

L'intelligenza artificiale

Francesca Rossi
Dipartimento di Matematica Pura ed Applicata
Universita' di Padova
Email: frossi@math.unipd.it

In questo breve articolo daremo una introduzione molto informale e necessariamente incompleta dell'intelligenza artificiale, una disciplina di ricerca scientifica nata negli anni '50 con l'intento di costruire macchine intelligenti. Chiariremo cosa si intende per macchina intelligente e delineremo una breve storia di questa disciplina, con particolare riferimento ad alcune pietre miliari importanti per il suo sviluppo. Descriveremo poi le principali linee di ricerca attive ad oggi nell'area dell'intelligenza artificiale, e concluderemo con alcune considerazioni etiche che spesso vengono sollevate quando si parla di macchine intelligenti.

Cos'e' una macchina intelligente? L'intelligenza artificiale forte e debole

L'intelligenza artificiale (IA) [5] puo' essere definita come la scienza che si propone di sviluppare macchine intelligenti.

Appena si introduce il concetto di macchina intelligente, vengono alla mente alcune domande, che sono state poste anche molto prima dell'avvento dei calcolatori. Primo, va chiarito cos'e' l'intelligenza, e per questo e' necessario capire come funziona la mente umana. Una volta data una definizione di intelligenza, e' possibile che una macchina agisca in modo intelligente? Se si, sarebbe cosciente come gli umani? E quali sono le implicazioni etiche dell'esistenza di macchine intelligenti e coscienti?

Per provare a rispondere almeno ad alcune di queste domande, dobbiamo prima chiarire la terminologia che useremo. Innanzitutto, per macchina qui intendiamo non necessariamente un sistema hardware, come un robot, ma anche un sistema software, come un programma o un insieme di programmi. Non e' pero' chiaro cosa significhi mostrare un comportamento intelligente. Si intende un comportamento simile a quello umano? O la capacita' di scherzare? O di risolvere problemi difficili? O di comporre musica? O di imparare dall'esperienza?

La maggior parte dei discorsi sull'intelligenza delle macchine possono essere ricondotti a due principali teorie sull'intelligenza artificiale: l'IA forte e l'IA debole.

- La teoria dell'**IA forte** si basa sulla convinzione che le macchine possano effettivamente essere intelligenti.
- La teoria dell'**IA debole**, invece, in modo piu' realistico e pragmatico, pensa che le macchine possano comportarsi come se fossero intelligenti.

Queste teorie sono collegate ad una visione della mente umana come un programma: dati degli stimoli (il suo input), la mente umana ragiona e produce un certo comportamento del corpo (il suo output). Il cervello, a sua volta, può essere visto come l'hardware parallelo (costituito da neuroni e connessioni tra loro) su cui viene eseguito questo programma. Quindi, se avessimo abbastanza risorse computazionali, il ragionamento umano potrebbe essere simulato da un calcolatore e una macchina potrebbe riuscire ad avere le stesse capacità di una mente umana. In realtà questa macchina non dovrebbe necessariamente simulare il cervello, ma solo i processi di ragionamento umano. Quindi l'hardware potrebbe anche essere diverso da quello direttamente ottenuto simulando il cervello.



Uno dei fondatori dell'IA, John McCarthy, nel 1950 scriveva: "Ogni aspetto dell'apprendimento o ogni altra caratteristica dell'intelligenza può essere descritta in modo talmente preciso da consentire la costruzione di una macchina in grado di simularla". Quindi l'IA fu fondata sulla assunzione che almeno l'IA debole fosse possibile.

Il test di Turing

E' veramente possibile simulare la mente umana tramite una macchina? Una prima cosa da notare e' che già negli anni '30 si era giunti alla conclusione che le macchine hanno dei limiti. Goedel nel 1932 ha dimostrato che, dato un qualunque sistema formale, ci sono affermazioni vere in questo sistema che non possono essere provate dalle regole logiche del sistema. Turing nel 1936 ha trasportato questo risultato nell'ambito delle macchine.

