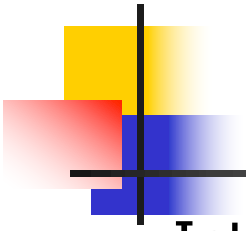
Ragionamento (artificiale)

Angelo Montanari
Dipartimento di Scienze Matematiche,
Informatiche e Fisiche
Università degli Studi di Udine
Udine, 6 aprile, 2024

Sommario

- Un breve inquadramento
- La nozione di intenzionalità
- La visione macchinista di Minsky
- Intelligenza, intenzionalità e macchine
- Il test di Turing
- Sistemi artificiali intelligenti: da Shakey a Justin
- La stanza cinese di Searle
- Il machine learning
- La nuova frontiera: la bionica

Un breve inquadramento



Intelligenza artificiale simbolica e sub-simbolica

Intelligenza artificiale simbolica: rappresentazione della conoscenza e ragionamento automatico

Intelligenza artificiale sub-simbolica: modelli e tecniche di machine learning

Trade-off tra efficienza / efficacia and **interpretabilità**

Explainable artificial intelligence e **trustworthy artificial intelligence**

Focus: intelligenza artificiale simbolica

Intelligenza e intenzionalità

Nella riflessione filosofica (Brentano, Husserl, Carnap), l'**intenzionalità** viene riconosciuta quale elemento distintivo della coscienza (in generale, di ogni fenomeno psichico)

Per Brentano, l'intenzionalità è il carattere costitutivo di ogni **fenomeno psichico**

Per Husserl, l'intenzionalità è il carattere costitutivo della **coscienza** e del rapporto soggetto (umano) - oggetto

Compito della filosofia è descrivere la struttura immanente con cui l'oggetto è intenzionato dalla coscienza

Intenzionalità: una definizione

In generale, possiamo definire l'**intenzionalità** come il riferimento interno di un atto o di uno stato mentale a un determinato oggetto, ossia la connessione che l'atto o lo stato hanno, in virtù della loro identità, con un certo oggetto, indipendentemente dalla sussistenza di questo eventuale oggetto nella realtà esterna

Esempio. Dell'identità di uno stato emotivo di speranza fa parte ciò che è sperato, indipendentemente dal fatto che si realizzi oppure no

Intelligenza artificiale e macchine

Il fascino delle macchine e la **visione "macchinista"** del mondo:

il cosmo e l'uomo come macchine

L'evoluzione nel tempo di tale visione:

- uomo come macchina di natura meccanica
- uomo come macchina termodinamica
- uomo come macchina chimica
- uomo come **macchina informazionale** (corpo/mente assimilati a hardware/software)

Marvin Minsky

La visione “macchinista”

Tale posizione è affermata in modo esplicito da Minsky (M. Minsky, *The Society of Mind*, Simon and Schuster, 1988):

> “Non vi alcun motivo per credere che il **cervello** sia qualcosa di diverso da una **macchina** con un numero enorme di componenti che funzionano in perfetto accordo con le leggi della fisica”

La mente come processo

E' una declinazione particolare della posizione materialistica classica, che si contrappone ad ogni dualismo mente/corpo

Per Minsky la mente è semplicemente ciò che fa il cervello (la **mente come processo**)

Tale interpretazione della mente stabilisce una stretta analogia tra la relazione tra la mente e il cervello e quella che intercorre tra le nozioni di **processo** (un programma in esecuzione) e di **programma** in informatica: per Minsky la mente è semplicemente il

"**cervello in esecuzione**"

Rapporto mente-cervello

- Per spiegare la mente evitando la circolarità, occorre descrivere il modo in cui le menti sono costruite a partire da materia priva di mente, parti molto più piccole e più semplici di tutto ciò che può essere considerato intelligente
- **Questione**: una mente può essere associata solo ad un cervello o, invece, qualità tipiche della mente possono appartenere, in grado diverso, a tutte le cose?
- Per Minsky, il **cervello** può essere visto come una **società organizzata**, composta da una molteplicità di componenti strutturate in modo gerarchico, alcune delle quali operano in modo del tutto autonomo, la maggior parte in un rapporto alle volte di collaborazione, più spesso di competizione, con altre componenti

