

# 1 Lezione del 07-10-24

## 1.1 Introduzione

Il corso si pone di fornire un'introduzione, almeno a livello culturale, alla storia dell'intelligenza artificiale dalle origini fino ad ora, e alcune specifiche sui modelli più usati oggi.

### 1.1.1 Cos'è l'intelligenza artificiale?

Con intelligenza artificiale ci riferiamo ad una serie di caratteristiche che vorremo ottenere da un particolare **agente**, cioè le capacità di:

- **Percepire**, attraverso sensori, informazioni dal suo ambiente esterno;
- **Imparare**, a partire da osservazioni, fino ad inferire leggi;
- **Ragionare e pianificare**, e quindi generare nuove informazioni;
- **Interagire**, e quindi scambiare informazioni, e modificare il suo ambiente.

I due modelli fondamentali che prenderemo in esempio sono il **cervello umano** e l'**agente razionale** (che potrebbe essere, volendo, anche più razionale della sua controparte umana). Faremo poi una distinzione fra **pensiero** e **comportamento**: non è sempre rilevante farci domande sulla coscienza o meno del modello preso in analisi nel prendere le sue decisioni. Possiamo quindi distribuire queste caratteristiche, fra di loro ortogonali, a formare quattro categorie:

- **Sistemi che pensano come gli umani;**
- **Sistemi che pensano razionalmente;**
- **Sistemi che si comportano come gli umani;**
- **Sistemi che si comportano razionalmente.**

Uno dei test più famosi per distinguere il comportamento di un programma *intelligente* è il cosiddetto **test di Turing**, il cui predicato è che per essere intelligente, un programma dovrebbe essere indistinguibile, a un osservatore esterno che comunica per via scritta, da un essere umano. Alcuni ricercatori inseriscono poi nel test di Turing altre facoltà, quali quelle di comunicazione in linguaggio naturale, interazione fisica con l'ambiente ecc... In ogni caso, per passare il test di Turing nella sua formulazione originale, un programma avrebbe bisogno di:

- **Elaborare il linguaggio naturale;**
- **Rappresentare l'informazione** che conosce;
- **Ragionamento automatico** sulle informazioni ottenute;
- **Apprendimento automatico** di nuove informazioni.

Questi rappresentano i rami principali dell'IA.

### 1.1.2 Pensare come umani

Per pensare come umani, dobbiamo prima capire come pensano gli umani. Alcune tecniche possono essere:

- Introspezione;
- Esperimenti fisiologici o sociologici;
- Imaging cerebrale (risonanza magnetica, ecc..)...

Ammesso di poter estrapolare qualcosa di utile da questi strumenti, dobbiamo poi ricordare che gli umani possono essere irrazionali, commettere errori, e generalmente non sono agenti perfetti. Potremmo quindi desiderare un'altro tipo di modello:

### 1.1.3 Pensare ragionevolmente

Invece di pensare come umani, potremmo decidere di pensare in maniera ragionevole. Questo richiederebbe scoprire **leggi di pensiero** definite a priori: quelle che definisce la logica.

Il problema di questo approccio però è che non è in grado di rispondere a informazioni incerte: ogni ambiente reale invece, avrà inevitabilmente un certo grado di incertezza nella sua rappresentazione. Inoltre, cercare sempre la risposta ottima potrebbe richiedere troppe risorse dal punto di vista computazionale per essere effettivamente utile.

### 1.1.4 Agire ragionevolmente

Un modello migliore potrebbe essere quello di un **agente razionale**, che si comporta in maniera tale da ottenere il miglior possibile risultato, ovvero *fa la cosa giusta*.

## 1.2 Agenti razionali

Un agente è un'entità che **percepisce** (attraverso sensori) e **agisce** (attraverso attuatori). L'agente **razionale** seleziona l'azione che massimizza la sua **utilità attesa**, cioè l'opzione localmente migliore. Le caratteristiche dello spazio percepito, e delle azioni possibili, dettano le tecniche per la scelta di azioni ragionevoli.

Un modello di questo tipo non usa sempre l'inferenza logica, ma a volte adotta linee di pensiero più induttive: conviene togliere la mano dalla pentola bollente prima di pensare alle conseguenze. Inoltre, a volte non c'è una cosa migliore da fare, ma bisogna comunque fare una scelta. Decidiamo quindi di interessarci solo a come il programma agisce, e non al modo in cui "pensa" ragionevolmente al modo di approcciare il problema.

### 1.2.1 Operazioni

Per permettere al programma di agire ragionevolmente, si stabilisce una **metrica di performance** fissa, che valuta la sequenza di azioni osservate sull'ambiente. Abbiamo quindi che un'operazione è descritta da:

- Una metrica di performance;
- L'ambiente;
- Attuatori e sensori.

Chiamiamo questo schema **PEAS**, dall'inglese *Performance measure, Environment, Actuators and Sensors*.

Ad esempio, per un modello di guida automatica, la metrica sarà la velocità, la sicurezza, il comfort dei passeggeri, ecc... l'ambiente sarà la strada, gli attuatori saranno i comandi della vettura e i sensori saranno telecamere o radar posti in maniera tale da consentire la vista (*computer vision*) della strada, oppure ancora saranno l'odometro della vettura, eventuali accelerometri, ecc...

### 1.2.2 Ambiente

L'ambiente può essere così caratterizzato:

- Osservabile / parzialmente osservabile;
- Deterministico / stocastico / strategico;
- Episodico / sequenziale;
- Statico / dinamico;
- Discreto / continuo;
- A singolo agente / multiagente.

Ad esempio, gli scacchi sono un'ambiente completamente osservabile, strategico, sequenziale, semistatico (diventa statico senza un orologio), discreto e multiagente. Il gioco del poker, invece, è solo parzialmente osservabile, strategico e stocastico, sequenziale, statico, discreto e multiagente. Infine, l'esempio del taxi di prima era parzialmente osservabile, stocastico, sequenziale, dinamico, continuo e multiagente.

### 1.2.3 Tipi di agenti

Gli agenti possono essere, ordinati per la loro complessità:

- **A riflesso semplice:** usano regole "if-then", solitamente non hanno memoria, e quindi hanno una visibilità limitata del passato e del futuro;
- **A riflesso con stati:** possono immagazzinare informazioni osservate precedentemente, e quindi ragionare su aspetti non osservabili dello stato corrente;
- **Basati su obiettivi:** dispongono di "obiettivi" che riflettono il loro "desiderio", ergo cercano di modellizzare il loro ambiente esterno e a testare soluzioni per raggiungere questi obiettivi;
- **Basati sull'utilità:** dispongono di una funzione che valuta l'utilità  $f(\text{stato}) \rightarrow \text{valore}$ , cercano di massimizzare quest'utilità reagendo di conseguenza alle fluttuazioni di questa misura;
- **Capaci di apprendere:** implementano cicli di feedback che li spingono a rispondere in modo dinamico alle variazioni dell'ambiente e dell'utilità percepita.