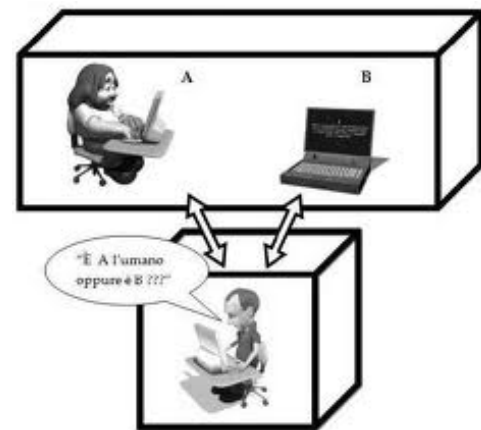
A quel tempo, i calcolatori non esistevano ancora, ma Turing aveva definito un modello di macchina, chiamata da allora macchina di Turing [1], che e' ancor'oggi un modello generale astratto di un calcolatore. La macchina di Turing riceve il suo input sotto forma di una stringa di simboli su un nastro, e le uniche operazioni che può fare e' scorrere il nastro in entrambe le direzioni, leggere la cella del nastro su cui e' posizionata, e scrivere tale cella. Non ci sono limiti alla lunghezza del nastro. Un tipico problema che si può chiedere ad una macchina di Turing di risolvere e' capire se la stringa in input ha una certa forma. Questa formulazione dei problemi e' apparentemente astratta ma riesce a modellare molti problemi pratici, come ad esempio cercare di capire se, dato un programma e un suo input, il programma con quell'input si ferma dopo un numero finito di passi o va in ciclo.

Analogamente a quanto dimostrato da Goedel per i sistemi formali matematici, ci sono problemi che una macchina di Turing non può risolvere, come il problema del programma appena citato, cioè non esiste nessuna macchina di Turing che

sappia sempre rispondere in tempo finito a questa domanda: dato un programma P, dato un input I, P con input I si ferma o va in ciclo? Dato che il concetto di macchina di Turing e' cosi' generale da modellare qualunque programma, per quanto sofisticato sia, questi limiti valgono anche per qualsiasi programma.

Visto che le macchine hanno dei limiti, riusciamo a costruire una macchina intelligente?

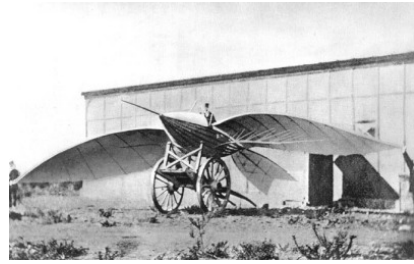
Lo stesso Turing, nel tentativo di rispondere a questa domanda, ha cercato prima di definire cosa vuol dire che una macchina e' intelligente. Nella sua idea, una macchina si puo' dichiarare intelligente se ha un comportamento che, visto da un umano, e' giudicato simile a quello di un altro umano. Questa sua idea l'ha esemplificata definendo nel 1950 il cosiddetto "test di Turing" [6], in cui (in una delle sue versioni) un umano ha una conversazione con un interlocutore e, senza vedere l'interlocutore e interagendo solo per mezzo di testo, deve capire se e' un programma o un essere umano. Il programma passa il test di Turing se riesce a ingannare il suo interlocutore almeno tre volte su dieci.



Una prima assunzione alla base di questo test e' che l'intelligenza possa essere misurata guardando il comportamento esterno, senza sapere niente del ragionamento che ha prodotto tale comportamento. Una seconda assunzione e' che un comportamento umano sia sempre intelligente, e viceversa, ma chiaramente ci sono comportamenti umani che non sono intelligenti, e anche comportamenti intelligenti che non sono di solito adottati dagli umani. Quindi viene spontaneo chiedersi se e' questa la nozione corretta di intelligenza artificiale, o comunque se e' questa la nozione che ci interessa.

Notare che questo test si basa sulla teoria dell' IA debole (la macchina non deve necessariamente essere intelligente, ma deve mostrare un comportamento intelligente) ma non pone vincoli sull'argomento della conversazione, quindi ci si aspetta che la macchina possa avere un comportamento intelligente in qualunque disciplina del sapere.

