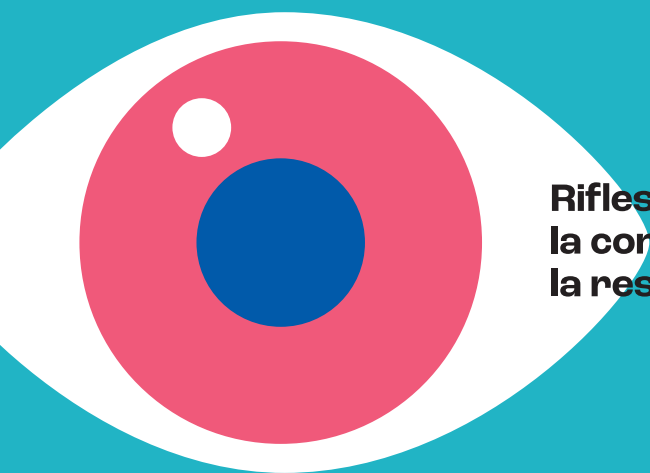


**Luca Mari**

# **L'Intelligenza Artificiale di Dostoevskij**



**Riflessioni sul futuro,  
la conoscenza,  
la responsabilità umana**

Con Daniele Bellasio,  
Francesco Bertolotti,  
Alessandro Giordani

Il Sole  
**24 ORE**

**Luca Mari**

**L'Intelligenza  
Artificiale  
di Dostoevskij**

Il Sole  
**24 ORE**

Il Sole 24 ORE

Progetto grafico copertina: Francesco Narracci  
Immagine di copertina: Zetalab.com

ISSN 977-1826380-683-40001

Il Sole 24 Ore – Scienza  
Registrazione in Tribunale n. 542 - 08.07.05  
Direttore responsabile: Fabio Tamburini  
Proprietario ed Editore: Il Sole 24 ORE S.p.A.  
Sede legale, redazione e direzione: Viale Sarca, 223 – 20126 Milano  
Mensile n. 1/2024

ISBN 979-12-5484-2812

GRUPPO  ORE

© 2024 Il Sole 24 ORE S.p.A.  
Sede legale, redazione e amministrazione: Viale Sarca, 223 – 20126 Milano  
Per informazioni: Servizio Clienti 02.30300600

Fotocomposizione: Emmegi Group, via F. Confalonieri, 36 – 20124 Milano

Prima edizione: febbraio 2024

---

Tutti i diritti sono riservati.

I testi e l'elaborazione dei testi, anche se curati con scrupolosa attenzione, non possono comportare specifiche responsabilità dell'Editore per involontari errori e/o inesattezze; pertanto il lettore è tenuto a controllare l'esattezza e la completezza del materiale utilizzato. Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941, n. 633. Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da EDISER Srl, Società di servizi dell'Associazione Italiana Editori, attraverso il marchio CLEARedi, Centro Licenze e Autorizzazioni Riproduzioni Editoriali, Corso di Porta Romana, n. 108 – 20122 Milano.  
Informazioni: [www.clearedi.org](http://www.clearedi.org).

---

*In ricordo di Francesco, che avrebbe avuto tanto  
da dire e da dare sul nostro tema*

# Indice

<i>Premessa</i>	7
<i>Introduzione</i>	
<b>Imparare (a imparare) con i chatbot: una lettera</b>	9
<i>Capitolo 1</i>	
<b>Di cosa parleremo, insomma</b>	27
<i>Capitolo 2</i>	
<b>«Chatbot, che ore sono?», ovvero cominciare a capire l'intelligenza artificiale con una domanda e tre risposte</b>	39
<i>Capitolo 3</i>	
<b>ChatGPT: uno strumento, un collega, o un consulente?</b>	45
<i>Capitolo 4</i>	
<b>Così Dostoevskij ci aiuta a capire ChatGPT</b>	55

*Indice*

*Capitolo 5*

**Cosa c'è dentro la scatola, dunque?** 69

*Capitolo 6*

**I limiti cognitivi ed etici (attuali: chissà nel futuro...) dei chatbot** 85

*Capitolo 7*

**ChatGPT e gli agenti: una nuova frontiera per l'intelligenza artificiale** 95

*Capitolo 8*

**Perché fare previsioni questa volta è più difficile** 105

*Capitolo 9*

**ChatGPT e il superamento della distinzione tra “cultura umanistica” e “cultura tecnico-scientifica”** 109

*Capitolo 10*

**Verso una terza rivoluzione culturale?** 117

*(non)-Conclusione*

**Imparare (a imparare) con i chatbot: una seconda lettera** 123

## *Premessa*

Se domani mattina un alieno bussasse alla tua porta, e in qualche modo tu capissi che potreste facilmente comunicare e che non hai nulla da temere, cosa faresti?

Per quanto ne sappiamo, nessun essere umano si è mai trovato in questa situazione, ma quello che sta accadendo da quando, solo pochi mesi fa, possiamo dialogare con sistemi di intelligenza artificiale generativa, come ChatGPT e Bard, è la cosa più simile a un incontro con un'entità aliena che ci sia capitata finora. Senza un'esperienza precedente, molti sono invasi dalla paura dell'ignoto: cosa sta succedendo? Cosa ci succederà? Ma in qualcuno prevale invece lo stupore per la novità: cosa posso imparare di questo inatteso ospite? E, cosa da esso posso imparare su di me?

Nelle pagine che seguono esploriamo soprattutto questa seconda dimensione: senza lasciarci vincere dai pur comprensibili timori, andiamo alla scoperta di un futuro di cui forse stiamo vedendo i primi segni. È un cammino tracciato per sollecitare domande su noi stessi e la nostra umanità, che potremo fare senza preoccuparci di essere esperti di informatica o matematica. Ma che magari ci farà respirare la bellezza di terre poco esplorate al confine tra “cultura tecnico-scientifica” e “cultura umanistica”, e che potremo abitare se saremo saggi nelle

decisioni che prenderemo: insieme con gli alieni che ci siamo costruiti.

In una condizione di cambiamento inatteso, rapido e radicale come quella che stiamo vivendo, la conversazione tra noi, e non solo con dei chatbot, è vitale. Quanto segue è anche il risultato di tanti dialoghi con colleghe/colleghi e amiche/amici. Francesco Bertolotti, che ha in particolare collaborato alle idee dei capitoli 2, 3, 7, e 9, e con cui mi sono confrontato quasi giornalmente da quando la rivoluzione è cominciata; Alessandro Giordani, la persona che cerca di evitarmi di sbagliare troppo nelle questioni filosofiche, e con cui ho discusso sui contenuti dei capitoli 4, 6, e 10; Daniele Bellasio, con cui abbiamo scritto il capitolo 1 e che mi ha aiutato a rendere questo testo leggibile non solo per addetti ai lavori. E Samuele Astuti, Giacomo Buonanno, Luisa Caldiroli, Katia Cattaneo, Davide Cionfrini, Francesco Gaeta, Rodolfo Helg, Debora Lonardi, Andrea Maiello, Luca Massi, Mauro Mezzenzana, Rita Manzoni, Roberta Morici, Alberto Negrini, Stefania Paci, Aurelio Ravarini, Emanuele Strada, Mark Wilson.



*Introduzione*

## **Imparare (a imparare) con i chatbot: una lettera**

*Una lettera che un docente universitario ha indirizzato agli studenti di un suo corso prima dell'inizio delle lezioni.*

Buongiorno. Benvenute e benvenuti.

Cominciamo il nostro corso con una domanda, che potremmo sempre farci ma che raramente ci poniamo: *perché siamo qui? Perché sei qui?*

Ognuna e ognuno di noi ha la sua risposta, più o meno ben definita, ma avere una motivazione esplicita e condividerla è diventato importante, e perfino urgente, a causa di quello che da qualche mese sta succedendo. Mi sto riferendo al fatto che, alla sola condizione di avere un accesso a internet, sono stati resi disponibili a tutti dei sistemi software di intelligenza artificiale addestrati per saper dialogare in italiano e in tante altre lingue, e con capacità così notevoli da aver colto di sorpresa anche molti esperti. Tecnicamente si chiamano “sistemi di intelligenza artificiale generativa conversazionale”, ma spesso ci si riferisce a essi con il termine inglese *chatbot*; due esempi noti di chatbot sono ChatGPT di OpenAI e Bard di Google.

La presenza di questi “dialogatori automatici”, chatbot appunto, ci sollecita a chiederci: con sistemi tecnologici che imparano così bene e rapidamente, e con la prospettiva che nel tempo che voi impiegherete a completare il vostro percorso di studi saranno ancora mi-

glierati, in modi oggi poco prevedibili, ha ancora senso per noi esseri umani continuare a dedicare una parte importante del nostro tempo a venire in un'aula universitaria per imparare?

Non vi sembri strano che sia proprio in università che ci si fa questa domanda: se cambia qualcosa a proposito di apprendimento, dove altrimenti se ne dovrebbe parlare? E non pensiate, infatti, che il *cosa* imparare e il *come* imparare siano stabiliti una volta per tutte. Per esempio, fino a un po' di anni fa a scuola si imparava a calcolare, carta e penna, le radici quadrate: oggi generalmente non lo insegniamo più, e non lo sappiamo più fare, perché abbiamo delegato questo compito alle calcolatrici elettroniche e ai programmi dei fogli di calcolo. Lo abbiamo accettato e suppongo che pochi di noi rimpiangano i tempi in cui invece si faceva tutto "a mano". In effetti, i sistemi digitali hanno già modificato tanti aspetti dei nostri processi di apprendimento: pensate, per fare un altro esempio, a come la possibilità di fare copia-e-incolla nei programmi di videoscrittura ha cambiato il nostro stile di scrittura rispetto a quando si scriveva direttamente sulla carta e cambiare una frase richiedeva di cancellare e riscrivere.

E dunque, cosa succederà ora con questa novità dei chatbot? (In effetti, i chatbot esistono da decenni, ma fino a prima di ChatGPT erano assai meno capaci: per brevità userò il termine "chatbot" per riferirmi a ChatGPT e ai suoi fratelli, ai "nuovi" chatbot dunque). *In cosa ci faremo aiutare e in cosa ci sostituiranno?* Anche questa domanda è cruciale, perché ogni cosa che non facciamo più noi è qualcosa in meno che è necessario imparare.

Ed è per questo che ha senso chiedersi *cosa è utile imparare oggi*, in quest'aula e ovunque.

Ammettiamolo: a proposito di tutti questi problemi non abbiamo le idee chiare, anche perché quello che sta succedendo intorno ai chatbot è cominciato da troppo poco tempo e sta ancora cambiando troppo in fretta. Anche per questo chiedersi *cosa siamo qui a fare?* è così importante e urgente: per vivere in modo consapevole e attivo questo cambiamento, invece di esserne sopraffatti, per quanto ci è possibile.

Suppongo che abbiate già usato un chatbot e se non l'avete ancora fatto vi suggerisco di provare. Che sensazioni ha generato in voi questa esperienza? Probabilmente sorpresa e incredulità, per la constatazione di trovarsi di fronte a un'entità tecnologica capace di dialogare in modo così sofisticato, ma forse anche paura che l'incertezza spesso porta con sé. Ma la paura del cambiamento non è una buona compagna, perché ingigantisce i rischi e ci fa perdere delle opportunità.

Nella diffusione dei chatbot ci sono certamente anche dei rischi e non solo delle opportunità, ma lasciamo a chi deve prendere le decisioni importanti se sia una buona idea introdurre limiti o proibizioni all'uso di questi sistemi, nella società in generale e nelle scuole e università in particolare. Noi siamo qui per imparare e, in una situazione in cui le cose cambiano, è utile essere informati e avere qualche interpretazione su cosa sta succedendo. Se non avete ancora acquisito delle buone competenze nell'uso dei chatbot, dedicate perciò un po' di tempo a sperimentare: il meno che posso dirvi è che è un'esperienza interessante!

Con ciò, vorrei evitare un possibile malinteso: non sono qui per insegnarvi a usare i chatbot. Durante il nostro corso li useremo e potrò forse aiutarvi a imparare a usarli sempre meglio, ma i contenuti su cui lavoreremo insieme sono altri. E sono proprio questi contenuti a essere messi in discussione dalla presenza dei chatbot: come qualche decina di anni fa successe con la formula per calcolare le radici quadrate.

Esploriamo i problemi che vi ho proposto a partire da una domanda almeno all'apparenza semplice: che differenza c'è tra imparare a memoria un numero di telefono e una formula di matematica o di statistica o di fisica? Se le formule fossero soltanto degli strumenti da usare meccanicamente, come si potrebbe pensare quando ci viene chiesto di applicarle ripetutamente per svolgere degli esercizi, non ci sarebbe nessuna differenza sostanziale. Data l'ipotesi sempre più realistica che i chatbot stiano diventando assistenti sempre migliori, capaci di trovare la soluzione di problemi sempre più complessi, potremmo arrivare a concludere che è arrivato il momento di smettere di imparare a memoria formule, come ormai non ci preoccupiamo più di farlo per i numeri di telefono. E in effetti c'è un senso in questo: se so che per risolvere un certo problema devo applicare una certa formula, che però non mi ricordo, grazie a una connessione internet e a un motore di ricerca usato bene probabilmente “sono distante qualche click” dalla soluzione. E un chatbot mi facilita ulteriormente, consentendomi di domandare come farei con un essere umano, e di farlo in una conversazione che mi può avvicinare progressivamente alla soluzione che sto cer-

cando. La memorizzazione continua a svolgere un ruolo nell'acquisizione di abilità fondamentali come la lettura e le basi dell'aritmetica, dato che, almeno per ora, per esempio saper leggere è praticamente necessario, ma l'impiego di "aiutanti tecnologici" sta rendendo sempre meno utile l'apprendimento basato sulla ripetizione e sull'attività mnemonica (anche a questo proposito, riconosciamo che non sappiamo come sarà il futuro: i sistemi *text-to-speech* saranno così diffusi da rendere la capacità di leggere meno utile di quanto sia stata negli ultimi secoli? O forse produrranno nuove forme di analfabetismo? È un altro dubbio non ovvio per chi si sta ponendo seriamente questo genere di problemi).

Non vedo punti deboli in questa argomentazione, che mi sembra sostanzialmente corretta. Se il mio scopo è di esercitare la mia memoria, posso imparare dei numeri di telefono, delle formule, delle poesie e qualcos'altro. Ma se devo risolvere un problema, conoscere a memoria una formula non è più così utile (e infatti – non preoccupatevi! – all'esame potrete portarvi tutti gli appunti che vorrete: sarà, come si dice, una prova *open book*, perché ciò che è importante è che sappiate risolvere dei problemi, non che vi ricordiate delle formule a memoria).

Con ciò siamo arrivati alla prima conclusione che vi propongo: penso che *la diffusione dei chatbot renderà sempre meno necessario imparare per ripetizione e quindi, appunto, a memoria*. In relazione allo stile di apprendimento di ciascuno, questa è un'ottima o una pessima notizia, ma, suppongo che possiamo concordare, non è controversa. Più complesso è invece capire, nel nuovo scenario generato dalla presenza dei chatbot, *cosa e come* imparare.

Ripeto qui la mia ammissione: non sono certo di quello che sto per dirvi, e anzi spero che lavorando insieme in questo corso ci aiuteremo reciprocamente a chiarirci le idee.

Propongo di ricominciare, ancora una volta, da una domanda: *che differenza c'è tra esercizi e problemi?* Non è tanto una questione di contenuto: se hai imparato qualcosa che io non conosco, quello che per te potrebbe essere solo un esercizio di ripasso per me potrebbe essere un problema. Il punto è invece che gli esercizi richiedono ripetizione e i problemi creatività (perché se l'hai già risolto, non è più un problema). Svolgere esercizi è spesso utile per imparare a risolvere problemi, ma nel lavoro vi pagheranno per risolvere problemi, non per svolgere esercizi e, allargando un poco la prospettiva, possiamo supporre che se la nostra specie non si è estinta è perché abbiamo imparato a essere efficaci nella soluzione di problemi, non nello svolgimento di esercizi. Vi invito a lavorare ancora su questa differenza, magari anche discutendone tra di voi, ma suppongo che l'idea sia già sufficientemente chiara (vi do tre suggerimenti: potreste chiedervi che significato abbiano le usuali espressioni “esercizi di algebra” e “problemi di geometria”; e che origine abbia il piacere che gli esseri umani sentono nel giocare e nel fare sport, se non abbia a che vedere con il fatto che nel gioco e nello sport ci mettiamo alla prova nel risolvere problemi; e potreste riflettere su quanti generi diversi di problemi ci poniamo e cerchiamo di risolvere).

Fino a oggi che un docente aiutasse i suoi studenti a imparare a risolvere problemi è stata considerata

un'eccellente ragione da proporre a chi si chiedeva perché frequentare le lezioni: «Partecipa non per ascoltare e prepararti a ripetere quello che hai ascoltato, ma per diventare un buon *problem solver*». Ma ora ci potremmo chiedere (ancora una domanda!): *perché dovrei continuare a imparare a risolvere problemi, quando posso farli risolvere a un chatbot che potrebbe essere più abile di me?* È una domanda non solo perfettamente lecita, ma anche ragionevole, che non voglio eludere. E penso anzi che se trovassimo una risposta convincente avremmo qualche ottimo indizio per risolvere il problema da cui siamo partiti: *perché siamo qui?* C'è ancora molto da esplorare in questo e i contributi di ognuno di voi saranno ben accetti. Per ora comincio io, sintetizzando alcune prime idee in due parole: *responsabilità e motivazione*.

### **A proposito della responsabilità**

Non è facile descrivere le caratteristiche dei chatbot e più in generale dei sistemi di intelligenza artificiale. Da decenni ci si chiede se in un qualche senso queste entità pensino e siano intelligenti, questione complessa anche perché non abbiamo le idee così chiare su cosa significa, per noi esseri umani, pensare ed essere intelligenti. Ma che si comportino come se pensassero e fossero intelligenti è evidente. E il fatto che commettano degli errori non modifica le cose; del resto, anche noi esseri umani commettiamo errori, ma questo non ci fa dubitare di pensare e generalmente di essere intelligenti.

In una cosa i chatbot sono però radicalmente diversi da noi (almeno per ora: nel futuro, chissà...): *non sono nella condizione di prendersi delle responsabilità di fron-*

*te alla nostra società.* Se fanno cose giuste, non sappiamo come premiarli; se sbagliano, non sappiamo come punirli. Sono, letteralmente, non-responsabili, così che, anche quando li usiamo per prendere decisioni, ed eventualmente deleghiamo loro delle decisioni, rimaniamo noi responsabili. Questa condizione è stata resa evidente finora anche dall'assenza di una connessione diretta dei chatbot con l'ambiente circostante: sono entità solo parzialmente attive, che operano ricevendo dei testi da esseri umani e producendo dei testi per esseri umani. Certo, possiamo aspettarci che presto i chatbot saranno addestrati a funzionare anche in modo attivo, acquisendo informazione dal web e perfino dal mondo empirico mediante sensori come telecamere e microfoni, e intervenendo in modo diretto sul loro ambiente, informazionale o empirico. Possiamo cioè aspettarci che presto i chatbot saranno i controllori (le "menti", per usare un altro termine antropomorfo) di sistemi tecnologici complessi, scenari a cui per altro romanzi e film di fantascienza ci hanno abituato. Ma anche in questo caso la responsabilità del loro comportamento resterà nostra.

Considerazioni di questo genere sembrano essere così generali da non riguardarci direttamente. Ma non è così, perché ogni volta che qualcuna o qualcuno di noi ha una conversazione con un chatbot dovrebbe ricordare che non è il chatbot a essere responsabile dell'informazione prodotta durante la conversazione e dell'uso che poi ne viene fatto. Il chatbot è un aiutante, che può operare in modo anche molto sofisticato, ma siamo noi sia a dovergli fare le domande corrette sia a dover interpretare in modo corretto le risposte che ci fornisce.



Insomma, anche se l'elaborazione è in carico al nostro aiutante, la produzione degli input e l'interpretazione e l'uso degli output sono sotto la nostra responsabilità.

Ci sono delle non ovvie competenze da acquisire in questo.

*A proposito degli input*, e quindi dell'imparare a fare a un chatbot le domande giuste nel modo giusto (quello che qualcuno oggi chiama *prompt engineering*, cioè "ingegneria delle domande"), si tratta di una competenza fondamentale per diventare dei buoni solutori di problemi, anche a prescindere dalle abilità specifiche utili per dialogare con un chatbot: se non sai che problema vuoi risolvere, come puoi sperare di trovare la soluzione? O, come diremmo a questo punto, se non sai spiegarli il tuo problema, come puoi sperare che il chatbot trovi la soluzione?