La società della mente

- Intelligenza umana frutto dell'interazione di un numero enorme di componenti fortemente diverse fra loro, i cosiddetti **agenti della mente**, componenti elementari ("particelle") di una (teoria della) mente
- **Questione**: come può l'opera combinata di un insieme di agenti produrre un comportamento che ogni singolo agente, considerato separatamente, non è in grado di fornire?
- Si tratta di una questione ben nota: comportamento emergente in sistemi complessi, dove la **nozione di emergenza** fa riferimento a proprietà di un sistema non derivabili dalle proprietà note dei suo componenti

La nozione di agenzia

- Per superare posizioni di riduzionismo ingenuo difficilmente sostenibili (Minsky contesta chi considera la fisica e la chimica modelli ideali di come dovrebbe essere la psicologia), Minsky introduce la **nozione di agenzia**

un'agenzia è un insieme di agenti collegati fra loro da un'opportuna rete di interconnessioni

- La **gerarchia delle agenzie**

La teoria degli agenti

La **teoria degli agenti** / i **sistemi multi-agente** in intelligenza artificiale possono essere visti come la controparte "applicativa" della società della mente di Minsky

Agente (**artificiale**) **intelligente**: un agente intelligente è un sistema in grado di decidere cosa deve fare e di intraprendere le azioni necessarie a realizzare quanto deciso

Sistemi multi-agente: sistemi costituiti da un insieme di agenti interagenti (interazione = scambio di messaggi tra agenti artificiali)

A. Montanari, **Intenzioni e fini nei sistemi artificiali intelligenti**, in: Scelte razionali, intenzionalità, fini, a cura di S. Rondinara, Città Nuova, 2014, pp. 156-181.

Naturale e artificiale

La prospettiva di Minsky non è l'unica possibile.
Punto di vista alternativo: le macchine sono ciò che vi è di più umano nella natura inanimata:

l'**artificiale** quale tratto distintivo dell'umano

Per il paleoantropologo Y. Coppens, la costruzione dei primi utensili (oggetti artificiali) segna l'inizio di una **storia culturale**, di tutto ciò che non è natura (*Storia dell'uomo e cambi di clima*, Jaca Book, 2006)

Oltre che nella tecnologia, tale storia si manifesta nelle dimensioni intellettuale, spirituale, morale ed estetica dell'uomo

L’uomo e le macchine

La questione fondamentale è quella del **rapporto** dell’**uomo** con le **macchine**, a fronte della crescita della **complessità** e del grado di autonomia di queste ultime

La distinzione tra chi progetta e costruisce (**progettista** / **costruttore**) una macchina e chi la utilizza (**utilizzatore**)

Le diverse modalità di progettazione, sviluppo e realizzazione di una macchina rispetto al passato: una singola persona non è in grado di controllare l’intero processo (conoscenza **distribuita**)

Intenzionalità e macchine

- Il legame tra **intenzionalità** (umana) e capacità di creare artefatti: la creazione di artefatti è per l'uomo un modo per estendere la propria intenzionalità

  "i nostri strumenti – osserva Searle -- sono estensioni dei nostri scopi/intenzioni"

- L'intenzionalità trova modo di esprimersi nella creazione di artefatti, dei quali costituisce, in un certo senso, la causa finale (**intenzionalità derivata**), e, successivamente e in modo diverso, nel loro utilizzo, ma non si trasferisce ad essi

L’intelligibilità delle macchine - 1

Il legame tra intenzionalità e intelligibilità delle macchine

L'**intelligibilità delle macchine**, ossia la possibilità di descriverne in modo comprensibile le caratteristiche strutturali e funzionali e le tecniche di costruzione, è condizione essenziale per il loro sviluppo e il loro utilizzo

