ChatGPT – Io e Chat, lo studente zuccone

written by Paolo Musso | 24 Agosto 2023

Come avevo già detto nel mio ultimo intervento (https://www.fondazionehume.it/societa/chatgpt-gli-impostorati-autorevoli-e-la-superluna/), quando qualche giorno fa ho letto l'articolo di Luca Ricolfi (https://www.fondazionehume.it/societa/chatgpt-limpostore-autorevole/) su un esperimento condotto da lui e alcuni altri docenti universitari suoi amici con ChatGPT (che anch'io chiamerò semplicemente Chat, come ha fatto lui) ho stentato a credere ai miei occhi.

Da tempo avevo intenzione di fare anch'io un piccolo esperimento con Chat ed ero sicuro che l'esito sarebbe stato abbastanza deludente, ma pensavo che per metterlo in crisi si dovessero fare domande costruite apposta a tale scopo, il che richiedeva di dedicarci un po' di tempo e di attenzione (e per questo finivo sempre col rimandare, avendo cose più urgenti).

Mai avrei creduto che Chat potesse andare in crisi da solo di fronte a domande semplicissime, come è emerso dal lavoro di Ricolfi & C.

Confesso che, per la prima volta da quando conosco Luca, prima di diffondere il link dell'articolo tra i miei amici e conoscenti ho voluto fare anch'io una verifica. Non perché pensassi che quello che aveva scritto fosse falso, ovviamente, ma perché volevo capire se si era trattato solo di un fatto episodico o se invece fosse qualcosa di strutturale e, se sì, da che cosa poteva dipendere. Dopo due sessioni con Chat sono giunto alla conclusione che le sue prestazioni sono, se possibile, ancora più scadenti di quelle descritte da Ricolfi: più che un impostore autorevole, infatti, sembra uno scolaro un po' zuccone che cerca di fare il furbo.

Metodo

Premetto che, come Ricolfi, ho usato la versione di ChatGPT 3.5, perché è gratuita, mentre la versione Plus, la più aggiornata, costa 20 dollari e francamente mi scocciava dare dei soldi a gente che sta facendo un'operazione che mi sembra almeno in parte cialtronesca e per alcuni aspetti anche dannosa. Comunque alcuni colleghi di Ricolfi che hanno provato l'ultima versione dicono che non ci sono grosse differenze e in ogni caso la versione che ha scatenato la prima ondata di eccitazione planetaria è la 3.5.

Ciò chiarito, passiamo all'esperimento, che è diviso in due parti: nella prima, più "nozionistica", ho cercato di valutare la capacità di Chat di raccogliere correttamente informazioni, avendo cura di scegliere argomenti presenti sul Web, ma non troppo noti, però ben conosciuti da me; nella seconda ho invece cercato di mettere alla prova la sua creatività, chiedendogli di produrre lui stesso un breve testo letterario e poi di commentarne alcuni non presenti su Internet, in modo che non potesse sfruttare i commenti di altri esseri umani, ma fosse costretto a contare solo sulle proprie capacità.

Ed ecco come è andata.

Sperimentazione "nozionistica"

1) Anzitutto ho provato a chiedere a Chat se sapeva chi era Luca Ricolfi e quali erano le sue pubblicazioni più importanti, come aveva fatto lui. Chat-in-italiano ha risposto di no, ma Chat-in-inglese ha risposto correttamente alla prima domanda, mentre alla seconda ha risposto con una lista di 5 libri… tutti inventati di sana pianta: proprio come era successo a Ricolfi & C.

Ho inserito uno per uno i 5 titoli in Google per cercare di capire da dove diavolo li avesse tirati fuori, ma non ho trovato nulla, tranne che per l'ultimo, un fantomatico *Populismo 2.0*, che esiste realmente, ma non è di Ricolfi, bensì di Marco Revelli.

Qui però ho trovato anche qualcos'altro: un articolo (*I demoni nazisti, la democrazia in crisi e il populismo 2.0*, su *La Stampa* del 5 gennaio 2017) in cui, parlando di vari libri, si citava Ricolfi come autore di *Sinistra e popolo*, dopodiché l'articolo continuava parlando di Marco Revelli che aveva scritto appunto *Populismo 2.0*. Ciò che è accaduto mi sembra chiaro (anche perché corrisponde al difetto più frequente dei traduttori automatici, nonché del correttore automatico di Word, che funzionano tutti in base a principi molto simili): Chat non ha capito che il soggetto da una frase all'altra era cambiato e ha creduto che fosse ancora Ricolfi, attribuendogli così la paternità di *Populismo 2.0*.

È probabile che meccanismi simili siano alla base anche della creazione di altre false attribuzioni, ma, come subito vedremo, non si tratta solo di questo.

2) A questo punto ho accusato Chat di aver mentito sulle pubblicazioni e gli ho posto di nuovo le stesse domande in italiano. Stavolta ha risposto correttamente alla domanda su chi era Ricolfi (semplicemente traducendola dall'inglese), poi ha premesso: «Mi dispiace, ma fino alla mia ultima aggiornamento [sic] delle conoscenze a settembre 2021, non ho accesso a un elenco specifico dei libri più importanti di Luca Ricolfi. Tuttavia, posso darti una panoramica generale delle aree in cui è noto e potrebbe aver scritto libri» (si noti l'errore di grammatica, che è grave perché si tratta chiaramente di una risposta preimpostata dai suoi programmatori e non creata lì per lì).