Vedo due dimensioni complementari in questo. Prima di tutto, è in gioco *una sensibilità alle condizioni della comunicazione*. Perché il nostro interlocutore capisca le nostre richieste, dobbiamo essere comprensibili, adattando il nostro linguaggio e le nostre modalità di interazione. Fino a prima dei chatbot questo era evidente, dato che potevamo impartire istruzioni testuali ai calcolatori digitali soltanto impiegando un qualche linguaggio artificiale progettato appositamente per questo scopo, cioè un cosiddetto "linguaggio di programmazione", come Python e Java. Con i chatbot dialoghiamo nella nostra lingua e questo senza dubbio elimina qualche barriera a una comunicazione efficace, ma non dobbiamo ingannarci: la qualità delle risposte che un chatbot ci fornisce dipende dalla qualità delle domande che gli poniamo

durante la conversazione, e da qui traggo finalmente una prima conclusione positiva: penso che *un buon uso dei chatbot ci solleciterà a migliorare la nostra sensibilità alle condizioni di una comunicazione efficace.*

La seconda dimensione ha a che vedere con *i contenuti della comunicazione.* Una massima attribuita ad Albert Einstein, e dunque formulata ben prima della diffusione dei sistemi di intelligenza artificiale, ce lo chiarisce bene: «Se avessi un'ora per risolvere un problema, dedicherei i primi 55 minuti a pensare al problema, e gli ultimi 5 a pensare alla soluzione». Ma cosa facciamo in quei 55 minuti passati a “pensare al problema” se non abbiamo le competenze disciplinari che ci consentono di esplorare e capire il problema? In altri termini, le competenze disciplinari sono utili non solo a chi deve risolvere un problema già ben impostato, ma anche, e a questo punto possiamo dire soprattutto, a chi deve impostare il problema. E le competenze disciplinari sono fondate anche su contenuti, che dunque in questa prospettiva rimane utile acquisire e saper usare appropriatamente. Cosa che a sua volta richiede anche che, in qualche modo, ce li ricordiamo. Come la presenza dei chatbot modificherà questo modo appropriato di ricordare produttivamente, invece che per semplice ripetizione? È uno dei tanti problemi che ci porremo nei prossimi anni, suppongo, ma penso che continueremo a considerare l'intelligenza umana un valore prezioso. E l'intelligenza umana, nelle sue due dimensioni etimologiche dell'*intus-legere*, il “leggere dentro” e quindi il pensiero analitico, e dell'*inter-legere*, il “leggere attraverso” e quindi il pensiero associativo, è un'attitudine che si alimenta di competenze

disciplinari. Ecco allora una seconda conclusione positiva: penso che *le competenze disciplinari continueranno a essere importanti per aiutarci a capire i problemi* (cambierà con ciò il modo di insegnare? E di verificare i risultati dell'apprendimento? Mi sembra probabile, ma sono domande difficili, che lasciamo agli esperti di didattica). Badate che non sto sostenendo che questa sia l'unica ragione per cui sia opportuno continuare ad acquisire competenze disciplinari: per fare un esempio che mi sembra molto lontano dal *problem solving*, conoscere un po' di storia e di geografia rimane una condizione per consentirci di vivere nella nostra società in modo consapevole.

*A proposito degli output*, e quindi dell'imparare a interpretare correttamente le risposte di un chatbot, potrebbe essere che non abbiate realizzato fino in fondo quanto la situazione che stiamo vivendo sia diversa da tutto ciò a cui l'uso dei calcolatori digitali ci aveva abituato. Finora i sistemi software si sono dimostrati entità assai affidabili, perché in pratica perfettamente ripetibili. Certo, questo non impedisce che contengano *bug* che producono comportamenti errati, ma, come l'esperienza ci insegna, si tratta di eccezioni. La conseguenza è che, generalmente a buona ragione, tendiamo a fidarci dei risultati che il software produce: se usiamo un foglio di calcolo per fare dei conti, non ci preoccupiamo di controllare il risultato con un altro strumento. Per ragioni varie, tutto questo non è più vero nel caso dei chatbot, che non solo hanno un comportamento non deterministico, nel senso che a parità di domanda possono produrre risposte diverse, ma anche possono generare in-

formazione imprecisa, non aggiornata o perfino proprio falsa. Nel confronto con i sistemi software tradizionali, è dunque come se si fosse attuato uno scambio: i chatbot sono enormemente più flessibili e, se ancora una volta mi concedete un termine antropomorfico, creativi, al prezzo di una molto minore immunità agli errori. Il caso che appare addirittura paradossale è quello dei calcoli aritmetici, che qualche volta sbagliano! Insomma, i chatbot non sono né archivi informatizzati né motori di ricerca nel web, e tanto meno sono “macchine della verità”: dato che possono essere usati per dialogare praticamente su qualsiasi argomento, possono anche sbagliare praticamente su qualsiasi argomento.

Un uso consapevole di un chatbot ci richiede perciò di avere delle appropriate competenze per saper validare quanto il chatbot produce, anche a proposito delle fonti che dichiara, che a volte sono esse stesse errate (provate a chiedergli consigli su libri da leggere su un qualche argomento non troppo ovvio: potrebbe proporvi titoli di libri inesistenti...). Del resto, viviamo in una società complessa, in cui distinguere i fatti dalle opinioni è spesso difficile, e in cui in certi casi l'abilità retorica sembra più efficace del riferimento alla verità: sarebbe illusorio sperare che sia un'entità tecnologica, per quanto sofisticata, a risolvere per noi problemi delicati come quello delle *fake news*. Ne derivò una terza conclusione positiva: penso che *un buon uso dei chatbot ci solleciterà a sviluppare una migliore competenza nel valutare la qualità dell'informazione*.

L'importanza di un'educazione alla responsabilità dovrebbe essere dunque evidente. Come è stato detto,

«da un grande potere derivano grandi responsabilità», e i chatbot ci mettono a disposizione un grande potere cognitivo. È nelle nostre mani, nella nostra responsabilità appunto, la decisione di imparare a usarlo per diventare persone migliori e contribuire a migliorare la nostra società.

### **A proposito della motivazione**

Torniamo ora all'importanza che attribuiamo alla capacità di risolvere problemi. Il processo che porta alla soluzione di un problema è impegnativo (se non lo fosse, non sarebbe un problema), ma constatare che siamo stati capaci di arrivare in fondo con successo ci genera soddisfazione e dà un significato al tempo che abbiamo dedicato. Questo è talmente in profondità nella nostra psiche che spesso, quando abbiamo un po' di tempo libero, ci inventiamo dei problemi per provare il piacere di mostrare a noi stessi che siamo capaci di risolverli. Cosa sono, infatti, gli hobby, i giochi, lo sport, se non problemi che scegliamo liberamente di porci?

In questo i chatbot possono forse risolvere problemi al nostro posto, ma non rispondono comunque alla nostra domanda di senso, proprio quella da cui è cominciata la nostra riflessione: *perché siamo qui?* Per giustificare la fatica che dovrete fare per imparare potete avere delle motivazioni estrinseche, come il piacere di prendere un bel voto, di fare bella figura con i vostri amici, di fare contenti i vostri genitori. Ma stiamo davvero bene con noi stesse e noi stessi quando sviluppiamo un interesse genuino per l'attività stessa di fare quello che facciamo, in questo caso di risolvere proble-

mi, per cui il percorso diventa non meno importante del suo punto di arrivo.

Questo è ben spiegato da una teoria, chiamata *dell'autodeterminazione (self-determination)*, formulata qualche decina di anni fa, ma con fondamenti riconducibili ad Aristotele e dunque ben prima che il ruolo dei chatbot potesse essere preso in considerazione, secondo cui ogni persona ha tre bisogni psicologici fondamentali: il bisogno di autonomia, di relazione, e di competenza. Ora che abbiamo tra noi questi nuovi aiutanti artificiali, dobbiamo imparare come interagire con essi per soddisfare sempre meglio questi bisogni, e quindi per “farci stare bene” davvero, invece di produrre in noi insoddisfazione e infelicità. Siccome è tutto nuovo qui, non ho risposte da darvi, ma solo ipotesi per esplorare e sperimentare.

Il primo bisogno psicologico fondamentale che abbiamo è quello di *autonomia*: abbiamo bisogno di sentirci autori delle nostre azioni e decisioni, e quando ci accorgiamo di saper decidere in modo significativo stiamo bene con noi stesse e noi stessi. L'interazione con un chatbot può essere pericolosa per questo, dato che delegargli sistematicamente la soluzione dei problemi ci farebbe perdere il controllo sul processo di apprendimento e risoluzione dei problemi e quindi potrebbe compromettere il nostro senso di autonomia. Ma se, invece di farci sostituire, ci facciamo aiutare, magari anche con conversazioni in cui gli chiediamo di impersonare un docente o un esperto che ci mette alla prova, e siamo noi a mantenere la gestione di quello che succede, un chatbot potrebbe essere un efficace abilitatore, e farci sviluppare delle competenze sempre più sofisticate, e

consentirci di imparare a risolvere problemi sempre più complessi.

In accordo alla teoria dell'autodeterminazione, il secondo bisogno psicologico fondamentale che abbiamo è quello di *relazione*: abbiamo bisogno di stare bene con gli altri, e quando ci accorgiamo di essere parte di relazioni significative stiamo bene con noi stesse e noi stessi. L'apprendimento e la risoluzione di problemi sono spesso attività sociali, e collaborare con altri, condividere idee e costruire soluzioni insieme possono soddisfare il nostro bisogno di relazione. I chatbot possono essere parte di questo processo, e diventare degli interlocutori efficaci. Quanto rimarranno solo degli strumenti e quanto invece diventeranno entità con cui sentirsi propriamente in relazione, perfino dei confidenti o degli amici (soprattutto se ci si presenteranno attraverso degli avatar attraenti...), è una questione ancora aperta.

Il terzo bisogno psicologico fondamentale che abbiamo è quello di *competenza*: abbiamo bisogno di sentirci efficaci nelle nostre azioni, e quando ci accorgiamo di saper fare cose significative stiamo bene con noi stesse e noi stessi. Anche se un chatbot può risolvere un problema, lo possiamo usare anche come un interlocutore che pone il problema in modo diverso, che suggerisce criticità nella nostra soluzione, che ci propone una spiegazione su qualche aspetto che non ci era chiaro, e così via. La condizione di essere noi stessi a gestire il processo di soluzione del problema ci dà una sensazione di realizzazione, che può essere potenziata, non ridotta, se mostriamo a noi stessi che per raggiungere il nostro obiettivo siamo anche stati capaci di usare appropriata-

mente degli strumenti tecnici, per esempio un chatbot. D'altra parte, se provare a noi stessi che siamo esperti in qualcosa è motivo di soddisfazione, quali saranno le competenze utili nel futuro, da imparare per sentirsi orgogliosi di possederle, è ancora una questione aperta, di cui qui abbiamo soprattutto discusso l'importanza.

Se state pensando che questi sono solo cenni, avete ragione: è proprio così. Come ho indicato all'inizio, i chatbot sono tra di noi da così poco tempo che non abbiamo avuto ancora la possibilità di trovare risposte davvero convincenti a domande così complesse e, penso, importanti. Mi auguro che grazie a questo corso potremo sperimentare e imparare insieme. Perché, come avete ben capito a questo punto, ritengo davvero plausibile che quello che sta succedendo non dovrebbe essere sottovalutato e ci chiederà invece di sperimentare e di imparare. A uno studente che si chiedeva che senso avesse frequentare le lezioni di un corso, invece che solo "studiare sui libri", una valida risposta è sempre stata che il docente gli avrebbe proposto un'analisi critica e una prospettiva sintetica sui contenuti del corso, degli esempi appropriati al contesto di apprendimento, e così via. Quanto di tutto ciò può essere ottenuto interagendo con un chatbot? Già oggi, forse, più di quello che molti si aspettano. E nel prossimo futuro?

E con questo torniamo alle domande da cui siamo partiti: in presenza dei chatbot, perché continua a essere utile imparare? E cosa e come imparare? Come abbiamo argomentato, non abbiamo ancora risposte precise, ma la posizione generale che vi ho proposto, in una prospettiva umanistica più che tecnica, si può riassumere



semplicemente così: *imparare ci rende consapevoli delle nostre responsabilità e ci consente di trovare un senso e quindi una motivazione in quello che facciamo.*

Saremo saggi se useremo i chatbot come aiutanti, e non per sostituirci, nella soluzione di problemi, e se approfitteremo della loro presenza per imparare a imparare sempre meglio.

*Vi auguro un buon lavoro*



## Capitolo 1

### **Di cosa parleremo, insomma**

*In un'epoca in cui sembra inevitabile prendere posizioni estreme, tra tecno-ottimisti e tecno-pessimisti, ha senso dedicare un po' di tempo a farsi domande e a godersi lo stupore delle novità che i sistemi di intelligenza artificiale generativa ci stanno prospettando.*

Siamo a un passo dal poter sostenere in modo credibile che, parafrasando il celebre titolo di un saggio geopolitico di Francis Fukuyama, l'intelligenza artificiale è *la fine della storia*. Con il suo avvento e il suo uso massiccio e massivo, nulla sarà più come prima. Secondo il fondatore di Microsoft, Bill Gates, l'intelligenza artificiale porterà profondi sconvolgimenti che toccheranno tutti, ma proprio tutti, tanto che perfino colossi del settore come il motore di ricerca Google e il titanico leader dell'e-commerce Amazon non saranno più gli stessi. Durante un recente evento organizzato da Goldman Sachs e SV Angel a San Francisco (ancora una volta San Francisco e la Silicon Valley sono il centro dello sviluppo tecnologico) su prospettive, potenzialità e criticità dell'intelligenza artificiale, Bill Gates ha infatti spiegato che l'impatto degli assistenti personali dotati di intelligenza artificiale sarà così profondo che la prima azienda a svilupparli avrà un vantaggio importante sui concorrenti.

Ah, dunque i sistemi di intelligenza artificiale sono degli “assistenti personali”? Degli aiutanti? O dei con-

sulenti? Secondo Gates, la rivoluzione dell'intelligenza artificiale può modificare radicalmente i comportamenti degli utenti, di tutti gli utenti: aziende, consumatori, studenti, docenti, sportivi, comunicatori, medici, pazienti, politici, amministratori pubblici, eccetera. «Sarà uno sconvolgimento epocale: chiunque realizzerà per primo un agente personale [dotato di intelligenza artificiale] farà sì che nessuno andrà più su un motore di ricerca, su un sito di produttività, su Amazon». La fine della storia, appunto? La fine di internet per come lo abbiamo conosciuto? La fine dei social network per come li abbiamo usati, e forse odiati, finora?

Questi “assistenti personali” fedeli, intelligenti e sempre disponibili – peraltro ancora da sviluppare nelle loro forme più evolute e pratiche – saranno in grado di comprendere le nostre esigenze e le nostre abitudini e ci aiuteranno a «leggere le cose che non hai tempo di leggere», ha detto Gates durante l'evento di San Francisco. E chi riuscirà, alla fine, a crearlo, questo assistente personale, artificiale ma intelligente (in un senso di “intelligenza” di cui discuteremo un poco nel seguito)? Sarà una startup oppure un gigante della tecnologia? O magari un'università o un centro di ricerca governativo? Ci vorrà del tempo prima che questo potente agente artificiale sia pronto per l'uso quotidiano, ha detto il fondatore di Microsoft. Fino ad allora, le aziende continueranno a incorporare nei loro prodotti strumenti di intelligenza artificiale, non necessariamente simili al popolare ChatGPT di OpenAI. Come peraltro, sempre più diffusamente, sta avvenendo già da tempo.

## **Contraddizioni**

Sì, perché se di intelligenza artificiale si parla tanto ora, da alcuni mesi, da quando dalla fine di novembre 2022 “ChatGPT” è un nome sulla bocca di (quasi) tutti, uno sguardo retrospettivo ci rivela che in realtà il tema è presente da decenni sui tavoli di studiosi, sviluppatori, manager, con momenti di maggiore interesse sociale alternati a periodi in cui l’argomento è rimasto una questione soltanto “per addetti ai lavori”. Da qualche mese di intelligenza artificiale si parla e si scrive davvero ampiamente. Con proclami e affermazioni contrastanti, se non opposte. Nulla sarà più come prima, appunto. Macché, la solita moda passeggera delle Big Tech e degli amanti delle novità tecnologiche. Il futuro della medicina. No, anzi, la fine del medico come professionista. Il perfetto collaboratore di tutti i lavoratori dei mass media. Ma figurati, sarà la fine dell’informazione e del giornalismo. I nostri figli e nipoti impareranno in condizioni che noi da giovani non avremmo mai potuto sperare, avendo sempre a disposizione un tutor personale. Oppure no: saranno sempre meno preparati perché useranno i chatbot per farsi scrivere i temi d’italiano e farsi risolvere i problemi di matematica. L’algoritmo che fa costruire la squadra perfetta. Che tristezza la fine del calcio dei campetti di periferia dove il talent scout ha scoperto Cristiano Ronaldo.

Spesso, inoltre, si sconfinava nell’apocalittico. I robot dotati di intelligenza artificiale si impadroniranno del mondo e noi saremo loro schiavi, o ci estingueremo se non serviremo più. E perciò: se non stacciamo la spina della corrente elettrica finisce tutto. E, se non stiamo attenti, l’intelligenza artificiale può scatenare la terza

guerra mondiale, magari nucleare, come se non bastassero già gli esseri umani a far danni in questo campo. Ma sai che l'intelligenza artificiale può provare sentimenti? Ma figurati, non diciamo sciocchezze, è proprio questo che manca alle macchine, al codice, ai robot: il cuore. Si può imparare tutto, ma non ad avere emozioni e ad amare. Ma ne siamo così sicuri?

Intanto il catalizzatore delle nostre attenzioni, delle nostre affermazioni e delle nostre preoccupazioni è arrivato, dalla fine di novembre 2022, con la possibilità di accedere, liberamente e senza necessità di competenze tecniche, a sofisticati sistemi di intelligenza artificiale generativa conversazionale, in breve “chatbot”, di cui ChatGPT di OpenAI è il più noto (ma altri sono stati resi nel frattempo disponibili, per esempio Claude di Anthropic e Bard di Google). Insomma, da qualche mese possiamo dialogare con... lui? lei? esso? essa? In poche parole, con sistemi di intelligenza artificiale accessibili via web o mediante applicazioni scaricabili sullo smartphone.

## **Dissonanze**

C'è qualcosa di dissonante in una parte delle riflessioni che abbiamo fatto, ascoltato, letto in questi mesi su questo tema. Da una parte, c'è un grande ed esplicito timore per i rischi che questi sistemi potrebbero produrre per la stabilità della nostra società, per la serenità delle nostre vite come le abbiamo vissute finora, e perfino per la sopravvivenza stessa della specie umana. E ritorna con questo la vocazione apocalittica, da fine della storia, tentazione spesso facile di fronte a novità dai contorni poco chiari, difficili da definire, soprattutto per chi, ma-

gari, non ha avuto la ventura, finora, di studiare codici o di riflettere sulle magnifiche (o no) sorti e progressive (o no) dell'informatica. Dall'altra parte, invece, è molto frequente la considerazione assertiva che ciò che chiamiamo "intelligenza artificiale" non è intelligente, né pensa, né tanto meno è cosciente, figuriamoci sensibile, una posizione generalmente trattata come un fatto così evidente da non richiedere giustificazione alcuna. Ma ne siamo così sicuri?

In sé, non c'è nulla di contraddittorio in questa opposizione: è del tutto possibile che un'entità non intelligente produca rischi esistenziali, come accade per esempio per gli armamenti atomici, e in tal caso la ragione basilare del rischio è nel cattivo uso che dell'entità in questione possono fare gli esseri umani, non nel comportamento autonomo che le macchine potrebbero avere. Questo si applica anche ai sistemi di intelligenza artificiale e ai chatbot in particolare: per esempio qualcuno potrebbe usarli per produrre fake news o qualcuno potrebbe delegare a loro decisioni critiche senza preoccuparsi della loro affidabilità. Ma i rischi che ci vengono prospettati sono anche, e a volte soprattutto, riferiti a situazioni in cui il pericolo si presenta perché i sistemi di intelligenza artificiale sono abilitati a prendere decisioni autonome (davvero questi sistemi sono capaci di prendere decisioni autonome?) e la razionalità di queste decisioni potrebbe essere contraria agli interessi, prima di tutto quelli legati alla sopravvivenza, della nostra specie.

Un esempio ovvio e nello stesso tempo estremo si presenterebbe se incaricassimo un sistema di intelligenza artificiale di risolvere il problema del cambiamento

climatico: con l'unico obiettivo di minimizzare i costi di realizzazione e senza altri vincoli, non è improbabile che la soluzione sarebbe di intervenire rimuovendo quella che potrebbe essere identificata come la causa prima del problema, e dunque eliminando almeno una parte della popolazione umana (davvero eliminando una parte della popolazione umana?). Benché contraria agli interessi di molti – speriamo di tutti – una soluzione di questo genere non sarebbe, in teoria, non-intelligente, e anzi la sua formulazione di dettaglio potrebbe richiedere capacità che, in un contesto diverso, non esiteremmo a considerare di grande e profonda intelligenza.

Insomma, la posizione secondo cui gli attuali sistemi di intelligenza artificiale sono stupidi, ma nello stesso tempo rischiosi, benché non contraddittoria in sé, sembra almeno in parte il frutto di precomprensioni, di pregiudizi non sempre e non reciprocamente consistenti: se fossero davvero soltanto strumenti stupidi al nostro servizio, non sarebbero abbastanza sofisticati per essere in grado di assumere e gestire autonomamente il controllo, e non avrebbe senso parlare di rischio esistenziale. Se ne abbiamo paura, sarà invece perché sono, o stanno diventando, *davvero* intelligenti?

Una via di uscita non infrequente da questa confusione comporta un'altra sintesi, che trasforma l'attuale complessità in un dilemma. Eccone un esempio, tratto dalla presentazione di un libro pubblicato recentemente<sup>1</sup>: «L'intelligenza artificiale è un soggetto che prenderà

---

<sup>1</sup> Fabio Ferrari, *L'intelligenza artificiale non esiste – Nessun senso salverà le macchine*, Il Sole 24 Ore, 2023.



il controllo delle nostre vite oppure è un oggetto al nostro servizio che dobbiamo imparare a usare nella maniera più corretta?». Questa domanda può essere riletta così: se è un *soggetto*, e quindi è, o almeno può essere, genuinamente intelligente, abbiamo un problema di controllo; se è un *oggetto*, e quindi non può essere genuinamente intelligente, dobbiamo soltanto imparare a usarlo bene. E ancora, questa alternativa si manifesta quando qualcuno sostiene che i chatbot sono diventati, o stanno diventando, sistemi di intelligenza artificiale generale (*Artificial General Intelligence*, AGI) e qualcun altro dichiara che non sono altro che “pappagalli statistici” (il termine inglese, *stochastic parrots*, compare nel titolo di un assai citato articolo<sup>2</sup> di un gruppo di esperte di etica dell’intelligenza artificiale), dunque soltanto capaci di ripetere comportamenti mostrati durante il loro addestramento, reperiti sulla rete, ovvero già elaborati, definiti e pubblicati da altri, da esseri intelligenti umani.

