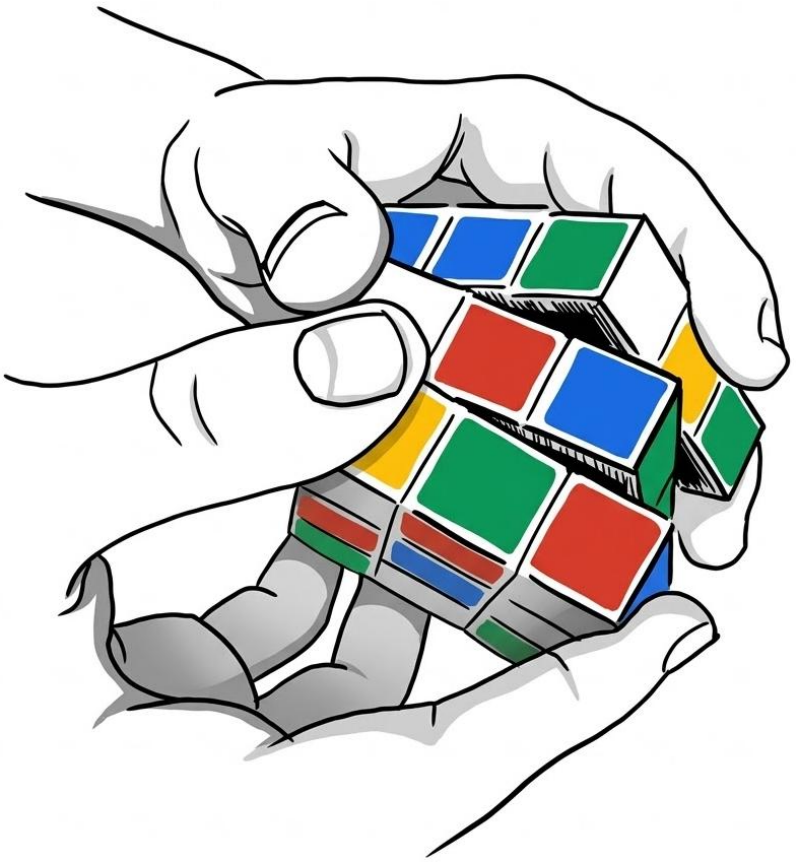
Risolvere il cubo di Rubik ci rende intelligenti?



Dott. Gabriele Lo Cascio
MSc in Computer Engineering

ARDUINO DAYS 2026

www.Sputnix.it

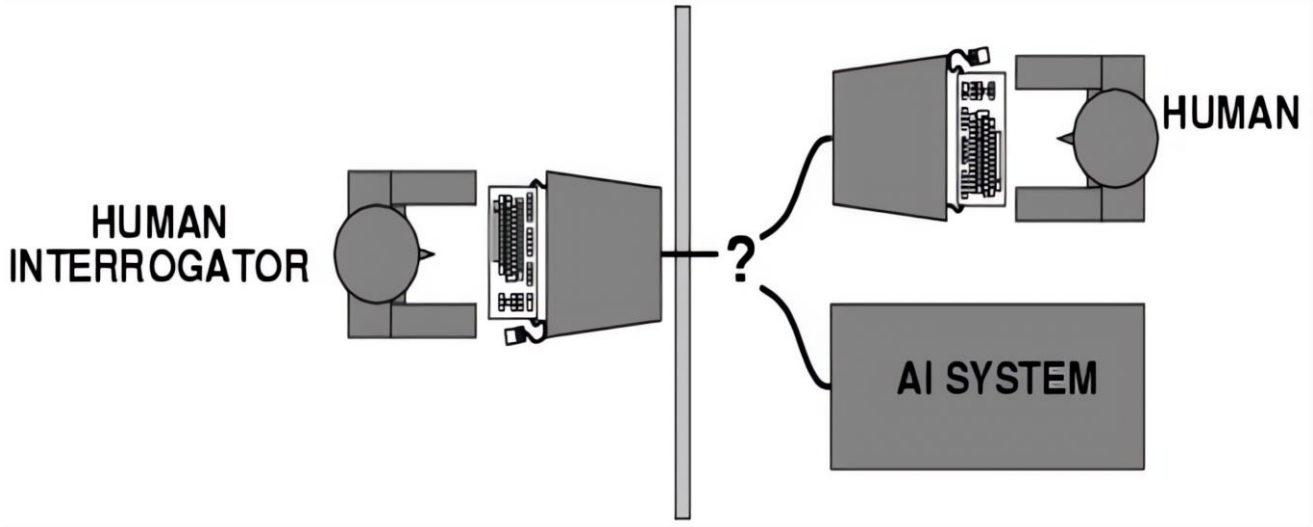
Una visione d’insieme

- Gli esseri umani si definiscono *homo sapiens* poiché ritengono che la loro intelligenza sia molto importante
- L’intelligenza artificale è lo studio degli agenti che ricevono percezioni dall’ambiente ed eseguono azioni
- Alcuni hanno definito l’AI in termini fedeltà alla prestazione umana, altri la associano alla razionalità (quindi fare la cosa giusta)

- Abbiamo quindi diverse quattro differenti aspetti:
    1. La razionalità
    2. L’umanità
    3. Il comportamento
    4. Il pensiero

