

LEZIONE 1

Cos'è l'intelligenza artificiale?

Qualcosa di complesso e articolato che ha un svolgimento lungo e approfondito.

È un progetto con un punto di inizio e un obiettivo: il punto di inizio è coincidente con l'inizio dell'informatica, infatti si considera l'iniziatore dell'informatica e dell'AI Alan Turing.

L'obiettivo del progetto è realizzare, quindi di ingegneria.

Ma realizzare cosa? Macchine intelligenti inteso in senso ampio.

Nonostante all'inizio l'intelligenza artificiale si occupasse di realizzare macchine intelligenti sia dal punto di vista del loro comportamento della loro interazione con il mondo che dal punto di vista degli strumenti per interagire con il mondo (hw per realizzarlo), poi le cose si sono specializzate.

Ad oggi si tende un po' di più a parlare del comportamento del sw —> macchine intelligenti inteso programmi da computer in grado di eseguire un qualche comportamento che per noi sarà ritenuto intelligente.

Chi si occupa invece di realizzare la parte hw (la parte di interazione con il mondo) non si dice più intelligenza artificiale ma di robotica.

Siccome l'AI è evoluta nel tempo (è un progetto che nasce negli anni 50) e ad oggi si occupa di:

- sw in grado di risolvere problemi complessi;
- Sw che espongono comportamenti razionali —> macchine che sanno cosa fare sulla base di obiettivi;
- Sw in grado di interagire con l'uomo in un linguaggio naturale;
- Conseguentemente macchine in grado di generare contenuti creativi (sintetizzare immagini sulla base di descrizioni testuali o costruire poesie, musica ecc...);
- Sw in grado di interagire con mondi complessi e dinamici (bot che cerca informazioni sul web, bot che suggerisce cosa fare riguardo al nostro profilo di utenza, ecc...).

Come tutti i progetti l'AI ha un punto di inizio: una lettera che John McCarthy ha mandato ai suoi colleghi a metà degli anni 50, in quel periodo abbastanza per caso si ritrova a lavare sul primo computer interattivo.

Alla domanda “cos'è l'intelligenza artificiale”, McCarthy risponde: è una scienza e un'ingegneria della

creazioni di macchine intelligenti. (...) È simile al capire l'intelligenza umana, ma AI non ha un confine ai metodi che sono biologicamente osservabili.

Per altri AI significa risolvere problemi complessi che umanamente possono essere risolti in maniera semplice, ma che dove non ci si può associare un algoritmo accanto.

Ad oggi i fenomeni che vengono considerati complessi sono ancora molto umani, per esempio camminare: umanamente è semplice, dal punto di vista algoritmico invece è difficile da descrivere.

L'AI la si può affrontare con due prospettive:

- FORTE: che dice che il progetto AI non finirà fintanto che non riusciremo a realizzare delle macchine intelligenti, qualsiasi cosa vuol dire intelligente;
- DEBOLE: AI che si concentra su problemi specifici, che cerca di risolvere questi problemi nel modo migliore possibile, scegliendo i problemi da risolvere sempre nell'ottica di realizzare un qualcosa che possiamo considerare intelligente, ma senza porsi l'obiettivo di arrivare a ricostruire le persone sintetiche. Quindi passi piccolo, che sempre più e involontariamente portano ad avvicinarsi all'AI forte.

Ad oggi si considera che il prossimo passo possa essere la Artificial general Intelligence, un tipo di intelligenza che non è più debole perché non è più legata a compiti specifici ma è generica, come quella che hanno le persone.

L'intelligenza artificiale ha avuto dei fenomeni ad ondate: sempre nel tentativo di raggiungere l'obiettivo, un po' grossolanamente sono state percorse 4 strade (Thinking humanly, thinking rationally, acting humanly, acting rationally).

Oggi si parla di più di reti neurali.

Il primo modo di fare AI è stato acting humanly: l'intelligenza artificiale la si può fare comportandosi come si comportano le persone. Questo modo viene da Touring: inventando un test (The Imitation game) si pone il problema serio di capire che cosa voglia dire realizzare una macchina intelligente e per farlo, parte dal presupposto che una macchina la si definisce intelligente se si comporta come una persona.