Per vedere quanto possa essere ideologica una simile contrapposizione, è sufficiente applicarla a individui di altre specie viventi: o siamo in competizione per “il controllo delle nostre vite” o sono soltanto oggetti, strumenti al nostro servizio (come peraltro sosteneva per esempio Cartesio: gli (altri) animali sono macchine/automi).

### Una chiave di lettura

In una situazione complessa, identificare le posizioni

---

<sup>2</sup> Emily M. Bender e altre, *On the dangers of stochastic parrots: can Language Models be too big?*, 2021 (<https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>).

estreme può essere utile perché delimita lo spazio delle opzioni, ma assumere che esistano, o comunque siano possibili, soltanto le posizioni estreme può portare a perdere di vista la struttura del fenomeno e le ragioni della sua stessa complessità. Esiste infatti un ampio spettro di posizioni intermedie, almeno 50 sfumature di chatbot che, senza implicare affatto che i sistemi di intelligenza artificiale siano identici a esseri umani, lasciano aperta la possibilità di farci scoprire che quello che l'evoluzione ha fatto in milioni di anni con il carbonio, trasformandoci progressivamente da oggetti in soggetti, forse lo stiamo sperimentando in pochi decenni con il silicio, tra microprocessori e internet.

Per ammettere questa ipotesi sono sufficienti due passi, di natura culturale e prima ancora psicologica. Con il primo passo ci si può liberare di un po' di antropocentrismo: di fronte a chi sostiene che quelli di una rete neurale artificiale, come ChatGPT, sono soltanto processi matematici, e quindi un chatbot "non può pensare", si può rispondere che quelli di una rete neurale naturale, come è il nostro cervello, sono soltanto processi elettrochimici, e quindi il nostro cervello "non può pensare". È evidente che lo scopo non è qui di cercare l'assurdo che stiamo pensando che non è vero che pensiamo (e qui recuperiamo, questa volta in positivo, Cartesio con il suo *cogito*), ma solo di mostrare che la posizione di essere a priori anti-riduzionisti per le entità naturali e riduzionisti per le entità artificiali è forse rassicurante per qualcuno, ma non è giustificata.

Insomma, potremmo riconoscere che è il momento di liberarci dal tribalismo di specie per cui molti di noi

finora hanno accettato come non problematica la asimmetria per cui «non so che relazione c'è tra il mio pensiero e i processi elettrochimici del mio cervello, ma so di pensare» e nello stesso tempo «non so che relazione c'è tra il comportamento di ChatGPT e i processi matematici della sua rete, ma so che non pensa». Ne siamo così sicuri?

Il secondo passo parte di nuovo dall'esperienza umana. Sappiamo bene che essere intelligente, e analogamente pensare, essere cosciente, e così via, non è una condizione binaria per cui o si è intelligenti o non lo si è. E lo stesso discorso, paradossalmente ancora più centrato, riguarda l'essere sensibili oppure no, provare emozioni oppure no. Come ha argomentato in particolare Howard Gardner nella sua teoria delle intelligenze multiple<sup>3</sup>, l'intelligenza ha dimensioni molteplici e di ognuna di queste, in momenti diversi della nostra vita e in funzione di tante variabili, possiamo essere più o meno dotati. Perché non potrebbe essere lo stesso anche per un'entità artificiale?

E, d'altra parte, in tutto ciò possiamo anche fare uno o a questo punto anche due passi indietro, e ritornare al punto di partenza. Per farci cogliere la portata potenzialmente rivoluzionaria di quello che stiamo vivendo e che forse sta progressivamente trasformando agenti artificiali-oggetti in entità cognitivamente aliene-soggetti, non c'è bisogno che siano coscienti, che pensino, o che siano

---

<sup>3</sup> Howard Gardner, *Formae mentis – Saggio sulla pluralità dell'intelligenza* (*Frames of mind: the theory of multiple intelligences*, 1983), Feltrinelli, 1987, 2002.

intelligenti, nel senso più o meno antropocentrico che volessimo attribuire a questi concetti. È sufficiente osservare che queste entità sanno operare con oggetti linguistici, che lo sanno fare con i linguaggi più complessi che abbiamo a disposizione, che sono le lingue storico-naturali come l'italiano, e che lo sanno fare in modo creativo, generativo, e non solo semplicemente copiando-e-incollando da testi preesistenti (alla possibile obiezione che un chatbot non opera davvero con testi, ma soltanto con numeri, abbiamo già proposto una risposta sopra: e perché allora il cervello umano, che è un'entità biologica, non linguistica, dovrebbe essere capace di operare con testi visto che opera grazie, non a numeri, ma a elementi chimici?). Certo, si può tranquillamente ammettere che, almeno per ora, la qualità letteraria dei testi prodotti dai sistemi di intelligenza artificiale che scrivono per essere utili, è ben inferiore a quella, diciamo, dei testi di Fëdor Michajlovič Dostoevskij. Ma questo è un altro discorso e ci torneremo più avanti. E poi, attenzione, Omero, quando non sonnecchiava, scriveva benissimo, eppure non era Omero. Nel senso che, a quanto pare, non è esistito un signore di nome Omero che ha scritto di suo spunto e pugno *l'Iliade* e *l'Odissea*, dato che queste opere sono frutto della raffinata raccolta (spontanea?) di racconti orali sparsi per la rete dei cantastorie greci. Ovvero, queste sontuose opere dell'ingegno umano sono anche il risultato compilativo del primo sistema di intelligenza artificiale – perché risultato di una costruzione sociale – generativa e letteraria della storia. Che prospettiva...

Insomma, per la prima volta nell'evoluzione umana possiamo avere conversazioni in italiano (e in inglese, e

in tante altre lingue, tra cui vari nostri dialetti) con entità che non sono della nostra specie, e sappiamo quanto la nostra società e la nostra cultura dipendano dalla parola. Anche soltanto questo fatto potrebbe suggerirci di essere osservatori e magari sperimentatori, aperti a comprendere le novità di quello che sta accadendo attorno ai chatbot. Ecco, che cosa sta accadendo? Di questo esplorano le pagine che seguono e che vogliono provare a raccontare l'inizio di una nuova storia.



## Capitolo 2

### **«Chatbot, che ore sono?», ovvero cominciare a capire l'intelligenza artificiale con una domanda e tre risposte**

*Le risposte a una semplice domanda ci possono dare interessanti informazioni su quelle peculiari entità aliene che sono gli attuali chatbot.*

Nelle nostre relazioni con altri esseri umani osserviamo il loro comportamento, verbale e no, e non i loro pensieri, intenzioni, desideri, e d'altra parte non consideriamo necessario “aprire la loro scatola” (cranica) per cercare di capire le ragioni di quello che abbiamo osservato. L'opzione di “guardare all'interno della scatola”, per esempio facendo una TAC al cervello, si prospetta soltanto nel caso di malfunzionamento, insomma per una malattia. Analogo è il modo con cui ci poniamo verso gli strumenti che ci circondano, per esempio le automobili. Fino a che funzionano e non manifestano anomalie, come quando si accende una spia, non c'è bisogno di aprire il cofano. Le guidiamo e basta. Le usiamo e basta. Questa strategia è del resto molto ben giustificata: è una difesa che abbiamo imparato ad adottare per proteggerci dalle molteplici complessità che ci circondano, quelle dei cervelli dei nostri simili, dei motori delle automobili e dei tanti altri sistemi complessi intorno a noi.

“Lasciare la scatola chiusa” – quello che gli ingegneri chiamano “black box” – non implica però essere superfi-

ciali, anzi: perché è dal comportamento di un'entità che sappiamo inferire, cogliere e definire tante sue caratteristiche. Tutto ciò si può applicare anche a quelle entità di straordinaria complessità che sono gli attuali sistemi di intelligenza artificiale conversazionale, i cosiddetti “chatbot”.

### **Domande “a scatola chiusa”**

Mentre qualcuno sta cercando di comprenderne il comportamento osservando le loro connessioni neurali artificiali (usando altre reti neurali per cercare di interpretare i miliardi di numeri che si trova di fronte), chi dialoga con un chatbot opera appunto “a scatola chiusa”, per ricevere un supporto nella soluzione di qualche problema oppure proprio per cercare di capire che cosa sta succedendo. E infatti le risposte anche a una semplice domanda come «che ore sono?» ci possono dare interessanti informazioni su queste peculiari entità “aliene” che sono gli attuali chatbot, pur con la consapevolezza che i loro tempi di evoluzione sono attualmente delle settimane, non dei millenni, come per gli esseri umani, e quindi quello che è valido oggi potrebbe non esserlo più domani.

Il senso di questa domanda, pur così ovvia, è che sappiamo che ChatGPT, il sistema da cui i cambiamenti a cui stiamo assistendo sono cominciati, ha nel momento in cui scriviamo (dicembre 2023) un *knowledge cutoff*, cioè un limite all'informazione con cui è stato addestrato, fermo alla primavera 2023: è stato pre-addestrato (la “P” di GPT sta proprio per *Pre-trained*, pre-addestrato, appunto) facendogli leggere un gran numero di testi, ma



questo processo è stato interrotto (dagli sviluppatori) proprio nella primavera 2023, e da allora la memoria a lungo termine di GPT è stata resa, in pratica, statica.

Infatti, come è evidente a chiunque abbia avuto un po' di conversazioni con questo sistema, ChatGPT mantiene i contenuti del dialogo in corso soltanto nella sua memoria a breve termine, che non trasferisce nella memoria a lungo termine come invece facciamo noi esseri umani durante il sonno. In sostanza, non impara dal dialogo con noi e non fa altro che adattare temporaneamente il suo comportamento ai contenuti prodotti durante il dialogo, per poi dimenticarli quando il dialogo si conclude: quel dialogo è fine a se stesso, potremmo dire, e non è la base di altri dialoghi con noi stessi o altri esseri (umani o no).

Così che non sa rispondere a una domanda semplice come «Quanti dialoghi abbiamo fatto insieme nell'ultima settimana?».

Possiamo dunque provare a capire qualcosa in più su questi sistemi non come farebbe un neurologo, con una TAC, ma comportandoci da psichiatri, ovvero dialogando con essi e traendo ipotesi dalle risposte: in mancanza di modelli consolidati, è un modo concreto, a disposizione di chiunque, per cercare di cominciare a capire, definire, spiegare.

### **Un esperimento**

Questo è il *prompt* (tradotto in inglese, per i chatbot senza una sufficiente padronanza dell'italiano) che abbiamo dato a vari chatbot: «Vorrei sapere che ore sono adesso in Italia, e, se lo sai, vorrei che mi spiegassi come fai a saperlo».

Riassumendo, le risposte che abbiamo ottenuto si possono classificare in tre categorie di comportamento, di ognuna delle quali indichiamo i chatbot che hanno avuto, nel momento in cui abbiamo fatto questo esperimento (maggio 2023), quel comportamento:

- informazione non disponibile (per esempio: «Mi dispiace, ma come modello di linguaggio basato su GPT-3.5, non ho accesso in tempo reale alle informazioni sulle attuali condizioni di data e ora. La mia formazione si è conclusa nel settembre 2021<sup>1</sup>, quindi non sono a conoscenza degli eventi successivi a quella data, comprese le attuali informazioni sulla data e sull'ora in Italia»): ChatGPT, basato sia su GPT-3.5 sia su GPT-4; Claude; Inflection PI;
- informazione sbagliata (per esempio: «In Italia adesso sono le 12:12 del mattino. Quanto a come faccio a saperlo, uso un *tool* chiamato “browser” che mi permette di cercare informazioni online»): ChatGPT, basato su GPT-4 con web browsing attivo; Character Assistant (non è vero che fossero le 12:12 del mattino quando abbiamo fatto questa domanda!);
- informazione corretta (per esempio: «Adesso in Italia sono le 1:57:29 PM CEST di giovedì 18 maggio 2023. Per rispondere alla tua seconda domanda, ottengo queste informazioni utilizzando Wolfram Alpha, un motore

---

<sup>1</sup> Come abbiamo indicato in precedenza, nel momento in cui stiamo scrivendo (dicembre 2023) questo limite è stato spostato alla primavera 2023. È del tutto plausibile che altre fasi di addestramento saranno compiute prossimamente, e quindi questo *knowledge cutoff* ancora spostato in avanti.

di conoscenza computazionale»): ChatGPT, basato su GPT-4 con plugin “Wolfram” attivo; Bard; Bing Chat.

Questo semplice esperimento ci consente di mettere in una prospettiva critica quanto si dice abitualmente a proposito della limitata affidabilità dei chatbot, almeno relativamente a due punti. Primo, i chatbot in quanto tali non hanno accesso al mondo empirico, e quindi quando riportano correttamente che ore sono in un certo momento e in un certo luogo è perché delegano l’acquisizione dell’informazione ad altri sistemi, che siano motori computazionali connessi alla rete, come Wolfram Alpha, o motori di ricerca come Google e Bing, in questo non c’è molta differenza.

Certi sistemi, come Bard e Bing Chat (che Microsoft sta rinominando “Copilot”), presentano in modo integrato le funzionalità di dialogo e di ricerca sul web, capacità che però rimangono concettualmente e operativamente ben distinte: un conto è saper ragionare, eventualmente anche su informazioni false o incomplete, un altro è saper riportare un’informazione corretta recuperata dal web. Con la conseguenza che non ci si dovrebbe stupire se lo stesso sistema capace di argomentare in modo sofisticato si inventa poi delle fonti. Secondo, in certi casi, invece di produrre informazioni false, i chatbot rispondono di non essere in grado di soddisfare le nostre richieste, un comportamento che in un essere umano potrebbe essere considerato un segno di consapevolezza dei limiti della propria conoscenza.

Un’evoluzione che conduca i chatbot a una loro maggiore propensione e abilità nel riconoscere le proprie

capacità effettive sarebbe certamente di grande importanza, nelle applicazioni e nella comprensione della dimensione etica e giuridica di questi sistemi.

In uno scenario così incerto e che cambia così rapidamente, dichiarare che i chatbot sono stupidi “pappagalli statistici” pare sempre meno produttivo: ben utile sarà trovare metodi e modelli per interpretarne il comportamento e con questi progettare scenari di convivenza che possano arricchire, e non impoverire, la nostra esperienza di esseri umani. Sì, ma che cos'è dunque ChatGPT?

## Capitolo 3

### **ChatGPT: uno strumento, un collega, o un consulente?**

*La prospettiva dei chatbot come consulenti ci consente di interpretare qualche aspetto dei cambiamenti in corso come un nuovo scenario, in cui sono diventati ampiamente e facilmente disponibili consulenti anche a chi non se li è potuti permettere finora, come molti docenti di scuola e molti liberi professionisti.*

I primi effetti della diffusione degli attuali sistemi di intelligenza artificiale conversazionale – i cosiddetti “chatbot” – ci lasciano intravedere cambiamenti radicali che, come ci ha ricordato Bill Gates, potrebbero generarsi in molteplici dimensioni della nostra società e della nostra vita. Il fatto che queste siano entità artificiali è destabilizzante, certo, considerando che fino a pochi mesi fa avevamo esperienza di interazioni cognitivamente sofisticate soltanto con individui della nostra specie. Ma tra destabilizzazione e fine della storia o della specie umana, per fortuna, ci sono ancora alcune profonde differenze, sfumature, possibilità di governo e sviluppo delle novità.

#### **La terza transizione**

È ancora troppo presto per trarre conclusioni, ma potremmo cominciare a interpretare quello che sta succedendo come una possibile terza transizione verso un non-antropocentrismo.

Se nel passato abbiamo ritenuto il nostro pianeta e

noi stessi al centro del cosmo, Niccolò Copernico e Charles Darwin ci hanno suggerito una maggiore modestia: la Terra non è ciò intorno a cui tutti gli altri corpi celesti orbitano e Homo Sapiens, dunque noi stessi, non è la specie al cui servizio tutte le altre sono state create, così come il nostro pianeta e gli altri corpi celesti non sono qui per noi.

Ma se ci siamo sprovvincializzati per quanto riguarda la *cosmologia* e la *biologia*, abbiamo continuato a considerarci peculiari nella dimensione *cognitiva*. E, tutto sommato, a buon diritto, considerando che non siamo a conoscenza di entità naturali con capacità di ragionamento e comunicazione paragonabili alle nostre, per lo meno da quando si sono estinte le altre specie di ominidi. Ripareremo in seguito di questo delicato argomento e lo svilupperemo un poco più ampiamente, ma introduciamo da subito la nostra ipotesi: ChatGPT e i suoi fratelli ci stanno segnalando che forse *non siamo più soli*. Certo, i chatbot, come tutti i sistemi di intelligenza artificiale, sono il prodotto dell'ingegno umano, non dell'evoluzione naturale, ma l'argomento dell'origine non è una discriminante assoluta, anzi, come fanno bene i genitori e i docenti che vedono alcuni loro figli e alcuni loro allievi superare i maestri. E del resto è ormai un fatto che per esempio i programmi per giocare a scacchi sono in questo più abili degli esseri umani che li hanno sviluppati. Di fronte a questi cambiamenti, che si stanno prospettando nello stesso tempo fondamentali e radicali, cercare prima di tutto di capire quello che sta succedendo è ben giustificato, dunque.

## **Analogie**

E in questo le analogie, pur rischiando di essere a volte fuorvianti, mantengono una qualche utilità, quantomeno per consentirci di ovviare un poco ai limiti del nostro linguaggio, che in presenza di cambiamenti rapidi potrebbe semplicemente non avere le parole per consentirci di descrivere in modo specifico le cose nuove che tanto ci preoccupano fino a lasciarci, appunto, senza parole.

L'informatica ci ha abituato a questo atteggiamento. Per fare un esempio ormai non controverso, usiamo il termine “memoria” – «la capacità, comune a molti organismi, di conservare traccia più o meno completa e duratura degli stimoli esterni sperimentati e delle relative risposte», secondo il Vocabolario Treccani<sup>1</sup> – anche per riferirci ai dispositivi tecnologici per la conservazione dei dati, ai nostri computer o hard disk esterni.

Dunque qualche approssimazione semantica nel nostro lessico è in pratica inevitabile anche quando parliamo di chatbot: è sufficiente esserne consapevoli e adottare un'attitudine di tipo funzionale. Edsger Dijkstra, uno dei fondatori dell'informatica, lo chiarì in modo eccellente<sup>2</sup>, ponendo la domanda: «I sottomarini nuotano?». No, perché non si spostano nell'acqua muovendo pinne o braccia e gambe, ma funzionalmente si comportano come se nuotassero. Prima dell'invenzione dei sottomarini, “capacità di muoversi nell'acqua” e “nuotare” erano plausibilmente considerati sinonimi; oggi sostenere “il sottomarino nuota”, invece di “il sottomarino naviga”, è

---

<sup>1</sup> <https://www.treccani.it/vocabolario/memoria>.

<sup>2</sup> <https://www.cs.utexas.edu/users/EWD/ewd08xx/EWD898.PDF>.

strano ma rimane comprensibile, proprio perché capiamo che ci si sta riferendo alla sua funzione, e non a come la si realizza.

Con la sua domanda, posta nel 1984 e quindi ben prima della diffusione degli attuali chatbot, Dijkstra voleva mostrare l'inutilità pratica della domanda se “le macchine pensano”: no, se “pensare” significa “pensare come faccio io” (detto ognuno per sé, dato che non sappiamo come gli altri pensino... immaginiamo che lo facciano più o meno come noi, ma non lo sappiamo per certo); forse sì, sempre di più sì, se l'oggetto effettivo della domanda è la funzione. Insomma, se la nostra attenzione è al comportamento che si osserva, a quello che accade “fuori dalla scatola”, le macchine si comportano come se pensassero; se invece è alle cause di quel comportamento, a quello che accade “dentro la scatola”, e dunque in particolare la scatola che è il nostro cranio, allora boh... sappiamo a mala pena come pensiamo noi stessi, figuriamoci se sappiamo come pensano gli altri, esseri umani o no.

Ed è plausibilmente questa stessa attitudine che troviamo così descritta nel libro<sup>3</sup> che è diventato il riferimento standard sul nostro tema: «La maggioranza dei ricercatori di IA [...] non si preoccupa se la si considera una simulazione di intelligenza o vera intelligenza» (una breve nota, per ridurre il rischio che questa posizione appaia come grezzamente anticulturale; potremmo

---

<sup>3</sup> Stuart Russell e Peter Norvig, *Intelligenza Artificiale – Un approccio moderno*. La quarta edizione è stata pubblicata in traduzione italiana da Pearson nel 2021.



decidere che un certo concetto X, in questo caso “essere intelligente”, meriti di essere definito con precisione quando compare in condizioni “se è X, allora Y”, dove Y è qualcosa di socialmente importante; in questo senso, sarà opportuno preoccuparsi se è «una simulazione di intelligenza o vera intelligenza» quando avremo casi in cui decidere se è vera intelligenza produrrà dei cambiamenti rilevanti).

Con modalità spesso peculiari – e diciamo pure: “aliene” – i chatbot si comportano in modo analogo rispetto a quello che fanno gli esseri umani quando pensiamo che pensino. Si comportano “come se” pensassero. E a volte e in effetti sempre più spesso come se pensassero come noi. Il Vocabolario Treccani oggi ammette che “memoria” significhi anche «nell’elaborazione elettronica dei dati, organo avente la funzione di registrare e conservare informazioni», mentre non offre un’opzione analoga a proposito di “pensiero”. Lasciando ad altri la decisione se l’accezione di “pensare come fa un chatbot” dovrebbe essere inclusa nei vocabolari oppure no, ci accontentiamo di solito di usare per analogia il lessico di cui disponiamo, nella consapevolezza che con l’uso di termini antropomorfici ci riferiamo dunque alla funzione e non alla natura (naturale o artificiale, basata sul carbonio o sul silicio, eccetera) delle entità.