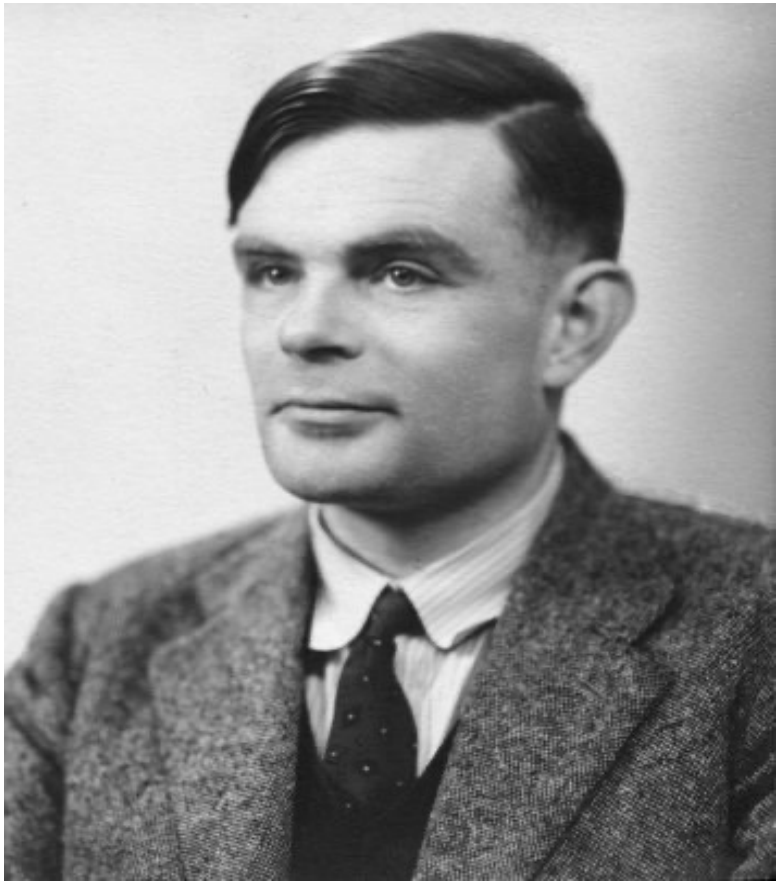
Solo l'esistenza di una **spiegazione adeguata** (razionale) del funzionamento di una macchina complessa consente, infatti, di predirne, nei limiti del possibile, il comportamento e di diagnosticarne eventuali guasti e malfunzionamenti

L’intelligibilità delle macchine - 2

La spiegazione mediante il **paradigma riduzionista**: l'analisi del sistema nel suo complesso viene ridotta all'analisi separata delle sue componenti elementari e delle loro interazioni

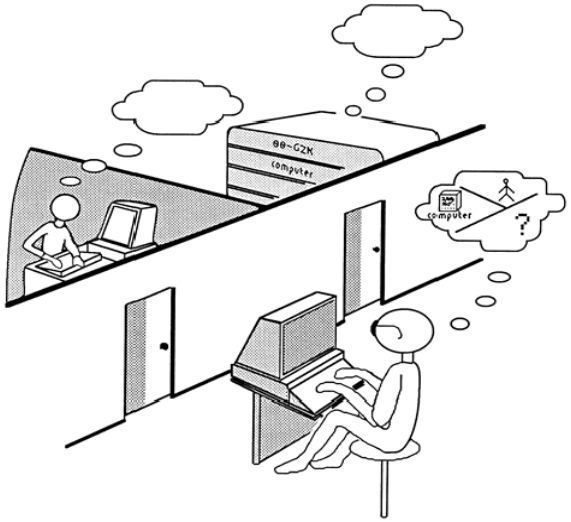
Efficace nel caso di macchine relativamente semplici, tale approccio diventa problematico in presenza di **meccanismi di controllo** (meccanismi di anticipazione e meccanismi di retroazione). Tali meccanismi possono essere visti come il tentativo di introdurre nella macchina un'opportuna **rappresentazione** dell'obiettivo (causa finale) per il quale la macchina è stata costruita

Alan Turing



Intelligenza e macchine

Il **test di Turing** (o gioco dell'imitazione): una macchina può essere definita intelligente se riesce a convincere una persona che il suo comportamento, dal punto di vista intellettuale, non è diverso da quello di un essere umano medio



Il test di Turing: dettagli - 1

Il test si svolge in 3 stanze separate. Nella prima si trova l'esaminatore umano (A); nelle altre due vi sono rispettivamente un'altra persona e il computer che si sottopone al test. Dei due A conosce i nomi (B e C), ma ignora chi sia la persona e chi il computer.

Sia B che C si relazionano separatamente con A attraverso un computer. Via computer A può porre domande a B e C e leggere le loro risposte. Compito di A è scoprire l'identità di B e C (**chi è la persona, chi è la macchina?**) entro un limite di tempo prefissato.

Il test di Turing: dettagli - 2

A può effettuare qualunque tipo di domanda; il computer ovviamente cercherà di rispondere in modo tale da celare la propria identità.