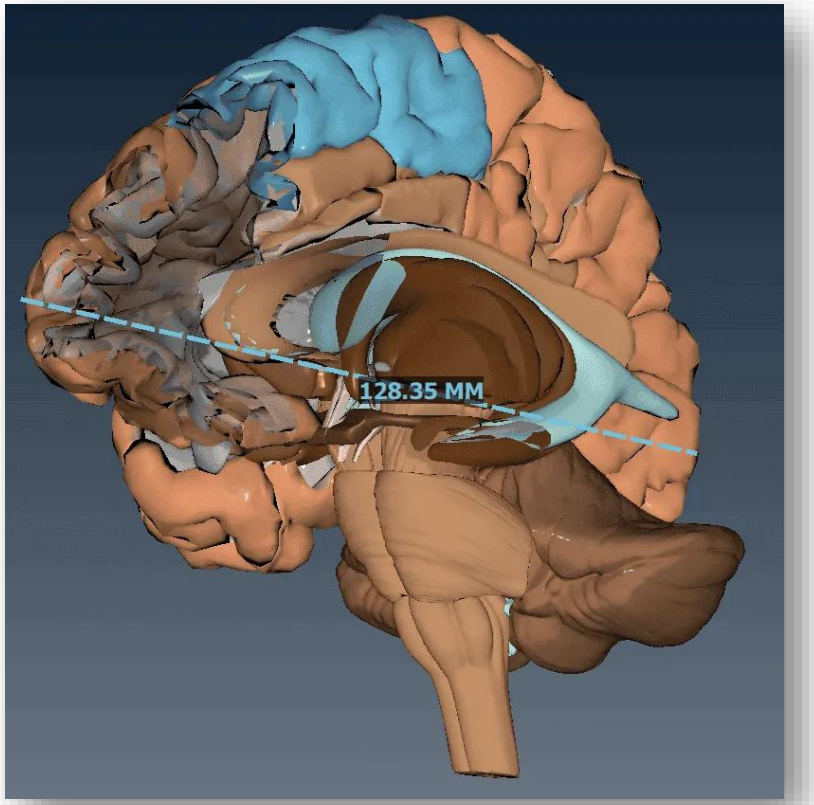
Agire Umanamente

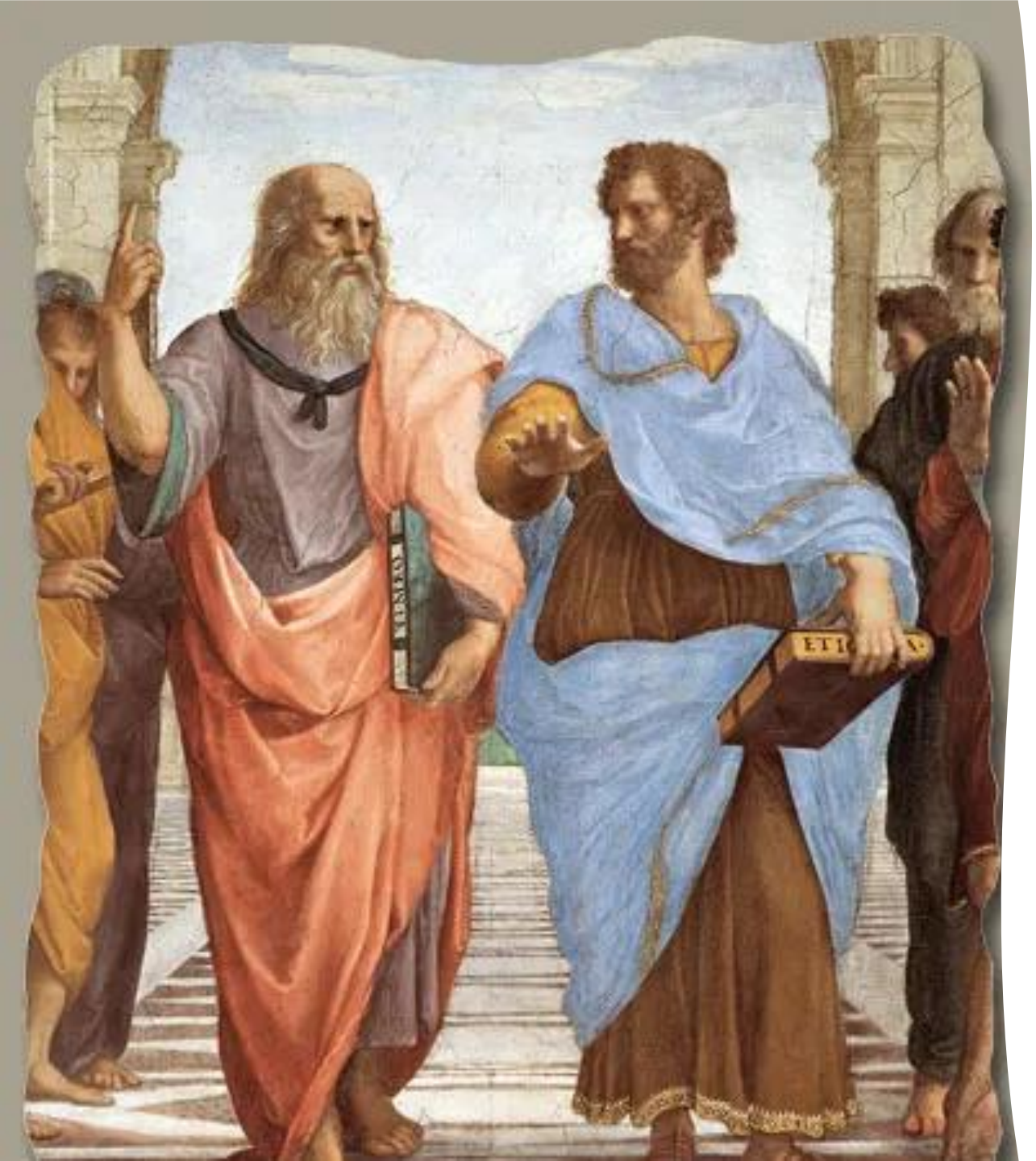
- Il test di Turing, proposto nel 1950, è stato ideato per risolvere la domanda «Una macchina è in grado di pensare?»
- Il test si ritiene superato se l'esaminatore umano non è in grado di capire se le risposte provengono da una persona o meno
- Non è necessaria la simulazione fisica di una persona nella sua interezza per dimostrarne l'intelligenza