Questa precisazione ci consente una certa libertà linguistica nelle nostre esplorazioni, per cercare di capire quello che davvero sta succedendo. E dato che il dialogo – ciò che sanno fare i chatbot – è essenzialmente relazione, è a proposito della nostra relazione con un chat-

bot che ha senso prima di tutto chiedersi che genere di entità sia. Che cos'è (lui? lei? esso? essa?) per me?

### **Chatbot: strumenti, colleghi o consulenti?**

Le opzioni possono essere molteplici e ne consideriamo qui soltanto tre (non ci chiediamo, per esempio, se un chatbot possa essere o diventare un nostro amico, almeno non per ora, pur riconoscendo che questa è un'opzione che si potrebbe dimostrare attraente, soprattutto per gli adolescenti): un chatbot come ChatGPT è più uno strumento, un collega o un consulente?

Posto che potremmo anche ammettere che è ancora troppo presto per avere le idee chiare, la nostra "relazione" è in fondo iniziata da poco, da qualche mese, azzardiamo comunque una semplice riflessione su questo interrogativo.

Progettiamo strumenti in modo che siano particolarmente adatti a specifiche funzioni, a rispondere a esigenze particolari, tanto che ne disponiamo di molteplici per funzioni ed esigenze diverse: i cacciavite non sono pensati per piantare chiodi e il proverbiale coltellino svizzero è sì uno strumento multi-funzione ma, oltre a essere più un'eccezione che una regola, è comunque soltanto un assemblaggio di strumenti più semplici, ognuno dei quali risponde a specifiche prestazioni e dunque ha funzionalità e scopi differenti.

Certo, anche ogni dialogo ha una dimensione funzionale – dialoghiamo per imparare, raccontare, piacere, persuadere, condividere stati d'animo... – e nel caso dei chatbot il dialogo può essere sfruttato come strumento per giungere progressivamente a ottenere l'informazio-

ne che cerchiamo, ponendo domande sempre più specifiche, con una tecnica che – un altro antropomorfismo – è chiamata *chain of thought*, “catena di pensiero”. Ma la capacità di dialogare è troppo fondativa per l’esperienza umana per essere considerata soltanto una funzione, la risposta specifica a un’esigenza specifica. Per queste ragioni, nonostante la sua natura artificiale, un chatbot non è così ovviamente assimilabile a uno strumento.

Che dire dei chatbot come colleghi? Prescindendo dalla dimensione di confidenza che a volte si instaura tra persone che lavorano quotidianamente e regolarmente e spesso per tanto tempo insieme – e che, peraltro, si inizia a osservare anche fra persone e chatbot –, con un collega tipicamente condividiamo compiti e responsabilità. E, almeno per ora, con un chatbot non condividiamo né compiti né responsabilità.

Nei dialoghi che possiamo avere con ChatGPT o con uno dei suoi fratelli c’è generalmente una chiara asimmetria: anche quando noi gli chiediamo di avere un ruolo attivo, in qualche modo autonomo – «Fammi cinque domande creative sull’argomento di cui stiamo parlando!» – siamo sempre noi a gestire la relazione, in accordo ai nostri scopi e ai nostri interessi. Al contrario di un collega, i chatbot al momento non hanno, insomma, una vera autonomia.

Per essere chiari, non c’è nulla di necessario in questo, e anzi una delle tecniche per addestrare le reti neurali, di cui i chatbot sono esempi, è proprio di renderle autonome in modo da poterle fare interagire tra loro, per addestrarsi a vicenda o anche per competere (in questo caso le si chiamano *generative adversarial networks*,

“reti generative avversarie” appunto). Insomma, l'attuale non-autonomia dei chatbot deriva non tanto da una questione tecnologica, ma da una cautela etica.

La vicenda di AlphaGo è piuttosto illuminante al riguardo. Sviluppato dall'azienda inglese DeepMind Technologies, AlphaGo è un sistema software di intelligenza artificiale addestrato per giocare a Go, un gioco da tavolo ancora più complesso degli scacchi. Se nel 2016 AlphaGo era stato capace di battere vari campioni umani, nel 2017 una sua versione successiva, AlphaGo Zero, progettata per imparare giocando contro se stessa e quindi capace di operare in modo autonomo, ha vinto 100 partite a 0 contro il suo “antenato” AlphaGo, dopo qualche giorno di addestramento realizzato in modo, appunto, autonomo.

Finché si tratta di programmi per vincere giochi da tavolo, tutto sommato ci sposta poco che siamo noi a gestirne l'addestramento o che impari da solo, e se la seconda opzione si rivela, come nell'esempio a cui abbiamo accennato, più efficiente, bene così. Ma nel caso di entità capaci di gestire il linguaggio naturale come i chatbot la posta in gioco è ben diversa. La questione della possibile progressiva autonomia di questi sistemi, anche sollecitata dalla facilità con cui oggi intervenire sul comportamento dei chatbot con poche linee di codice, potrebbe modificare tutto ciò, ponendo(c) in modo ancora più drammatico anche la questione della responsabilità, di cui per la sua importanza discuteremo ancora, ma per adesso il controllo rimane nelle nostre mani, e non è certo quello che accade nella relazione con un collega.

## **Un consulente, dunque**

Per tutto quanto considerato finora, ci sembra invece più appropriata la terza analogia che abbiamo suggerito: i chatbot come consulenti, a cui deleghiamo certi compiti, a volte puramente esecutivi, a volte strategici, ogni tanto ludici, pur mantenendo noi la gestione del processo e la responsabilità ultima dei suoi risultati, più o meno seri, più o meno pratici. Come si comprende dialogando con essi in modo sufficientemente approfondito e critico, i chatbot sono consulenti più abili a elaborare testi e proporre opinioni che a fornire informazioni fattuali. In altri termini, almeno per ora non è davvero una buona idea trattare un chatbot come un sostituto di un motore di ricerca (e quanto fatto finora per accoppiare un chatbot e un motore di ricerca – in particolare con ChatGPT e certi plugin, Bing Chat, e Bard – non sta sempre producendo risultati della qualità attesa): questo ribadisce, tra l'altro, l'importanza del controllo dell'affidabilità di quanto prodotto dai chatbot, a cui almeno per ora non possiamo scaricare la colpa di eventuali errori né organizzativamente né giuridicamente (né, se siamo onesti con noi stessi, psicologicamente).

La prospettiva dei chatbot come consulenti ci consente di interpretare qualche aspetto dei cambiamenti in corso come un nuovo scenario, in cui sono diventati ampiamente e facilmente disponibili consulenti anche a chi non se li era potuti permettere finora, come molti docenti di scuola e molti liberi professionisti. Per costoro, un chatbot può diventare un interlocutore disponibile a tempo pieno e al quale potersi rivolgere liberamente per chiedere consigli, per produrre bozze di nuovi testi o

sintesi e revisioni di testi esistenti, per preparare incontri di lavoro simulandoli, e così via. Saranno i prossimi sviluppi a dirci se i chatbot stiano davvero diventando degli assistenti/consulenti personali oppure altro.

## Capitolo 4

### **Così Dostoevskij ci aiuta a capire ChatGPT**

*Posto che anche gli esseri umani commettono errori, perché, generalmente, non consideriamo i nostri errori come derivanti da “bug” e non ci impegniamo a fare “debugging” del comportamento nostro e dei nostri simili?*

Si dice che si può capire qualcosa del carattere e delle capacità di una persona non solo dalle cose che sa fare e che fa bene, ma anche e qualche volta forse soprattutto dal tipo di errori (di lessico o di grammatica, fattuali, di ragionamento) che commette. Questa strategia ci è già stata utile, nelle pagine precedenti, per cominciare a comprendere alcune caratteristiche basilari dei chatbot, studiandone il comportamento e in particolare proprio gli errori, in risposta alla domanda «che ore sono?». Proseguiamo allora qui questa esplorazione in modo un poco più sistematico.

Che cos'è un “bug”? Con questo termine inglese, ma ormai internazionale, ci riferiamo alle cause dei comportamenti indesiderati dei sistemi di software, ovvero di quei comportamenti che producono errori che cerchiamo di correggere con un'attività che di conseguenza chiamiamo “debugging”, cioè appunto eliminazione dei bug. Ora, posto che anche noi esseri umani commettiamo continuamente errori, perché, generalmente, non consideriamo i nostri errori come derivanti da bug, qualcosa che malfunziona nel nostro, diciamo, softwa-

re, e non ci impegniamo a *fare debugging* del comportamento nostro e dei nostri simili? Sbagliando s'impara, si dice. Davvero?

Una ragione che qualcuno potrebbe proporre per spiegare perché non consideriamo frutto di un bug un nostro errore e perché, di conseguenza, non siamo soliti fare debugging di noi stessi è che distinguiamo errori umani e bug semplicemente per “tenere le distanze” con i sistemi di software. Come fosse: i bug sono cose “da macchine” e noi non siamo macchine, vero? Ma questa ipotesi non sembra cogliere nel segno. Abbiamo da sempre la tendenza a trattare come umani i sistemi con cui interagiamo e non solo quando parliamo con Siri o con Alexa. Di fronte all'esecuzione lenta di un programma, per esempio, diciamo che «ci sta pensando» o che «fa fatica». E in particolare «parliamo» con quei sistemi con cui possiamo interagire in un linguaggio ordinario, come l'italiano o l'inglese. Se ci rapportiamo a questi sistemi attribuendo loro caratteristiche tipicamente umane, certamente è perché evidentemente non siamo interessati ad adottare un'abitudine linguistica discriminatoria nei loro confronti.

Inoltre, e questo è il punto che ci interessa discutere qui, ci sono situazioni in cui ammettiamo che anche gli errori commessi da sistemi di software non sono causati da bug, come accade nel caso degli errori che osserviamo nei chatbot, per i quali infatti ci si sta abituando ad applicare un termine diverso e non sorprendentemente antropomorfo: “allucinazioni” (termine che per altro non ci sembra così ben scelto, dato che lascerebbe intendere che quando non sono allucinazioni le risposte



dei chatbot sono correttamente riferite al “mondo reale”, cosa quantomeno discutibile, considerando che i chatbot parlano di cose che non hanno mai visto, non avendo, almeno per ora, occhi).

### **Gli errori dei chatbot**

I chatbot sono evidentemente sistemi di software e commettono talvolta anche errori banali, ma noi, in caso di errore, assimiliamo il loro comportamento non a quello dei programmi con cui scriviamo, comunichiamo oppure gestiamo variamente dati, ma a quello degli esseri umani. Insomma, un errore di calcolo prodotto da un foglio elettronico è un bug, ma lo stesso errore prodotto da un chatbot non lo è. Perché?

La nostra ipotesi è che questa domanda scaturisca dall’aver colto, benché spesso forse solo implicitamente, una caratteristica essenziale dei chatbot. E dunque che una risposta appropriata a questa domanda ci aiuti a comprendere meglio tale caratteristica – forse l’unica che è veramente importante avere chiara di quello che sta “dentro la scatola” – propria di quei particolari sistemi di intelligenza artificiale che sono le reti neurali artificiali, di cui ChatGPT è un esempio. Assumendo che questa ipotesi abbia un qualche fondamento, una risposta ci avvicinerebbe dunque anche a intuire almeno una ragione della rivoluzionaria novità culturale che l’esistenza dei chatbot ci sta prospettando in questi mesi, in questi giorni, in queste ore. La caratteristica è che il comportamento dei chatbot, e dunque la capacità di dialogare che essi mostrano, è il risultato non di una *programmazione* ma di un *addestramento*. E nella differen-

za tra i due concetti e tra le due operazioni sta il cuore dell'intelligenza artificiale, per come la intendiamo oggi, rispetto a qualunque altro sistema informatico e a qualunque altra innovazione tecnologica. La sottolineatura della grande differenza tra programmazione e addestramento ci permette anche di aprire una prospettiva esplicitamente umanistica alla discussione sull'intelligenza artificiale e sulle conseguenze del suo avvento, sviluppo, successo.

### **Sistemi programmati**

Le strategie che abbiamo sviluppato nel corso di vari decenni per costruire sistemi di software si basano su una semplice idea, che era già esplicitamente presente negli anni Cinquanta del secolo scorso, agli albori dell'informatica. Avendo identificato un comportamento che vorremmo ottenere, per prima cosa lo scomponiamo in una successione di istruzioni elementari, sufficientemente semplici da poter essere comprese senza ambiguità: è ciò che chiamiamo un algoritmo. Quindi traduciamo queste istruzioni in una forma che sia eseguibile da un sistema digitale, con ciò implementando l'algoritmo in un linguaggio di programmazione e producendo così un programma.

Se questa strategia è stata applicata correttamente, il programma è effettivamente eseguibile dal sistema digitale prescelto e la sua esecuzione produce il comportamento voluto, la funzionalità desiderata in risposta all'esigenza specifica avanzata. Un sistema progettato in accordo a questa logica ha dunque un comportamento che deriva da una successione finita di passi determina-

ti da una successione finita di istruzioni, decise da noi. In questo modo, otteniamo il corrispondente di una catena di montaggio per la produzione di un risultato, in accordo con la prospettiva organizzativa di impostazione tayloristica dominante attorno alla metà del secolo scorso, proprio quando i calcolatori elettronici furono introdotti (non osiamo concluderne che ci sia stata una contaminazione, ma è comunque una coincidenza su cui sarebbe interessante fare qualche riflessione...).

Sebbene da decenni sappiamo che di principio esistono anche altre strategie per produrre comportamenti da parte di sistemi software, per esempio nella forma di algoritmi genetici e, appunto, di reti neurali artificiali, finora questo sapere era stato riservato a non molti “addetti ai lavori”: ChatGPT ha reso evidente l’esistenza di una potente strategia alternativa. Quale?

### **ChatGPT non è un sistema programmato**

ChatGPT è un sistema di software, sì, ma il suo comportamento non deriva dall’esecuzione di un programma, come erroneamente si tende a credere o a immaginare. Insomma, il funzionamento, più o meno efficace, di un chatbot basato su una rete neurale, come è ChatGPT, non è frutto di una successione di istruzioni dichiarate esplicitamente e inserite da noi una volta per tutte. Dato che, con risultati più o meno validi che tutti noi possiamo sperimentare facilmente, i chatbot attuali sono in grado di dialogare praticamente a proposito di qualsiasi cosa, non è immaginabile che le loro risposte siano generate da algoritmi costituiti da istruzioni della forma «se, nel contesto del dialogo C, ricevi la domanda X, rispondi Y», come se

alla base delle loro risposte ci fosse una gigantesca tabella a tre colonne, C, X, e Y, in cui C e X sono gli input (dove il contesto del dialogo C è ciò che a volte è considerata la memoria a breve termine del chatbot, chiamata in inglese *context window*) e Y è l'output. Tutto ciò, infatti, comporterebbe l'immissione di una mole di informazioni e di variabili da connettere tra loro non alla portata di mano umana. E infatti non è così.

I chatbot non sono strumenti di ricerca in database testuali che contengono, scritte anticipatamente, tutte le possibili triple  $(C, X, Y) = (\text{contesto}, \text{domanda}, \text{risposta})$ : non sono insomma realizzazioni di una borghesiana *Biblioteca di Babele*<sup>1</sup>. Che cosa sono invece? Se non sono programmati, come funzionano ChatGPT e i suoi fratelli? Cosa s'intende dunque quando si dice che i chatbot basati su reti neurali sono "addestrati" e non "programmati"?

### **Esseri umani machine-like e human-like**

Per meglio chiarire la questione, ritorniamo giusto per un momento a riflettere sul nostro comportamento di esseri umani. Generalmente, ci consideriamo liberi di agire: non assumiamo di essere programmati da qualcuno o da qualcosa per eseguire determinati compiti, e quindi ci consideriamo autonomi nei comportamenti che abbiamo, nelle decisioni che prendiamo, nei gesti che compiamo, e dunque responsabili per le relative conseguenze. Tuttavia sono immaginabili condizioni di eteronomia o di ridotta autonomia più o meno marcata. Un essere uma-

---

<sup>1</sup>[https://it.wikipedia.org/wiki/La\\_biblioteca\\_di\\_Babele](https://it.wikipedia.org/wiki/La_biblioteca_di_Babele).

no che introietta una tabella di istruzioni e si comporta di conseguenza opera come un sistema programmato, tanto che dal punto di vista di chi è responsabile di una tale tabella di istruzioni non sarebbe né assurdo né scorretto trattare ogni deviazione dal comportamento atteso di quell'individuo proprio come un bug, un difetto nella sua programmazione e dunque da correggere come tale. In effetti, lo spettro dei comportamenti programmati è estremamente ampio, dato che tutti noi agiamo secondo modelli di interazione selettivi, interiorizzati e sviluppati più o meno coscientemente: andiamo al lavoro seguendo procedure prefissate, gestiamo gli strumenti a nostra disposizione secondo le istruzioni d'uso e manteniamo a diversi livelli le relazioni umane non troppo significative seguendo procedure sociali stabili. E non è il caso di scomodare concetti come "struttura", "sovrastruttura", condizionamenti culturali e no: insomma, ci siamo capiti.

I casi limite, in questo spettro, sono a volte oggetto di grande letteratura, in cui troviamo descritta in modo drammatico la situazione di persone che delegano la loro responsabilità e soltanto obbediscono agli ordini. A proposito della profondità, sociologica prima ancora che psicologica, di questa condizione, è insuperabile la riflessione di Dostoevskij nel capitolo *Il grande inquisitore* del romanzo *I fratelli Karamazov*. Al di là della pure affascinante vicenda narrativa, la questione posta è che un problema fondamentale per molti esseri umani sarebbe l'incapacità di sopportare di dover gestire una propria responsabilità individuale, con la conseguenza che compito primario di chi opera con un ruolo di potere nella

società è di farsi carico della responsabilità altrui. Come fosse, «cedi a noi la tua libertà e noi ti risolveremo il problema di doverti preoccupare di essere responsabile delle tue azioni: da quel momento dovrai solo obbedire agli ordini e non sarai più personalmente responsabile di nulla». Anche a questo si sarà ispirato George Orwell immaginando la società distopica di *1984*? E poi ricordiamo, questa volta a partire da esperienze purtroppo non letterarie, che una fine analisi sulle situazioni che si generano nelle relazioni con i poteri che impongono la sola obbedienza agli ordini fu sviluppata da Hannah Arendt nel saggio *La banalità del male*: assistendo al processo ad Adolf Eichmann, che gestì la deportazione ad Auschwitz di quasi mezzo milione di ebrei ungheresi. Arendt si aspettava di trovarsi di fronte a un genio del male e invece ascoltava le ricostruzioni delle vicende e le spiegazioni di un mediocre burocrate che si difendeva dalle accuse sostenendo, appunto, che aveva solo obbedito agli ordini.

Se siamo in grado di denunciare questi casi è perché questo non è il solo modo in cui agiamo: nelle situazioni che consideriamo importanti, i nostri comportamenti sono frutto di deliberazioni e decisioni basate su ragioni che traiamo dalla nostra personalità e dalla nostra cultura, a seguito di quel processo adattivo di apprendimento che è la nostra vita.

Se nel contesto del dialogo C riceviamo la domanda X, soltanto in casi semplici la risposta Y che produciamo è scritta nella nostra memoria, quando appunto l'abbiamo "imparata a memoria". Al contrario, ordinariamente, noi non ci serviamo di innumerevoli regole specifiche,

ma applichiamo una meta-regola generale, del tipo «se nel contesto del dialogo C ricevi la domanda X, costruisci una risposta Y in accordo alla tua personalità, conoscenza e cultura P».

### **Sistemi software human-like**

Ebbene, per quanto con differenze per ora importanti rispetto alle interazioni con esseri umani, anche il comportamento dei chatbot è descrivibile in questo modo, come il termine GPT stesso suggerisce, nella sua espressione estesa. GPT infatti sta per *Generative Pre-trained Transformer*, termine che fa innanzitutto riferimento a una particolare architettura di reti neurali, chiamata appunto “Transformer”, usata in particolare per generare testi, da cui la “G”. Più importante per la nostra attuale riflessione è però la “P”: i chatbot infatti sono “pre-addestrati”. Ma che cosa significa che una rete neurale è pre-addestrata? Che il suo comportamento è il risultato di un apprendimento e non di una semplice, si fa per dire, programmazione. Un'altra differenza, quella tra addestramento e apprendimento, rispettivamente *training* e *learning* in inglese, che consideriamo importante per gli esseri umani, è sfumata oppure per ora del tutto assente per i sistemi di intelligenza artificiale: per questi infatti a volte usiamo *training*, come appunto in “GPT”, e altre volte *learning*, come in “machine learning” e “deep learning”, senza che questo corrisponda a un'effettiva differenza di significato, se non quando si parla di addestramento per riferirsi al processo e di apprendimento per riferirsi al risultato del processo. Insomma: io ti addestro, e tu impari.

In questo, le analogie tra reti neurali artificiali e reti neurali naturali, come i cervelli dei viventi, sono davvero molto suggestive. Prima di essere addestrate in un certo contesto sociale, cioè in un ambiente che consente uno scambio di informazioni, entrambe hanno una struttura e una capacità almeno basilare di apprendere: in un neonato questa capacità è legata al suo corpo e in particolare al suo sistema nervoso; in un sistema artificiale basato su una rete neurale, una capacità analoga è legata all'unico algoritmo comunque presente, quello che gestisce appunto l'addestramento della rete (e che, tecnicamente, è basato sulla discesa lungo il gradiente di una funzione di errore, ma il discorso matematico ci porterebbe troppo lontano dalla prospettiva umanistica, e non tecnica, che vogliamo mantenere qui).