La **macchina supera il test** se A non riesce a identificarla nel tempo prefissato. Il test verrà ripetuto più volte, coinvolgendo anche esaminatori diversi, in modo da ridurre i margini di soggettività.

Turing e il comportamentismo

E’ evidente l'influenza del **comportamentismo** in auge nella prima metà del novecento sul gioco dell’imitazione proposto da Turing per stabilire se una macchina possa essere definita intelligente

Per il comportamentismo (metodologico), l'introspezione non è uno strumento adatto allo studio della mente in quanto non può fornire alcun dato affidabile. L’unica alternativa praticabile è lo studio delle misurazioni degli stimoli/percezioni forniti ad un uomo/animale e delle risposte/azioni risultanti

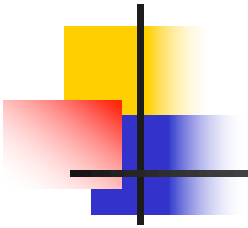
Intelligenza e linguaggio

Il Test di Turing assume un legame molto stretto tra intelligenza e **linguaggio** (naturale): l'intelligenza si manifesta nell' interazione/comunicazione attraverso il linguaggio.

Il linguaggio occupa un ruolo fondamentale nel rudimentale **sistema esperto ELIZA** (psicanalista digitale) proposto da Weizenbaum negli anni '60.

La comprensione di testi in linguaggio naturale è al centro del famoso **esperimento mentale della stanza cinese** di Searle (impossibilità per una macchina di manifestare l'**intenzionalità** che caratterizza gli esseri umani e, sia pure in forme diverse, gli animali).

All'ambito dell'elaborazione del linguaggio naturale appartiene anche l'universo dei sistemi conversazionali (Chatbot), dei quali **ChatGPT** è uno dei rappresentanti più noti.



Intelligenza e corporeità

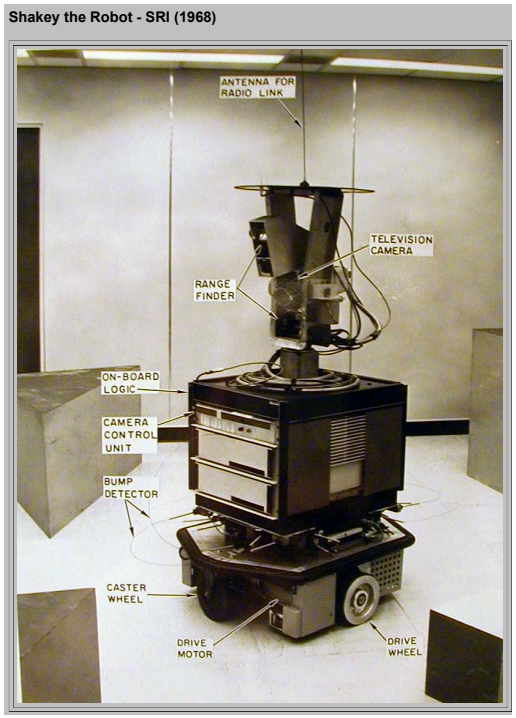
Il modello di intelligenza sotteso al Test di Turing è un modello astratto/disincarnato dell'intelligenza.

Una delle acquisizioni più importanti della ricerca in IA degli ultimi decenni è la consapevolezza del ruolo cruciale che gli organi di senso svolgono nell'interazione dell'uomo col mondo e della conseguente impossibilità di un'intelligenza (artificiale) priva di “corporeità”. Ciò ha portato allo sviluppo di un rapporto sempre più stretto tra IA ("cervello senza corpo") e robotica ("corpo senza cervello").

Per paradossale che possa suonare, per avvicinarsi all'intelligenza umana l'IA deve diventare un'**intelligenza incarnata**.