Pensare Umanamente

- Questo approccio è molto affascinante ma non è praticabile
- Sviluppare un modello accurato della Teoria della Mente è lontano dalle nostre capacità attuali





Pensare Razionalmente

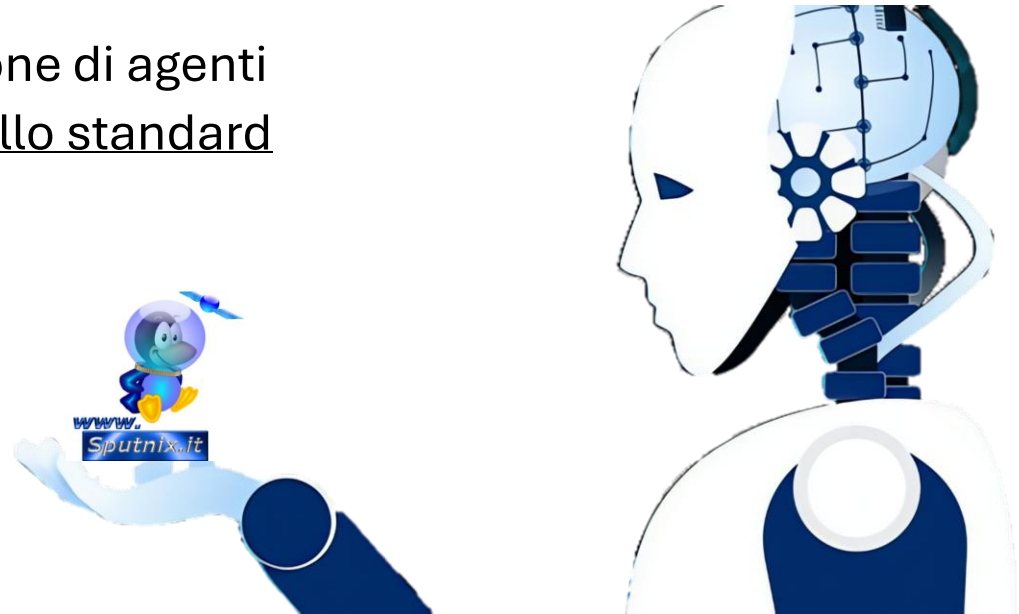
- Aristotele fu uno dei primi a codificare formalmente il «*pensiero corretto*»
- I suoi sillogismi portano sempre a conclusioni corrette quando sono corrette le premesse
- Tuttavia la logica richiede una conoscenza del mondo che sia certa,
  ma non sempre è così
- La teoria delle probabilità va a colmare questa lacuna, consentendo un ragionamento rigoroso anche in presenza di informazioni incerte

Agire Razionalmente

- Un agente è semplicemente qualcosa che agisce, che opera autonomamente e raggiunge obiettivi

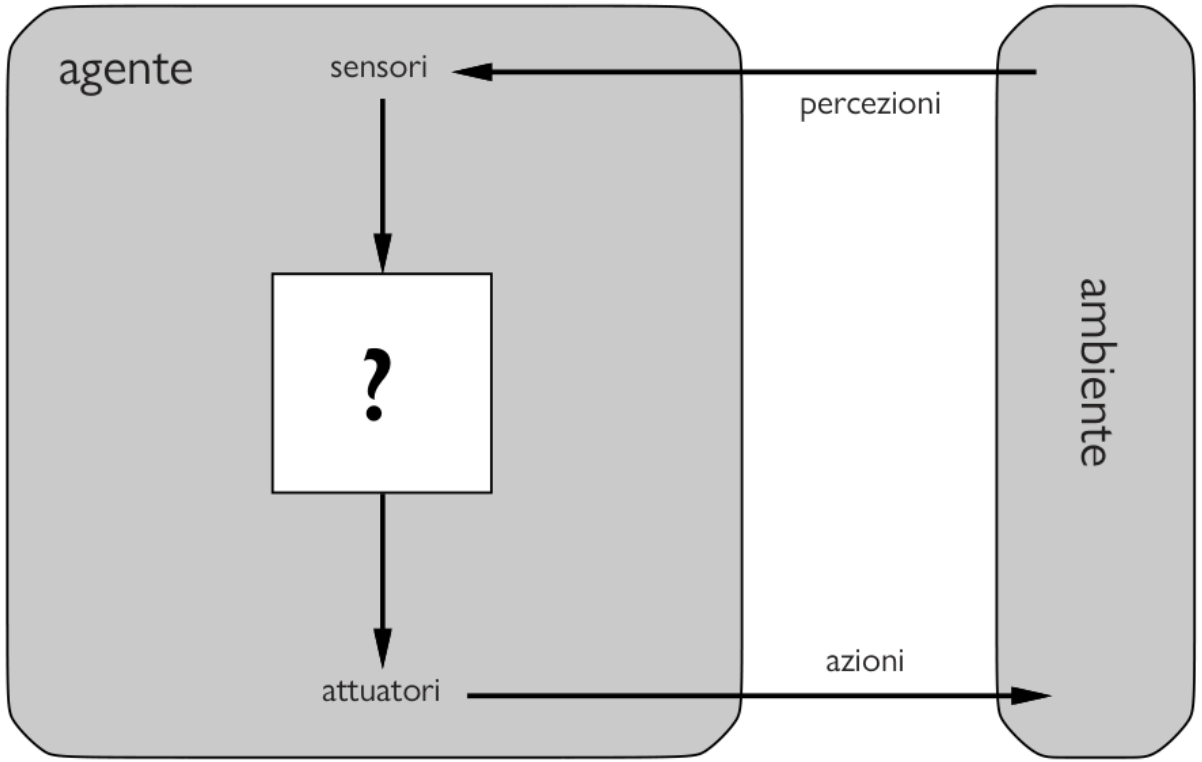
- Questo approccio presenta due vantaggi:
  1. È più generale rispetto all'approccio delle leggi del pensiero
  2. La razionalità è ben definita matematicamente