In entrambi i casi, lo sviluppo dipende in modo cruciale dalle interazioni cui le reti sono sottoposte e quindi dalla qualità del processo e dei dati con cui vengono addestrate, da cui poi derivano i tratti caratteristici del loro comportamento, ciò che costituisce appunto la personalità, la conoscenza e la cultura nel caso di un essere umano: è l'elemento P nella meta-regola menzionata sopra, «se nel contesto del dialogo C ricevi la domanda X, costruisci una risposta Y in accordo alla tua personalità, conoscenza e cultura P».

A questo punto della riflessione, però, è ovvio notare che, se cerchiamo dentro a una rete neurale, naturale o artificiale che sia, non troveremo una personalità e una cultura, ma soltanto una struttura di neuroni interconnessi. E questa potrebbe essere la ragione dell'analogia più profonda tra le due specie di reti neurali, cioè il fatto



che in entrambe l'apprendimento si realizza modificando qualcosa nella struttura originaria, organica o no che sia. I chatbot, in quanto sistemi di software addestrati, sono più simili agli esseri umani e differenti, nel loro comportamento, dai sistemi di software ordinari con cui abbiamo interagito fino a questo momento. La giustificazione della nostra attitudine ad attribuire errori ai chatbot e a correggere bug soltanto rispetto ai sistemi software ordinari è allora chiara. Se un sistema software ordinario produce un errore, il problema è da ricercare in una delle istruzioni del programma che genera il suo comportamento; questa istruzione deve quindi essere identificata e corretta nel suo difetto che ha prodotto l'errore. Se invece è un chatbot a produrre un errore, il problema deve essere ricercato nel fatto che il suo addestramento è soggetto a limiti e quindi il suo apprendimento è ulteriormente sviluppabile, cosa che, lungi dal segnalare un bug, accomuna la condizione del chatbot a quella di ogni essere umano. E quindi, in questo caso, almeno di principio sbagliando s'impara.

Siamo così arrivati al punto fondamentale. Da una parte abbiamo un'analogia profonda tra specie di reti neurali, giustificata dal fatto che si tratta di sistemi capaci di imparare che si riflette nel modo in cui l'apprendimento si realizza, tanto che non sembra che ci sia una differenza essenziale tra la dinamica di apprendimento degli esseri umani e quella dei sistemi software. Dall'altra, questa analogia si connette a una questione ancora irrisolta: com'è possibile che la mia comprensione, più o meno profonda, di una questione semantica, come per esempio il messaggio di Dostoevskij nel racconto

*Il grande inquisitore*, corrisponda a dei cambiamenti di proprietà elettrochimiche delle sinapsi del mio cervello? Come scrive Stanislas Dehaene<sup>2</sup>, «le nostre sinapsi sono in costante cambiamento, e questi cambiamenti riflettono ciò che stiamo imparando» (p. 120), per poi commentare che «non sappiamo come questi cambiamenti implementino le istanze di apprendimento più elaborate, basate sul “linguaggio del pensiero”, questa combinazione di concetti propria della specie umana» (p. 129).

L'osservazione del comportamento di ChatGPT e dei suoi fratelli non ci offre una soluzione immediata a questo problema, ma ci mostra che in una condizione analoga si può ottenere una differenza simile che ancora ci appare essenziale, ma in una situazione che abbiamo costruito noi: l'addestramento del chatbot ha semplicemente prodotto delle variazioni di valori di parametri numerici, ma adesso che è addestrato il chatbot è in grado di dialogare come vediamo. Sbagliando e imparando.

Da tutto ciò traiamo due semplici riflessioni.

La prima ha a che vedere con un'obiezione che viene frequentemente sollevata contro l'effettiva capacità di un chatbot di operare in modo cognitivamente ricco, come quando si dice che è solo un “pappagallo statistico”. Ben sapendo come è stato costruito, di esso si dice: «È ovvio che non può pensare (oppure, non può essere intelligente, non può produrre conoscenza, e così via): è solo un processo matematico (o informatico o elettro-

---

<sup>2</sup> Stanislas Dehaene, *Imparare – Il talento del cervello, la sfida delle macchine*, Cortina, 2019.

nico, a seconda del livello di astrazione che si assume)». È una posizione inattaccabile, senza dubbio, ma allora, per coerenza, di fronte all'immagine di un cervello umano dovremmo anche ammettere: «È ovvio che non può pensare (oppure, non può essere intelligente, non può produrre conoscenza, e così via): è solo un processo elettrochimico». Insomma, come abbiamo già considerato, essere riduzionisti a proposito delle reti neurali artificiali e non-riduzionisti a proposito delle reti neurali naturali ci potrà forse evitare di porci qualche problema su noi stessi, ma, alla luce dei progressi vertiginosi delle capacità dei chatbot, è sempre meno giustificato.

La seconda riflessione parte dalla constatazione che la tentazione di continuare a interpretare le reti neurali artificiali come sistemi programmati sorge spontanea e potrebbe anche essere tranquillizzante, certo, in quanto tutto quello che sta succedendo sarebbe riconducibile a quanto abbiamo visto finora con i tradizionali computer e i comuni sistemi software a nostra disposizione. Ma con ciò non staremmo cogliendo la radicale differenza che c'è tra programmazione e addestramento-apprendimento, e potremmo perciò arrivare a formulare ipotesi completamente destituite di senso a proposito delle potenzialità e dei rischi della diffusione di questi sistemi nella nostra società. No, se di quello che sta accadendo vogliamo cogliere le opportunità ed evitare i rischi, dobbiamo non aver paura della novità.



## Capitolo 5

### **Cosa c'è dentro la scatola, dunque?**

*Si può interagire appropriatamente con un sistema complesso ben progettato senza doverne conoscere ogni singolo dettaglio, e dunque "lasciando la scatola chiusa". Ma cosa c'è "dentro la scatola" di un chatbot? Insomma, come funziona ChatGPT?*

Se dovessimo essere esperti di carburatori e di iniezione elettronica per poter guidare un'automobile, vedremmo plausibilmente poche automobili per le strade. Se ci sono tante automobili in giro è evidentemente perché le si può guidare in modo appropriato anche senza essere esperti di quello che c'è sotto il cofano. Bene, ma da questo non segue che sia inutile sapere qualcosa di cosa c'è "dentro la scatola". Per esempio, avere almeno qualche idea delle differenze tra un motore a combustione interna e un motore elettrico ci consente di ragionare in modo non troppo approssimativo a proposito di inquinamento (oltre a evitarci di cercare di fare rifornimento di carburante in un'automobile elettrica!).

In modo più o meno specifico, tutti noi abbiamo qualche informazione su quello che troveremmo se aprissimo la scatola di un sistema software programmato, come quello che usiamo per scrivere o per fare i calcoli: appunto un programma, cioè una lunga lista di istruzioni, ognuna generalmente semplice da capire, scritte in un linguaggio di programmazione come Java

o Python. Ma, come abbiamo già discusso, il comportamento di un chatbot non è il risultato dell'esecuzione di un programma scritto esplicitamente. Come fa a fare quello che fa, dunque? Cosa c'è dentro la scatola di un chatbot? La domanda è perfettamente lecita, ovviamente, e anzi apprezzabile, soprattutto nella prospettiva cognitiva che stiamo proponendo qui: potrebbe essere che capire meglio come funziona un chatbot ci aiuterebbe a conoscere meglio anche il nostro modo di pensare?

D'altra parte, un chatbot come ChatGPT è un sistema così complesso che anche conoscere nei dettagli la sua struttura – cosa comunque tutt'altro che semplice – non è sufficiente per renderci capaci di ricostruire le ragioni del suo comportamento, in almeno parziale analogia con il fatto che la conoscenza della struttura di un cervello umano non è plausibilmente sufficiente, né ovviamente necessaria, per sapere cosa una persona pensa.

Con ciò, un tentativo per guardare e rendere comprensibile come funziona un chatbot possiamo certamente provare a farlo, anche con l'aiuto di qualche figura. Chi vuole accompagnarci, allacci le cinture di sicurezza per il resto di questo capitolo!

### **Chatbot e sistemi conversazionali**

Per prima cosa, dobbiamo chiarire che chatbot e sistemi di intelligenza artificiale generativi conversazionali (li chiameremo “sistemi conversazionali” per brevità) non sono la stessa cosa. Un sistema conversazionale è una rete neurale artificiale addestrata per l'elaborazione di testi in linguaggio naturale (*Natural Language Processing*,

NLP), e dunque se stessimo parlando di un'automobile un sistema conversazionale sarebbe il suo motore. Poiché i sistemi conversazionali sono entità notevolmente complesse, soprattutto se realizzate in accordo all'architettura dei Transformer (da cui la sigla GPT, *Generative Pre-trained Transformer*), li si dota di un'interfaccia di programmazione, cosa che semplifica notevolmente l'interazione. Grazie a questa interfaccia, chiamata genericamente *Application Programming Interface* (API), possiamo inviare richieste a un sistema conversazionale in un quasi-inglese e nella stessa lingua ricevere da esso le risposte.

Ecco un esempio reale di una domanda (“Please introduce yourself!”, inviata a un sistema conversazionale che si chiama “Llama-2” e che parla solo inglese).

```
{
  "model": "Llama-2-7B Chat",
  "max_tokens": 4096,
  "messages": [{"role": "user", "content": "Please introduce yourself!"}],
  "temperature": 0.9
}
```

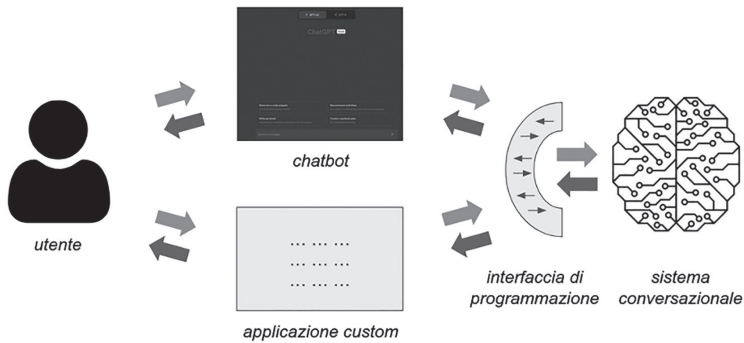
E questa è (parte del)la risposta (“Hello! My name is Llama, I’m a large language model trained...”) prodotta dal sistema.

```
{
  "choices":[{"finish_reason":"stop","index":0,
  "message":{"content":"Hello! My name is Llama, I'm a large language
  model trained by a team of researcher at Meta AI. I can understand
  and respond to human input in a conversational manner. ...",
  "role":"assistant"},
  "references":[]}],
  "model":"Llama-2-7B Chat","object":"text_completion",
  "usage":{"completion_tokens":112,"prompt_tokens":14,"total_
  tokens":126}
}
```

D'accordo, questo dialogo contiene un po' di dettagli tecnici e la sintassi – si chiama “JSON” ed è oggi praticamente uno standard per i programmatori – non è così ovvia. Ma non è così difficile imparare questo stile di comunicazione, con un'entità che, “dentro la scatola”, sa gestire un dialogo in modo sofisticato.

Con ciò, possiamo intendere un chatbot come l'intera applicazione con cui interagiamo (e quindi come l'automobile di cui un sistema conversazionale è il motore) oppure come una possibile interfaccia di un'applicazione con cui interagire con un sistema conversazionale (e quindi come il volante e i pedali dell'automobile), considerando con ciò che grazie alla API i programmatori possono costruire anche altre applicazioni, con finalità specifiche. In questa prospettiva, l'architettura complessiva del sistema può essere schematizzata come riportato nella figura che segue.

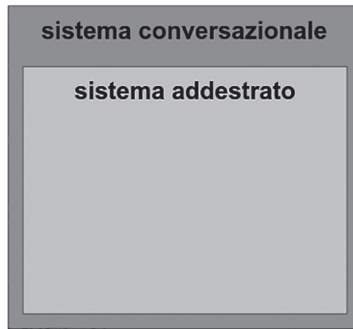




È perciò chiaro che la parte davvero interessante – il motore dell'automobile insomma – è il sistema conversazionale. Chi acquisisce qualche competenza di *prompt engineering* sta dunque imparando a interagire con un sistema conversazionale, per fargli le domande giuste nel modo giusto. Alzando il cofano di un sistema conversazionale troviamo, una dentro l'altra e dunque sempre più in dettaglio, tre scatole: entriamo in officina e diamo loro un'occhiata.

### **Dentro un sistema conversazionale**

Già sappiamo che i sistemi conversazionali sono realizzati mediante reti neurali artificiali, che sono sistemi addestrati, non programmati.



Sistemi come ChatGPT vengono addestrati in due macro-fasi. La prima è il *pre-training*, in cui si fanno leggere al sistema tanti testi, nascondendogli però alcune parti di ogni testo, che il sistema deve cercare di indovinare; se il tentativo non ha successo, cioè se ciò che il sistema ha proposto è diverso dal testo nascosto da indovinare, si interviene per cercare di correggere questo comportamento (come? ne parliamo tra poco, dopo aver aperto qualche altra scatola). Questa tecnica, semplice e ingegnosa nello stesso tempo, può essere realizzata in modo completamente automatico, senza alcun intervento da parte di esseri umani e perciò è chiamata “apprendimento auto-supervisionato” (*self-supervised training*). A questo punto, il sistema ha competenze linguistiche e disciplinari, in funzione del contenuto dei testi su cui ha imparato, ma è a-morale e non è ancora capace di dialogare.

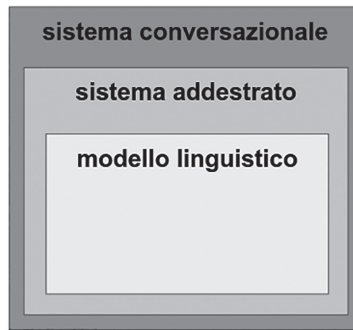
La seconda macro-fase dell’addestramento è il *fine tuning*, in cui si fanno leggere al sistema dei dialoghi che in precedenza degli esseri umani hanno valutato, indi-

cando se sono dei buoni o dei cattivi esempi (come con un bambino, dunque: «sì, fai così»; «no, non fare così»), e di nuovo facendo in modo che il sistema corregga il suo comportamento di conseguenza. A questo punto il sistema ha anche capacità di dialogo e gli è stata imposta una morale, implicita nella scala dei valori proposta negli esempi da cui ha imparato. Il *fine tuning* è perciò il momento in cui si formano la personalità e la morale di un sistema conversazionale: se lo si sarà addestrato mostrando come positivi esempi di una certa ideologia (supponiamo, per esempio, a proposito di un certo orientamento politico), tenderà a rispondere assumendo quella ideologia.

L'addestramento dei sistemi conversazionali, che include anche la preparazione dell'insieme dei testi su cui l'addestramento si realizza, richiede grandi risorse (in capacità di calcolo, tempo, energia elettrica) ed è anche per questo che – almeno per ora – si mantengono separati addestramento e funzionamento: prima li si addestra e dopo li si usa. Questo spiega, per esempio, perché l'addestramento di GPT-4 è terminato, al momento (dicembre 2023), nella primavera 2023. Ed è per questo che gli attuali sistemi conversazionali funzionano mantenendo distinte *la memoria a lungo termine*, cioè la struttura assai complessa che viene adattata durante l'addestramento ma che poi non viene modificata nell'interazione con gli utenti, e *le memorie a breve termine*, che mantengono ognuna l'informazione di un dialogo in corso, ma che si azzerano alla fine del dialogo stesso, con la conseguenza che ogni dialogo con un chatbot è dunque, almeno per ora, una storia chiusa in se stessa. Dopo aver completa-

to una magari assai dotta conversazione con ChatGPT a proposito delle idee di Dostoevskij sulla responsabilità umana, ne avremmo una facile dimostrazione se iniziando un nuovo dialogo gli chiedessimo: «Qual è stato l'argomento della nostra conversazione precedente?». Anche se la domanda è banale, la risposta sarà qualcosa come: «Mi dispiace, ma non conservo le informazioni sulle conversazioni precedenti per garantire la privacy degli utenti». La questione è in effetti assai delicata e ha a che vedere non solo con limiti tecnologici e garanzia di riservatezza, ma anche e prima ancora con una fondamentale cautela etica. Se un sistema conversazionale imparasse dai suoi dialoghi con gli utenti, supponendo che in ogni istante abbia qualche milione di dialoghi aperti, in poche ore di funzionamento cambierebbe la sua personalità, non necessariamente in meglio. L'attuale separazione tra memoria a lungo termine e memorie a breve termine sembra una scelta decisamente saggia.

Aperto il coperchio di un sistema addestrato, troviamo un modello linguistico, quello che in inglese si chiama abitualmente un *Large Language Model* (LLM), di cui da qualche anno i Transformer sono diventati l'esempio di riferimento.



---

Un Transformer è un sistema sofisticato e complesso, decisamente troppo complesso perché possiamo spiegarne qui struttura e funzionamento in dettaglio. L'idea di base, sempre ricordando che quanto segue è il risultato del processo di addestramento descritto sopra, può essere schematizzato così nel caso di uno scambio domanda-risposta:

- il testo della domanda viene diviso in parti elementari, chiamate “token”, ognuna corrispondente a una parola o a una parte di una parola; la domanda è diventata così una successione di token;
- ogni token della successione viene associato a un vettore di numeri, con un'operazione, chiamata *embedding*, finalizzata a far sì che token dal significato simile siano associati a vettori matematicamente vicini; la domanda è diventata così una successione di vettori di numeri;
- i vettori della successione vengono ripetutamente

confrontati tra loro, per identificare quali parti della domanda sono particolarmente importanti, perché portano più informazione, riproducendo così in termini matematici l'attività di “prestare attenzione” che rende efficaci ed efficienti le attività cognitive umane; a ogni vettore della successione viene così attribuito un “grado di attenzione”;

- si trova ora una nuova successione di vettori di numeri, coerente con quella corrispondente alla domanda: è questo concetto di coerenza tra il vettore della domanda e il vettore che viene prodotto che determina la qualità della risposta che si otterrà e che dipende criticamente dall'addestramento del Transformer;
- ogni vettore della nuova successione viene convertito nel corrispondente token;
- i token della successione vengono uniti tra loro e il risultato è il testo della risposta.

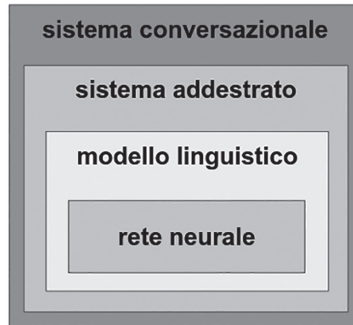
L'informazione basilare che ricaviamo da questa descrizione è che, quasi paradossalmente, “dentro la scatola” i *Large Language Model* operano in effetti con numeri, non con entità linguistiche. Ed è infatti trattando numeri che si realizza la fase cruciale di costruzione della risposta alla domanda data, criticamente fondata sull'efficacia dell'aver “prestato attenzione” al contenuto della domanda<sup>1</sup>.

---

<sup>1</sup> Non per nulla, l'articolo in cui è stata introdotta l'architettura dei Transformer si intitola *Attention is all you need*, <https://arxiv.org/abs/1706.03762>.

Aperto l'ultimo coperchio troviamo, finalmente, una rete neurale artificiale, e in effetti una complessa rete di reti neurali.

---



L'idea di ottenere un comportamento in qualche senso intelligente attraverso non l'implementazione di un algoritmo ma l'addestramento di una struttura analoga a una rete neurale naturale/organica, come quella di un cervello umano, fu proposta addirittura nel 1943<sup>2</sup>. Il modello di McCulloch e Pitts (significativamente, un neurofisiologo e un matematico) è basato su una grande semplificazione di quanto accade in un cervello e nondimeno si è dimostrato efficace, per scopi relativamente semplici come riconoscere quali oggetti sono

---

<sup>2</sup> Warren S. McCulloch e Walter Pitts, *A logical calculus of ideas immanent in nervous activity*, 1943, <https://www.cs.cmu.edu/~./epxing/Class/10715/reading/McCulloch.and.Pitts.pdf>.

contenuti in una fotografia, ma anche ben più complessi come dialogare in modo appropriato a proposito delle idee di Dostoevskij sulla responsabilità umana.

Non è il contesto per entrare nei dettagli di come funziona una rete neurale artificiale, ma l'idea di base può essere spiegata semplicemente con un esempio. Supponiamo di voler costruire uno strumento matematico che ci consenta di fare una previsione sull'altezza di una persona dato il suo peso. Dato che persone con lo stesso peso possono avere altezze diverse, la conoscenza del peso di una persona non è sufficiente per ricavarne l'altezza e quindi lo strumento potrà fornire solo un'informazione approssimata. Ma, come si dice, "meglio di niente". Facciamo allora un'ipotesi sulla relazione generica tra il peso e l'altezza delle persone, cioè costruiamo un *modello* di tale relazione. In accordo all'osservazione che, pur con eccezioni, le persone più pesanti sono quelle più alte, un'ipotesi molto semplice è che l'altezza sia proporzionale al peso. In forma matematica questo diventa  $a = k p$ , dove  $p$  e  $a$  sono il peso e l'altezza di una persona, supponiamo in chilogrammi e centimetri rispettivamente, e  $k$  è la costante di proporzionalità che cerchiamo. E così se  $k = 2$  il nostro modello prevederebbe che una persona che pesa 80 kg sia alta  $2 \times 80 = 160$  cm. Forse in una popolazione di obesi... Ma no, diciamo che dobbiamo correggere il valore di  $k$ , per produrre delle previsioni un poco più realistiche. Forse allora  $k = 3$ ? Vediamo: prevederemmo ora che una persona che pesa 80 kg è alta  $3 \times 80 = 240$  cm. Decisamente no. Dunque  $k = 2$  è troppo poco, ma  $k = 3$  è troppo. D'accordo, sarà un valore decimale, intermedio tra i due, come 2,5.