Shakey – anni ‘60/‘70

Tanto è stato fatto da quando, nella seconda metà degli anni '60, a Stanford, è stato sviluppato Shakey, il primo robot mobile general-purpose, capace di "ragionare" sulle proprie azioni



TMSUK 04 – anni ’80/’90

A partire dagli anni '80, prevalentemente in Giappone, sono stati prodotti diversi robot umanoidi incaricati di svolgere compiti domestici (uno dei più famosi è TMSUK 04)

TMSUK 04

Il robot viene controllato a distanza da un sistema che, attraverso una rete di telefonia cellulare, acquisisce dal robot informazioni sul suo stato corrente e sull'ambiente in cui opera (feedback visuale) e invia ad esso le necessarie istruzioni di controllo

TMSUK 04 ha 27 gradi di libertà. E' stato venduto a diverse università e istituti di ricerca impegnati in studi di ingegneria robotica. Nel 2001 è stata sviluppata una versione a 6 ruote (TmSuk04-2), prototipo di un robot per attività di ispezione

Justin – dal 2009 (1)

Negli ultimi anni, il Centro Aerospaziale Tedesco ha sviluppato una nuova famiglia di robot umanoidi programmabili (Justin e le sue varianti) in grado di operare con una significativa autonomia

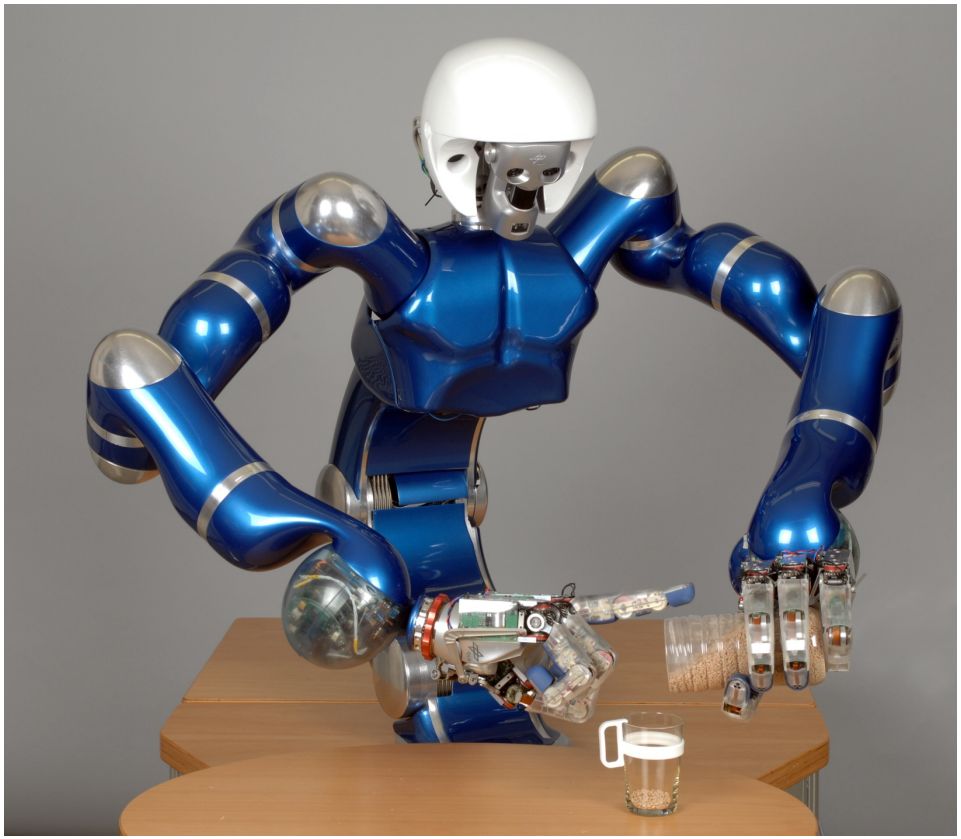
Justin – dal 2009 (2)

Dotato di braccia e ruote per il movimento, controllabile da remoto attraverso la telepresenza, Justin è stato progettato per essere installato su un satellite, al fine di modificarne, qualora necessario, il comportamento (direzione e modalità di movimento) e di consentire eventuali interventi di riparazione su altri satelliti

Justin può essere utilizzato anche sulla terra per eseguire alcuni semplici compiti. Sono disponibili su youtube dei video che mostrano come Justin sia in grado di preparare in completa autonomia un tè o un caffè

Justin – dal 2009 (3)

Justin impegnato nella preparazione del caffè



Qualità di un sistema artificiale

La qualità di un sistema artificiale dipende

- dalla qualità delle **conoscenze** a disposizione del sistema

  - conoscenze fattuali di senso comune sul dominio di interesse

  - conoscenze, codificate in forma procedurale, alla base delle funzionalità di comunicazione ed elaborazione delle informazioni