Un modo per convincersi che la teoria dell'IA forte pone richieste irragionevoli per dichiarare che una macchina e' intelligente consiste nel far riferimento all'esempio del volo artificiale: chiedersi se una macchina puo' essere intelligente e' come chiedersi se un aereo puo' volare. Mentre all'inizio si cercava di ottenere una macchina capace di volare imitando gli uccelli, ci si e' presto resi conto che questo non era ragionevole e si e' quindi passati a concepire macchine il cui volo si basasse su altre leggi fisiche. L'IA debole in questo esempio corrisponderebbe a creare aerei che volino in modo cosi' simile ad un uccello da ingannare altri uccelli, per fare un'analogia con il test di Turing. Il fatto che gli aerei di oggi ci abbiano permesso di raggiungere il sogno del volo artificiale fa ben sperare per l'IA, che puo' creare macchine intelligenti anche senza che siano coscienti o che sappiano mostrare



Oggi, quasi nessun ricercatore di IA si pone il problema di creare una macchina intelligente (nel senso dell'IA forte), e solo alcuni cercano di creare una macchina che abbia un comportamento intelligente su tutti i campi dello scibile umano (cioe' che passi il test di Turing). Quello su cui la maggioranza dei ricercatori si concentra e' la possibilita' di costruire macchine con comportamento intelligente in campi specifici.

Gli inizi dell'intelligenza artificiale come scienza, e alcune pietre miliari

Per capire cos'e' l'intelligenza artificiale, in quanto disciplina di ricerca scientifica, occorre risalire ai suoi inizi, che di solito vengono datati al 1956, anno in cui venne organizzato un incontro estivo a Dartmouth (New Hampshire, USA) tra tutti quei ricercatori che si occupavano di capire come simulare alcuni aspetti del ragionamento umano tramite macchine. Durante questo incontro, John McCarthy introdusse il termine "intelligenza artificiale", che viene usato da allora.

Nell'ambito dell'incontro di Dartmouth, fecero molto scalpore alcuni lavori che miravano a riprodurre il processo di prova di alcuni teoremi, che veniva ritenuto come un tipico esempio di comportamento intelligente. In particolare, fu molto apprezzato il lavoro di Simon per simulare la prova di teoremi: il suo Logic Theorist [4,10] riusci' a provare 38 dei 52 teoremi dai Principia Mathematica di Russell e Whitehead. Nella stessa linea, fu anche presentato il General Problem Solver del Carnegie Institute, che riusci' a provare molti teoremi e puzzles, e anche un teorema a aperto.

In quegli anni, McCarthy definì anche un linguaggio di programmazione specifico per problemi dell'intelligenza artificiale, chiamato LISP (MIT, 1958) [3]. Questo linguaggio permise di creare moltissimi programmi di IA in domini molto diversi tra loro. Un esempio è Student, che risolveva problemi algebrici in forma di frase, e un altro esempio è Dendral, che risolveva problemi di chimica.

```
FreeStylLISP Translator
(clears)
(namedef X #xletter)
(repeat i 0 6 1
  (repeat j 0 6 1
    (loc 0 37 3 (loc i j 0 (byname X)) )
  )
)
(view)
(set z 1)
(repeat k 0 73 1
  (repeat i 2 6 1
    (repeat j 0 6 1
      (loc 0 k 2 (loc i j 0 (sprite ".") )
    )
  )
)
(loc 4 z 1 (sprite ">"))
(addto z 3)
(delay 8000)
(view)
)
Linking: xletter.fsl
(sprite 2)
FreeStylLISP Runtime
NULLCAR
```

Programa scritto in LISP

Un'altra pietra miliare dell'IA è il programma ELIZA (1965) [7,9]. Questo programma simulava l'interazione tra un terapeuta rogeriano e un paziente. Molti pazienti erano convinti di parlare con un vero terapeuta, e non una macchina, quindi il programma passava il test di Turing. D'altra parte, le frasi che ELIZA diceva ai pazienti erano solo frutto di un sapiente matching di alcune parole chiave nelle frasi dette dal paziente, e non di un ragionamento complesso. Quindi, pur superando il test di Turing, era difficile dichiarare questo programma come intelligente



Esempio di interazione tramite ELIZA

Negli anni '70 vennero creati molti programmi chiamati "sistemi esperti". L'idea era di focalizzarsi su un dominio specifico e di simulare il comportamento di un esperto di quel dominio. Un tipico esempio è stato MYCIN (1976) [11], un sistema esperto di grande successo che sapeva effettuare diagnosi per malattie ematiche infettive.

