Quando si scambia un motore linguistico per altro

Qualche tempo fa spiegavo agli studenti a cosa si va incontro con un esempio pratico usando gli LLM per fare analisi dei dati. Ho preso una serie di dati epidemiologici dal 2010 al 2014 e ho chiesto al modello di costruire un grafico. Il risultato conteneva anche valori del 2020–2021, cioè il periodo in cui la disponibilità di dati esplode per via della pandemia. Non ha “rispettato i dati”; ha seguito il punto di massima densità informativa. Questi sistemi non “leggono” il tuo dataset, si muovono in uno spazio linguistico che è già stratificato intorno alle zone dove il dato è più abbondante, più ripetuto, più recente, più statisticamente conveniente. Io gli chiedo 2010–2014; lui mi porta comunque a 2020–2021 perché lì il terreno è più fertile, più denso, più “sicuro” dal punto di vista della plausibilità. Se metto accanto a questo esperimento quello che via via raccolgo e annoto, la trama è sempre la stessa. C’è chi prova a usare un modello per generare codice o markup ripetitivo.

All’inizio la procedura sembra funzionare, poi, pagina dopo pagina, l’uscita si sfalda, la coerenza si perde, le stesse istruzioni producono varianti divergenti, e il costo di correzione supera quello di fare il lavoro a mano. C’è chi usa

il modello su contenuti specialistici (musica, armonia, analisi di testi complessi) e ottiene risposte formalmente ben confezionate ma concettualmente vuote. Dalle frasi lunghe con lessico corretto, ma regole sbagliate in modo grossolano, come se il sistema imitasse il rumore di fondo della disciplina senza averne mai incontrato la struttura. Il meccanismo è sempre lo stesso, in ambito tecnico succede lo stesso: conversioni numeriche semplici, corrette per le prime richieste, cominciano a deragliare non appena si aumenta leggermente la complessità o la quantità di esempi.

La procedura non si stabilizza, non si "irrobustisce", si ridispone ogni volta come se fosse la prima. Quando si passa al dominio fattuale, la cosa diventa più inquietante: cronologie storiche riscritte con sicurezza, programmi esistiti dichiarati inesistenti o viceversa, riferimenti geografici inventati, dettagli biografici attribuiti a persone reali senza alcuna base; solo chi conosce già l'argomento ha gli strumenti per riconoscere l'invenzione. Chi non sa, prende atto. E integra. Nel dominio medico il pattern è ancora più evidente (alla faccia dei racconti metaforici e appezzotati fatti a botte di epistemia). Un referto viene interpretato con apparente competenza, alcuni dettagli sono spiegati in modo plausibile, poi una sigla viene proiettata in un contesto completamente incompatibile (ostetrico in presenza di un apparato genitale maschile), e il sistema razionalizza l'errore invece di riconoscerlo. Non dice "non lo so", dice "ho capito, è un refuso".

Nella produzione di testi culturali (guida turistica, analisi letteraria, citazione poetica) si vede l'altro lato della stessa cosa, ovvero la capacità di generare un testo perfettamente leggibile, tonalmente adeguato, ritmato nel modo "giusto", ma privo di informazioni. Pagine intere che potresti spostare da un luogo all'altro del mondo senza che cambi nulla. Infine ci sono i casi di delega integrale con strumenti configurati per monitorare notizie, che ripropongono come

“nuovi” articoli vecchi di mesi, oppure ne saltano di rilevanti senza criterio apparente; manager convinti di risparmiare tempo affidando a un modello la ricostruzione di cifre complesse, che si ritrovano con numeri sbagliati di ordini di grandezza, ma esposti con tale sicurezza lessicale da passare il primo vaglio superficiale.

Quello che tiene insieme tutte queste situazioni non è il singolo errore, ma la combinazione di tre elementi: la fluidità del linguaggio, la pressione verso le aree ad alta densità di dato e l’assenza di una rappresentazione del mondo che faccia da vincolo. L’idea bislacca, ma molto diffusa, che “basterebbe cambiare la base di conoscenza” per risolvere il problema è la versione aggiornata della vecchia fede nella fonte giusta. Come se il difetto fosse “cosa ha letto il modello”, e non il modo in cui funziona. Aggiungere più dati, o dati migliori, può ridurre alcuni errori di superficie, ma questi sistemi non operano su un modello del mondo, non possiedono strutture interne che garantiscano coerenza temporale, causale o concettuale. Operano su distribuzioni di probabilità condizionate: massimizzano la plausibilità linguistica locale, frase per frase, token per token. Quando c’è molta informazione su un certo periodo, un certo evento, un certo modo di parlare, il gradiente le spinge lì, anche se tu stai chiedendo altro. Quando una spiegazione è stilisticamente convincente, la produce, anche in assenza di un criterio che la colleghi a qualcosa di vero.

L’EpistemIA nasce esattamente in questo punto di contatto: dove un meccanismo cieco rispetto al mondo incontra un utente che cerca conoscenza, non testo. L’utente vede coerenza grammaticale, tono competente, riferimenti plausibili, e scambia tutto questo per prova. Ma il modello non “sa” se ciò che dice è vero; non ha un luogo interno dove la verità possa essere rappresentata o controllata. La verifica, se avviene, è sempre esterna: siamo noi. E proprio mentre ci affidiamo al sistema per risparmiarci la fatica della verifica, la verifica

stessa scompare dall’orizzonte cognitivo: non viene più concepita come fase necessaria del processo, perché è delegata. Infatti il numero di ricercatori indipendenti, dotti-immaginari e tutto l’insieme di gente che dice menate cresce ogni giorno di più. Per questo i vostri esempi sono così importanti: mostrano che non siamo davanti a una collezione di bug da correggere con l’aggiornamento successivo, ma a una trasformazione strutturale del rapporto tra linguaggio e conoscenza.

Non è un problema di “accuratezza percentuale”, è lo slittamento da un ecosistema informativo basato sulla ricerca e sul confronto di fonti a uno basato sulla simulazione continua di risposte plausibili. Se non teniamo fermo questo punto, ci ritroviamo a discutere di queste tecnologie come se fossero motori di ricerca più evoluti o assistenti infallibili, mentre sono un’altra cosa. Sono interfacce che ricostruiscono il mondo a partire da come ne abbiamo scritto, non da com’è.

Continuare a raccogliere casi non serve a fare l’elenco degli errori, ma a mappare il perimetro di questa nuova condizione: un ambiente in cui la facilità di produzione di testo può dare l’impressione che la conoscenza sia a portata di chat, mentre in realtà si è solo spostato, e reso invisibile, il lavoro più importante: quello della verifica.