammesso che tale obiettivo sia conseguibile. Infatti, la circostanza che l'esaminatore non rilevi alcuna differenza tra le risposte ottenute dai due interlocutori non sta necessariamente a significare che non esistono differenze; può anche voler dire che l'esaminatore non è stato abbastanza abile nel formulare le domande e nel valutare le risposte.

Si può migliorare la situazione, ma solo parzialmente, sostituendo il singolo esaminatore con una giuria composta di più persone. In tal modo si hanno migliori garanzie che almeno gli errori e le sviste più grossolane vengano neutralizzati. Siamo tuttavia ancora ben lontani dalla certezza che il verdetto raggiunto sia totalmente affidabile: il giudizio rimane comunque legato alla soggettività, anche se a una soggettività relativa a più persone, trasformata in una più rassicurante *intersoggettività*.

Potrebbe sembrare che, dotando gli esaminatori di ben definiti criteri e procedure da applicare nella valutazione o addirittura incaricando della valutazione stessa un sistema computazionale appositamente programmato, costituisca un passo decisivo verso l'oggettività e quindi verso la validità dei risultati raggiunti. Ma non è così.

La verità è che non esistono criteri assolutamente affidabili per distinguere l'intelligenza dell'uomo dall'intelligenza artificiale. Del resto, se ci fossero, non staremmo qui a disquisire sul test di Turing. Qualsiasi criterio di valutazione, per quanto complesso o articolato, utilizzato per discriminare l'intelligenza umana da quella di una macchina, una volta definito e reso pubblico, potrebbe facilmente venir aggirato da un bravo programmatore, fornendo il computer di adeguate tipologie di risposte in grado di trarre in inganno gli esaminatori.

E' una simile possibilità che, senza dubbio, aveva in mente il grande epistemologo Karl Popper quando scrisse: «Turing disse qualcosa del genere: specificate il modo in cui credete che un uomo sia superiore a un computer e io costruirò un computer che confuti la vostra credenza. La sfida di Turing non dovrebbe essere raccolta, perché in linea di principio non si potrebbe usare alcuna specificazione sufficientemente precisa per programmare un computer». (2)

A queste considerazioni fa eco ai nostri giorni il filosofo tedesco Thomas Metzinger, riferendosi al più ampio contesto delle capacità mentali: «O non potete identificare quale aspetto della coscienza umana e della soggettività non può venire

implementato in un sistema artificiale, oppure, se lo potete fare, allora la questione si riduce alla scrittura di un algoritmo che può essere implementato in un software». (3)

Possiamo rappresentarci in maniera semplificata quello che, in definitiva, rappresenta un vero e proprio limite della computazione con il seguente esempio. Immaginiamo che degli esaminatori, che basano i loro giudizi sul test di Turing, e dei programmatori di grande esperienza intraprendano una sfida a oltranza per vedere dove essa conduce. Da una parte, verranno, di volta in volta, esplicitati i nuovi principi a cui devono sottostare le risposte fornite dall'elaboratore in esame per poter essere considerato indistinguibile da un essere umano; dall'altra, si apporteranno le opportune integrazioni a livello software così da consentire di prendere in considerazione anche questi principi.

Questa strategia, classificabile come *induttiva*, può effettivamente permettere, seguendo un percorso asintotico, di giungere sempre più vicino al confine ideale che renderebbe nulle le differenze tra macchina e uomo. Confine che delimiterebbe l'insieme di tutte le situazioni affrontate dal genere umano nel corso della sua storia. Una macchina computazionale dotata di uno sterminato database in cui siano memorizzati i criteri di valutazione per poter discriminare tra una vastissima tipologia di casi possibili, offrendo le relative risposte, è molto difficile da realizzare nella pratica, ma non impossibile da immaginare in linea di principio, a livello teorico. Una simile macchina, sottoposta a una qualsiasi delle domande o dei problemi che gli uomini si sono trovati ad affrontare fino a quel momento, saprebbe in effetti rispondere appropriatamente, traendo in inganno qualsiasi esaminatore, umano e non.

Senonché, tale macchina presenterebbe un grave difetto, che renderebbe le risposte da essa fornite facilmente riconoscibili rispetto a quelle di un uomo: quella macchina non saprebbe affrontare situazioni *nuove*, per le quali non siano stati ancora messi a punto appropriati algoritmi; in particolare, non saprebbe pronunziarsi adeguatamente su questioni relative a ipotesi inedite o situazioni immaginarie appositamente inventate dagli esaminatori per smascherare la macchina.