- dalla qualità degli **organi di senso** artificiali (percezioni visive, acustiche, tattili) e delle componenti deputate all'esecuzione delle azioni pianificate

Alcuni limiti di Justin

Dal punto di vista dell'IA, Justin presenta alcuni limiti significativi:

- Grado di **autonomia**: il comportamento di Justin è sofisticato, ma molto vincolato (pre-programmato)
- **Struttura gerarchica**: le conoscenze a disposizione di Justin sono organizzate in modo fortemente gerarchico (dal livello della percezione al livello della rappresentazione e del ragionamento simbolico e da tale livello al livello dell'azione)

I rapporti tra il livello della percezione/azione e il livello della rappresentazione e del ragionamento simbolici sono gestiti da un "magico" livello intermedio)

Occorre un'organizzazione molto più flessibile (si veda il modello della società della mente di Minsky)

John Searle

La stanza cinese di Searle

Tesi fondamentale: impossibilità per una macchina di manifestare l'**intenzionalità** che caratterizza gli esseri umani e, sia pure in forme diverse, gli animali

Intenzionalità: caratteristica che contraddistingue certi stati mentali, quali le credenze, i desideri e le intenzioni, diretti verso oggetti e situazioni del mondo

Per Searle, l'intenzionalità è un dato di fatto empirico circa le effettive relazioni causali tra mente e cervello, che consente (unicamente) di affermare che certi processi cerebrali sono sufficienti per l'intenzionalità

Intenzionalità e programmi

Per Searle, l'esecuzione di un programma su un dato input (**processo** nel linguaggio informatico comune) non è mai di per se stessa una condizione sufficiente per l'intenzionalità

La "dimostrazione"

Searle immagina di sostituire un agente umano al calcolatore nel ruolo di esecutore di una specifica istanza di un programma e mostra come tale esecuzione possa avvenire alcuna forma significativa di intenzionalità

L’esperimento mentale di Searle

La struttura dell'**esperimento mentale**: una teoria della mente può essere confermata/falsificata immaginando che la propria mente operi secondo i principi di tale teoria e verificando la validità o meno delle affermazioni/previsioni della teoria

L'**esperimento mentale di Searle**:

Searle prende in esame i lavori sulla simulazione della capacità umana di **comprendere narrazioni**

Caratteristica distintiva di tale abilità: la capacità di rispondere a domande che coinvolgono informazioni non fornite in modo esplicito dalla narrazione, ma desumibili da essa sfruttando conoscenze di natura generale

L'esperimento in dettaglio - 1

Searle immagina che una persona venga chiusa in una stanza e riceva **3 gruppi di testi** scritti in una lingua a lei sconosciuta (**cinese**), interpretabili (da chi fornisce i testi) rispettivamente come il testo di una narrazione, un insieme di conoscenze di senso comune sul domino della narrazione e un insieme di domande relative alla narrazione

Immagina, inoltre, che tale persona riceva un **insieme di regole**, espresse nella propria lingua (**inglese**), che consentano di collegare in modo preciso i simboli formali che compaiono nel primo gruppo di testi a quelli che compaiono nel secondo e **un altro insieme di regole**, anch'esse scritte in una lingua a lei nota, che permettano di collegare i simboli formali che compaiono nel terzo gruppo di testi a quelli degli altri due e che rendano possibile la produzione di opportuni simboli formali in corrispondenza di certi simboli presenti nel terzo gruppo di testi

L'esperimento in dettaglio - 2

Le **regole** vengono interpretate (da chi le fornisce) come un **programma** e i **simboli prodotti** come **risposte** alle domande poste attraverso il terzo gruppo di testi. Quanto più il programma è ben scritto e l'esecuzione delle regole spedita, tanto più il comportamento della persona sarà assimilabile a quello di un parlante nativo (un cinese)

Immaginiamo ora uno scenario in cui la persona riceva il testo narrativo e le domande ad esso relative nella propria lingua (**inglese**) e fornisca le risposte in tale lingua, sfruttando la propria conoscenza di senso comune. Tali risposte saranno indistinguibili da quelle di un qualunque altro parlante nativo, in quanto la persona è un parlante nativo

L'esperimento in dettaglio - 3

Dal punto di vista esterno, le risposte fornite in lingua cinese e quelle fornite in lingua inglese saranno egualmente buone; il modo in cui vengono prodotte è, però, radicalmente diverso.