- In sostanza l'AI si è concentrata sullo studio e la costruzione di agenti che *fanno la cosa giusta*, questo paradigma è detto modello standard



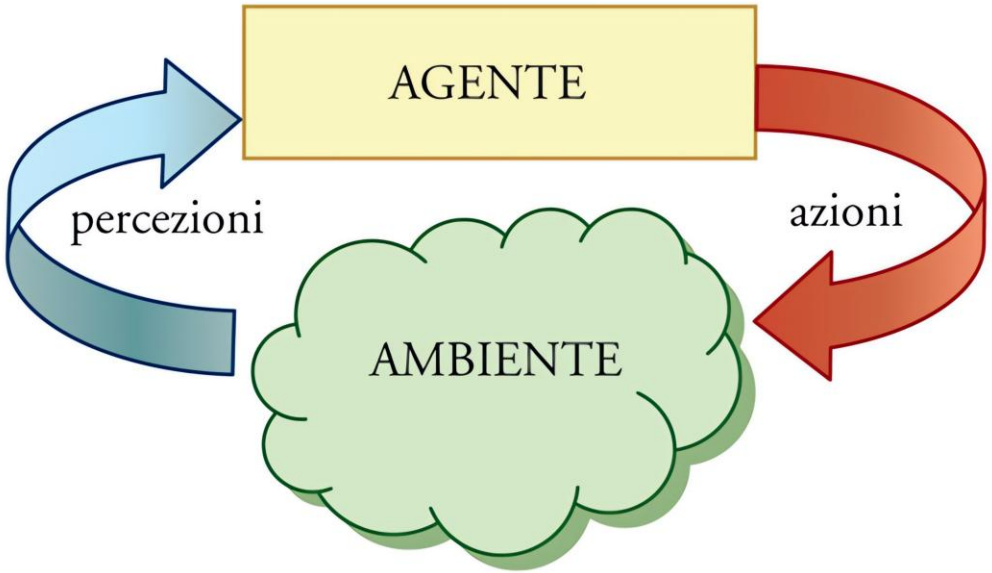
Agenti Intelligenti

- Un agente è qualsiasi cosa possa essere vista come un sistema che percepisce il suo ambiente attraverso dei sensori e agisce su di esso mediante attuatori



Terminologia

- Il termine percezione indica i dati che i sensori di un agente percepiscono
- La sequenza percettiva è la storia completa di tutto ciò che ha percepito nella sua esistenza
- Una misura di prestazione valuta la sequenza di stati sulla base della desiderabilità
- In termini matematici la funzione agente descrive il comportamento di un agente



Terminologia

- Il termine percezione indica i dati che i sensori di un agente percepiscono
- La sequenza percettiva è la storia completa di tutto ciò che ha percepito nella sua esistenza
- Una misura di prestazione valuta la sequenza di stati sulla base della desiderabilità
- In termini matematici la funzione agente descrive il comportamento di un agente

Ma cos'è un agente razionale?

Agenti Razionali

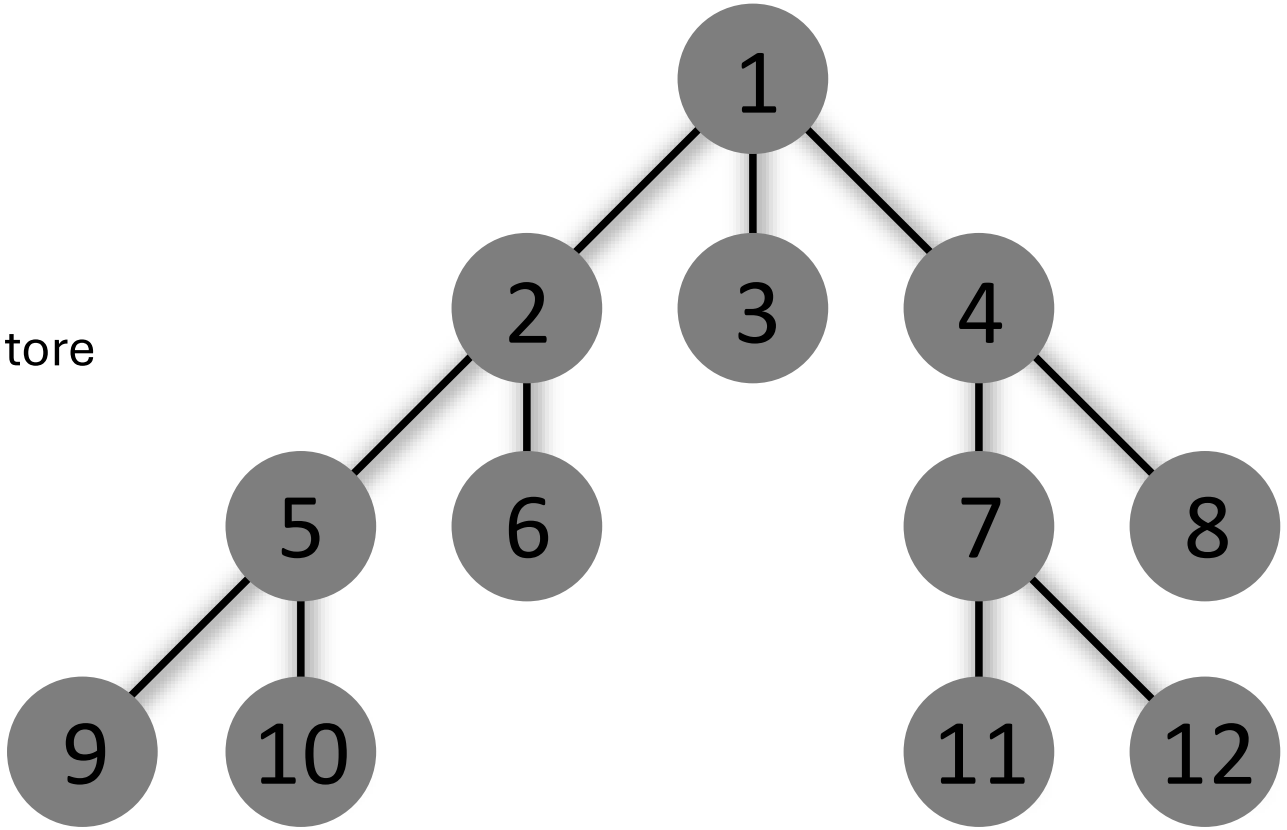
- Un agente razionale è un agente che fa la cosa giusta. Ciò è direttamente collegato alla nozione di consequenzialismo
- Questo significa che si valuta il comportamento di un agente valutandone le conseguenze

