Quarta tornata accademica

L'intelligenza artificiale generativa. Parliamone per dominarne le parole

Firenze, Accademia della Crusca
7 luglio 2026[1]

I grandi modelli linguistici capiscono il linguaggio?

Gino Roncaglia (Università Roma Tre)

Nella prima parte, l'intervento ha proposto una sintetica spiegazione dei meccanismi di funzionamento dei sistemi di IA generativa. Partendo dalle aspettative – disattese – dell'intelligenza artificiale logico-simbolica delle origini, e illustrando il passaggio delle reti neurali da strutture logiche deterministiche a sistemi statistico-probabilistici, sono stati discussi i concetti di addestramento di un LLM, di *tokenizzazione*, di *embedding*, di attenzione. Su queste basi, è stato illustrato il funzionamento dei LLM come predittori di *token*, sono stati spiegati i concetti di *bias* e di *allucinazione* (illustrandone l'origine rispettivamente nei *bias* presenti nei corpora di addestramento e nel carattere probabilistico della predizione di *token*), ed è stata presentata l'idea di proprietà emergenti di basso livello: situazioni in cui il testo costruito dai modelli ci "stupisce" per le apparenti capacità di pianificazione nella predizione di *token* e di comprensione del contesto. Nella seconda parte è stata discussa la questione della comprensione del linguaggio, partendo dall'idea di significato come uso e dalla considerazione dei corpora di addestramento come 'precipitati' di usi linguistici. Si è visto come il tema risulti estremamente divisivo per la comunità di ricerca che lavora in questo campo: comunità che sembra non aver raggiunto, almeno finora, un consenso ragionevolmente ampio attorno a una cornice interpretativa comune al cui interno collocare i risultati empirici ottenuti. In particolare, il mondo della ricerca sembra diviso fra interpretazioni 'deflazioniste', che negano l'esistenza di proprietà emergenti 'interessanti' nel comportamento linguistico dei modelli e tendono a riprendere la metafora dei 'pappagalli stocastici' proposta originariamente da Emily Bender e dalle sue collaboratrici, e interpretazioni 'inflazioniste', che tendono a sopravvalutare le capacità dei grandi modelli linguistici e a ritenere assai vicino il traguardo dell'intelligenza artificiale generale, almeno rispetto al tema della comprensione linguistica. La tesi di fondo proposta è che sia ragionevole adottare una posizione intermedia, riconoscendo come l'avvicinamento da parte dei migliori modelli linguistici a livelli di competenza linguistica produttivamente assai vicini a quelli umani renda utile riflettere sulla possibilità di attribuire ad essi una forma di comprensione linguistica generale – almeno se si

[1] L'ordine dei riassunti segue quello in cui si sono effettivamente susseguiti gli interventi.

adottano interpretazioni del concetto di ‘comprensione’ legate alla competenza linguistica manifestata nell’uso – senza ritenere però che tale attribuzione sia già oggi giustificata rispetto ai sistemi attualmente disponibili.

Quanto è naturale l’italiano sintetico prodotto dall’IA generativa?

Anna-Maria De Cesare (Technische Universität Dresden)

L’italiano sintetico – ovvero l’italiano prodotto dall’intelligenza artificiale generativa di testo – può sembrare a prima vista indistinguibile dall’italiano naturale, scritto da mano umana. Le ricerche condotte negli ultimi anni, in particolare incrociando le metodologie proprie della linguistica dei corpora e della linguistica contrastiva, mostrano tuttavia che i testi prodotti dall’IA generativa presentano peculiarità a tutti i livelli linguistici (grafia, ortografia, punteggiatura, morfologia, sintassi, testualità). Questi testi si caratterizzano più in particolare per la presenza di tre tipi di tratti: tratti *esclusivi* (presenti solo nei testi sintetici), tratti *tipici* (più frequenti nei testi sintetici che in quelli naturali) e tratti *assenti* (che si trovano solo nei testi naturali). L’intervento illustra i due primi tipi di tratti soffermandosi sulla categoria delle *impronte algoritmiche dell’inglese*, così chiamate perché i tratti in questione possono essere concepiti come tracce lasciate dalla lingua inglese (sovrappresentata nei dati di addestramento) nei testi generati in italiano; l’aggettivo *algoritmico* fa invece riferimento al processo statistico-probabilistico alla base della generazione dei testi sintetici. La descrizione – qualitativa e in alcuni casi anche quantitativa – delle impronte algoritmiche dell’inglese verte su dati inclusi in corpora (creati ad hoc, composti da testi sintetici e naturali) rappresentativi del genere testuale della biografia breve.