Oltre al desiderio di simulare la conoscenza di esperti umani in domini specifici, ci furono anche tentativi che volevano dotare una macchina di una conoscenza di tipo generale e di "senso comune". Questo è il caso di CYC [8], un progetto nato nel 1983 e tuttora in corso, che ha accumulato molti milioni di fatti nella propria base di conoscenza, al fine di generare una macchina che sappia ragionare almeno come un bambino di 6 anni.

Un'altra pietra miliare dell'IA e' relativa a macchine in grado di giocare in un certo gioco come i migliori giocatori umani al mondo. L'esempio piu' eclatante e' Deep Blue (1997) [2], che riusci' a vincere il campione del mondo Kasparov a scacchi.



Kasparov contro Deep Blue

L'intelligenza artificiale oggi

Col passare degli anni, e con la constatazione che non aveva molto senso impegnarsi a creare una macchina intelligente secondo i dettami dell'IA forte, o che superasse il test di Turing, prese piede una teoria alternativa, che intende l'IA come la disciplina che vuole risolvere specifici problemi, o anche effettuare ragionamenti che non possono essere compresi pienamente dalle capacita' cognitive umane, ma in domini specifici. Quindi non ci sia aspetta che la macchina sia in grado di avere l'ampiezza di abilita' cognitive propria dell'uomo, o che sia cosciente di cio' che fa, ma solo che sappia risolvere in modo efficiente e ottimale problemi, anche difficili, in specifici campi d'azione.

Negli attuali convegni di IA, sono questi gli scopi che la maggior parte dei lavori presentati hanno. Praticamente nessuno si preoccupa di definire macchine che superino il test di Turing, mentre tutti si occupano di individuare problemi difficili in domini specifici, che le capacita' cognitive umane fanno fatica a risolvere. Mentre questa teoria dell'IA puo' sembrare riduttiva, e' in realta' di grande respiro, e ha prodotto molti risultati importanti.

Oggi giorno, le principali aree di ricerca nell'ambito dell'IA si possono raggruppare in alcuni filoni coerenti al loro interno. Ad esempio, il principale convegno internazionale di IA (chiamato IJCAI, per International Joint Conference on Artificial Intelligence), che nel 2011 si tiene a Barcellona (Spagna), include risultati nei seguenti filoni:

- **Sistemi multi-agente:** come si possono descrivere e risolvere problemi relativi alla collaborazione o competizione tra piu' agenti;
- **Vincoli e ricerca:** come si possono modellare problemi rle cui soluzioni devono rispettare delle restrizioni (i vincoli) e come si puo' cercare una loro soluzione ottima in modo efficiente;
- **Rappresentazione della conoscenza:** come si puo' modellare la conoscenza in un certo campo applicativo, e come si puo' usare tale conoscenza per agire in modo intelligente in tale campo;

- **Apprendimento automatico:** come le macchine possono apprendere nuovi e migliori comportamenti dall'esperienza;
- **Elaborazione del linguaggio naturale:** come le macchine possono comprendere e ragionare su un testo in linguaggio naturale;
- **Schedulazione e pianificazione:** come si possono schedulare o pianificare varie attività in modo ottimo in presenza di risorse limitate;
- **Robotica e visione:** come si può costruire un robot adattivo e flessibile che sappia vedere e percepire il mondo esterno, e sappia decidere il comportamento ottimo da adottare sulla base della sua percezione del mondo;
- **Gestione dell'incertezza:** come si possano prendere decisioni ottime anche se alcuni dati sono mancanti o vaghi;
- **Web:** come si possano applicare tecniche di IA al mondo del web.