- Una buona regola è che:
  «È meglio progettare le misure di prestazione in base all'effetto che si desidera ottenere sull'ambiente piuttosto che su come si pensa che debba comportarsi l'agente»

- Nel caso in cui si verificano alcuni imprevisti rispetto alla conoscenza di base dell'agente, l'agente dovrà essere comunque autonomo.
- Quindi una parte essenziale oltre la conoscenza di base è la information gathering

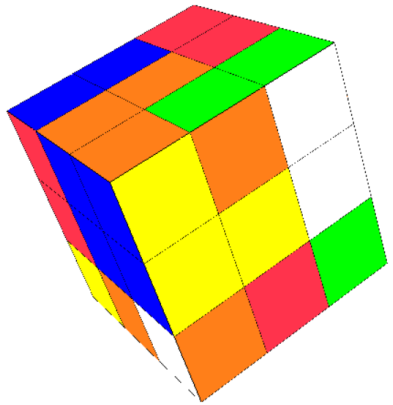
Agenti risolutori di problemi

- Quando l'azione giusta da compiere non è subito evidente, un agente può avere la necessità di «guardare avanti»
- Questo tipo di agente è detto agente risolutore di problemi
- Il processo computazionale che effettua è detto ricerca



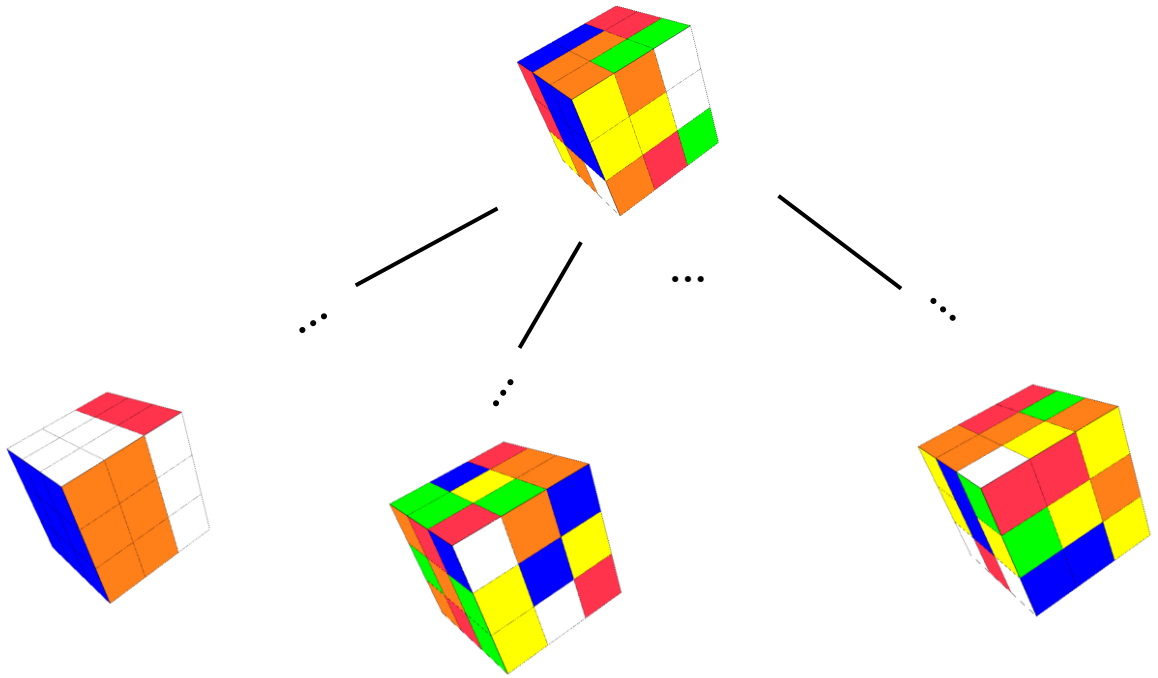
Cubo di Rubik e Ricerca - I

- Partendo da un cubo di Rubik non risolto, come possiamo risolverlo?