Test di Touring:

Prendiamo due persone C e B che giocano a scacchi, ma c'è un muro che separa i due giocatori. I

due iniziano a giocare a scacchi, nessuno dei due è un campionissimo ma sono entrambi in grado di esibire un'intelligenza sensata. Dopo un po' senza nessun avvertimento, il giocatore B viene sostituito da un programma per computer. L'obiettivo del test è misurare quanto sia intelligente nel comportamento del gioco degli scacchi il programma A, per farlo però ci serve una misura, un'unità, un numero: numero di mosse che A è in grado di giocare senza che C si renda conto dello scambio. Se il numero di mosse è basso, non va bene; al contrario con il numero di mosse elevato allora A si comporta bene nel task.

Siccome contare vuol dire generare numeri, siamo in grado di mettere in scala l'intelligenza di A: se le mosse sono poche, l'intelligenza di A sarà bassa, al contrario se le mosse sono tante, l'intelligenza di A è elevata.

Il problema però sta in C, che ha un'intelligenza a sè, nel cambiare il giocatore C, probabilmente cambierà la scala perchè cambia la capacità valutativa di C.

Queste macchine che *imitano* l'azione umana, erano dette macchine da imitation game perchè appunto il loro obiettivo non è capire non è fare, ma imitare.

Nota: le persone non agiscono così perché non sono in grado di fare meglio ma perchè hanno un pensiero che le spinge ad agire (il pensiero genera l'azione, non viceversa); quindi se vogliamo fare un passo avanti rispetto ad agire come persone, possiamo porci il problema di pensar come persone, fare quindi sistemi sw in grado di pensare come le persone = simulare l'organo fisico in cui il pensiero si forma -> il cervello.

I computer sono in grado di simulare qualsiasi sistema fisico, nell'ipotesi che esistano note delle leggi in grado di descrivere il sistema fisico, nell'ipotesi di avere abbastanza potenza di calcolo per poter fare una simulazione in un tempo ragionevole, però note queste due cose conosciamo abbastanza nel dettaglio il sistema che stiamo simulando e abbiamo abbastanza potenza di calcolo per simularne un comportamento interessante, il sistema fisico è simulabile e in maniera semplice. (Es. previsioni del tempo).

Il cervello non è nient'altro che una rete formata da neuroni, simuliamo il comportamento simulando una rete di struttura di calcolo che chiameremo neuroni. I neuroni ricevono un input tramite segnale elettrico attraverso delle propagazioni tramite la rete, che prendono il nome di dendriti, quindi

collegheremo l'input non con dei segnali elettrici, ma con dei numeri che indicheranno l'intensità.

Ci viene detto che il neurone prende i segnali elettrici e li propaga verso l'esterno attraverso delle propagazioni che prendono il nome di assone e il tutto finisce in collegamenti che prendono il nome di sinapsi.

In astratto è una formula: entrano numeri, escono numeri, quello che succede nel mezzo dipende dalla somma dei numeri che entra, sapendo che qualche numero conta più di altri, ma in modo non lineare produce questi numeri in uscita.

Quindi quello che facciamo è prendere i numeri in ingresso, attribuire un peso a ciascuno dei numeri in ingresso, sommare tutto e questo ci darebbe una struttura totalmente lineare, ma non basta, passiamo per una funzione non lineare e andiamo a produrre un uscita: per ognuna di queste facciamo un calcolo, per ognuna di queste scegliamo i pesi.

La rete neurale non è nient'altro che un simulatore di cervello, che oggi riusciamo a far funzionare bene (cosa non possibile negli anni 70) perché abbiamo computer sufficientemente potenti e con un hw sufficientemente raffinato da analizzare bene queste cose.

Quello che andiamo a costruire è un rete neurale (in italiano un rete neurale è la simulazione di una rete neuronale = rete composta di neuroni), se noi prendiamo in generale un rete neurale, che cosa fa? Prende i numeri in input, produce numeri in output e di conseguenza ho qualcosa che lega gli input agli output: in matematica è una funzione da x variabili ad y variabili, in informatica è una funzione che calcola l'output sulla base dell'input.

Il calcolare l'output sulla base dell'input però vuol dire andare a scegliere i pesi da mettere sui collegamenti tra neuroni e neuroni, a questo punto ci vuole un programmatore che non scriva un programma ma che vada a definire i pesi e se ne abbiamo 10^{12} , diventa complicato; quindi anziché programmare la rete neurale, la si addestra: si forniscono degli input, si forniscono gli output corretti, si cercano dei pesi in grado di mappare l'input nell'output, poi si fornisce un altro input, un output corretto, si forniscono due pesi in grado di mappare i due output corretti, ecc... si continua così fino a quando tutti gli input che poniamo non vengono mappati in modo abbastanza preciso con tutti gli output corretti, se non ci riesce vuol dire che abbiamo bisogno di più neuroni, se la contrario si riesce subito vuol dire che ce ne sono troppi. Il problema è poi avere un algoritmo di apprendimento, cioè

un algoritmo in grado di calcolare i pesi, che sono numeri reali.