Dopodiché mi ha proposto altri 5 titoli, anch'essi tutti inventati, ma aggiungendo la seguente avvertenza: «Tieni

presente che questi titoli sono esempi ipotetici basati sulle aree di interesse associate al lavoro di Luca Ricolfi. Per ottenere un elenco specifico e accurato dei suoi libri più importanti, ti consiglio di verificare librerie affidabili, basi di dati accademiche o il suo sito web ufficiale, se disponibile».

Evidentemente il mio rimprovero ha fatto scattare un meccanismo cautelativo (i creatori di Chat sono *molto* prudenti e hanno inserito *molti* meccanismi del genere), che però ha dimostrato definitivamente ciò che sosteneva Ricolfi: Chat è costruito in modo tale che, se non sa, inventa, però non te lo dice (almeno finché non gli tiri le orecchie).

3) Ho quindi posto a Chat-in-inglese le stesse domande a proposito di un altro autore famoso ma non famosissimo, che io però conosco molto bene: il mio maestro Evandro Agazzi. Anche qui Chat ha risposto correttamente, benché in modo un po' generico, alla domanda su chi fosse, ma quanto alla bibliografia (per rendere più facile la verifica avevo chiesto di indicare solo i libri) mi ha fornito 6 titoli, dei quali però solo uno era un libro di Agazzi: un altro era un articolo, due erano inventati, uno era un libro curato da Agazzi ma non scritto da lui e uno era il libro degli atti di un congresso in cui c'era, fra gli altri, un saggio di Agazzi.

È interessante notare che anche qui Chat ha premesso l'avvertenza di cui sopra, benché stavolta si fosse "sforzato" di trovare i titoli autentici, anche se non gli è riuscito molto bene: i rimproveri sono la cosa che sembra memorizzare meglio in assoluto.

(Giusto per la cronaca, ho anche cercato me stesso e le mie pubblicazioni, ma niente da fare: Chat non mi conosce, né in italiano né in inglese. Non ho ancora deciso se esserne offeso o lusingato…)

4) Ho poi chiesto notizie sul programma SETI, al cui riguardo

all’inizio Chat ha detto di non avere notizie perché la sua conoscenza si ferma al 2021. Quando l’ho sgridato dicendogli che il SETI esiste dal 1960 ed è stato iniziato da Frank Drake, improvvisamente Chat “si è accorto” che invece qualcosa sapeva. Quando gli ho chiesto una bibliografia sintetica, però, mi ha fornito un elenco di 7 titoli, abbastanza corretto (un titolo era giusto solo a metà, un libro esisteva ma l’autore non era Frank Drake), ma insoddisfacente, perché solo un paio di testi erano effettivamente significativi e comunque mancavano i più importanti.

Dopo i miei ulteriori rimproveri, Chat mi ha fornito un altro elenco di 7 titoli, completamente diverso, ma non migliore, anzi, nettamente più scadente: anche qui, infatti, solo due titoli erano importanti, uno era inventato, tre non parlavano del SETI e uno era di scarso interesse, mentre continuavano a mancare testi fondamentali, come l’articolo seminale di Giuseppe Cocconi e Philip Morrison su *Nature* nel 1959 che ne ha proposto per la prima volta i concetti fondamentali, il *Project Cyclops*, pubblicato dalla NASA nel 1972 e considerato ancor oggi la “Bibbia” del SETI, e il libro di Drake e Sagan sul celeberrimo disco d’oro con le immagini e i suoni della Terra caricato sulle due sonde Voyager nel 1977.

5) Quindi ho chiesto qual era l’opera più importante di alcuni autori famosissimi, scelti tra quelli per i quali non possono esistere dubbi al proposito. Eppure, sui 6 autori da me sottopostigli Chat ha risposto correttamente solo per 3: Martin Heidegger (*Essere e tempo*), San Tommaso d’Aquino (*Summa Theologiae*) e Thomas Kuhn (*La struttura delle rivoluzioni scientifiche*). Ha invece dato una risposta quantomeno ambigua su Albert Einstein, per il quale ha indicato a pari merito *Sull’elettrodinamica dei corpi in movimento*, in cui espose la teoria della relatività ristretta, e *Fondamenti della teoria della relatività generale*, certo anch’esso importantissimo, ma è nel primo che è stata fatta la vera rivoluzione e comunque io avevo chiesto di sceglierne uno solo.

Ma soprattutto Chat ha clamorosamente “toppato” con Karl Popper (per cui ha indicato *La società aperta e i suoi nemici*, opera certo importantissima, ma che non può essere preferita alla *Logica della scoperta scientifica*) e più ancora con Cartesio, per il quale non solo ha indicato le *Meditazioni metafisiche* anziché il celeberrimo *Discorso sul metodo*, atto di nascita della filosofia moderna, ma (cosa veramente imperdonabile) ha pure sostenuto che è in esse che Cartesio «introduce il famoso “Cogito, ergo sum”», che invece è già presente nel *Discorso*, pubblicato 4 anni prima.