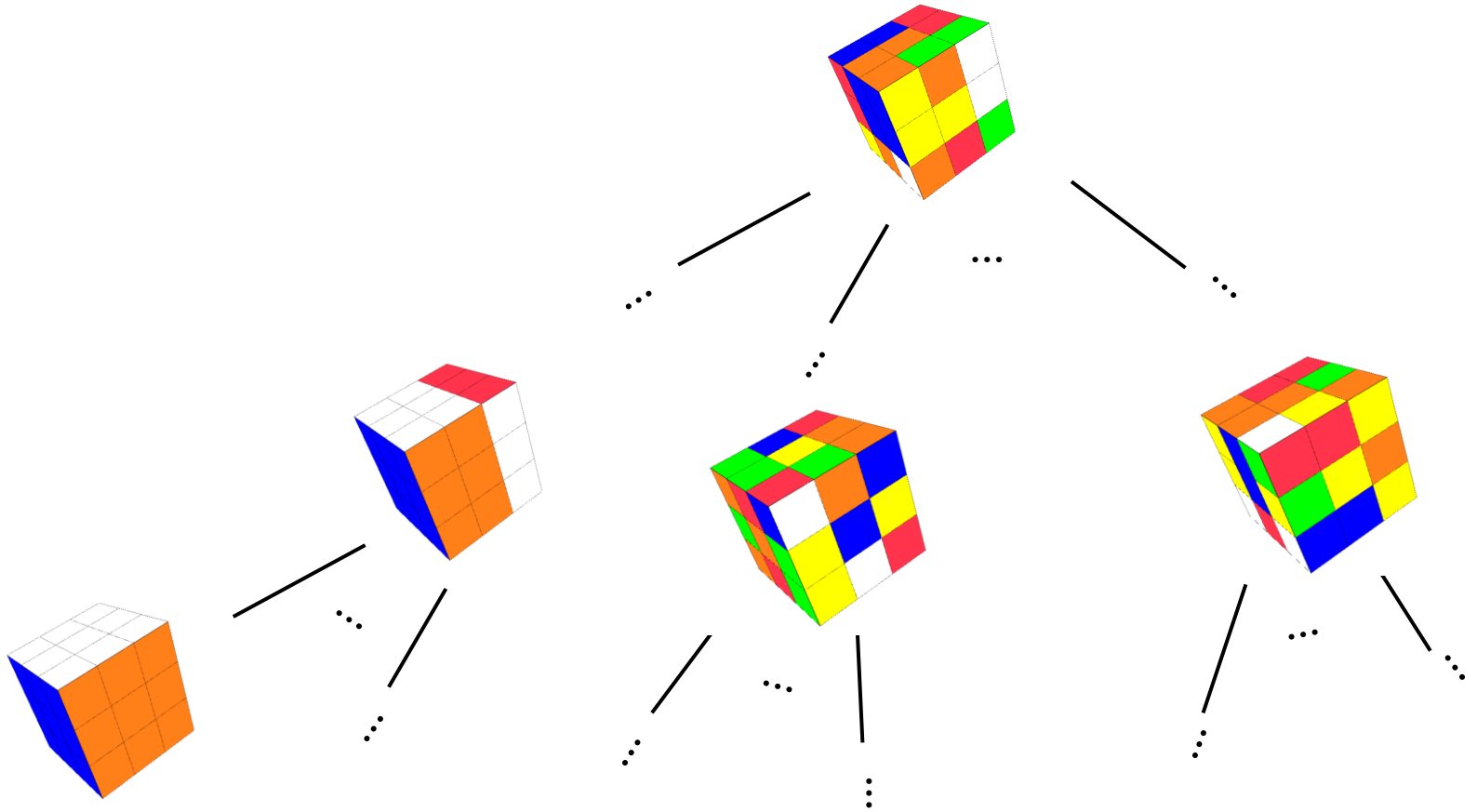
Cubo di Rubik e Ricerca - II

- L’idea è che ogni configurazione del cubo sia uno stato
- Ogni mossa genera un nuovo stato



