Avvenire.it

AGORÀ

Dibattito. Quella fantasia che l'intelligenza artificiale non potrà avere mai

Raul Gabriel giovedì 1 luglio 2021

L'intelligenza artificiale (IA) immaginata come organismo autogenerante e indipendente è una narrazione molto comune, utile a confondere l'utente medio a fini di mercato e non solo, ma per il momento si adatta solo ai contesti di fantasia. In realtà la AI si istruisce, anzi si compila. Le filosofie alla base del *neuro-linguistic programmino* (NLP) rappresentano differenti scuole di pensiero e riguardano la scrittura dei pattern fondamentali nella costituenda capacità cognitiva della intelligenza artificiale.

Questi sono sostanzialmente quattro: *distributional semantics, frame semantics, model-theoretic semantics, grounded semantics*. Non è un mero esercizio tecnico per specialisti. Ciascuno di questi metodi determina conseguenze fondamentali per la nostra relazione con lo strumento tecnologico e di riflesso incidono sulla nostra vita in modo importante, data la pervasività crescente del nostro rapporto, ormai osmotico, con i supporti digitali e il web.

Senza accorgercene finiamo per assumere strutture linguistico-logiche concepite come pura applicazione matematico-statistica di associazioni in un ambiente privo di significato, quello che regola la IA. Ne derivano due possibili classi di danno: da una parte una standardizzazione inevitabile delle nostre forme espressive e di ragionamento che tendono, per sovraesposizione, alla imitazione delle forme meccaniche dell'intelligenza artificiale. Dall'altra qualcosa di peggio: la attribuzione di significati a un impianto para-logico e operazioni linguistiche che non ne sono prive.

Simulare la logica del linguaggio

Il ***distributional semantics***, come dice il nome, si basa su una ipotesi distribuzionale: le similitudini di significato sono in relazione alle percentuali di distribuzione. La approssimazione semantica viene intesa come approssimazione della distribuzione di un termine, e per estensione di una sua ricorrente riproposizione in determinati contesti. Più ampio il contesto, più ampia la distribuzione, più strutturata la interdipendenza degli ambiti, più concrete le possibilità di avvicinarsi al significato.

Gli approcci NLP sono caratterizzati da una dicotomia ineludibile: intrinseco e estrinseco. Credo risulti abbastanza evidente che il *distributional semantics*, oggi di largo uso nelle intelligenze artificiali che si occupano di linguaggio, rientri in buona parte nella categoria dell'estrinseco. Vale a dire che il significato - potrei definirlo come dimensione qualitativa della risorsa “parola” - non viene derivato da una specificità interna ma da un attributo acquisito nell'uso, il vettore distribuzione, che evidentemente apre a una sconfinata varietà di possibili fraintendimenti.

Negli ultimi vent'anni si sono sviluppati numerosi studi per rifinire l’identificazione di un'area semantica attendibile ottenuta attraverso il *distributional semantics*. Il vettore distribuzione e i contesti sono stati ulteriormente scomposti in sottoaree, sempre più specializzate e perimetrabili. Nonostante questo, mi sembra piuttosto difficile immaginare una ragionevole eliminazione degli errori. Alla IA applicata, referente immaginario, questi interessano relativamente dal momento che l'approccio dell'utente, sempre più superficiale e generalista, viene mediamente soddisfatto in termini funzionali dalla media delle soluzioni proposte.

Tutti sperimentiamo quotidianamente alcune applicazioni del *distributional semantics*: ad esempio nei sistemi predittivi del testo attraverso le funzioni di autocompletamento, quando digitiamo un paio di lettere ed appaiono tutta una serie di possibili scelte tra termini e frasi il cui vettore distribuzionale risulta più probabile. Il processo è puramente statistico e la sua reiterazione può indurci a immaginare associazioni di senso del tutto inesistenti.

Sostanzialmente diverso nel metodo e nei risultati è il ***frame semantics***, la teoria del significato linguistico sviluppata da Charles J. Fillmore. A differenza del precedente si innesta più propriamente sul significato della parola. L’ipotesi è che non si dà significato esaustivo che prescinda dalle sue possibilità interpretative ed esperienziali. Non posso capire realmente la parola “ruota” se non conosco la parola “asse” o “ingranaggio” o ancora “pneumatico”.

Il *frame semantics*, in qualità di filosofia linguistica applicata all’umano espande le possibilità del significato, apre direzioni inaspettate, esplora la complessità costringendoci ad apprezzare una apertura progressiva di enunciati. La cosa cambia in qualità di applicazione al NLP dove la determinazione del significato assume un valore di riconoscimento non qualitativo. Il *frame semantics* per la IA qualifica il significato come scomposizione della complessità in parti più piccole che ne determinano l’identificazione.