Rete convoluzionale: è un tipo di rete neurale artificiale spesso utilizzato per l'elaborazione delle immagini. Si basa sul concetto di convoluzione, che è un'operazione matematica che consente di estrarre caratteristiche da delle immagini.

Se si forza una rete ad avere una struttura convoluzionale/convolutiva, allora va fatta con più di 2 strati (input, output + intermedi) e allora si parla di rete profonda (deep network) perchè ha più di due strati e un addestramento di una rete profonda prende il nome di deep learning.

Una volta che abbiamo capito che agire come una persona ci porta ai giochi di imitazione, pensare come una persona ci porta a simulare il cervello, ci rendiamo poi conto che le persone non sono interessanti. In relata dell'esperienza umana, quello che ci piace di più è la parte razionale: agire razionale e pensare razionale.

Il pensiero razionale vuol dire la logica: ne conosciamo due logiche tra le più semplici: logica delle proposizioni e la sua estensione temporale LTL, ovviamente non sono sufficienti per poter descrivere la logica razionale, ci servirà altro più raffinato (logica del primo ordine).

Il pensiero razionale è descritto bene dalla logica, quindi nessuno ci vieta di utilizzarla per descrivere il pensiero che vogliamo abbiamo i nostri sistemi sw.

Seguire la logica passa per un'idea che è quella di descrivere lo stato del nostro programma tramite un insieme di fatti che ci dicono cosa riteniamo vero e tramite un insieme di regole di deduzione (regole di inferenza), ci dicono cosa sarà vero una volta che noi abbiamo certi fatti veri

– Facts:

$mother_of(ann, bob), sister_of(claire, ann)$

– Inference rules:

$mother_of(X, Y) \wedge sister_of(Z, X) \rightarrow aunt_of(Z, Y)$

Questo è un sistema esperto perchè conosce come un esperto, un insieme di fatti, un insieme di regole, ha una descrizione del mondo e in base a quella sa produrre nuove verità logiche incontrovertibili.

Pensare non basta, i sistemi di AI agiscono: nel momento in cui noi abbiamo un essere razionale

vogliamo immergerlo in un mondo e permettergli di muoversi nel tentativo di raggiungere i propri obiettivi. Un sistema sw immerso in un mondo in grado di ascoltare lo stato del mondo, in grado di compiere azioni sul mondo, che in modo razionale cerca di portare il mondo verso uno dei suoi stati di goal prende il nome di agente intelligente.

Un agente vuole il mondo (agente e ambiente) in certi stati, e agisce autonomamente per portarceli. L'agente è caratterizzato dal suo comportamento ma non da cosa lo ha spinto a comportarsi in quel modo. Gli agenti sono spesso basati su linguaggi logici e, di conseguenza, usano la deduzione per decidere quale azione eseguire sull'ambiente e dato che sono razionali, non gli è richiesto imitare il comportamento umano.

Agenti intelligenti che sanno cosa fare in modo eventualmente stocastico e sono in grado di motivare le decisioni.

Se prima si considerava un singolo agente, ad oggi si considerano sempre più spesso sistemi multi-agente.

GPT (Generative Pre-Trained Transformers) sono delle reti neurali convoluzionali profonde ricorrenti che sono state addestrate per prevedere il prossimo testo a fronte di un testo parziale. (CHAT GPT, il completatore di parole sulla keyboard).

Gli LLM (modelli di linguaggio di grosse dimensioni) sono stati addestrati per fare questo compito: completare quello che l'utente ha lasciato incompleto, tranne che l'utente non fornisce qualche lettera o qualche parola, fornisce un testo e il completamento non è di un o qualche parola, ma è potenzialmente anche di un testo anche lungo un centinaio di parole; quindi in realtà ci stiamo solo illudendo che GPT stia rispondendo alla nostra domanda ma in realtà non è così, sta solo completando un dialogo ipotetico tra un personaggio che sta dicendo quello che state digitando voi e un altro personaggio che risponde.

GPT e LLM vengono addestrati con una quantità di materiale enorme che l'effetto sembra che GPT sappia scrivere.

Nei linguaggi naturali c'è anche la pragmatica (in aggiunta a semantica e sintassi), cioè come le cose vengono davvero utilizzate nelle conversazioni.

Tutti gli LLM (GPT incluso) sono esempi di agenti generativi di AI.