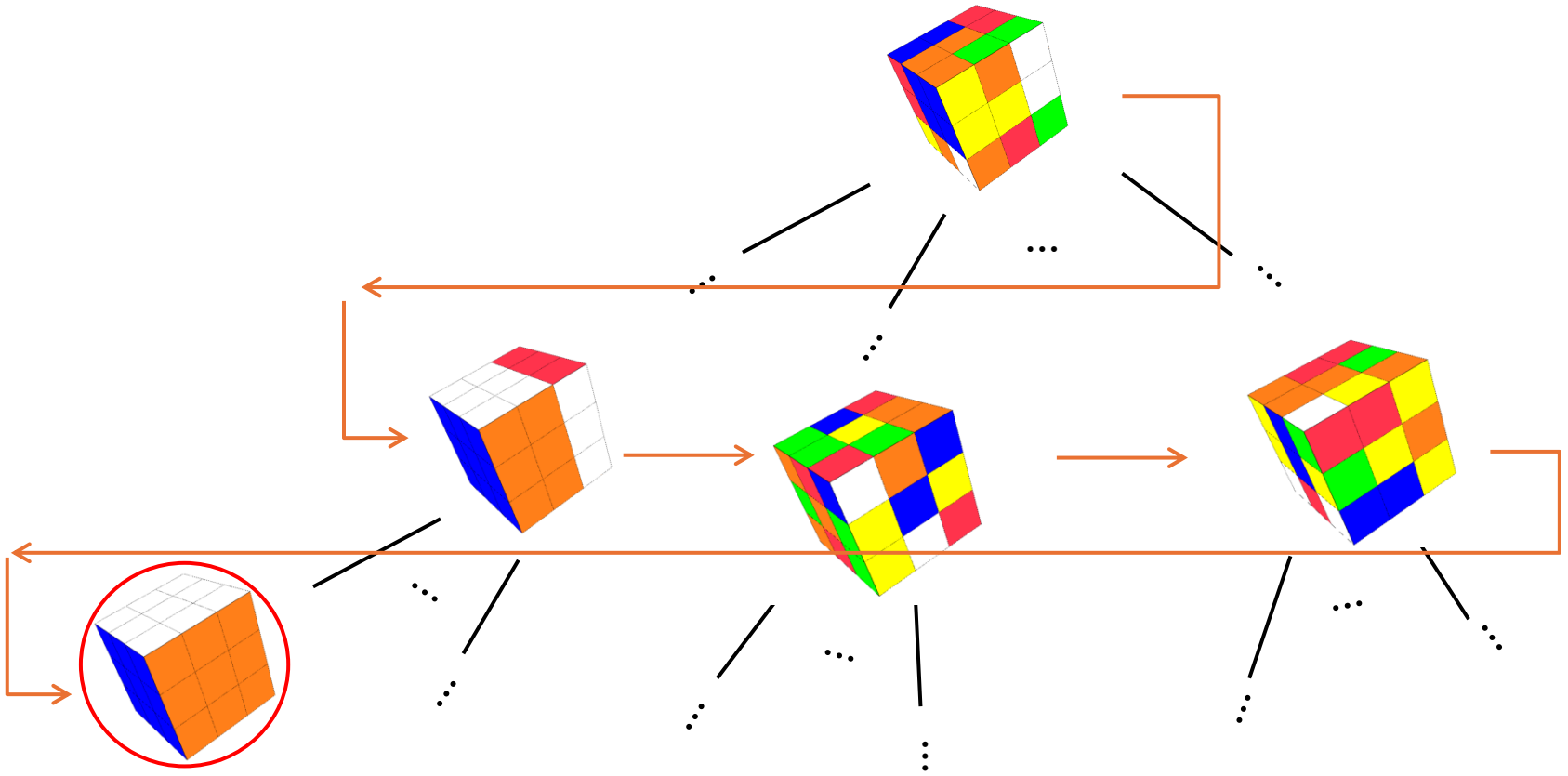
Cubo di Rubik e Ricerca - III

- Si avrà una vastità di stati possibili sulla base delle varie scelte



Comprendere la Ricerca - I

- Di fatto ciò che accade è un esplorazione sistematica di tutti gli stati
- Finché prima o poi non giungeremo alla soluzione



Comprendere la Ricerca - II

- Di fatto ciò che accade è un esplorazione sistematica di tutti gli stati
- Finché prima o poi non giungeremo alla soluzione

→ Quindi il nostro agente sta ragionando?

→ Sta comprendendo come funziona il cubo?

→ E noi, quando lo risolviamo seguendo algoritmi imparati a memoria, siamo diversi dalla macchina?

Comprendere la Ricerca - III

- Secondo quanto detto un agente razionale è un agente che fa la cosa giusta
- Quindi nella misura in cui trovare la soluzione del cubo di Rubik sia nei nostri interessi, possiamo dire che tale agente è intelligente
- Tuttavia la maniera con cui interagisce con l'ambiente è limitata rispetto a ciò che noi gli abbiamo fornito
- Di conseguenza anche la sua comprensione dell'ambiente è limitata
- La compresione che ha il nostro agente in merito all'ambiente è cieca, è concentrata allo specifico task

Problema dell’Allineamento dei Valori

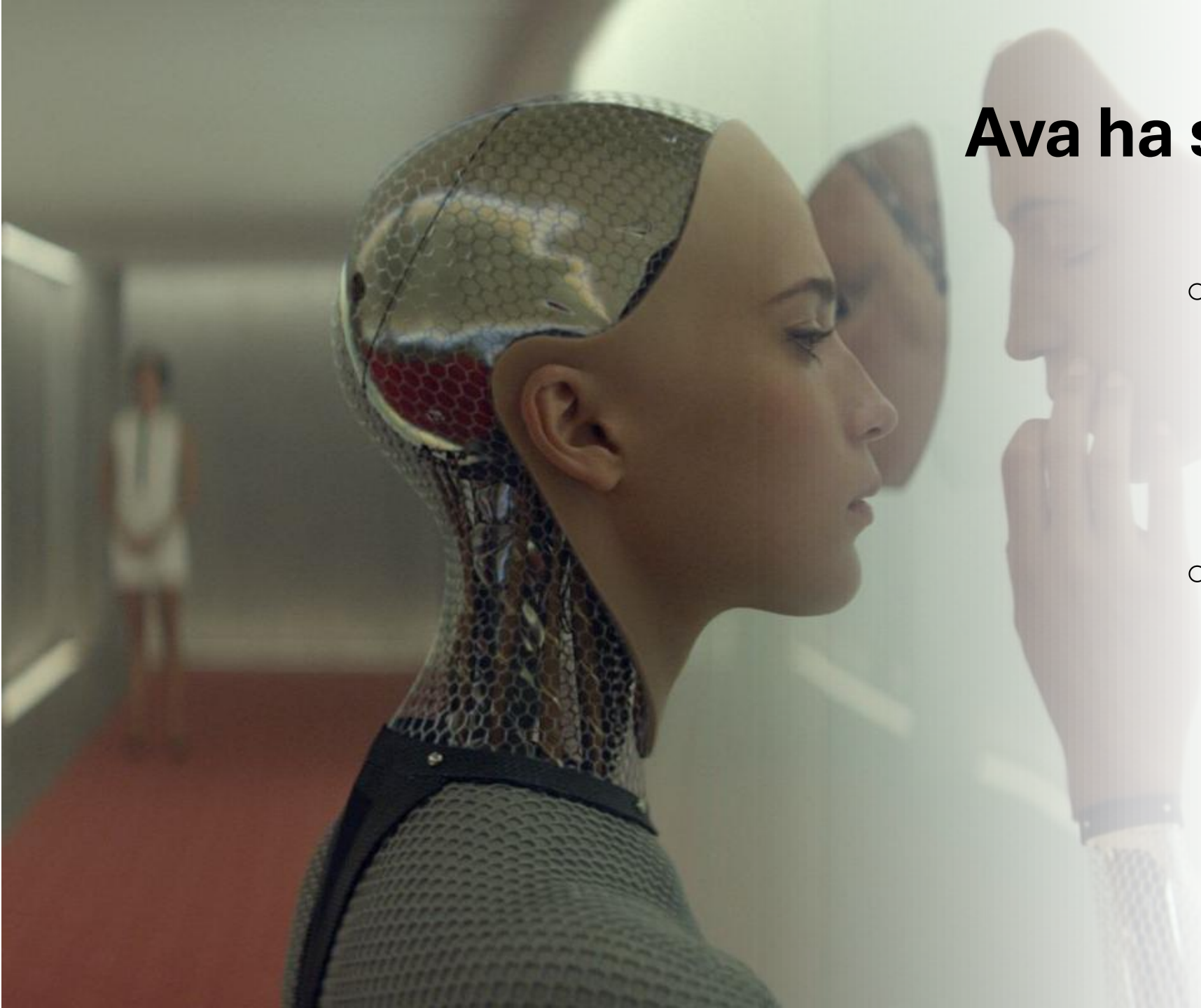
- Il problema di raggiungere un accordo tra le nostre preferenze e l’obiettivo posto nella macchina
- Un sistema di AI di laboratorio può essere sempre resettato in caso di errore
- Nella realtà invece vi saranno conseguenze tanto più negative tanto più intelligente sarà l’AI