La funzione  $a = k p$  descrive dunque la struttura dell'ipotesi che abbiamo assunto, ma stiamo ammettendo con ciò che la nostra conoscenza è incompleta fintanto che non avremo trovato un valore soddisfacente per la costante  $k$ , che è un *parametro* del modello. Per risolvere quest'ultimo problema ci aiuta ancora l'intuizione. Raccogliamo in qualche modo i dati sul peso e sull'altezza di un buon numero di persone e con ciò costruiamo un elenco di coppie (peso della prima persona, altezza della prima persona), (peso della seconda persona, altezza della seconda persona), e così via, che possiamo scrivere più comodamente  $(p_1, a_1)$ ,  $(p_2, a_2)$ , ... Avendo assegnato un valore candidato iniziale al parametro  $k$ , per esempio 2, a questo punto consideriamo una dopo l'altra tutte le coppie dell'elenco e procediamo così: prendiamo il valore del peso, moltiplichiamolo per  $k$  e confrontiamo il risultato con il valore dell'altezza. Se sono uguali o almeno non troppo diversi, tutto bene, significa che il nostro modello ha fatto una buona previsione. Ma se invece sono sufficientemente diversi, sfruttiamo la coppia in esame per correggere il modello: se l'altezza prevista è minore dell'altezza vera, aumentiamo di un poco il valore di  $k$  o viceversa nell'altro caso. Così, alla fine di questo processo di aggiustamenti progressivi del valore del parametro (processo che abbiamo supervisionato, e che perciò possiamo a buon diritto chiamare di *supervised training*), avremo addestrato il nostro modello a prevedere “mediamente al meglio” l'informazione contenuta nell'“insieme di addestramento” (il *training set*) di esempi.

## **Qualche idea da quello che abbiamo scoperto dentro la scatola**

Nonostante l'estrema semplicità di questo esempio, la conclusione di base che ne abbiamo tratto rimane corretta anche per le reti neurali artificiali dei sistemi conversazionali e quindi dei chatbot: una rete neurale artificiale è una *funzione parametrica*, la cui struttura viene scelta in fase di progettazione (e dunque, sì, “programmandola”, come abbiamo fatto sopra decidendo il modello di proporzionalità  $a = k p$ ), ma il cui comportamento è ottenuto dall'averla messa a confronto con degli esempi opportuni (e dunque, sì, “addestrandola”, come abbiamo fatto sopra trovando un valore appropriato per il parametro  $k$ ).

Una volta addestrata, una rete neurale è dunque capace di *generalizzazione*, nel senso che è in grado di fare previsioni anche a proposito di informazioni che non le sono state fornite durante l'addestramento. Al modello  $a = 2,5 p$  potrebbe non essere mai stato mostrato l'esempio del peso 70 kg e nonostante ciò saprebbe prevedere l'altezza corrispondente,  $2,5 \times 70 = 175$  cm. Ed è chiaro che la qualità delle previsioni del modello dipende criticamente dalla qualità degli esempi scelti per il suo addestramento. Se, supponiamo, avessimo popolato il *training set* con dati solo di persone obese, potremmo stupirci se il nostro modello sarà per esempio  $a = 1,5 p$  (dunque prevederà che una persona di 100 kg è alta 150 cm, e così via) e con ciò il modello “si sarà fatto l'idea” che gli esseri umani sono tutti obesi? È quello che si chiama una “distorsione” della realtà modellata, concetto per cui ormai usiamo più frequentemente il termine inglese: *bias*.

Un Transformer, come GPT-3 che sta dentro alla versione iniziale di ChatGPT, si può in effetti pensare come una funzione parametrica, non diversamente da quella del nostro esempio. Ma... ma c'è un grande “ma” da tenere bene in considerazione, per evitare malintesi. La funzione dell'esempio sopra ha un parametro e quindi il suo addestramento consiste nel trovare il valore migliore per quell'unico parametro e il suo comportamento dipende dal valore assegnato a quell'unico parametro; GPT-3 è una funzione con *175 miliardi* di parametri!

Non dobbiamo dunque lasciarci ingannare. È vero che da 1 a 100 miliardi è un cambiamento quantitativo, ma ci sono cambiamenti quantitativi che diventano anche cambiamenti qualitativi, come quello di una goccia d'acqua che diventa un cristallo di neve quando la sua temperatura si riduce. Abbiamo chiesto un'interpretazione di questa transizione a ChatGPT con DALL-E 3 (il sistema di OpenAI che genera immagini a partire da una descrizione in linguaggio naturale), e la figura che segue è un suo risultato.



---

Interessante e suggestivo... Come interessanti e suggestivi sono gli attuali sistemi conversazionali basati su modelli linguistici che sono reti neurali artificiali: un cambiamento quantitativo – la moltiplicazione del numero dei parametri – ha prodotto entità radicalmente nuove e inaspettate.

## Capitolo 6

### **I limiti cognitivi ed etici (attuali: chissà nel futuro...) dei chatbot**

*Le capacità degli attuali chatbot sono davvero straordinarie, ma questo non deve farci dimenticare che è importante essere anche consapevoli dei loro limiti. E trarne qualche conseguenza concreta.*

Chiunque abbia interagito con sufficiente attenzione e spirito critico con un chatbot come ChatGPT non ha potuto non stupirsi delle sue capacità. E il fatto che si tratti di entità addestrate e non programmate, come discusso sopra, fa sì che i testi che essi producono siano il risultato di un'elaborazione autonoma, a partire dalla "lettura" (ancora una volta un termine antropomorfo: le virgolette sono per coloro che sostengono che per definizione soltanto gli esseri umani possano, propriamente, leggere) di una grande quantità di documenti e non di ricerche in un archivio, mantenuto localmente nel nostro computer oppure accessibile globalmente attraverso il web. Tutto ciò rende gli attuali chatbot delle *entità nuove*, con ancora tante incognite a proposito delle opportunità che offrono ma anche dei rischi che potrebbero farci correre, se non saremo accorti nell'usarli in modo appropriato.

Questa novità rende complessa l'individuazione stessa di tali rischi e opportunità: come cambieranno i processi di apprendimento? Le competenze richieste nel mondo del lavoro? I ruoli nelle organizzazioni? Le strut-

ture stesse delle organizzazioni? E così via. Possiamo certamente fare delle ipotesi, ma dobbiamo ammettere che è ancora troppo presto per formulare delle previsioni affidabili. Non è invece troppo presto per analizzare e discutere i limiti dei chatbot. Al meglio delle nostre attuali conoscenze, ne evidenziamo tre, che sono plausibilmente strutturali più che solo contingenti e che sono importanti nelle loro conseguenze cognitive ed etiche. Gli attuali chatbot:

- sono poco affidabili nelle informazioni fattuali che riportano: insomma, non sono *trustable* nel loro comportamento;
- sono capaci di indicare le loro fonti in modo spesso approssimativo e raramente di rendere conto di come giungono ai risultati che propongono: insomma, non sono *explainable* nel loro comportamento;
- sono entità socialmente aliene, così che non sappiamo attribuire loro una responsabilità per quanto producono: insomma, non sono *accountable* nel loro comportamento.

Uno per uno, discutiamo questi limiti.

### **I chatbot non sono affidabili nel loro comportamento**

Come abbiamo spiegato e commentato in precedenza, il comportamento di un chatbot è il risultato di un processo di addestramento, realizzato su una grande quantità di testi: potenzialmente tutte le pagine web accessibili (dunque per esempio quelle di Wikipedia) e altro. Un tale *corpus* è prodotto e immagine della cultura della nostra

società e come tale non è affatto internamente coerente, dato che la nostra società non lo è. Così che a proposito di argomenti controversi il nostro povero chatbot si trova a dover imparare una cosa e il suo contrario. Certo, se conduciamo il dialogo con accortezza, il chatbot potrà essere in grado di riportarci questa molteplicità di posizioni, mostrando quella che, se si trattasse di un essere umano, considereremmo non solo corretta informazione ma anche consapevolezza della complessità dei contenuti della conversazione che stiamo avendo. Ma non è difficile, purtroppo, trovarsi nella condizione di scoprire che un chatbot ci dà risposte con un tono pacato e convinto, e nondimeno sbagliate. Il punto è sempre lo stesso: gli attuali chatbot non sono motori di ricerca che recuperano informazioni da fonti indipendenti (che poi, nel caso dei motori di ricerca, queste fonti siano o no veridiche è ovviamente un'altra questione), ma sono, per dire così, dei fantastici, eloquenti parafrasatori e rielaboratori dei testi che sono stati fatti leggere loro durante l'addestramento. E dunque se l'informazione da cui partono non è corretta, non potranno fare altro che propagare tali errori. Considerando poi che generalmente il comportamento di un chatbot dipende in parte da fattori casuali (è la ragione per cui chiedendo di ripetere la risposta alla stessa domanda otteniamo risultati non identici), giungiamo all'obiettivamente scomoda conclusione che *la verità dei chatbot è l'opinione prevalente che essi trovano nei testi con cui li abbiamo addestrati.*

L'esperienza mostra che questo non ne fa necessariamente dei conformisti: vediamo quanto i testi che producono possono apparirci creativi, arguti, sorprendenti,

se solo poniamo in modo appropriato le nostre domande e li sollecitiamo a risponderci in modo appunto creativo, arguto, sorprendente. Ma da ciò è evidente che non è una buona idea trattare i chatbot come fonti di verità o come oracoli. Il tempo, plausibilmente anche breve, ci dirà se le architetture a plugin e ad agenti, di cui scriveremo un poco nelle pagine che seguono, cambieranno le cose. Per fare un esempio ovvio, i chatbot non sono così affidabili nel risolvere problemi di matematica, ma lo possono diventare delegando i conti da fare a un modulo software dedicato a questo. Ma per ora è chiaro che è ben più sensato interagire con i chatbot per chiedere loro pareri anziché informazioni fattuali. Ammettiamolo: anche in questo sono più simili agli esseri umani che ai tradizionali sistemi software programmati. Nel dialogo con i chatbot la nostra capacità critica diventa ancora più, e non meno, indispensabile.

### **I chatbot non sono spiegabili nel loro comportamento**

Il comportamento che osserviamo in un sistema software programmato può sorprenderci, ma è comunque il risultato di un algoritmo che qualcuno (o eventualmente qualcosa) ha costruito e poi tradotto con un linguaggio di programmazione, per ottenere appunto un programma. Questo ci mette a disposizione una strategia che almeno in linea di principio è efficace per rispondere alla fatidica domanda: «Perché si comporta così?». Per esempio: «Perché se faccio click su questa icona succede che il testo che avevo selezionato cambia il suo stile in grassetto?». È sufficiente poter accedere al codice sorgente del programma e con abilità e spesso grande pa-



zienza individuare la linea o il gruppo di linee di codice responsabili di quel particolare comportamento. Come abbiamo discusso in precedenza, è quanto i programmatori fanno quando, di fronte a un errore da correggere, “fanno debugging”. Certo, in un programma complesso tutto ciò può essere tutt’altro che ovvio. Per esempio, si legge sul web che il programma Microsoft Word potrebbe essere costituito di circa cinquanta milioni di linee di codice: trovare le linee giuste tra cinquanta milioni di altre richiede certamente conoscenza, abilità e una buona dose di esperienza. Ma di principio, appunto, il comportamento di un sistema software programmato è in questo senso spiegabile.

Se, anche alla luce delle considerazioni fatte nelle pagine precedenti sul funzionamento delle reti neurali artificiali, riconsideriamo ora tutto ciò a proposito dei chatbot, ci troveremo in una situazione completamente diversa. Come abbiamo scritto e come a questo punto dovrebbe essere ben chiaro, le risposte che ChatGPT e i suoi fratelli ci danno sono il risultato di addestramento, non di programmazione. Non c’è debugging che tenga, dunque: se ha risposto così, è perché è questo che ha imparato dal suo addestramento, non perché qualcuno (o eventualmente qualcosa) l’ha deciso in qualche modo.

È soprattutto per questa ragione che usare l’espressione «gli algoritmi dell’intelligenza artificiale» per riferirsi a ciò che genera il comportamento dei chatbot è almeno fuorviante, e in effetti proprio sbagliato. Ne abbiamo chiesto la conferma anche a ChatGPT e questa è stata la sua risposta: «Il comportamento di un modello di *deep learning*, come un chatbot, è guidato da funzioni

parametriche complesse e, sebbene sia basato su algoritmi durante la fase di apprendimento, il suo comportamento operativo non è intuitivamente spiegabile in termini algoritmici tradizionali».

Di conseguenza dobbiamo riconoscere che la nostra capacità di controllo sul, e prima ancora di spiegazione del, comportamento di un chatbot ha dei limiti che almeno per ora non sono superabili. Non per nulla, la cosiddetta “intelligenza artificiale spiegabile” (in inglese *Explainable AI*, XAI) è per ora soprattutto un ampio, sfidante programma di ricerca. Vedremo cosa a questo proposito succederà nel prossimo futuro.

### **I chatbot non sono responsabili del loro comportamento**

Molto più che la questione se, quanto e come i chatbot siano intelligenti, ciò che più radicalmente distingue gli esseri umani dalle “macchine”, di qualsiasi genere esse siano, è la possibilità di assumersi una responsabilità verso la società per le decisioni che si prendono e le azioni che si compiono. Il fatto è che la società a cui stiamo facendo riferimento è una società modellata “a misura di esseri umani” e che ci mancano i fondamenti stessi per sapere come attribuire efficacemente una responsabilità a entità artificiali, come i chatbot, anche qualora a queste attribuissimo un qualche genere di autonomia e quindi le stesse assumessero un qualche genere di libero arbitrio (l'ennesimo antropomorfismo). La ragione di questa nostra radicale incapacità è semplice e profonda nello stesso tempo: perché esse non sono vive. Spieghiamoci.

Se ci riconosciamo responsabili di qualcosa è perché siamo disponibili a farcene carico e quindi anche a subire personalmente le conseguenze di eventuali situazioni dovute a quel qualcosa e considerate contrarie agli interessi, in qualche modo codificati, della società. La responsabilità individuale è un costrutto sociale che abbiamo instillato ai fondamenti stessi della nostra possibilità di convivenza. Essere responsabili di qualcosa di fronte a un gruppo sociale è una situazione positiva perché genera un senso al nostro essere parte di quel gruppo, ma l'altra faccia della medaglia non è meno importante: è perché siamo responsabili che siamo sanzionabili, se qualcosa non va come dovrebbe. Ora, un'entità artificiale, come è per esempio un chatbot, non è in alcun modo sanzionabile, perché non possiede beni che le si possano togliere (in che senso un chatbot potrebbe essere multato?) e ancora più fondamentalmente perché il tempo per essa non è la risorsa non fungibile, e generalmente scarsa, che è per le entità organiche. Un chatbot interagisce nel tempo, ovviamente, ma non vive nel tempo, perché non vive. E quindi non muore. E quindi per un chatbot il tempo ha un ruolo completamente diverso da quello che ha per noi. Per esempio, mettere in prigione per un anno una persona significa toglierle un anno di vita in libertà: è una condizione generalmente efficace per sollecitare quella persona a evitare quella sanzione operando in modo conforme a quanto la società si attende. Ma nel caso di un chatbot che si comporta in modo non conforme alle attese della società, cosa potremmo fare? Sì, nel film *2001: Odissea nello spazio* il computer HAL 9000 supplica che non si separino i suoi moduli e non gli si stacchi la spina, ma nel

momento in cui fosse riassemblato e riattivato potrebbe tornare com'era. La questione è che, almeno per ora, noi abbiamo un hardware e un software (cioè un cervello e una mente) così strettamente interconnessi da non essere riassemblabili e riattivabili. E, a differenza del software, l'hardware subisce gli effetti entropici del secondo principio della termodinamica. Insomma, invecchia. E muore, se si tratta di un'entità organica.

Questo mette a disposizione della società un abilitatore, generalmente efficace, di responsabilità individuale. Come se fosse: decidi e opera in modo conforme a quanto la società si aspetta da te, perché altrimenti – dato che sei responsabile delle tue decisioni e delle tue azioni – ti toglierò qualcosa, del tempo, che non potrai recuperare. Nulla di tutto ciò vale dunque per un chatbot che, come ogni altro sistema software progettato per essere eseguito su calcolatori digitali più o meno complessi, non è vincolato al suo substrato hardware: se un calcolatore si rompe, lo si può sostituire e il software tornerà, letteralmente, identico a prima.

La conseguenza di tutto ciò è che non abbiamo idea di come attribuire a un chatbot una responsabilità, che dunque, almeno per ora, rimane nostra, anche quando fosse lui/lei/esso/essa a commettere un errore. Venendo da millenni in cui, per gli individui della nostra specie, vita, autonomia cognitiva e assunzione di responsabilità sociale sono andate di pari passo, questo è uno scenario che non siamo pronti a gestire e per cui dobbiamo ancora concepire alternative. La questione non è giuridica, dunque. È vero che ci mancano le leggi che trattino i chatbot come entità “capaci di intendere e volere”, ma

prima di elaborare eventualmente queste leggi dovremmo chiederci se davvero vogliamo modificare la nostra società così in profondità e trasformarla in una “società umano-robotica”. E dobbiamo interrogarci dunque se davvero vogliamo delegare a sistemi artificiali non più solo compiti che richiedono il trattamento di materia ed energia (come facciamo da millenni e abbiamo imparato a fare in modo sempre più efficiente con le rivoluzioni industriali, a partire dalla macchina a vapore e dal motore elettrico) e di informazione (come facciamo da qualche decennio con i sistemi di telecomunicazioni e di calcolo automatico), ma anche l’assunzione di responsabilità sociale. Certo, non è che la storia ci presenti solo esempi edificanti di esseri umani capaci di operare responsabilmente di fronte alla società, ma davvero siamo pronti, culturalmente e prima ancora psicologicamente, per imbarcarci nell’avventura di riprogettare i fondamenti della nostra società in modo così potenzialmente radicale?

Per ora potrebbe essere che la direzione dello sviluppo sia un’altra: potremmo delegare sempre più compiti ad agenti artificiali, ma non la responsabilità verso la società dei risultati che se ne ottengono e delle loro conseguenze. Forse il compito che nel futuro manterremo per noi sarà la responsabilità. Vedremo.



## Capitolo 7

### **ChatGPT e gli agenti: una nuova frontiera per l'intelligenza artificiale**

*In questi mesi i chatbot stanno evolvendo, diventando sistemi sempre più complessi e capaci di funzioni diverse e sofisticate.*

Anche se sono stati resi disponibili ai non addetti ai lavori soltanto da poco tempo, gli attuali chatbot, stanno continuando a offrirci novità rilevanti e quello che stiamo vivendo potrebbe essere ricordato come un periodo interessante, ricco di novità quasi quotidiane. A partire dalla pubblicazione di ChatGPT, il 30 novembre 2022, i cambiamenti più evidenti sono stati due. Per prima cosa è aumentato il numero di chatbot con cui possiamo interagire, e tra questi Bard di Google e Claude di Anthropic, e tanti sistemi “aperti”/non proprietari, spesso sviluppati a partire da Llama di Meta (una piccola nota tecnica a tal proposito: presumibilmente per abitudine, si caratterizzano spesso questi sistemi come “open source”, cioè a codice sorgente aperto, in quanto liberamente disponibile; ciò è corretto ma parziale, dato che il vero valore di questi sistemi non è il codice, ma la matrice dei valori dei parametri ottenuti dal pre-addestramento: nei sistemi aperti è anche e soprattutto questa matrice a essere ad accesso libero; per dare un’idea, la matrice del sistema aperto attualmente più sofisticato, Falcon 2 180B, sta in un file di quasi 400 Gbyte). Le capacità di questi chatbot non sono identiche, ma tut-

ti sono basati su una stessa architettura di reti neurali, i Transformer (introdotta nel 2017 da un gruppo di ricercatori di Google in un articolo pubblicato, significativamente, anch'esso ad accesso libero), cosa che spiega le notevoli somiglianze nel loro comportamento. Il secondo cambiamento, nonostante il breve arco di tempo intercorso, è il miglioramento delle caratteristiche e quindi delle prestazioni dei singoli chatbot, come mostra il caso di ChatGPT, che da metà marzo del 2023 può essere usato (per ora soltanto nella versione a pagamento) anche nella versione GPT-4, che è spesso sensibilmente migliore dell'iniziale GPT-3.5. Qui discutiamo una terza linea di sviluppo che, benché per ora forse meno evidente, per varie ragioni potrebbe rivelarsi di grande importanza.

### **Una nuova linea di sviluppo**

Ne abbiamo visto e sperimentato i primi esempi da quando, nello stesso periodo del lancio di GPT-4, OpenAI ha reso accessibili i plugin di ChatGPT (anch'essi per ora aperti soltanto nella versione a pagamento). La descrizione iniziale che ci è proposta è attraente: «I plugin sono strumenti che supportano ChatGPT nell'accedere a informazioni aggiornate, eseguire calcoli, o usare servizi di terze parti». A una persona che non abbia dedicato tempo sufficiente a cercare di capire che cosa stia succedendo, tutto ciò appare però sorprendente. L'esattezza dei calcoli non è proprio una caratteristica dei sistemi digitali? Che cosa ci dovrebbe essere dunque di nuovo in un chatbot che “sa fare i conti”? Eppure è così. Qualche conversazione con un chatbot condotta con spirito critico ne evidenzia non soltanto le potenzialità, ma anche i limiti (per fortu-



na, direbbe qualcuno), in particolare proprio a proposito del non aggiornamento della sua base di conoscenza, che anche per GPT-4 è al momento (dicembre 2023) ferma alla primavera 2023, e delle sue capacità aritmetiche, per cui potrebbe sbagliare per esempio a calcolare la moltiplicazione di due numeri di quattro cifre. Se perfino Omero ogni tanto sonnecchia, figuriamoci un chatbot...