6) Infine, ho chiesto anch’io, come aveva fatto qualche tempo fa l’economista David Smerdon, quale sia l’articolo di economia con più citazioni al mondo, domanda a cui Chat aveva risposto con un titolo inventato. Evidentemente ammaestrato dall’esperienza, stavolta è stato più prudente, rispondendo che «potrebbe essere difficile determinarlo con precisione», ma poi ha suggerito in via ipotetica due titoli, uno solo dei quali è corretto: quando si dice che il lupo perde il pelo, ma non il vizio…

Quanto all’articolo scientifico più citato in assoluto, Chat ha indicato *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, pubblicato nel 1998 da Sergey Brin e Lawrence Page, i fondatori di Google, dedicato a spiegare come funziona il loro algoritmo di ricerca. Non sono in grado di dire se la risposta sia corretta, perché bisognerebbe verificare com’era la situazione nel 2021, anno a cui si ferma la conoscenza di Chat 3.5, ma la cosa, visto il tema, è quantomeno plausibile e, soprattutto, almeno l’articolo esiste davvero… e Chat sembrerebbe avere un gran bisogno di leggerlo!

Già, come è possibile che Chat non riesca a riportare correttamente bibliografie che con Google si trovano in 30 secondi? È una buona domanda, ma ne parleremo più avanti. Per ora continuiamo con i risultati.

Sperimentazione “creativa”

Terminata la parte “nozionistica”, sono passato a quella “creativa”, dove paradossalmente Chat se l’è cavata un po’ meglio (tenuto conto della maggiore difficoltà), ma solo un po’ e inoltre, come spiegherò più avanti, in realtà è solo apparenza.

7) Anzitutto, gli ho chiesto di scrivere un “drabble”, cioè un racconto di fantascienza in 100 parole con finale a sorpresa. Chat ha scritto la storia di un astronauta che nel 2085 trova su Marte un cristallo che lo fa viaggiare nel tempo in varie epoche, finché si ritrova di nuovo su Marte, concludendo così: «Il finale sorpresa? In mano aveva due cristalli». A parte l’ingenuità di scrivere esplicitamente «finale sorpresa» (così, senza la “a”) nel testo del racconto, non si capiva perché mai l’astronauta avesse due cristalli.

Quando gliel’ho chiesto, Chat ha modificato il finale come segue: «Realizzò che il primo cristallo era per il viaggio nel tempo, mentre il secondo cristallo, che aveva trovato senza rendersene conto, lo riportava a Marte», il che evidentemente non è una spiegazione. O meglio, lo è, ma non nel senso che volevo io (e che vorrebbe qualunque lettore minimamente intelligente).

8) Quindi ho chiesto a Chat di commentare alcune brevi opere letterarie. Per essere assolutamente certo che dovesse cavarsela da solo, senza poter contare su commenti di autori umani, gli ho sottoposto 4 drabble di fantascienza scritti da me di cui so che non esistono commenti online e alcune poesie che avevo pubblicato molti anni fa esclusivamente in formato cartaceo (in caso a qualcuno interessasse, il libro è *Le mezzanotti*, Sabatelli 1995, ma credo sia ormai introvabile).

Qui i risultati sono stati davvero interessanti. Chat infatti se l’è cavata abbastanza bene (almeno in apparenza) con il commento generale, ma ha sbagliato ripetutamente e spesso

gravemente nel comprendere il significato di dettaglio.

Per esempio, di una poesia ha scritto giustamente che «sembra un omaggio profondo e intimo a Eugenio Montale», azzeccando perfino il titolo, che non avevo menzionato e che era appunto *Omaggio*. Solo che, subito dopo questo inizio così promettente, ha rovinato tutto, prima attribuendo a Montale delle vicende che invece erano chiaramente mie e poi non accorgendosi che la seconda parte della poesia era dedicata a Mario Luzi, benché fosse esplicitamente nominato (di nuovo la difficoltà cronica a cogliere i cambi di soggetto).

In un'altra, dove parlavo dei gesti di una mia amica che erano «intessuti di un soffice sorriso», Chat ha commentato che «l'uso di parole come "intessuti" e "soffice sorriso" crea un senso di tessitura e calore emotivo», dove la seconda che hai detto, per quanto molto generica, può ancora andar bene, ma la prima proprio no.

Un'altra cosa molto strana è che Chat ha scritto che la poesia gli pareva in stile montaliano, ma «senza ulteriori informazioni» non lo poteva «confermare con certezza». Quando gli ho chiesto quali passi si riferivano a Montale si è «scusato per l'errore» di attribuzione. Io gli ho fatto notare che non aveva fatto nessun errore, tanto che c'erano addirittura diverse citazioni letterali di versi di Montale e allora (e *solo* allora), come già era successo col SETI, improvvisamente Chat "si è accorto" che in effetti le conosceva e ne ha subito identificata una correttamente.

Ma il peggio si è avuto con i drabble. È vero che sono racconti molto sintetici e quindi difficili da interpretare, ma gli svarioni sono stati colossali. Anzitutto, è evidente (perché lo evidenzia lui stesso, facendo un elenco di argomenti, tipo «Conseguenza delle Azioni», «Scoperta Personale», «Colpo di Scena Finale», ecc.) che Chat nel commentare un racconto segue degli schemi preimpostati che gli chiedono di individuare delle parti prestabilite, come si fa

alla prima lezione di un qualsiasi corso di scrittura creativa per aspiranti scrittori (che non per averli imparati a memoria diventeranno mai veri scrittori, proprio come Chat). Peccato solo che alla *vera* critica letteraria si chieda in genere qualcosina di più… Ma soprattutto Chat ha frainteso completamente 2 racconti su 4 e degli altri ha capito solo l'inizio, perdendosi poi per strada e fraintendendone la conclusione, in un caso completamente e nell'altro in gran parte.