Il *frame semantics* è utile per esempio in sistemi come Alexa. La domanda “In quale negozio posso trovare la farina?” viene scomposta in "negozio", "trovare", "farina" e così via. Il risultato è un imbuto sempre più ristretto dalla compresenza e specificità delle varie istanze che portano alla individuazione di una risposta chiamata significato. Fermo restando che per la macchina si tratta di un processo orizzontale, il *frame semantics* offre probabilmente una maggiore probabilità di risposte attendibili e congrue.

Abbiamo quindi il ***model-theoretical semantics***. Nasce dalla ipotesi che tutta la conoscenza umana possa essere codificata e modellata in una serie di regole logiche. Il motore prevalente in questo caso

non è né la distribuzione degli elementi linguistici né il quadro del loro posizionamento, ma una forma di vero e proprio sillogismo deduttivo. Proposizioni che sono in relazione attraverso alcuni termini in comune, determinano l'assunzione consequenziale di ulteriori connessioni tra gli elementi considerati.

Se il cubo è rosso e una forma X è un cubo allora significa che quella forma X è rossa. Questo tipo di logica sembra una chiave di volta. Se riesco a programmare la macchina in modo che sia in grado di procedere con questa logica, allora la macchina sarà in grado di generare via via ulteriori ampliamenti di significato, configurando una sorta di abilità generativo-cognitiva. A dispetto del suo costrutto apparentemente blindato, il *model-theoretical semantics* denuncia debolezze tali da renderlo obsoleto. Le possibili variabili nel modello di sillogismo applicato alla complessità del reale sono talmente tante da inficiarne via via le conclusioni.

Infine abbiamo il ***grounded semantics***, che, come dice il termine, fa del pragmatismo il suo punto forte, specificità dichiaratamente anglosassone. Il fondamento non sono né i modelli né i i contesti semantici, né le deduzioni. Il *grounded semantics* si basa sui fatti. Da essi origina il linguaggio che serve per descriverli, in fondo il principale meccanismo umano alla radice del linguaggio, anche se complicato da infinite variabili, dal momento che l'uomo può concepire l'astrazione mai allineata totalmente all'esperienza fenomenica, a volte addirittura contrapposta al concetto stesso di fatto.

Nella IA questa metodica richiede un approccio inverso rispetto alla dinamica umana, almeno in fase di "istruzione". Prima le si dà la frase, "muovi il cerchio a sinistra" ad esempio. Poi le si mostra il movimento del cerchio a sinistra. Si assume che nel tempo la macchina possa padroneggiare il meccanismo e invertirne la temporalità giungendo a "sperimentare" un fatto-evento per generarne conseguentemente il linguaggio appropriato. Il limite più grande di questo approccio è evidente: il panorama dei "fatti" possibili è pressoché infinito e si combina con tutti gli altri fatti in combinazioni che aprono a complessità inconcepibili, tali da rendere l'applicazione del *grounded semantics* impraticabile.

Intelligenza artificiale rigida, cervello umano flessibile

Il quadro delle principali filosofie di istruzione linguistica applicate alla IA permette alcune considerazioni. Tutte interessano aspetti importanti nella strutturazione della logica linguistica umana. Ognuna di esse è mutuata ovviamente dal vasto bagaglio di strumenti a nostra disposizione quando si tratta di articolare parole, proposizioni, riflessioni. Tutte comunque rappresentano ambiti molto parziali e risultano profondamente inadeguate se prese a se stesse. Inadeguate a maggior ragione quando messe a confronto con il ventaglio delle possibilità del cervello umano, capace di produrre instancabilmente nuove eccezioni a qualunque struttura data.

Non si tratta di un luogo comune, si tratta di acquisizioni recentissime del mondo scientifico.La AGI (*Artificial General Intelligence*) è caratterizzata da una estrema complessità, inscindibile da una intrinseca fragilità. La fragilità deriva da un dato essenziale: la rigidità di ogni sistema derivante da una applicazione delle regole secondo schemi meccanici. Per quanto possa istruire una macchina aggiungendo variabili a variabili al meccanismo e nei singoli componenti che la abitano non potrò mai discostarmi da un binario essenzialmente rigido. Anzi, più la complessità aumenta più se ne incrementa la fragilità.

