STEFANO BARTEZZAGHI

CHATGPT. NON È DETTO CHE SIA VERO MA È VERO CHE LO SI È DETTO.

ERRARE HUMANUM EST

Il modello di intelligenza artificiale più famoso e più inquietante è stato molto a lungo costituito da Hal 9000 di *2001 Odissea nello spazio* di Stanley Kubrick. Questo computer, reputato infallibile, infallibile non è: gli incidenti che succedono sull'astronave sono dovuti a suoi errori che lui cerca di dissimulare o sminuire, venendo infine smascherato.

Il modello di intelligenza artificiale più famoso e più inquietante che oggi abbiamo (non più nel mondo immaginario ma in quello reale) è ChatGPT. Una studentessa inglese che l'aveva impiegato è stata smascherata perché aveva consegnato un paper sintatticamente perfetto, mentre il suo curriculum evidenzia che lei non è in grado di scrivere in inglese senza commettere errori di grammatica. Nella realtà le cose vanno dunque alla rovescia rispetto a come se le era immaginate Kubrick. Il problema reale con le macchine non è più dato dai loro errori, ma dal fatto di non farne. Il problema si è posto anche nella sperimentazione della guida totalmente automatica (di quelle che chiamavamo "automobili" quando non si immaginava che avrebbero potuto ambire a un'autonomia totale). È difficile fare in modo che il sistema preveda che i comportamenti degli altri automobilisti, quelli "umani", possano essere scorretti e irrazionali.

Roger Caillois diceva che il destino degli esseri umani è stato quello di mettere un po' di gioco nell'immenso ingranaggio che è l'Universo. L'Universo si comporta senza intenzione, produce quel che produce, gli ecosistemi si stabilizzano e si evolvono secondo le logiche inesorabili della materia. L'essere umano può introdurre il brutto, l'errore, l'insensato. Ma se alla fine l'essere umano diventasse capace di produrre grandi, se non immensi, ingranaggi che gioco non hanno? La specie umana come consumatrice dello spazio di manovra che lei stessa aveva introdotto.

ATTRIBUZIONE

Un problema di ChatGPT a suo modo affascinante, ma affascinante nel senso del perturbante, è che produce testi "lisci". Non offrono appigli per comprendere se la loro fonte diretta sia umana o no. Sono stati approntati sistemi di controllo per stabilirlo ma hanno margini d'errore enormi e quindi non sono attendibili. Miglioreranno, ma a quel punto è probabile che il software generativo sia a sua volta migliorato e che dunque si inneschi la stessa corsa paradossale tra Achille e tartaruga che avviene in materia di doping e antipoding sportivo. Su quest'ultimo

problema c'è chi dice: legalizziamo il doping. Per il corpo degli atleti è molto più pericoloso che la ricerca e la somministrazione si svolga in modo clandestino. Per ChatGPT verrà certamente prospettata una soluzione del genere e si riterrà che non ci sia nulla di male a impiegare i testi che produce. Ho già sentito in università e in editoria qualcuno dire: "io lo uso, è comodissimo, si fa molto prima. Che male c'è?".

Già ma si fa prima a fare cosa? E nel tempo guadagnato cosa si farà mai?

AUTORIALITÀ PRESUNTA

Un'altra questione interessante che ho discusso con una collega è che ChatGPT non plagia, non copia, non cita neppure. Rielabora, cioè fa quello che nel corso della nostra formazione ci hanno sempre invitato a fare tutti i migliori maestri che abbiamo avuto. Se vuoi essere sicuro di aver appreso qualcosa che hai studiato da un libro, chiudilo e ripeti con parole tue. I paper universitari sono prove scritte in cui gli studenti dimostrano di aver appreso qualcosa. Non si fanno in aula, li si prepara a casa, potendo consultare i libri. Tanto se sono copiati chi li corregge se ne accorge. ChatGPT non copia ma propone la stessa forma di rielaborazione, ed *è proprio per quello* che dà l'impressione "di aver capito qualcosa". La rielaborazione causa di per sé la stessa illusione antropomorfica che possono dare gli avatar di certe comunicazioni virtuali e immersive, figure che sembrano umane e invece sono puramente pixel. L'analogia credo che regga: la mia immagine in uno schermo è fatta di pixel come l'immagine di un avatar virtuale, solo che quella è proprio soltanto una composizione di pixel mentre la mia è la riproduzione in pixel di un corpo di carne, ossa, sangue.

Io sopporto senza sforzarmi il fatto che la mia immagine sia fatta di pixel come un'immagine costruita dal computer. Ci tengo invece a ribadire la differenza di quel che sta, per così dire, *sotto* l'immagine (l'"oggetto dinamico"di Ch. S. Peirce): cioè al fatto che *sotto* la mia fotografia ci sia qualcosa a cui faccio la barba la mattina.