La capacità di affrontare il nuovo costituisce, a ben guardare, uno dei principali fattori discriminanti tra l'intelligenza umana e l'intelligenza artificiale. Tale capacità è riconducibile, in senso lato, all'intuizione, spesso messa in relazione con la *creatività*, ovvero all'attitudine di porsi *al di là* dei principi e delle regole preesistenti.

Mentre la computazione è capace di restituire, pur se in forme variamente elaborate, soltanto ciò che era stato precedentemente immesso in essa, l'intelligenza umana mostra di saper andare oltre i principi regolativi e le informazioni disponibili a un determinato istante. I risultati ottenuti non offrono garanzie assolute di validità, ma la percentuale di successi è così largamente superiore rispetto alla mera probabilità statistica, da lasciare pochi dubbi circa la sua efficacia. (4)

La psicologia ha da tempo riconosciuto questo aspetto fondamentale dell'intelligenza, definendola, ad alto livello, come *capacità di risolvere problemi*. Si presuppone, ovviamente, che i problemi siano nuovi o, perlomeno, che contengano elementi di sostanziale originalità rispetto ad analoghi problemi affrontati con successo in passato; oppure si richiede che i problemi stessi vengano risolti in maniera innovativa. In generale, bisognerebbe considerare un problema come autentico quando non sono note le regole e i procedimenti che conducono alla sua soluzione. Tale riconoscimento della

psicologia viene purtroppo offuscato e, di fatto, relegato al margine della riflessione sulla mente, dal fascino irresistibile esercitato dalla concezione del cervello come elaboratore d'informazione, comunemente nota come *cognitivismo*. Una simile concezione fa sì che certi studiosi siano come ciechi, non solo di fronte alle evidenti difficoltà di emulare il pensiero umano, non solo davanti all'incapacità di immaginare come un computer potrebbe essere reso creativo o cosciente, ma anche rispetto a numerose obiezioni sollevate sul piano teorico da alcuni autori negli ultimi decenni. (5)

In conclusione, si può dire che il modello di intelligenza identificata con le capacità logico-matematiche appare oggi largamente insufficiente, anche se esso conta ancora molti sostenitori per mancanza di valide alternative. Il test di Turing, utilizzato per dimostrare la totale sovrapposibilità tra l'intelligenza dell'uomo e l'intelligenza artificiale, mostra i suoi limiti proprio per la necessità di ricorrere a criteri di intervento e di valutazione non formalizzabili a priori, affidati all'iniziativa e all'estro del valutatore.

Questa necessità, che in ultima analisi presuppone il ricorso a doti creative, non costituisce forse una ulteriore prova della diversità radicale della nostra intelligenza rispetto a quella delle macchine?

NOTE

(1) Alan Turing, "Computing Machinery and Intelligence", in *Mind*, 59, 1950, pagg. 433-460; trad. it., "Calcolatori e intelligenza" in Daniel Dennett - Douglas Hofstadter, *L'io della mente*, Adelphi, Milano, 1985, pagg. 61-74.

(2) Karl Popper – John Eccles, *L'io e il suo cervello*, Armando, Roma, 1981, pag. 252.

(3) Thomas Metzinger, *Il tunnel dell'io. Scienza della mente e mito del soggetto*, Raffaello Cortina, Milano, 2010, pag. 223

(4) Per un approfondimento sulle particolari caratteristiche della creatività umana e sulla loro sostanziale irriducibilità ai modelli offerti dalla teoria dell'informazione e a quello, più generale, della spiegazione scientifica, si veda Astro Calisi, *Oltre gli orizzonti del conosciuto. La sfida cruciale della mente alla scienza del XXI secolo*, Uni Service, Trento, 2011, pagg. 25-38.

(5) Una panoramica a livello divulgativo su questo argomento si trova in David Harel, *Computer a responsabilità limitata. Dove le macchine non riescono ad arrivare*, Einaudi, Torino, 2002. Una trattazione più complessa, e anche assai più specifica, è quella di Roger Penrose in *Ombre della mente*, Rizzoli, Milano, 1996.

[luglio 2012]