A differenza del secondo caso, nel primo caso le risposte vengono ottenute attraverso un'opportuna manipolazione algoritmica di simboli formali ai quali la persona non associa alcun significato (simboli non interpretati).

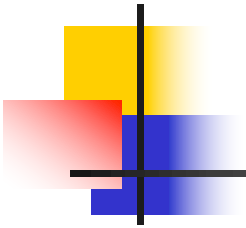
Il **comportamento della persona** è, in questo caso, del tutto **assimilabile** all'**esecuzione di un programma** su una specifica istanza da parte di un sistema artificiale.

Esito dell'esperimento

Esito dell'**esperimento**: la capacità (di un uomo/una macchina) di manipolare le informazioni ricevute secondo regole formali ben definite non è sufficiente a spiegare il processo di comprensione (non vi è nemmeno alcuna evidenza che essa debba essere una condizione necessaria) – "carattere non intenzionale, e, quindi, semanticamente vuoto, dei simboli elaborati da un sistema artificiale" (Diego Marconi)

Conclusione: i processi mentali non possono essere ridotti a processi di natura computazionale che operano su elementi formalmente definiti

Osservazione: confutazione della validità del cosiddetto test di Turing



Conseguenze

L'affermazione dell'**irriducibilità** dell'intenzionalità all'esecuzione di programmi su particolari input ha alcune importanti conseguenze:

- impossibilità di spiegare le modalità con le quali il cervello produce l'intenzionalità attraverso il meccanismo della istanziazione di programmi
- ogni meccanismo in grado di produrre intenzionalità deve avere abilità di tipo causale pari a quelle del cervello

Problemi (irrisolti): cosa differenzia il caso in cui la persona comprende il testo (inglese) da quello in cui non vi è alcuna comprensione (cinese)? Questo qualcosa può (se sì, come) essere trasferito ad un macchina?

Il problema della sintesi

Le ragioni dell'**inadeguatezza** dei **sistemi artificiali:** l'affermazione circa l'inadeguatezza dei sistemi di manipolazione di simboli formali va inquadrata nella problematica generale dell'incompletezza, e in subordine indecidibilità, dei sistemi formali di calcolo sufficientemente espressivi

Il fatto che (l'uomo visto come) un sistema di manipolazione di simboli formali non realizzi alcuna forma di comprensione potrebbe essere spiegato con l'impossibilità di realizzare un sistema di calcolo corretto e completo in grado di catturare il processo di comprensione (lo stesso vale per le altre capacità cognitive). Se un tale sistema non può esistere, ne segue che non può essere attraverso un tale sistema che l'uomo realizza la comprensione e le altre sue capacità cognitive

Le tecniche di machine learning

Dal momento che non possiamo fornire ad un sistema un insieme di conoscenze completo e immodificabile, a partire dal quale ogni altra conoscenza di interesse possa essere derivata mediante opportune procedure, in un unico passo iniziale, il sistema deve essere in grado di **estendere** in modo automatico, ed eventualmente **rivedere**, la propria conoscenza

Diversi approcci all'acquisizione della conoscenza (tecniche di apprendimento simbolico, reti neuronali, algoritmi evolutivi). Tutte fallibili (definizione del bias induttivo, scelta del training set)

Il problema del **machine** learning (il problema dell'induzione): possono i sistemi indurre regole generali, da usare per future predizioni, sulla base delle regolarità rilevate fino ad un certo punto?

La bionica

La frontiera forse più interessante è, però, quella della **bionica**

La bionica muove da una prospettiva diversa: non più la sostituzione dell'uomo col robot nell'esecuzione di compiti sempre più sofisticati, ma uomo e macchina come **sistema integrato**

Essa integra conoscenze di biologia, neuroscienza, elettronica e informatica con l'obiettivo di impiantare all'interno del corpo umano dei dispositivi artificiali

Fra i dispositivi correntemente in uso o in avanzata fase di sperimentazione, finalizzati al recupero di capacità percettive o motorie, vi sono i dispositivi per la stimolazione riabilitativa per la terapia del dolore cronico, le protesi utilizzate per compensare anatomicamente i canali neurali, gli impianti per la neurostimolazione, gli impianti cocleari, gli impianti retinici