PIÙ UMANO PIÙ VERO

Negli anni Novanta ho notato che nei loro giochi e discorsi i bambini avevano un concetto che nei miei giochi infantili non ricordavo di aver avuto: quello di "umani". Pensandoci, mi sono fatto l'idea che avevano sviluppato quella nozione perché i loro racconti si erano popolati di esseri animati di diverso tipo, robotico-androide o fantastico-mutante, da Ufo Robot a Pikachu.

In semiotica tendiamo a vedere ogni cosa attraverso quella a cui si oppone. ChatGPT è qualcosa di nuovo, che crea una nuova opposizione e quindi un nuovo concetto: quello di testo linguistico generato da esseri umani. Prima questo concetto non esisteva perché *tutti* i testi

linguistici non potevano che essere generati da esseri umani (e i fedeli delle religioni del Libro mi perdonereanno il presupposto ateo di questa affermazione).

Tra le altre cose ChatGPT è quindi anche l'occasione per capire cosa abbiamo inteso sinora per "testo". Il discrimine fra le sue produzioni e le nostre non è certo la correttezza sintattica, perché vediamo che già queste versioni che presto ci sembreranno antiquate possono produrre testi più corretti di quelli di un'universitaria inglese. Allora è una questione semantica? Anche qui ChatGPT fa errori ma gli esperti ci dicono che ne farà sempre meno. E anche gli errori che fa sono errori *plausibili* , parola che se interpretata letteralmente dovrebbe significare "che può essere approvato". Quindi il sistema fornisce testi corretti sintatticamente e plausibili semanticamente. Ciò soddisfa la nostra nozione di testo?

VOLER DIRE

In un testo noi non cerchiamo la correttezza sintattica e la plausibilità semantica. Se ci sono, bene; se non ci sono possiamo anche farcene una ragione. Noi comprendiamo le frasi sgrammaticate, leggiamo poesia, apprezziamo James Joyce anche se negli ultimi decenni si denota una certa predilezione maggioritaria per testi meno problematici, diciamo meno implausibili. Ai testi noi chiediamo di esprimere un'autorialità intenzionata, nei testi cerchiamo cioè un "voler dire", che non a caso usiamo anche come sinonimo di "significare".

Qual è il "voler dire" di un testo composto per via statistica? Nella massima efficienza che oggi possiamo immaginare per sistemi come ChatGPT noi avremo testi ottenuti da una macchina che a richiesta sa elaborare e assemblare sintagmi corretti per sintassi e plausibili per semantica. La dimensione pragmatica risulta assente, perché non c'è atto linguistico. L'enunciazione corrisponde alla risposta della macchina al comando di generare il testo *oppure* all'appropriazione di chi eventualmente lo firmerà. Se il testo prodotto dalla macchina è X e io lo firmo noi diremo che "Stefano Bartezzaghi ha detto X" (anche se non l'ho composto io ma me ne sono appropriato: la mia enunciazione non è neppure la citazione di una fonte che lascio occulta); se lo presento come un testo di ChatGPT l'enunciazione è puramente basata sulle fonti di cui è alimentato il database. Questa è una specie di serbatoio di senso comune. Ma non un senso comune a cui possiamo fare riferimento con il "Si dice" che usiamo per le chiacchiere correnti, perché la rielaborazione di ChatGPT è ogni volta diversa (anche se non la possiamo dire "personale").

Quindi l'enunciazione del testo è qualcosa del tipo "Una macchina chiamata ChatGPT dice che si dice X"; o: "... dice che è plausibile dire X".

STATI FINITI E CADAVERI SQUISITI

Non so se il funzionamento di tale sistema possa essere avvicinato a quelle "macchine a stati finiti" di cui parlava la teoria dell'informazione negli anni Cinquanta o Sessanta del Novecento. Macchine che elaboravano segmenti di output uno dopo l'altro, ho ricordi vaghi. Ma andando più indietro troviamo il caso dei "cadaveri squisiti" del primo Surrealismo. Breton e i suoi amici cercavano modi "automatici", quindi non consapevoli (non autoriali, diremo ora), di elaborare i testi. Volevano ottenere testi senza invenzione. Inventarono un gioco. Su un biglietto ripiegato il primo scriveva un sostantivo, il secondo un aggettivo, il terzo un verbo transitivo, il quarto un sostantivo... Ogni riga veniva coperta a mano a mano e chi scriveva non sapeva cosa avessero scritto gli altri. Il gioco prende il nome dalla prima frase risultante, che fu: "Il cadavere squisito berrà vino novello". Frase che a me è sempre parsa strana, perché fra il verbo "berrà" e il complemento oggetto "vino" e tra il sostantivo "vino" e l'aggettivo "novello" corrono relazioni di pertinenza (e plausibilità) un po' troppo strette per essere generate a caso. Ma chissà come era davvero andata.