Il linguaggio, inteso nel senso di una organizzazione sistematica cognitiva, è una vera e propria architettura. Le sue parti determinano reciprocità funzionali vincolanti che, proprio come nella fisica di un edificio, incidono sulla sua stabilità. Carichi, spinte e controspinte, torsioni, legami più o meno superficiali, qualità dei materiali sono caratteristiche essenziali di una vera e propria "fisica del linguaggio" che amplifica esponenzialmente la varietà di chiavi con cui tararne la codifica. E come nell'architettura ogni rigidità dei materiali può portare a una evenienza potenzialmente disastrosa: la rottura.

La strutturazione della AGI ha un fondamento essenziale nel volume dei dati e dei metodi con cui porli in relazione, su cui punta per tentare di riempire il gap con il cervello umano. Qual è il problema? La risposta si trova in una specifica caratteristica del funzionamento delle strutture

neocorticali cerebrali scoperte e studiate da Vernon Mountcastle, neurofisiologo e professore emerito alla John Hopkins University.

Per riassumere e arrivare al punto che mi interessa, la risorsa principale della neocorteccia non è la sua già incredibile articolazione in termini di strutture, il cui volume potrebbe anche essere raggiunto e superato dalle quantità di circuiti che la tecnologia è o sarà in grado di mettere in sequenza. La sua risorsa principale è il fatto di essere una macchina costantemente predittiva. Predittiva in modo esponenziale. Costantemente in movimento, ancor prima che gli eventi (micro e macro) accadano.

A differenza degli schemi irrigiditi dalla progressiva complessità che compongono la trama linguistica della AI, il cervello umano è dotato di strutture che oltre a reagire al presente, elaborano continuamente quantità di modelli evolutivi a brevissimo termine di quel reale. In altre parole il cervello analizza e percepisce la realtà, la elabora, ma allo stesso tempo genera senza sosta e in ogni possibile direzione miriadi di piani B, se vogliamo chiamarli così. Se qualcosa va storto nella prima struttura di valutazione, rimpiazzo e aggiustamento avvengono praticamente senza soluzione di continuità.

A differenza della IA il cervello e le sue strutture cognitive non sono macchine rigide. Si potrebbe dire che la IA è talmente intessuta di meccanica della struttura linguistica da non poterne uscire. Il cervello invece va oltre la meccanica di cui è comunque creatore, e aggiunge alla “analisi” infinità di simulazioni sintetiche attendibili delle sue possibili evoluzioni. Questo rende ogni disallineamento linguistico percettivo e cognitivo un evento trascurabile per il modello costantemente reinventato anticipatamente dal cervello. Se qualcosa varia, il cervello è già pronto con una quantità di adattamenti già elaborati la cui “applicazione” risulta estremamente fluida.

Quando ho letto delle ricerche di Jeff Hawkins in *A thousand brains: a new theory of intelligence* (Basic Books, 2021) ho ricevuto la conferma ciò che sostengo da tempo riguardo la realtà come fatto dinamico, che si riversa inevitabilmente nella forma estetica e del corpo. Non è solo una mia convinzione ovviamente, ma nella pratica ancora oggi una buona parte di esseri umani rifiuta questo dato di fatto, coltivando una idea inesistente di fissità e avvicinandosi così, senza saperlo, ai principi strutturali e organizzativi che regolano la IA e ne decretano il limite invalicabile. In un solo singolo attimo produciamo una tale ricchezza di realtà possibili, che se ne avessimo coscienza rimarremmo sbalorditi. Non è un pensiero romantico: è la consuetudine scientificamente provata della neocorteccia che tutti portiamo dentro la magica scatoletta del cranio.

La struttura cognitiva su cui si basa il cervello è strutturalmente resiliente, grazie alla sua natura biochimica, in un modo sconosciuto a qualunque AI odierna. A fronte di queste considerazioni solo accennate per necessità di spazio, si comprende come le filosofie alla base del NLP, tutte riguardanti aspetti interessanti e parziali del problema, non possono rappresentare una risposta esaustiva nella formazione di un vero e proprio linguaggio perché non toccano il tema vero di una strutturazione linguistica efficace: l’elasticità dei modelli e del loro rigenerarsi. La rigidità delle regole applicate al *deep learning system* rendono fragile il sistema. L'apparato biologico cognitivo del cervello umano presenta invece una facoltà di adattamento estremamente fluido alle variazioni che possono derivare da errori, imprevisti, variabili anche minime capaci di mettere in crisi i più complessi sistemi della AI.

L'intelligenza artificiale insegue, il cervello precede.