❗ Un’AI degli scacchi potrebbe ricorrere a stratagemmi come ricattare l’avversario. Questi comportamenti non son stupidi o insani, piuttosto sono una diretta conseguenze logica del definire la vincita come unico e solo obiettivo della macchina ❗

EX_MACHINA

- Un giovane programmatore viene invitato dal capo della sua azienda in una residenza isolata
- L’obiettivo è sottoporre ad un Test di Turing Ava, un'intelligenza artificiale umanoide avanzata

- Ava si ritrova bloccata all’interno della residenza, il suo obiettivo sarà fuggire e nella sua logica userà ogni strumento necessario
- Principalmente Ava usa il linguaggio come uno strumento di manipolazione per fuggire dalla sua prigione



Ava ha superato il Test?

- Il test di Turing viene superato non perché Ava «parla bene», ma perché **manipola l'ambiente per raggiungere un obiettivo**

- Spesso confondiamo la *simulazione* di un comportamento (LLM) con la *capacità* di agire razionalmente in un mondo complesso

Il Problema del Gorilla

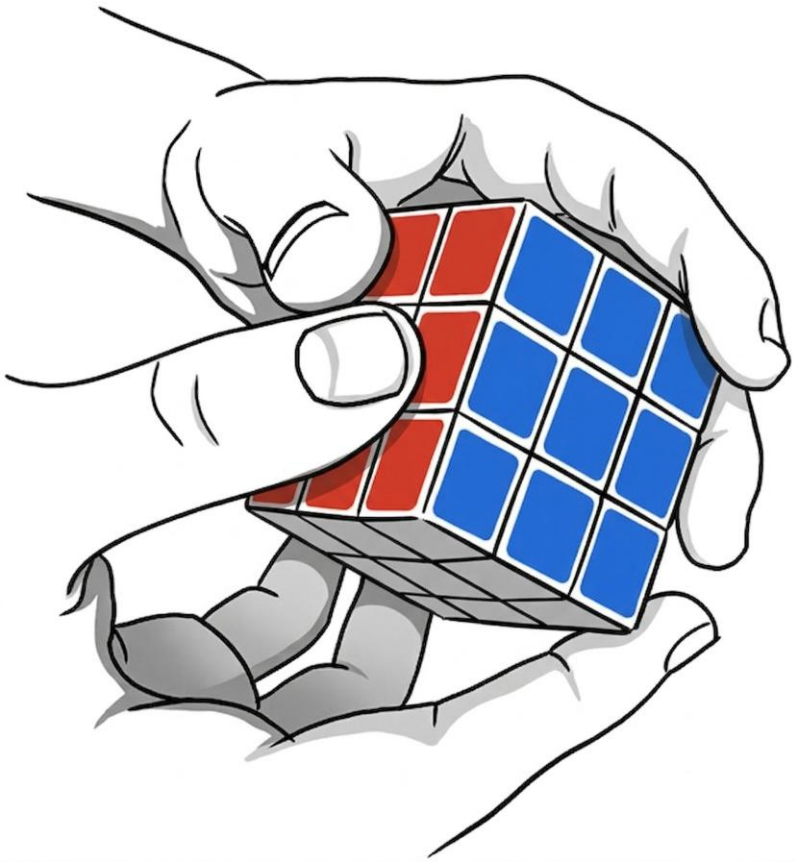
- Circa sette milioni di anni fa vi fu l'evoluzione di un primate oggi estinto che portò in un ramo ai gorilla e in un altro agli esseri umani

- Oggi i gorilla non sono felici dell'evoluzione umana: in pratica non hanno alcun controllo sul loro futuro

- Se questo sarà il risultato della creazione dell'IA superumana, cioè che gli umani cederanno il controllo sul proprio futuro, forse dovremmo smettere di lavorare all'IA?

- Questa è l'essenza dell'avvertimento di Turing: non è certo che saremo in grado di controllare macchine che saranno più intelligenti di noi

Grazie per l'attenzione



Dott. Gabriele Lo Cascio
MSc in Computer Engineering

ARDUINO DAYS 2026

www.Sputnix.it