Per come lo prevedono le sue regole, e per dirla in termini saussuriani, in un *Cadavere squisit*o noi abbiamo un'intenzione sull'asse della selezione (ogni autore infatti sceglie che parola usare) ma nessuna intenzione sull'asse della combinazione (nessuno conosce la scelta di nessun altro). Da questa graziosa macchinetta uscivano messaggi bizzarri e i surrealisti si sbizzarrivano a interpretarne la poeticità. Era una macchinetta per produrre testi implausibili.

IL VERSIFICATORE

Nel 1960, per un radiodramma e un racconto Primo Levi immaginò una macchina chiamata "Il versificatore", capace di comporre poesie a richiesta, con la possibilità di agire su certi comandi simili ai registri dell'organo musicale per impostare argomento, genere, stile, prosodia del componimento.

Una prova in registro lirico-filosofico, terza rima, endecasillabi, secolo VII, sull'argomento "Limiti dell'ingegno umano" dà:

Cerèbro folle, a che pur l'arco tendi?
A che pur, nel travaglio onde se' macro
Consumi l'ora, e dì e notte intendi?
Mentì, mentì chi ti descrisse sacro
Il disio di seguire conoscenza,
E miele delicato il suo succo acro.

Quindi dalla macchina surrealista (e realmente funzionante) per produrre implausibilità siamo passati a una macchina immaginaria per produrre plausibilità.

Paragonato al Versificatore, Chat GPT sembra un prosificatore installato nel mondo reale e non più nel futuro immaginato da Levi. O se vogliamo dire diversamente il futuro immaginato da Levi si è realizzato, come altre sue profezie tecnologiche o no (e speriamo non si realizzino tutte).

LECTOR EXTRA FABULA

Ora però mettiamoci dalla parte di chi legge. Cosa chiediamo a un testo? Ci basta che sia corretto sintatticamente e plausibile semanticamente? E detto fuori dai denti: chi leggerà la roba che verrà prodotta da questo sistema? A che scopo la leggerà?

Quando io leggo i versi di "Limiti dell'ingegno umano" mi divertono e incuriosiscono perché so che li ha scritti l'autore del racconto, apprezzo che si riferiscano proprio all'ingegno umano. Ma se li leggo pensando che li ha composti il versificatore che senso attribuirò loro? Mi interesseranno solo come esempi di linguaggio di origine non umana, ma al loro senso mancherà sempre l'intenzione enunciativa.

Il che potrebbe non essere male per testi che vogliono essere informativi.

ChatGPT diventerà allora una specie di Wikipedia generativa? Non so molto della Rivoluzione francese, dell'allevamento dei tacchini o della meteorologia thailandese e ho un sistema che mette assieme qualcosa che si può plausibilmente dire di questi argomenti. Non è detto che sia vero ma è vero che lo si è detto.

È insomma il trionfo dello stereotipo e io non ho proprio nulla contro gli stereotipi, almeno sino a che siano riconoscibili come tali. Temo solo che la macchina finisca per farci compiere l'ultimo passo che manca per congiungere apparenza di verità e verità, cioè per rendere vero il plausibile quando più e quanto più è condiviso.

ESPRESSIONE E CONTENUTO

L'importante sarebbe sapere che questo non è affatto il modo in cui funziona il linguaggio umano, che solo apparentemente è un montaggio di una parola dietro l'altra. Ogni elemento linguistico, diciamo pure ogni parola, ha infatti un senso attributivo e un senso direzionale. Cioè è un'espressione in relazione a un contenuto (senso attributivo, di pertinenza della semantica) ma dà anche indicazioni di pertinenza della sintassi su come legarla al contesto, le parole che precedono e quelle che la seguono. Le concordanze per numero e genere, per esempio.

Senza un rapporto tra espressione e contenuto non c'è lingua. Non c'è discorso senza una costruzione in cui i due piani sono in relazione. Lo pseudodiscorso di ChatGPM è costruito come un cadavere squisito, salvo il fatto che la selezione dei segmenti del sintagma non segue criteri da poesia surrealista ma da algoritmo statistico e quindi è volta alla plausibilità anziché all'implausibilità. Ma ottiene la plausibilità semantica soltanto attraverso il montaggio statisticamente sorvegliato dell'espressione.