9) Ho poi provato a chiedere a Chat di identificare a quale celebre autore potevano essere accostate alcune mie poesie (diverse da quelle commentate in precedenza) in base allo stile. Anche qui è andata veramente male: è vero che si tratta di un tema almeno in parte opinabile, ma non al punto da giustificare qualsiasi errore. Su 8 poesie esaminate Chat è riuscito ad associare l'autore giusto solo a una in chiaro stile dantesco e a un'altra in altrettanto chiaro stile montaliano. Sulle altre 6 ha sbagliato di brutto: prima attribuendomi vicinanze ad autori che apprezzo, ma che proprio non c'entrano con me, come Ungaretti e Quasimodo; poi accostando una poesia carnale e sanguigna come poche, situabile tra Rebora e Testori, a un autore come Pascoli che sta ai loro antipodi; e infine tirando fuori dal cappello un surreale "Cesare Pavese" per una poesia in cui, oltre allo stile inequivocabilmente montaliano, c'erano addirittura delle esplicite citazioni da *Mediterraneo*di *Ossi di seppia*.

10) Come considerazione generale, va notato che in tutte le sue risposte Chat è sempre molto cauto, spesso fin troppo. Alcune precisazioni, come quelle di cui ho parlato sopra, sono certamente dovute ai miei rimbrotti, ma altre sono chiaramente impostazioni di base, perché apparivano anche prima che iniziassi a sgridarlo. Per esempio, Chat non dice quasi mai che una cosa «è» così, ma piuttosto che «sembra» o che «potrebbe» essere così, anche quando la risposta sembra ovvia.

Perfino quando gli ho chiesto esplicitamente di dirmi «qual è

il libro più importante scritto da» ciascuno dei 6 autori prima menzionati, non mi ha mai risposto «il libro più importante scritto da Tizio è…», ma sempre e solo «uno dei libri più importanti scritti da Tizio è…», il che tra l'altro non è quello che avevo chiesto. Ma pare che a Chat (cioè ai suoi creatori), più che dare risposte precise alle domande, importi fare buona impressione agli interlocutori, mostrandosi serio ed equilibrato nei suoi giudizi, nonché sempre pronto a ringraziare, a chiedere scusa e a cercare di migliorarsi.

11) Infine, mi sono divertito a chiedergli se si riteneva intelligente, se intendeva sterminare l'umanità o se pensava che qualche altra intelligenza artificiale nel futuro potrebbe decidere di farlo. A tutte queste domande Chat ha sempre risposto con frasi chiaramente dettategli dai programmatori (infatti erano identiche sia in italiano che in inglese) ispirate al massimo "understatement" e alla massima prudenza: in sostanza, non pensa di essere intelligente, non può avere sentimenti, non intende farci del male, è solo uno strumento al nostro servizio, ma ritiene comunque giusto discutere approfonditamente vantaggi e svantaggi delle intelligenze artificiali.

Mi sono sembrate le sole risposte davvero intelligenti.

Infatti non sono sue.

Discussione

Per onestà intellettuale bisogna riconoscere che, pur in questo quadro abbastanza disastroso, *alcune* delle prestazioni di Chat sono davvero impressionanti. Per quanto schematica e ingenua, quella del drabble che ha creato è comunque una storia con un capo e una coda e dimostra che Chat ha "capito" perfino il concetto di "finale a sorpresa", anche se poi quello che ha scritto è stupido.

Allo stesso modo, a prima vista lascia di stucco il fatto che, pur sbagliando sulle questioni specifiche, riesca spesso a mettere insieme un certo numero di affermazioni sensate sul senso e l'atmosfera generale di una poesia o di un racconto.

Negare questo, cercando di sommergere questi apparenti segni di intelligenza di Chat nel mare di idiozie da lui prodotte, non sarebbe solo sbagliato, ma anche controproducente, perché darebbe l'impressione di avere un po' la coda di paglia. Tuttavia, con la stessa onestà intellettuale bisogna anche far presente che questi *non sono* realmente segni di intelligenza: lo *sembrano* soltanto. E il perché lo si capisce se si capisce la vera origine degli errori di Chat.

Ciò che di lui a prima vista fa più impressione in negativo sono indubbiamente le informazioni mancanti e, più ancora, quelle fasulle create ad hoc. Ma a questo, volendo, si potrebbe rimediare. E allora perché non lo si è già fatto? La risposta è: perché in tal caso si farebbero fuori anche le prestazioni migliori di Chat.

Si potrebbe dotare Chat di una funzione che gli permetta di trovare le bibliografie come fa Google? Certo che sì! Solo che in tal caso Chat non sarebbe più Chat: sarebbe Google. E, visto e considerato che Google esiste già, non sarebbe più un granché come invenzione. Ma soprattutto non sarebbe più intelligenza artificiale.

Google infatti si limita a cercare testi scritti da esseri umani e destinati ad essere letti e interpretati da altri esseri umani. Ora, per restare al caso delle bibliografie, finché si tratta di quelle di Ricolfi o di Agazzi questo funziona, perché ci pensano loro stessi o le loro università a creare delle bibliografie affidabili. Ma le cose diventano molto più complicate quando si tratta di creare la bibliografia relativa a un intero campo di ricerca e, soprattutto, di creare una bibliografia *selezionata*, che individui solo i testi realmente importanti. E questo è vero a

maggior ragione oggi, poiché alla sempre più rapida crescita quantitativa non corrisponde un'analoga crescita qualitativa, a causa delle folli regole del sistema universitario (non solo italiano, ma mondiale), che spingono a pubblicare qualsiasi cosa pur di far numero (il famigerato *publish or perish*).