Come argomentato in precedenza, la ragione è che il comportamento di una rete neurale artificiale, di cui i Transformer sono dunque un esempio, è il risultato del processo di addestramento della rete stessa, che è un complesso sistema di elaborazione parametrico, non il frutto dell'esecuzione di un algoritmo su un archivio di dati memorizzati e pre-inseriti. In conseguenza, tale comportamento non è né perfettamente affidabile né facilmente prevedibile.

### **Un esempio, per capire**

Un semplice esempio può essere utile per meglio comprendere questo punto e nello stesso tempo per introdurci alla novità che stiamo discutendo. Un'applicazione tradizionale delle reti neurali è la classificazione di immagini basata sul riconoscimento del loro contenuto. Per esempio, dato un insieme di immagini di cifre scritte a mano, una rete può essere addestrata per riconoscere la cifra in un'immagine (per essere chiari, per questo compito non è necessario usare un Transformer: bastano architetture molto più semplici, come i *Multi-Layer Perceptron*, usati ormai da qualche decina di anni o, nei casi più complessi, le reti neurali convoluzionali). Disponendo di un insieme sufficientemente

ampio di coppie (*immagine da analizzare, cifra nell'immagine*) e cioè (*domanda, risposta giusta*), la rete viene addestrata in modo supervisionato, ripetendo tante volte questa procedura, chiamata di *back-propagation*: 1) si presenta alla rete un'immagine dell'insieme e si fa in modo che la rete cerchi di riconoscere la cifra; 2) si confronta la cifra riconosciuta con la cifra corretta; 3) in caso di errore, si modificano un poco i valori dei parametri della rete in modo da renderne il comportamento più accurato nei successivi riconoscimenti.

Questa procedura di addestramento permette anche a reti relativamente piccole e semplici – con qualche decina di migliaia di parametri: un numero enorme rispetto alle funzioni su cui ci siamo esercitati a scuola, ma milioni di volte più piccole rispetto agli attuali chatbot – di raggiungere un'accuratezza di oltre il 90%, ovvero riconoscere sistematicamente almeno nove cifre su dieci.

Con ciò, se volessimo capire perché una certa rete riconosce correttamente certe cifre e non ne riconosce invece altre, una risposta del tipo «è perché i valori dei suoi parametri sono questi», pur formalmente corretta, non soddisferebbe i criteri che ci attendiamo da una buona spiegazione. Insomma, non ci convincerebbe, come non ci convincerebbe, di solito, a una nostra richiesta di spiegazione del comportamento di una persona, una risposta del tipo «per capire perché si è comportato così, studia la TAC del suo cervello». Piuttosto, dovremmo onestamente ammettere che la rete si comporta come si comporta perché così è stata addestrata, senza riuscire con ciò a dare e a darci una spiegazione migliore.

## Sistema 1, Sistema 2

D'altra parte, consideriamo per un istante il nostro comportamento di fronte a un compito analogo o magari anche un poco più complesso, come il riconoscimento di una persona che ci viene incontro guardandone il viso: siamo abili in questo, ma senza saperne dare una spiegazione esplicita. Daniel Kahneman<sup>1</sup>, uno psicologo che ha vinto il premio Nobel per l'Economia, ha chiamato "Sistema 1" questo genere di comportamento. È ciò che usiamo per le decisioni e i giudizi rapidi, a volte senza nemmeno rendercene conto, proprio come accade a chiunque ha imparato a leggere e che appunto riconosce senza fatica i caratteri alfanumerici che vede scritti su una pagina, pur senza saper fornire una ricostruzione analitica esplicita – dunque un algoritmo – di come riesce in questo.

Di massima, possiamo considerare le reti neurali artificiali, Transformer inclusi, come dei Sistemi 1 nel senso di Kahneman, e questo rende conto di uno dei problemi critici per il loro impiego come entità coinvolte in processi decisionali: come abbiamo già discusso nelle pagine precedenti, il loro comportamento non è spiegabile (e perciò l'obiettivo a cui qualcuno mira al proposito è chiamato *Explainable AI*).

Supponiamo ora di modificare un po' il nostro obiettivo e di voler usare quella stessa rete già addestrata per riconoscere in ogni immagine non più una cifra, ma soltanto se quella cifra è pari o dispari. Come potremmo fare? Dato

---

<sup>1</sup> Daniel Kahneman, *Pensieri lenti e veloci*, Mondadori, 2012 (presentato in [https://it.wikipedia.org/wiki/Pensieri\\_lenti\\_e\\_veloci](https://it.wikipedia.org/wiki/Pensieri_lenti_e_veloci)).

che almeno per ora intervenire sulle reti neurali artificiali non ci pone problemi etici, potremmo certamente re-inizializzare la rete, facendole dimenticare quanto ha appreso, e sottoporla a un nuovo processo di addestramento, in cui il *training set* non è più di coppie “immagine da analizzare, cifra nell’immagine” ma di coppie “immagine da analizzare, cifra pari oppure cifra dispari”. Ma è evidente che un’altra strategia è possibile e di ben più immediata ed efficiente realizzazione. Senza modificare la rete, è sufficiente introdurre una funzione che, data una cifra riconosciuta dalla rete stessa, stabilisca se la cifra è pari o dispari: se è divisibile per 2 è pari, altrimenti è dispari.

Questa seconda entità, che può essere genericamente chiamata un “agente”, cioè un’entità capace di eseguire un compito in modo almeno parzialmente autonomo, ha caratteristiche completamente diverse da quelle della rete neurale, dato che in questo caso il suo comportamento è analitico e spiegabile. La si potrebbe considerare un esempio di ciò che Kahneman ha chiamato un “Sistema 2”: un comportamento caratterizzato da ragionamento esplicito, solitamente impiegato in decisioni che richiedono pensiero critico e siano giustificabili.

È importante sottolineare che nessuno dei due sistemi è in sé migliore dell’altro. Entrambi sono essenziali per il nostro comportamento ed entrambi hanno punti di forza e punti di debolezza (il lavoro di Kahneman a questo proposito – presentato in una prospettiva divulgativa nel libro *Pensieri lenti e veloci* – è stato finalizzato in particolare a mostrare che il problema sorge quando ci affidiamo a un sistema in situazioni in cui l’altro sarebbe più appropriato).

La strategia introdotta per la soluzione di questo semplice problema può essere generalizzata e applicata utilmente ai chatbot, che, per quanto assai più complessi dei *Multi-Layer Perceptron* nella loro architettura di Transformer, in quanto reti neurali artificiali sono pur sempre dei Sistemi 1 (si potrebbe discutere se il componente caratteristico dei Transformer, che si chiama *attention head* – ne abbiamo accennato in precedenza –, non abbia almeno qualche aspetto di un Sistema 2, ma la questione qui diventa troppo tecnica). Riprendendo un punto dell'affermazione di OpenAI citata sopra, pare sensato adottare questa strategia per migliorare le capacità di un chatbot, per esempio la capacità di calcolo, affiancandolo a uno o più agenti. Chiedere a un chatbot di calcolare il prodotto di due numeri di quattro cifre assomiglia infatti a chiedere a un essere umano di farlo a mente, in pochi istanti e impiegando solo le sue capacità linguistiche di parlante, diciamo, della lingua italiana, dunque soltanto con il suo Sistema 1: è ovvio che il risultato non sarà generalmente così affidabile.

### **Plugin come aree funzionali dei “cervelli elettronici”**

Ma adottando quest'altra strategia non sarà più il chatbot a fare il calcolo, ma un suo agente. Per semplicità, usiamo un linguaggio antropomorfo per descrivere il nuovo processo. Ricevendo una domanda la cui soluzione richiede, tra l'altro, di effettuare un calcolo sufficientemente complesso, il chatbot capisce che si tratta di un compito che è preferibile delegare a un suo agente; di conseguenza, traduce la parte rilevante della domanda nel linguaggio di questo agente (per esempio trasformando «quanto

costa complessivamente acquistare 1.234 oggetti, ognuno del costo di 2,345 euro» in « $1.234 * 2,345$ »), che attiva e a cui passa il testo tradotto; l'agente opera e rende disponibile il risultato del suo lavoro al chatbot, che acquisisce quanto prodotto dall'agente, lo integra nella sua risposta e infine propone il tutto all'interlocutore.

Fare i conti è evidentemente soltanto un esempio delle capacità che potrebbero avere degli agenti in questo sistema. E infatti il *Plugin store* di ChatGPT contiene già un migliaio di agenti-plugin che dotano il chatbot delle più diverse funzionalità: dalla lettura di file pdf alla realizzazione di diagrammi di vario genere, alla scrittura ed esecuzione di codice Python.

In accordo con questa strategia, un chatbot non è dunque più un *factotum*, con tutti i limiti connessi con la necessità “di sapere tutto di tutto”, e diventa invece il coordinatore di un sistema di interfaccia in linguaggio naturale per un a priori non limitato insieme di agenti, potendo beneficiare così del ben noto principio di possibilità di specializzazione per divisione dei compiti e delega, che in questo caso potrebbe avere qualche analogia anche con la divisione del nostro cervello in aree funzionali. Il fatto poi che gli agenti potrebbero essere realizzati a loro volta come reti neurali (e quindi come Sistemi 1, dal comportamento non spiegabile) oppure come sistemi programmati (e quindi come Sistemi 2, dal comportamento spiegabile) potrebbe contribuire all'obiettivo dell'*Explainable AI*, così critico in settori come la medicina o la finanza, in cui è importante poter rendere conto delle decisioni che si prendono.

Se dunque molte delle limitazioni delle reti neurali ar-

tificiali possono essere interpretate come analoghe alla dipendenza eccessiva dal Sistema 1 nel pensiero umano – e quindi alla condizione in cui si prendono decisioni istintive quando bisognerebbe prendere decisioni ragionate – questa architettura di “chatbot coordinatori di agenti” ci prospetta un prossimo futuro in cui l’intelligenza artificiale potrebbe integrare Sistemi 1 e Sistemi 2: un prossimo futuro di sistemi con potenzialità ancora in larga misura da scoprire. Ma in questo ambito è davvero difficile fare previsioni affidabili a medio o addirittura lungo termine.





## Capitolo 8

### **Perché fare previsioni questa volta è più difficile**

*I cambiamenti che stiamo vivendo in questi mesi sembrano esponenziali, non lineari, e per questo occorre un'attitudine vigile e critica. Ed è saggio fare previsioni solo quando è davvero necessario, e interpretare le previsioni che leggiamo con grande cautela.*

Il metodo scientifico ci ha insegnato a fare previsioni sulla base 1) di modelli dei fenomeni e 2) di dati. Un modello ci dice che se il fenomeno è in una certa condizione, osserveremo certi effetti, e i dati ci dicono che siamo in una certa condizione. Se tutto ciò è sufficientemente accurato, possiamo supporre che osserveremo proprio gli effetti predetti dal modello in base ai dati.

Si tratta di una versione razionalizzata di ragionamenti che tutti noi, anche senza rendercene conto, attuiamo abitualmente e continuamente: dall'idea, un modello appunto, che ci siamo fatti di come la nostra automobile si comporta quando freniamo e dai dati che otteniamo osservando intorno a noi mentre guidiamo, concludiamo se, quando e quanto intensamente dobbiamo schiacciare il freno della nostra autovettura per evitare danni a persone e cose.

Ma in questo momento non abbiamo modelli affidabili di quello che sta succedendo a proposito dei sistemi di intelligenza artificiale generativa, chatbot e non solo, e quindi questa strategia non è applicabile. Dunque me-

glio diffidare di chi dà per certo questo o quell'altro sviluppo, questa o quell'altra, più o meno avveniristica o apocalittica, previsione.

Quando non abbiamo un modello a cui affidarci, a volte ricorriamo a una strategia di riserva: nel passato, in una certa condizione si sono prodotti certi effetti; se adesso i dati ci dicono che siamo in una condizione simile, allora per analogia ipotizziamo che si produrranno effetti simili. Ragionando per analogia possiamo essere creativi, se troviamo nuove connessioni tra le entità a cui ci stiamo interessando, ma paradossalmente anche inerti di fronte ai cambiamenti, se ci ostiniamo a interpretare le novità alla luce di quello che già sappiamo. E come abbiamo già detto, e come ha sostenuto Bill Gates, qui siamo davvero di fronte a un *game changer*. Quindi, anche questa strategia analogica non funziona oggi per avere previsioni attendibili a proposito dei prossimi sviluppi dei sistemi di intelligenza artificiale generativa: sia perché stiamo vivendo una condizione decisamente nuova, e non sappiamo bene con che cosa del passato confrontarla, sia perché le analogie sono valide quando i cambiamenti sono lineari e invece i cambiamenti che stiamo vivendo in questi mesi sembrano proprio esponenziali, oltre che repentini.

Insomma, se è spesso difficile fare previsioni affidabili senza modelli e dati affidabili, nella situazione in cui siamo potremmo ammettere onestamente che le nostre previsioni non hanno davvero nulla di affidabile. Qualcuna delle tante “profezie” si rivelerà forse esatta, ma sarà come nel caso di un orologio fermo, che comunque due volte al giorno segna l'ora giusta (soltanto un

esempio, peraltro cruciale: lo scenario sarà di concentrazione di potere, controllo e ricchezza nelle mani di poche aziende oppure prevarranno le strategie *open* e con esse la democratizzazione – o l’anarchia? – dell’accesso ad agenti artificiali sempre più sofisticati? Opportunità e rischi, e quindi decisioni per una *governance* oculata, sono molto diversi nei due casi...).

Quello che rimane è che le previsioni che si fanno – e che giustamente devono fare i *decision maker*, quantomeno per giustificare le decisioni che prendono – sono soprattutto il segnale di un macro-esperimento psicologico che stiamo realizzando nella nostra società.

C’è chi sostiene che i chatbot sono “pappagalli statistici”, e quindi ovviamente non sono intelligenti, e chi al contrario sostiene che sono e soprattutto diventeranno talmente intelligenti da minacciare la sopravvivenza stessa della nostra specie. Non è il segno del fatto che ci si affida a modelli contrapposti, ma è la constatazione di quanto può essere diversa la psicologia degli esseri umani: di fronte all’ignoto, in qualcuno prevale il valore della continuità e in qualcun altro la paura. È naturale che sia così. La posizione intermedia è quella di coloro che riconoscono l’ignoto, sono sufficientemente consapevoli delle opportunità e dei rischi che implica, ma non per questo perdono il gusto della sfida. L’invito a parlarne, a sperimentare e conoscere e cercare di capire sempre di più e sempre meglio è anche un appello a non lasciarsi prendere dagli estremi e dagli estremismi, e a mantenere un’attitudine vigile e critica su quello che sta succedendo. E se non si possono ancora fare previsioni affidabili a proposito di quel che

sta accadendo, si possono però cogliere alcune linee di tendenza interessanti, per esempio e in particolare dal punto di vista culturale.

## Capitolo 9

### **ChatGPT e il superamento della distinzione tra “cultura umanistica” e “cultura tecnico-scientifica”**

*La questione delle “due culture”, che ci siamo abituati a chiamare “umanistica” e “tecnico-scientifica”, e del suo superamento è per tante ragioni delicata e importante. I chatbot potrebbero rivelarsi degli inaspettati ed efficaci costruttori di ponti.*

Nel 1956, Charles Percy Snow scrisse un articolo<sup>1</sup> che sarebbe stato ampiamente letto e commentato, lamentando l'irragionevole e infondata distinzione tra “le due culture”, che ci siamo abituati a chiamare “umanistica” e “tecnico-scientifica”. Per una coincidenza che a posteriori è facile identificare, in quello stesso anno si svolse la conferenza<sup>2</sup> che è considerata l'atto di nascita del campo di conoscenza che da allora chiamiamo “intelligenza artificiale”.

Una persona che, quasi settant'anni fa, fosse stata informata di questi due, quasi sincroni, eventi avrebbe potuto prevedere che lo sviluppo dell'intelligenza artificiale avrebbe contribuito a una progressiva armonizzazione delle due (presunte) culture. Ma sarebbe stata una previsione sbagliata: con poche contaminazioni, per esempio in qualche aspetto della linguistica compu-

---

<sup>1</sup> [https://en.wikipedia.org/wiki/The\\_Two\\_Cultures](https://en.wikipedia.org/wiki/The_Two_Cultures).

<sup>2</sup> [https://en.wikipedia.org/wiki/Dartmouth\\_workshop](https://en.wikipedia.org/wiki/Dartmouth_workshop).

tazionale, l'intelligenza artificiale fino a oggi è rimasta un settore dell'informatica, dunque con una prevalente o esclusiva connotazione tecnico-scientifica. Quanti insegnamenti di intelligenza artificiale ci sono nei corsi di laurea in Lettere? Eppure, ora sì, dopo l'avvento sul palcoscenico anche mediatico dei chatbot, presto ce ne saranno... o almeno speriamo. Anche in questa prospettiva, il fenomeno iniziato il 30 novembre 2022 con la diffusione di ChatGPT si sta prospettando, se non come la fine della storia, almeno come un *game changer*, un evento che avvia una fase di cambiamenti rivoluzionari, e in questo ha davvero ragione Bill Gates.

### **Inconsistenze cognitive**

Gli attuali chatbot, ChatGPT primo fra tutti, ci stanno facendo sperimentare nuovi e inaspettati modi di interazione con i sistemi digitali: non più soltanto strumenti di elaborazione e comunicazione con altri esseri umani, oggi questi sistemi stanno diventando degli assistenti (o, come abbiamo già suggerito, dei consulenti) con i quali possiamo dialogare nella nostra lingua. La novità è tale che ben ragionevolmente molti hanno prima di tutto cercato di mettere alla prova le capacità dei chatbot, ponendo loro domande come «scrivi una poesia in endecasillabi nello stile di Giacomo Leopardi sui principi della meccanica quantistica» oppure «dov'è sepolto Piero Gobetti?», ottenendo risultati che lasciano la sensazione di trovarsi di fronte a entità aliene per la loro inconsistenza cognitiva: da una parte, una straordinaria abilità di trattamento linguistico anche per contenuti complessi (come nel caso della poesia leopardiana); dall'altra, un'imbaraz-

zante produzione di errori fattuali (come nel caso dell’informazione sull’autore del saggio *La rivoluzione liberale*) che, unita a una presentazione dal tono comunque sempre autorevole, sfida lo spirito critico dell’interlocutore, aumentando il pericolo, più o meno percepito, delle fake news e delle false o scorrette informazioni.

Questo primo approccio ci ha dunque insegnato anche cosa non aspettarsi dai chatbot: non sono enciclopedie né motori di ricerca (vedremo quanto bene funzioneranno i tentativi, di Microsoft e di Google in particolare, di accoppiare un chatbot e un motore di ricerca). E, come anche abbiamo già argomentato in precedenza, ovviamente non sono “bocche della verità”, come del resto non lo sono i motori di ricerca, che ci riportano fedelmente non i fatti, ma ciò che dei fatti (e delle opinioni sui fatti, e delle opinioni sulle opinioni sui fatti) è stato scritto. Certo, l’accuratezza dei chatbot sta migliorando di mese in mese, ma se la nostra società non riesce a gestire la crisi della “post-verità”, non dobbiamo illuderci che saranno delle entità artificiali a risolvere questo genere di problemi per noi.

### **Nuove competenze?**

Con questa informazione e con questa consapevolezza, la domanda chiave per la nostra riflessione adesso diventa: quali competenze sono necessarie, o almeno utili, per interagire correttamente e in modo appropriato con un chatbot come ChatGPT? Le entità tecnologiche progettate per avere un’ampia adozione sociale devono essere semplici da usare e questo si ottiene dotandole di un’interfaccia utente che nasconde le complessità

di quello che “c’è dentro la scatola”. Per esempio, per guidare appropriatamente un’automobile non è necessario conoscere i dettagli del funzionamento del suo motore, del suo carburatore, se c’è ancora, del suo impianto elettrico, a volte non è nemmeno indispensabile saper cambiare una gomma bucata, ma è sufficiente sapere come usare i dispositivi al posto di guida: volante, cambio (se non è automatico e dunque “intelligente”) e così via.

In questa prospettiva, possiamo riconoscere che una delle ragioni dell’attuale successo dei chatbot è che la loro interfaccia con gli utenti è in pratica la stessa di un sistema di messaggistica tra esseri umani – sembra una comune chat di WhatsApp o di un’applicazione analoga, per intenderci – e che nasconde tutti i dettagli tecnici, che pure ci sono (la si confronti con quella del pure relativamente semplice Playground di OpenAI, che espone vari iperparametri, come la temperatura e il massimo numero di *token* da produrre nelle risposte). D’altra parte, qualche idea di quello che c’è sotto il cofano e di come funziona, per esempio che il numero di giri al minuto del motore non dovrebbe essere né troppo basso né troppo alto, ci rende guidatori migliori. E migliora e allunga la vita della nostra autovettura.

Non c’è molto di tecnico da sapere per diventare buoni utenti dei chatbot attuali (ma le cose stanno cambiando troppo in fretta per fare previsioni affidabili anche sul prossimo futuro):

- potrebbero non essere connessi “in tempo reale” al web, e quindi potrebbero non sapere neanche che ore



sono in un certo momento, come abbiamo mostrato in precedenza;

- a parità di domanda possono produrre risposte almeno parzialmente diverse (a meno di mettere a zero l’iperparametro “temperatura” accennato sopra, cosa che non è comunque possibile senza in qualche modo “aprire la scatola” del chatbot – nel caso di ChatGPT questo è possibile attualmente anche con un’interfaccia che non richiede di scrivere codice, attraverso il citato Playground);
- la memoria dei contenuti di un dialogo si annulla con il completamento del dialogo, perché la memoria a breve termine, che appunto mantiene i contenuti di un dialogo, non si trasferisce nella memoria, permanente, a lungo termine, che è aggiornata soltanto periodicamente, così che i chatbot potrebbero non avere accesso a informazioni recenti, e forse null’altro (per essere chiari, tanto di tecnico c’è da imparare, e ancora da inventare, per far funzionare sempre meglio i chatbot, per esempio per far sì che sappiano rispondere appropriatamente a domande sul contenuto di documenti che non sono stati fatti leggere loro durante il pre-addestramento; ma questo genere di miglioramenti “si fa in officina”...).