CONTENUTO

Ora pensiamo a come usiamo oggi in italiano la parola "contenuto". Tutta la fatica della semiotica per dare rigore al lessico con cui si parla di comunicazione sembra essere andata sprecata. Nella lingua comune ma anche nei gerghi tecnici della comunicazione per contenuto si intende infatti non il piano semantico in relazione con l'espressione, bensì testo veicolato da un supporto. Si dice: "il contenuto" di un'email. Si dice che la piattaforma ha "contenuti" interessanti, per esempio un buon catalogo di film.

In questo slittamento semantico il contenuto non è più quel che viene significato, come si fa in semiotica dopo Louis Hjelmslev e Roland Barthes (senso numero uno), ma in quel che viene recapitato (senso numero due). Si tratta di un modello comunicativo che sul senso attributivo privilegia quello direzionale e che al centro del processo comunicativo, e del relativo interesse imprenditoriale, non mette il messaggio, bensì il contatto.

Il senso attributivo di questi tipi di comunicazione coincide perfettamente con l'idea di testo che esce da Chat GPT. Come destinazione deve essere plausibile; come origine non può che essere stato già detto. Ancora una volta: non è detto che sia vero ma è vero che è stato detto.

MODULI DA RIEMPIRE

Quindi ChatGPT produce contenuti (senso numero due) che non hanno contenuto (senso numero uno), se non come contenuto plausibile, non intenzionato, senza enunciazione. Non è l'impressione che ci viene comunicata anche da altre cose chiamate "contenuti" sulla Rete o no? I bot, i conduttori virtuali, gli articoli non firmati di certe testate online... Correttezza sintattica, plausibilità semantica, sostanziale carenza di senso.

Sono contenuti, questi, che riempiono spazi modulari, caselline da annerire, blocchi di materia riversata in uno stampo. La loro vera caratteristica è l'*indifferenza*: non intendono fare differenza.

Ennio Flaiano aveva compilato un "Frasario per passare inosservati in società"; questi sono testi che vogliono passare inosservati. C'era anche stata una collana editoriale inglese che si intitolava "Bluff your way". Erano piccoli manualetti che insegnavano cose da sapere per sembrare esperti di qualcosa. Avevo acquistato "Bluff your way in chess". Non ho mai capito in cosa si differenziassero dai normali manualetti e bignamini che danno infarinature. Che differenza passa tra avere una conoscenza parziale e fingere di averla ?

SENSO

Prima di preoccuparci tanto potremmo esprimere una speranza: la speranza che questi sistemi generativi di testi indifferenti ci insegnino qualcosa a proposito del senso. Dato che *per noi* un testo ha senso, dato che *noi* quando produciamo un testo non ci basiamo su un algoritmo statistico, dato che *noi* conosciamo le risorse espressive dell'agrammaticalità e dell'ambiguità semantica, allora dobbiamo concepire la comunicazione non come un vuoto da riempire, non come un "contenuto" imballato da recapitare a un destinatario, ma come la costruzione di un testo. Il testo è un tessuto che intreccia espressione e contenuto. La sua costruzione è un processo che include le posizioni di chi lo enuncia e di chi lo legge. La differenza sta qui e quando c'è l'intreccio e ci sono le istanze dell'enunciazione allora ci sarà il testo.

Non sarà detto che sia vero ma sarà vero che non sarà mai stato detto.

NONSENSO

La preoccupazione è che la hýbris tecnologica umana arrivi al punto di vantarsi avere tolto gioco al grande se non immenso meccanismo cailloisiano. Vogliamo inventare una macchina che sappia parlare il nostro linguaggio. Ci accorgiamo che il nostro linguaggio è troppo complesso e irriducibile all'automazione. Non potendo complicare la macchina allora semplifichiamo il linguaggio. Per poterci gloriare di aver inventato la macchina che sa parlare, limitiamo il linguaggio verbale al grammaticale e al plausibile.

Lo facciamo già quando ci rivolgiamo agli assistenti virtuali e magari chiamiamo "interpretare" l'operazione con cui tali dispositivi reagiscono ai nostri comandi vocali. Ecco. Se cominciamo a confondere il riconoscimento con l'interpretazione saremo già a metà dell'opera, ben incamminati sulla strada di una demenza programmatica e volontaria.

(Questo testo è la rielaborazione di un intervento tenuto alla tavola rotonda "ChatGPT: promesse e illusioni", partecipanti Stefano Bartezzaghi, Elena Esposito, Roberto Navigli e Daniela Tafani; moderatore Guido Boella; Circolo dei Lettori, Torino, 17 aprile 2023).