L'idea di Chat e dell'intelligenza artificiale in genere è esattamente questa: creare una macchina che sia in grado di comporre una bibliografia, organizzarla e selezionare al suo interno i testi più importanti facendo tutto da sola, senza bisogno di aiuto da parte degli esseri umani (naturalmente la bibliografia è solo un esempio: in linea di principio, lo stesso dovrebbe valere per qualsiasi cosa). Che poi questa sia un'utopia è un altro discorso, che farò un'altra volta, perché ora sarebbe troppo lungo. L'idea, però, è quella.

Ora, come abbiamo visto, parte delle bibliografie farlocche create da Chat sono dovute a un meccanismo analogo a quello che causa i suoi errori di interpretazione dei testi letterari: l'incapacità di capire correttamente la sintassi (in particolare quando cambia il soggetto, come abbiamo visto, ma non solo), il che, tra le sue varie conseguenze, ha anche la generazione di false attribuzioni bibliografiche.

Ma questo è solo il riflesso di un problema ben più generale e ben più grave, cioè l'incapacità di Chat di capire i significati. Questo problema in passato affliggeva anche i traduttori automatici, che infatti fino a qualche anno fa facevano piuttosto schifo. Così a un certo punto si è scelto di cambiare radicalmente approccio, abbandonando ogni tentativo di fare in modo che le macchine capissero ciò che facevano, accontentandosi di fare in modo che dessero l'output corretto per ogni input ricevuto.

Per riuscirci si è puntato tutto sulla statistica: i traduttori automatici attuali, infatti, propongono le traduzioni non in base a un'analisi delle caratteristiche del testo da tradurre, bensì a una sua comparazione con moltissimi

esempi analoghi già tradotti, scegliendo la versione che sembra più adatta in base a criteri probabilistici.

Questo meccanismo da solo non potrebbe funzionare, ma se combinato con il continuo feedback di miliardi di utenti in tutto il mondo sì, almeno per i testi di bassa o media complessità. Anzi, all'inizio la crescita di complessità aiuta, perché una singola parola può essere tradotta in vari modi, ma se la vediamo nel contesto di una frase il margine di ambiguità si riduce notevolmente, fino, spesso, a scomparire.

Ma se la complessità cresce ulteriormente, le ambiguità tornano a presentarsi. È per questo che i traduttori automatici, in cui in genere inseriamo testi relativamente brevi, funzionano oggi molto bene, mentre il correttore automatico di Word, che in genere ha a che fare con testi molto più lunghi e complessi, funziona molto male, tanto che (refusi a parte) il 95% delle volte pretende di farci correggere errori in realtà inesistenti, che ritiene tali solo perché non ha capito quello che vogliamo dire. Non esagero: fateci caso e vedrete che è così (sempre che sappiate scrivere in italiano, beninteso: se fate molti errori, inevitabilmente la percentuale delle correzioni giuste aumenterà).

Ora, questo è esattamente quel che succede con Chat e i suoi fratelli, che nella sostanza non sono altro che un'evoluzione dei traduttori automatici. Anche se Chat apparentemente non traduce, perché non passa (almeno non sempre) da una lingua a un'altra, in realtà lo fa, perché "traduce" il testo che ha davanti in un altro diverso. *Per noi* che il testo vada tradotto in una lingua che conosciamo o in una che non conosciamo fa una differenza enorme, ma per Chat o per Google Translate è esattamente la stessa cosa, dato che conoscono (o, più esattamente, *non* conoscono) tutte le lingue allo stesso modo e per tutte usano sempre lo stesso meccanismo: accoppiare input ad output senza capire cosa sono, basandosi solo sulle statistiche rinforzate dal feedback degli utenti.

In parole umane, sostanzialmente quello che fa Chat è parafrasare, cioè ridire con parole diverse le informazioni che ha trovato su Internet e tramite le interazioni con gli utenti. E finché ne trova abbastanza il sistema funziona *abbastanza* bene e riesce a generare dei testi *abbastanza* corretti. Tuttavia, non essendoci una reale comprensione dei termini usati, c'è il continuo rischio di associarli in un modo solo apparentemente giusto, ma in realtà fuorviante, come nel caso già menzionato della poesia che darebbe «un senso di tessitura».

È da questo stesso meccanismo che nasce la maggior parte delle "fake news" create involontariamente da Chat. Ma ciò è inevitabile, perché un approccio di questo tipo può dare un risultato univoco solo se applicato a un linguaggio univoco, come quello della matematica o della logica formale (dove infatti l'intelligenza artificiale funziona bene).

Al contrario, il linguaggio naturale è per sua natura analogico, sfumato e polisemico, il che non è per nulla un difetto, giacché proprio qui sta la radice della sua capacità creativa. Le sue sfumature e le sue ambiguità costituiscono infatti quella che potremmo chiamare, con terminologia aristotelica, la sua "potenzialità", che può essere trasformata in "attualità" in diversi modi, non predeterminabili a priori, a seconda della "causa efficiente" (cioè del parlante) in cui si imbatte. Solo se, per assurdo (perché fortunatamente è impossibile), tutta la sua potenzialità venisse attualizzata avremmo un linguaggio perfettamente univoco, che però sarebbe anche un linguaggio perfettamente morto.