Non si vuole quindi sostituire un essere umano in tutte le sue capacità, ma invece si intende affiancare o aumentare l'intelligenza umana in alcuni domini specifici. L'aumento può essere basato sulla potenza di calcolo o sulla quantità di memoria a disposizione.

Un tipico esempio di IA applicata con successo ad un dominio specifico è il caso degli algoritmi usati da Google per indicizzare le pagine web e quindi per permettere agli utenti di cercare efficientemente informazioni sul web. Altri domini in cui l'IA ha prodotto sistemi molto utili sono, ad esempio, i seguenti:

- Sistemi esperti (medicina e altri campi dove serve un esperto umano)
- Scacchi e altri giochi
- Sistemi di schedulazione di aerei
- Gestione dei turni del personale
- Robot per ospedali
- Lettori per ciechi
- Traduzione e comprensione del linguaggio naturale
- Biologia e genomica
- Visione artificiale
- Ricerca su web, aste online

Tutte le macchine che forniscono servizi importanti nei campi sopra citati falliscono il test di Turing. Infatti, non presentano un'intelligenza generale. Però aiutano gli umani a compiere in modo intelligente lavori specifici che altrimenti sarebbero eseguiti meno bene o non eseguiti affatto. Pur non prevedendo macchine che abbiano un comportamento intelligente in ogni campo, è ragionevole prevedere che in futuro l'IA migliorerà la qualità della nostra vita in ogni attività umana.



Esempi di uso dell'IA nella vita di tutti i giorni

Le macchine intelligenti non devono far paura

L'idea di macchine intelligenti può portare alla paura che tali macchine possano agire in modo indipendente dagli esseri umani e uscire dal loro controllo. Dato che le questioni morali ed etiche sono difficili da formalizzare, si ritiene difficile che si possano dotare le macchine intelligenti di regole precise di comportamento che rispettino certi codici etici e morali. Questo porta alla paura che le macchine intelligenti possano danneggiare gli esseri umani sulla base di errori o sbagliate interpretazioni della moralità. Nonostante questo, anche senza rendercene conto, siamo già oggi circondati da macchine intelligenti, che ci supportano in molte attività della nostra giornata, e non siamo tuttavia governati da tali macchine, ma solo aiutati nelle nostre attività.

Un'altra paura tipica nei confronti dell'IA è che le innovazioni fornite dalla creazione di macchine intelligenti possano togliere lavoro agli esseri umani. In realtà, fino ad oggi l'automazione raggiunta attraverso la tecnologia dell'informazione in generale, e l'IA in particolare, ha creato più posti di lavoro di quanti ne abbia eliminati, e i nuovi lavori per gli esseri umani sono più interessanti. Oggi che l'IA si focalizza su macchine intelligenti che sappiano assistere gli esseri umani, la perdita di posti di lavoro preoccupa meno di quando si cercava di creare macchine intelligenti che sostituissero gli uomini.

Riferimenti bibliografici

[1] B. Jack Copeland ed., *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*", Clarendon Press (Oxford University Press), 2004.

[2] Hsu, Feng-hsiung. "Behind Deep Blue: Building the Computer that Defeated the World Chess Champion". Princeton University Press. 2002.

[3] John McCarthy. "Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I", *Communication of the ACM*, 1960.

[4] Allen Newell and Herbert A. Simon, "The logic theory machine", *Information processing*, 1956.

[5] Russell, Stuart J.; Norvig, Peter, "Artificial Intelligence: A Modern Approach", Prentice Hall, 2003.

[6] Turing, Alan, "Computing Machinery and Intelligence", *Mind*, (236): 433–460, 1950.

[7] Weizenbaum, Joseph, "ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine", *Communications of the ACM* 9 (1): 36–45, 1966.

[8] Cyc Wikipedia page: <http://en.wikipedia.org/wiki/Cyc>

[9] ELIZA Wikipedia page: <http://en.wikipedia.org/wiki/ELIZA>

[10] Logic Theorist Wikipedia page:
http://en.wikipedia.org/wiki/Logic_Theorist

[11] Mycin Wikipedia page: <http://en.wikipedia.org/wiki/Mycin>