L’intervento si chiude con una serie di interrogativi più generali. Le specificità linguistico-testuali dei testi sintetici scritti in italiano aprono infatti una serie di domande importanti in particolare in ambito sociolinguistico: dato che i testi sintetici si usano oramai anche nei domini dell’ufficialità (pensiamo al loro impiego in ambito educativo), l’italiano sintetico potrà svolgere un ruolo modellizzante e cambiare l’assetto dell’italiano scritto dei prossimi anni? Quali tratti (ma anche quali livelli linguistici) sono più soggetti a cambiamento? Quelli già in espansione nell’italiano neo-standard? Più in generale, l’italiano sintetico generato dai modelli linguistici rappresenta una (nuova) varietà di lingua? E se è il caso, con quale (nuova) dimensione di variazione abbiamo a che fare? Anche a livello empirico si aprono domande cruciali, per esempio sulla natura dei corpora di cui abbiamo bisogno per descrivere in modo appropriato l’italiano scritto dal 2022 (data che coincide con il rilascio della piattaforma ChatGPT): dobbiamo usare (o creare) corpora in cui sono presenti campioni di testi sintetici ‘puri’ o/e ibridi (composti da testi sintetici e naturali)? Quanto peso dovrebbero avere questi campioni e a che genere testuale dovrebbero appartenere?

Oltre il significato letterale. Le abilità pragmatiche dei modelli linguistici generativi

Alessandro Panunzi (Università di Firenze)

L’intervento ha affrontato il tema delle capacità pragmatiche dei grandi modelli linguistici generativi, partendo da una disamina delle caratteristiche specifiche del linguaggio umano a confronto con i meccanismi che regolano gli attuali Large Language Models (LLM).

L’intervento si è strutturato in quattro parti. Nella prima parte è stata proposta una descrizione sintetica del funzionamento dei LLM, intesi come sistemi di predizione del token successivo semanticamente informati su base distribuzionale. Sono stati poi introdotti i principali meccanismi di manipolazione degli output dei modelli pre-addestrati, e in particolare le tecniche di *prompt engineering* e di *fine-tuning* (post-addestramento su dati qualitativamente controllati).

La seconda parte è stata dedicata alla pragmatica e al suo rapporto con l’intelligenza artificiale. Dopo aver ripreso alcuni dei concetti fondamentali della prospettiva griceana sulle dinamiche conversazionali, è stato posto il problema dell’antropomorfizzazione e della mentalizzazione delle IA generative. La principale contraddizione messa in risalto è la seguente: queste tecnologie (o meglio i prodotti commerciali da esse derivati) sono spesso presentate come agenti conversazionali, ma di fatto sono basate unicamente su meccanismi probabilistici di continuazione testuale. In che modo, quindi, un LLM tratta le implicature soggiacenti al discorso? In che senso rappresenta i processi inferenziali legati alla comprensione delle intenzioni comunicative del suo utente?

Nella terza parte sono stati discussi alcuni studi recenti sulla valutazione delle abilità pragmatiche dei LLM in lingua inglese. Sono stati presentati lavori basati su task binari e a risposta multipla, riguardanti la risoluzione di implicature. È stato quindi mostrato che le prestazioni dei modelli variano in modo significativo in funzione del design sperimentale, e che la loro valutazione è spesso limitata a dialoghi stereotipati, molto lontani dalla realtà linguistica concreta.

L’ultima parte dell’intervento ha presentato ricerche recenti sull’italiano, condotte in particolare sul corpus IMPAQTS, una risorsa multimodale di parlato politico annotato per contenuti impliciti discutibili. Il parlato politico costituisce un terreno particolarmente rilevante per lo studio della pragmatica, poiché è ricco di impliciti con funzione persuasiva e potenzialmente manipolatoria. Questi studi si distinguono da quelli presentati nella parte precedente per almeno tre motivi sostanziali che riguardano: i dati di partenza (discorsi reali in contesto ecologico), il task richiesto ai modelli linguistici (generare testo esplicativo), la valutazione dei risultati (valutazione qualitativa condotta da esperti del settore).

I risultati mostrano che le capacità pragmatiche dei LLM non possono essere interpretate semplicemente come una conseguenza automatica dell’aumento di scala dei modelli, ma che dipendono sostanzialmente dalla qualità dei dati (e relativo *fine-tuning* dei modelli), dalle procedure di valutazione e dalla disponibilità di informazione pragmatica derivata da annotazione umana.