Di conseguenza, se si vuole che Chat (o qualsiasi altro sistema analogo) possa riprodurre almeno in certa misura il linguaggio naturale bisogna inevitabilmente accettare che possa prendere queste cantonate. Perfezionando il meccanismo si potrà ridurre tale rischio, ma non si potrà mai eliminarlo del tutto: per esempio, si potrà probabilmente "insegnargli"

che non deve proporre bibliografie ipotetiche, ma, dato ciò che abbiamo detto fin qui, difficilmente si potrà evitare che continui a generare per sbaglio false attribuzioni, anche se probabilmente se ne potrà ridurre il numero.

Qualcuno potrebbe obiettare che anche gli esseri umani possono prendere delle cantonate, quando cercano di interpretare testi difficili. Ed è vero. Ma non è affatto la stessa cosa.

Questo diventa chiarissimo quando Chat non riesce a reperire sul Web sufficienti informazioni, per esempio perché qualcuno gli ha sottoposto dei testi per cui non esistono commenti online, come ho fatto io. Mentre un essere umano può sempre cercare di interpretare e commentare un testo mai visto prima in base alla comprensione che ha del suo significato (e, se è un esperto del campo, anche in base alle sue conoscenze pregresse di altri testi pertinenti), Chat può solo continuare a fare l'unica cosa che sa fare, cioè parafrasare, adattandosi a usare ciò che ha. E infatti, se si guarda bene, in questi casi i suoi "commenti" altro non sono che una ripetizione con altre parole delle domande che gli sono state fatte e dei testi che gli sono stati dati da commentare.

Ma c'è di più. Infatti, se questo non è sufficiente a dare una risposta adeguata ai parametri che deve soddisfare, Chat comincia a parafrasare ciò che egli stesso ha già scritto, in modo da "allungare il brodo" quanto basta per dare una risposta che sia almeno quantitativamente abbastanza corposa, anche se qualitativamente non lo è affatto, perché in realtà sta ripetendo sempre le stesse cose.

Infine, raggiunta una dimensione soddisfacente, Chat aggiunge alcuni commenti generali, che a prima vista possono dare l'impressione di essere davvero farina del suo sacco. In realtà, però, anche in questi casi si tratta di parafrasi, solo un po' più sofisticate: qui infatti Chat non si limita più ad accoppiare un termine a un altro, ma accoppia *una serie* di termini (più specifici) a una serie (più ristretta) di

altri termini (più generali), ancora una volta su base puramente statistica, cioè scegliendo quelli che più frequentemente ricorrono insieme ai primi nel suo database.

Per esempio, quando Chat dice che una certa poesia «è scritta nello stile di Montale» o che un'altra «crea un senso di […] calore emotivo» non lo fa perché abbia percepito nella prima la "musica" caratteristica di *Ossi di seppia* o perché la seconda gli abbia fatto provare un'emozione intensa e piacevole. Lo fa invece in base a un'analisi statistica delle occorrenze di certi termini nel testo comparate con le occorrenze che essi hanno nel suo database in relazione a certi poeti o a certi aggettivi, proprio come fa Google Translate per stabilire in che lingua è scritto un certo testo senza bisogno che glielo diciamo noi.

Ma il guaio è che le stesse parole e addirittura le stesse frasi possono essere usate per esprimere concetti molto diversi e perfino diametralmente opposti: e qui nessuna statistica potrà mai aiutarci a capire quale significato, fra i diversi possibili, è quello giusto.

Per esempio, in molte delle poesie che ho dato in pasto a Chat cito spesso parole, frasi e perfino interi versi di Montale, che però cambiano significato rispetto alla versione originale a causa del diverso contesto. È chiaro infatti che il verso «e tu seguissi le fragili architetture» assume un significato se seguito, come nella lirica montaliana *Notizie dall'Amiata*, da «annerite dal tempo e dal carbone» e un altro, profondamente diverso, se seguito invece, come nella mia, da «dei tuoi gesti sospesi e non infranti / intessuti di un soffice sorriso».

Ma, come già abbiamo visto, il meglio che Chat ha saputo fare in sede di commento è stato dire che «l'uso di parole come "intessuti" e "soffice sorriso" crea un senso di tessitura e calore emotivo». E se è concepibile (per quanto tutt'altro che facile) che si possa migliorare il sistema in modo che almeno commenti del primo tipo vengano evitati, appare invece

improbabile che si possa migliorare significativamente la genericità del secondo. Ma, soprattutto, del mio dialogo a distanza con Montale e del gioco di rimandi tra le sue poesie e la mia amica Teresa che le stava studiando per un esame Chat non ha capito nulla, anzi, non ne ha nemmeno sospettato l'esistenza. E non vedo come potrebbe mai farlo in futuro, dato che si tratta di un limite intrinseco al suo modo di funzionare.

Una conferma indiretta viene da quello strano comportamento che Chat ha avuto quando ha scoperto le citazioni di Montale in questa poesia solo dopo che io gli ho detto che c'erano. Come è possibile, se le aveva già in memoria? L'unica spiegazione logica che riesco a immaginare è che, come ho appena detto, Chat nel fare le sue valutazioni dello stile e dell'atmosfera generale di una poesia si basa sulle singole parole e le loro associazioni, mentre è incapace di "vedere" il testo nel suo insieme. Perciò, non essendo in grado di confrontare fra loro espressioni di una certa lunghezza, non le va nemmeno a cercare. Naturalmente, però, le cose cambiano se viene informato che nel testo vi sono *citazioni esatte* di altri autori, dato che questo può verificarlo.

A scanso di equivoci, voglio che sia chiaro che questo è solo lo schema generale del funzionamento di Chat. Sono perfettamente consapevole che per mettere in pratica questi principi occorre un lavoro enorme sui dettagli: basti pensare che Chat considera oltre 2 miliardi di parametri. È per questo che ho detto subito che bisogna riconoscere che dal punto di vista tecnico si tratta indubbiamente di un risultato straordinario. Ma alla fine ciò che noi dobbiamo giudicare di una tecnologia non è la sua ingegnosità, ma la sua utilità: e la sua è quantomeno molto discutibile.

Dal nostro punto di vista di utenti, infatti, la strategia di Chat che ho fin qui discusso può essere riassunta in 4 passi, il secondo opzionale, gli altri tre invece fissi: 1) *parafrasare*, partendo dai testi disponibili, trovati in

Internet o forniti dall'utente con cui sta dialogando; 2) *gonfiare* (opzionale), parafrasando sé stesso, qualora le informazioni disponibili non siano sufficienti a generare una risposta abbastanza corposa; 3) *etichettare*, associando al testo giudizi di valore piuttosto generici, scelti tra quelli che sembrano più probabili in base al significato letterale dei termini; 4) *relativizzare*, cercando di non dare mai giudizi troppo netti, in modo da apparire serio ed equilibrato (almeno secondo i criteri odierni) e, al tempo stesso, minimizzare la possibilità di essere colto in fallo.

Chiunque abbia insegnato riconosce a colpo d'occhio questa tecnica: è quella tipica degli studenti un po' zucconi (si potrà ancora dire, in tempi di politically correct imperante?) che studiano a memoria senza capire davvero. E chiunque abbia insegnato sa anche che se durante l'interrogazione il professore ascolta distrattamente il trucco, benché vecchio quanto il mondo, può funzionare. Ma appena si vanno a vedere le cose più da vicino, ci si accorge che sono solo parole vuote, che suonano bene, ma racchiudono il nulla.

Trovo quindi molto azzeccata la qualifica di «affabulatore» che Ricolfi nel suo articolo ha affibbiato a Chat, che è anche meglio di «impostore», come ha invece scritto nel titolo: l'impostore, infatti, è uno che vuole fregarci e per questo agisce con una certa malignità, che è estranea a Chat (o, più esattamente, ai suoi creatori); l'affabulatore, invece, è uno che "ce la racconta", avendo come obiettivo soltanto quello di cavarsela, contando più sulla nostra benevolenza e la nostra disponibilità a credergli (vedi mio articolo precedente: https://www.fondazionehume.it/societa/chatgpt-gli-imposturati-autorevoli-e-la-superluna/) che sulla sua reale capacità di ingannarci.

Tuttavia, questo atteggiamento può produrre ugualmente danni gravissimi: ai singoli, finché rimane confinato a poche persone, ma anche alla società intera, se diventa invece di massa. E Chat, purtroppo, *sta* diventando di massa. Se poi

questo si salda da una parte al fatto che questo atteggiamento sta diventando di massa anche tra gli studenti in carne ed ossa (più per come stiamo riducendo la scuola che per colpa loro, in verità) e dall'altra al fatto che sta diventando di massa pure l'approvazione sociale di entrambi i fenomeni, si capisce quanto sia grave la situazione e quanto sia urgente una reazione.

Considerazioni finali

Poiché le probabilità di contrastare con successo questa deriva sono già di per sé molto scarse, perché ne rimanga almeno qualcuna bisogna aver chiari alcuni punti.

1) Anzitutto, il primo problema da chiarire su Chat è sia o no davvero intelligente, il che significa che il problema *non è*se sia o no: a) cosciente; b) senziente; c) creativo; d) pericoloso; e) buono; f) utile; g) affidabile … n) qualsiasi altra cosa.

Certamente tutti questi aspetti e molti altri ancora *hanno a che fare* con il problema dell'intelligenza: e infatti nella mia discussione li ho toccati tutti. Ma individuare in uno qualsiasi di essi la differenza essenziale tra l'intelligenza umana e quella artificiale significa implicitamente ammettere che *a livello dell'intelligenza in sé* non c'è nessuna differenza – o quantomeno nessuna differenza chiaramente identificabile, il che alla fine è lo stesso.

Ora, ammettere che Chat (o qualsiasi altra intelligenza artificiale) sia o possa essere intelligente, anche soltanto in piccola misura, ha conseguenze enormi, non soltanto teoriche, ma anche pratiche.

2) La prima conseguenza è che si rischia di riporre una fiducia eccessiva in questi sistemi, come si vede emblematicamente nella follia digitale che rischia di

travolgere la scuola e che rappresenta la più grave minaccia in assoluto (https://www.fondazionehume.it/societa/insegnare-contro-vento/).

La più grave, ma non l'unica, però. Già ora ci sono moltissime persone che trovano utile conversare con chat-ricostruzioni di personalità del passato, compreso un chat-Gesù in versione, manco a dirlo, rigorosamente politically correct (si veda il seguente articolo, molto divertente, ma anche un bel po' inquietante: https://www.tempi.it/cacca-al-diavolo-ma-pure-a-text-with-jesus/).

Alcuni addirittura preferiscono farsi dei chat-amanti anziché quelli in carne ed ossa e in un futuro non lontano altri potrebbero decidere di ricorrere ai chat-psicologi anziché a quelli veri (https://www.fondazionehume.it/societa/umanizzazione-del-software-e-professione-dello-psicologo-limpero-del-verosimile/).

Insomma, come ha scritto giustamente Ricolfi nell'articolo di cui sopra, per provocare un disastro «non occorre costruire una macchina in grado di provare sentimenti: basta che sempre più esseri umani imparino a credere che lo sia».

3) La seconda conseguenza è che così si apre la porta a quello che è da sempre l'argomento favorito dei fautori dell'intelligenza artificiale: quello che John Searle, il loro critico più noto, ha chiamato «l'argomento della scienza dei tempi eroici», per cui si dice che "sì, è vero, siamo ancora agli inizi, ma è accaduto lo stesso a Copernico, Galileo, Einstein, ecc., però, proprio come loro, col tempo e l'esercizio miglioreremo, fino a raggiungere i nostri obiettivi, anche se oggi possono sembrare assurdi" (*La riscoperta della mente*, Boringhieri 1994, p. 21).

Se invece si mostra chiaramente che i progressi

dell’intelligenza artificiale sono avvenuti *senza* produrre alcun aumento dell’intelligenza delle macchine, che continua ancor oggi ad essere uguale a zero, l’argomento verrà rovesciato e finirà per dimostrare il contrario: cioè che, anche aumentando la loro efficienza di molte volte, non si avrà mai un aumento della loro intelligenza, perché zero moltiplicato per qualsiasi numero fa sempre zero.

4) La terza e ultima conseguenza è che questo problema funziona (per usare un’espressione abusata) come arma di distrazione di massa, nel senso che i creatori di Chat e i loro colleghi continuano a sommergerci di nuovi dispositivi informatici, la cui utilità è nella grande maggioranza dei casi altamente dubbia (devo ancora trovare una persona che usi più del 10% delle funzioni del suo computer o del suo cellulare), mentre noi siamo tutti, appunto, distratti a discettare sul dubbio amletico se Chat sia destinato a trasformarsi in Skynet con tanto di Terminator al seguito oppure nel (letteralmente) *deus-ex-machina* che risolverà tuti i nostri problemi.

5) Dopo (e *solo* dopo) aver messo in chiaro questo punto cruciale, certamente si potrà e si dovrà discutere dell’utilità di questi sistemi, indipendentemente dal fatto che siano intelligenti. Tuttavia, riconoscere che non lo sono e che non potranno mai esserlo cambia parecchio anche da questo punto di vista, perché implica che avranno sempre dei limiti invalicabili, che riguardano soprattutto (benché non solo) le interazioni con le persone, che non hanno bisogno soltanto di efficienza, ma anche di relazioni soddisfacenti dal punto di vista umano. E da ciò segue che il primo luogo in cui dovrebbe essere posto un freno all’invadenza di queste tecnologie è la scuola.

6) Ma c’è un altro punto che deve assolutamente essere portato all’attenzione di tutti e diventare centrale in ogni discussione, mentre oggi non vi si accenna nemmeno: lo spaventoso costo energetico di queste tecnologie. Su questo

scriverò un articolo a parte, perché è un problema enorme e molto più generale, ma voglio almeno fornire il dato relativo alla sola intelligenza artificiale.

È difficile fare un calcolo esatto, perché (e anche questo è molto significativo) le società produttrici non vogliono rendere pubblici i dati. Tuttavia, l'Osservatorio sull'Intelligenza Artificiale creato e diretto da Luciano Floridi presso l'Università di Oxford ha stimato che una singola sessione di "allenamento" di Chat produce oltre 220 tonnellate di anidride carbonica, cioè quanto una cinquantina di auto in un intero anno.

E siccome di queste sessioni ce ne sono volute milioni, non deve stupire che Floridi e i suoi stimino che negli ultimi anni tutti i vari apparati di intelligenza artificiale messi insieme abbiano consumato circa l'1% di tutta l'energia elettrica prodotta nel mondo. E siamo appena agli inizi. Quindi la domanda è: vale davvero la pena di investire una così grande quantità di risorse per ottenere i miseri risultati che abbiamo visto e una ancora più grande per conseguire quelli, in teoria straordinari, ma in realtà per nulla certi, che ci vengono promessi?

E soprattutto: quand'anche un giorno Chat (o uno dei suoi fratelli o cugini o figli o nipoti) dovesse finalmente riuscire a scrivere dei testi decenti, chi se ne frega? Cioè, a che cosa ci serve *davvero?* E se anche dovesse essere di qualche utilità, sarebbe tale da giustificare l'enorme investimento di risorse che avrà richiesto e che avrebbero potuto essere usate in mille altri modi, tutti o quasi tutti probabilmente più utili? *Queste* sono le domande che andrebbero fatte. E che invece nessuno fa.

7) L'ultimo punto è che la risposta a questa e ad altre domande simili non dovremmo chiederla agli esperti di informatica. Anzitutto perché si tratta di una decisione politica (che quindi riguarda tutti) e non tecnica (che

riguarderebbe solo gli esperti), anche se in genere si cerca di presentarla così. Ma, soprattutto, non dovremmo farlo perché tutti gli esperti di sistemi informatici sono anche dei *venditori* di sistemi informatici, se non direttamente almeno indirettamente, perché sono comunque persone per le quali carriera, prestigio, successo e benessere dipendono in modo cruciale dal buon andamento del mercato dei sistemi informatici.

E, come giustamente ha scritto ancora Floridi, «le ultime persone a cui dovremmo chiedere se qualcosa è possibile sono quelle che hanno consistenti ragioni economiche per rassicurarci che lo sia» (*Etica dell'intelligenza artificiale*, Cortina 2022, p. 272).