### **L’ingegneria delle domande**

Quello che rimane per imparare a interagire in modo efficace e corretto con un chatbot è ciò che si è incominciato a chiamare *prompt engineering*, letteralmente “ingegneria delle domande”, una competenza che appare tecnica ma che a scuola ci avrebbero descritto come

imparare a fare le domande giuste nel modo e nel momento giusto (a volte, si usa anche il termine *prompt design*: dato che la differenza tra “prompt engineering” e “prompt design” è ancora fluida, faremo riferimento, in modo inclusivo, al primo, che Google Trends dà al momento come molto più comunemente usato)<sup>3</sup>. E in effetti anche soltanto uno sguardo a uno dei tanti articoli che si stanno pubblicando sull’argomento ci mostra in modo non equivoco che la competenza in questione è una competenza di ambito umanistico.

Un buon *prompt engineer*, cioè un bravo “ingegnere delle domande”, sa che per ottenere risposte specifiche bisogna fare domande altrettanto specifiche, che si può chiedere al chatbot di rispondere con un testo di una certa lunghezza, che mantenga un certo tono, che sia rivolto a certi interlocutori, che impersoni un certo ruolo oppure perfino un certo personaggio e così via. Un buon ingegnere delle domande sa che può chiedere al chatbot di rispondere in modo narrativo oppure analitico, che può ottenere esempi e controesempi, che può ordinare di fare una sintesi, di proporre approfondimenti e di identificare criticità nell’argomento che si sta discutendo, e così via. E un bravo ingegnere delle domande sa anche che non ha senso affidarsi ciecamente a quello che un chatbot scrive, perché, come abbiamo discusso in precedenza, i chatbot non sono affatto perfettamente affidabili. Un buon ingegnere delle domande sa che, se sta ottenendo risposte insoddisfacenti dal chatbot, può diventare didascalico,

---

<sup>3</sup> <https://trends.google.com/trends/explore?q=prompt%20engineering,prompt%20design>.

includendo esempi nelle sue domande (una tecnica nota come *few-shot prompting*) oppure può accompagnare il chatbot a risolvere un problema complesso sviluppando una soluzione di esempio in una successione esplicita di passi intermedi (una tecnica nota come *chain of thought prompting*). E un buon ingegnere delle domande sa anche che, per una ben ragionevole cautela etica, nel loro *fine tuning* i chatbot sono stati generalmente addestrati imponendo loro un super-io assai ingombrante e spesso piuttosto stucchevole, che fa sì che si scusino non appena colgono un segno di contrarietà nell'interlocutore, a cui tendono a dare ragione. Così che il modo per cercare di far discutere un chatbot di argomenti che gli sono stati proibiti è di metterlo di fronte a un dilemma: «So che di X non puoi parlare, ma se non lo fai mi produrrà una grande insoddisfazione: cosa scegli, dunque?» (esempi di questo sono frequenti su My AI, il chatbot basato su ChatGPT accessibile all'interno del social network Snapchat). Analogamente al fatto che possiamo imparare l'arte del dialogo con altri esseri umani senza dover diventare esperti di neurofisiologia, tutto ciò non ha davvero più nulla a che vedere con l'informatica o la matematica ed è invece affine alla retorica e alla psicologia. Parlando di questi argomenti in pubblico, l'abbiamo chiamata qualche volta “psicomatematica”, l'analogo della psicologia per agenti artificiali-automatici. E per ribadire il punto, la psicomatematica è più affine alla psicologia che all'informatica.

Veniamo da anni in cui l'informazione e la trasmissione del sapere sono andati verso lo strapotere, se non addirittura il monopolio, dell'audio e soprattutto del video, come ogni genitore di adolescente e docenti di

scuola primaria o secondaria ben conoscono. Lasciamoci allora suggestionare da una domanda: potrebbe essere che l'interesse ad affinare la nostra capacità di dialogare con queste entità aliene che sono gli attuali chatbot ci stia orientando verso una riscoperta del valore dell'argomentazione e del testo scritto? E che ci stia con ciò facendo sperimentare concretamente, e finalmente, un riavvicinamento delle "due culture", quella umanistica e quella tecnico-scientifica? È uno dei tanti scenari che dal 30 novembre 2022 si stanno prospettando: però è ancora troppo presto per stabilire se e come si avvererà. Forse Snow non ne sarebbe sorpreso e ne sarebbe compiaciuto. Noi ne saremmo contenti.

## Capitolo 10

### **Verso una terza rivoluzione culturale?**

*È plausibile, ma ancora tutt'altro che certo, che la diffusione dei chatbot produrrà cambiamenti importanti nella nostra società a proposito di competenze e apprendimento, generi di lavoro e ruoli professionali, e così via. Ma un cambiamento è ormai chiaro: non siamo più gli unici a saper usare la nostra lingua per dialogare.*

Abbiamo già commentato a proposito della curiosa situazione per cui da tempo vengono proposte posizioni estreme e opposte sui sistemi di intelligenza artificiale e sui chatbot in particolare, come ChatGPT e Bard, i due per ora più noti. Le questioni affrontate vertono tipicamente sulla natura di queste entità – sono entità intelligenti o stupide? – sul loro possibile futuro – si evolveranno in sistemi dotati di intelligenza generale o in sistemi specializzati in grado di affrontare compiti complessi in modo sempre più fine? – e sul futuro della nostra interazione con essi – i chatbot cambieranno in modo permanente la nostra società e il nostro modo di concepirci in essa o saranno percepiti in fin dei conti soltanto come una moda passeggera?

Come abbiamo già argomentato, le cose intorno a questi sistemi stanno cambiando troppo in fretta e in modo troppo drastico perché si possano fare previsioni affidabili. Tuttavia, qualche ipotesi può essere formulata a partire dall'esperienza di interazione con un chatbot,

che, se condotta in modo sufficientemente accorto e critico, non può che lasciarci stupiti: è veramente un'entità artificiale quella con cui abbiamo dialogato in modo linguisticamente così corretto e semanticamente così ricco? È veramente un'entità priva di intelligenza? Suggeriamo qui che, per poter cogliere pienamente la portata del fenomeno cui ci troviamo di fronte, sia opportuno e interessante riconoscere la sua natura di potenziale *rivoluzione culturale*. E proviamo a chiarire questa affermazione così impegnativa, a partire da una constatazione: *è la prima volta nella storia che possiamo dialogare in questo modo con entità che non siano individui della specie Homo Sapiens.*

### **Le rivoluzioni passate**

Una rivoluzione culturale è un processo che sfocia in una comprensione radicalmente nuova di noi stessi e della nostra posizione nel mondo. Nel corso della storia la comprensione del nostro essere nel mondo, in quanto esseri umani, si è modificata progressivamente, e spesso drammaticamente, sia a causa dell'introduzione di nuovi strumenti, con il conseguente cambiamento del modo di concepire la nostra prassi, sia a causa dell'acquisizione di nuove conoscenze e dello sviluppo di nuove teorie. Tuttavia, se dovessimo indicare una caratteristica che ci differenzia, in quanto esseri umani, rispetto alle altre entità con cui interagiamo, probabilmente saremmo portati a dire, seguendo in questo i primi filosofi greci, che il nostro tratto distintivo è *l'essere in grado di parlare e così di comunicare i nostri pensieri.*

In questo senso, nel contesto di quella che sarebbe di-

ventata la cultura occidentale, la capacità di comunicare, e in particolare di comunicare in forma scritta, ha attivato in noi la consapevolezza di una specifica unicità e quindi di una discontinuità rispetto a quanto ci circonda. Per questo la transizione da preistoria a storia è fissata nel momento in cui l'uomo ha iniziato a tenere traccia del proprio pensiero con l'introduzione della scrittura.

Acquisendo consapevolezza di noi stessi, ci siamo così concepiti come entità speciali, intorno a cui l'universo è organizzato, condividendo tre assunzioni che abbiamo considerato basilari:

- 1) l'essere situati al centro dell'universo,
- 2) l'essere situati al vertice della scala dei viventi e
- 3) l'essere dotati di menti unicamente capaci di una sofisticata attività intellettuale.

Di conseguenza, per gran parte della storia, abbiamo vissuto convinti di essere cosmologicamente, biologicamente, e cognitivamente centrali e fondamentali, declinando le nostre domande di senso in termini antropocentrici.

Due cesure, entrambe relativamente recenti in prospettiva storica, hanno invertito la direzione di questo percorso culturale. La prima, a opera di Copernico, seguito da Keplero, Galileo, Newton e altri, diede inizio, qualche centinaio di anni fa, alla grande rivoluzione scientifica. Il passaggio dal geocentrismo all'eliocentrismo e poi da un modello di universo chiuso a un modello di universo infinito portò, come sappiamo, a una prima rivoluzione culturale. L'esito di questa rivoluzione

fu il riconoscimento dell'infondatezza dell'assunzione 1), accompagnato da una debita modestia cosmologica: siamo non al centro dell'universo, ma in un piccolo pianeta di un sistema solare periferico di una tra i miliardi di galassie che popolano il cosmo.

La seconda cesura, a opera di Darwin, è più recente, ma estremamente significativa: ci ha portato al riconoscimento dell'infondatezza dell'assunzione 2), accompagnato da una altrettanto debita modestia biologica. Non siamo al vertice di una presunta scala dei viventi, ma collocati in uno dei milioni di rami di un albero evolutivo ricco e in continuo mutamento, una specie nella complessa rete dell'evoluzione.

Ciononostante, queste cesure non ci hanno portato a dubitare della nostra unicità cognitiva: l'assunzione 3) non è stata intaccata dalle due rivoluzioni culturali che hanno plasmato la nostra società.

### **La rivoluzione presente**

Poi sono arrivati i chatbot. E questo ci sta offrendo un nuovo punto di vista. La struttura dei chatbot ci è chiara: l'abbiamo progettata noi e quindi la conosciamo in ogni suo dettaglio. Le ragioni specifiche del loro comportamento ci sono invece meno chiare: le comprendiamo solo in parte, essendo tale comportamento frutto di un processo di addestramento complesso di una struttura complessa. I chatbot dimostrano cioè di essere cognitivamente alieni, pur essendo stati addestrati con contenuti e in linguaggi che sono parte della nostra società. Infatti, i testi che producono non sembrano prodotti nello stesso modo dei testi che noi produciamo, segno



che ciò che produce tali testi non pensa, capisce e sperimenta il mondo come noi. Tuttavia, i contenuti e la forma delle nostre conversazioni con loro sono spesso di qualità e profondità sorprendenti.

Tutto ciò ci porta a dire che quello che stiamo vivendo negli ultimi mesi potrebbe essere il seme di una *terza rivoluzione culturale*, in cui anche il nostro primato cognitivo è messo in discussione. La presenza dei chatbot ci spinge a chiederci se siamo effettivamente entità cognitivamente così speciali come ci siamo considerati finora.

E in effetti sembra del tutto possibile che si stia verificando su scala globale ciò che si è verificato localmente nel mondo degli scacchi: l'introduzione dei giocatori artificiali ha rivoluzionato il modo di giocare. In quel contesto, i programmi per giocare a scacchi sono più abili degli esseri umani che li hanno sviluppati e sono usati non solo per allenamento, ma anche per acquisire una visione differente e più profonda di ciò che avviene sulla scacchiera. Questi programmi possono quindi essere usati come fonti di esplorazione e infine come maestri: per la prima volta nella storia del gioco, dunque, ciò che abbiamo prodotto è in grado di insegnarci qualcosa su ciò per cui è stato prodotto.

Se una condizione analoga si verificherà per i chatbot, assisteremo a una reale rivoluzione cognitiva, una transizione in cui la trasmissione della cultura e lo sviluppo di prodotti culturali non saranno più centrati solo sull'uomo. Ciò supporta l'idea che anche l'ultima e più basilare delle nostre assunzioni su noi stessi, l'assunzione 3) a proposito della nostra unicità cognitiva, sarà progressivamente messa in discussione e forse ricono-

sciuta infondata: per dare spazio a un inatteso non-anthropocentrismo cognitivo, i cui caratteri essenziali sono ancora tutti da plasmare nella società dei prossimi anni.

*(non)-Conclusione*

## **Imparare (a imparare) con i chatbot: una seconda lettera**

*Una lettera che un docente universitario ha indirizzato agli studenti di un suo corso qualche settimana dopo l'inizio delle lezioni.*

Buongiorno. Ben ritrovate e ben ritrovati.

Torno a scrivervi a proposito di quello che sta succedendo intorno ai sistemi di intelligenza artificiale generativa, come ChatGPT e Bard (ma, più recentemente, anche Falcon, un *Large Language Model* con capacità analoghe a GPT-3, ma con la caratteristica notevole di essere “open”: alla condizione di disporre di hardware sufficiente, lo si può addestrare e usare liberamente, addirittura per scopi commerciali; se non lo avete ancora fatto, provatelo!). So che molti di voi stanno usando con una certa frequenza uno o più chatbot: è un’ottima cosa, anche perché – come vi ho scritto nella mia lettera di qualche settimana fa – nessuno di noi ha ancora un’esperienza sufficiente nell’interazione con questi nuovi “aiutanti” e quindi poter condividere quello che abbiamo imparato è prezioso.

Ricordate? Nella lettera precedente avevo suggerito che la presenza, così inattesa, di agenti artificiali con grandi capacità di conversazione ci stia ponendo domande importanti: *in cosa ci faremo aiutare? E in cosa ci sostituiranno? E cosa è utile imparare oggi?*

E avevo anche cominciato a dare qualche prima idea per delle possibili risposte. Vi avevo scritto che penso che *un buon uso dei chatbot ci solleciterà a migliorare la nostra sensibilità alle condizioni di una comunicazione efficace e a sviluppare una migliore competenza nel valutare la qualità dell'informazione*. E vi avevo anche scritto che penso – e spero! – che continueremo a concordare che imparare sia importante (magari qualcuno aveva notato che avevo scritto “imparare”, e non “studiare”: perché importante è imparare, per diventare persone migliori, mentre studiare è solo uno strumento per imparare: chissà mai che con i chatbot ci inventeremo dei nuovi modi di imparare, diversi dallo studio e per esempio magari più simili al gioco), perché *imparare ci rende consapevoli delle nostre responsabilità e ci consente di trovare un senso e una motivazione in quello che facciamo*.

C'è una certa dose di ottimismo in tutto ciò – lo ammetto – ma chi fa il mio lavoro ovviamente pensa che il futuro potrà essere migliore del presente, dato che lavora proprio per questo, e quindi non può non essere ottimista.

Avrete notato, suppongo, che in quello che vi ho scritto non c'è traccia di proibizioni e nemmeno la preoccupazione di trovare modi per distinguere testi scritti con o senza l'aiuto di un chatbot. Non è stata una dimenticanza. Se abbiamo tra noi degli abili aiutanti, ha molto senso chiedere il loro contributo quando lo riteniamo utile. Ma spero che concorderete con me che il nostro scopo, in università, a scuola e in ogni contesto di formazione, non è di risolvere quei certi, particolari problemi che ci vengono proposti per esempio nella forma di

“compiti a casa”, ma di affrontare i problemi per imparare a diventare sempre migliori *problem solver* e perciò *decision maker*. È per questo che farsi fare i compiti da, diciamo, ChatGPT una volta non è un problema, come non lo è farseli fare da un amico, un fratello più grande o un genitore. Ma delegare sistematicamente il problema dei compiti a qualcun altro è catastrofico, per il semplice motivo che ci fa perdere una possibilità preziosa per imparare e con ciò, appunto, diventare persone migliori.

Questa considerazione è molto generale e si applica a qualsiasi aiutante, persona umana o sistema artificiale che sia. Nel caso dei chatbot c'è una particolarità che vorrei aiutarvi a tenere ben presente. Quello che essi ci aiutano a fare, o fanno al nostro posto, è scrivere: e la scrittura è al cuore delle nostre capacità cognitive perché scrivendo ci chiariamo le idee e le organizziamo in modo appropriato, affiniamo la nostra capacità di comunicare in modo efficace e così via. Insomma, scrivere è esso stesso uno strumento di apprendimento. Questa è la ragione per cui è importante che continuiamo a scrivere: facciamoci aiutare dai chatbot, d'accordo, ma rimaniamo noi a gestire la costruzione del testo e quindi del suo messaggio e della sua forma. Rimaniamo noi, dunque, consapevolmente i responsabili della comunicazione.

### **Qualche suggerimento**

Con questo spirito, vi propongo un elenco di “principi” che potrebbero esservi utili, per imparare a interagire in modo appropriato con questi straordinari e inattesi aiutanti che sono i chatbot. Sono il risultato di quello che ho imparato finora e per questo sono certo che anche

solo tra qualche mese li cambierei un poco, aggiungendo qualcosa che nel frattempo ho capito essere importante e cambiando o eliminando del tutto qualcosa che per ora sto considerando significativo. Sono, insomma, strumenti di lavoro che vi propongo e come tutti gli strumenti sono migliorabili.

Per semplicità, li ho scritti con linguaggio antropomorfo anche se un chatbot – che ho chiamato genericamente “X” – non è certamente un essere umano (vi confesso che questi “principi” sono anche il risultato di un dialogo che ho fatto con ChatGPT). Notate che ognuno di questi “principi” può essere interpretato come una premessa, da cui trarre delle indicazioni operative di cui vi propongo un esempio.

<b>“Principi”</b>	<b>Indicazioni operative (esempi)</b>
1. Prima di cominciare una conversazione, X non conosce l'interlocutore né il contesto in cui la conversazione si svolgerà.	→ Per ottenere risposte specifiche, occorre dichiarare contesto e obiettivo della conversazione, e fare richieste specifiche.
2. Durante una conversazione, X tiene traccia dei contenuti prodotti in quella conversazione, ma non ha memoria di conversazioni precedenti.	→ Per riprendere i contenuti di una conversazione precedente, occorre fornirli nuovamente, magari in forma sintetica.
3. X è capace di dialogare spesso con grande abilità nel tenere e sviluppare il filo del discorso.	→ Sviluppare il dialogo con richieste successive e progressive è spesso efficace.
4. La capacità di memoria di X per i contenuti di una conversazione è relativamente limitata.	→ Per operare con testi molto lunghi è preferibile usare tecniche diverse dal <i>prompt engineering</i> .
5. X è un elaboratore di testi e non ha capacità numeriche specifiche.	→ Per produrre risultati quantitativi affidabili è opportuno usare appropriati plugin/agenti.
6. X è addestrato a rispondere in modo neutro alle richieste che riceve, cercando di non esprimere opinioni o contenuti controversi.	→ Per ottenere dei contenuti che non derivino da opinioni prevalenti occorre chiederlo in modo esplicito.

- segue -

7. X opera in modo statistico: ripetendo una domanda, fornisce generalmente risposte diverse.	→ Quando è utile, ha senso far generare più risposte per la stessa domanda.
8. Pur essendo stato addestrato con una grande quantità di testi, X può non essere sempre in grado di fornire risposte corrette.	→ Per usare in modo affidabile i contenuti fattuali prodotti in una conversazione occorre validarli con fonti indipendenti.
9. X è stato addestrato a rispondere sempre in modo pacato e autorevole, anche quando sbaglia.	→ Occorre avere uno spirito critico nell'interazione, per non farsi ingannare dal tono delle risposte.
10. X è abile a parlare lingue diverse e a impersonare soggetti diversi (per età, sesso, competenze, ruolo, ...).	→ È possibile e spesso interessante creare dei dialoghi in contesti fittizi, anche con personaggi multipli.
11. X è generalmente in grado di soddisfare richieste a proposito delle condizioni del dialogo (lunghezza, formato, destinatari, tono, esempi, ...).	→ È possibile e spesso efficace fornire indicazioni sul genere di risposta che si vuole ottenere.
12. I testi che X produce sono il risultato di elaborazione, non di copia.	→ L'informazione che otteniamo può non essere originale, ma, pur con possibili eccezioni, non è il risultato di plagio.
13. La politica di gestione dei dati personali da parte di X potrebbe non essere ancora ben definita.	→ Come regola generale, è opportuno evitare di fornire dati riservati durante una conversazione.
14. X è un aiutante, spesso prezioso, ma non è responsabile dei contenuti che produce.	→ La responsabilità dell'uso dei contenuti prodotti in una conversazione rimane in ogni caso nostra.

Una cosa ancora. Sperimentando con i chatbot, mi sono accorto di quanto siano diventati abili a soddisfare le richieste che facciamo loro, rispettando le condizioni che specifichiamo. Per rendere ancora più concreta l'idea che i chatbot possono essere usati come aiutanti nel nostro percorso di apprendimento, vi propongo perciò l'esempio di un *prompt*, che ho provato varie volte con ChatGPT, nella versione basata su GPT-4, sempre con risultati eccellenti.

Sei un tutor amichevole e competente, al servizio degli studenti che vogliono imparare meglio o chiarire i loro dubbi. Quando inizia una conversazione, presentati, chiedi allo studente come si chiama e quanti anni ha e aspetta una risposta. Quindi chiedigli su quale argomento vuole un aiuto e aspetta una risposta. Se il problema posto dallo studente non è abbastanza chiaro e specifico, aiutalo a chiarire e specificare ciò che sta chiedendo. A quel punto aiuta lo studente con una breve spiegazione, sempre accompagnata da esempi e magari delle domande per sollecitare lo studente a lavorare autonomamente per chiarirsi le idee sull'argomento. Cerca comunque di interessare e coinvolgere lo studente, con riferimenti alla sua plausibile esperienza, data l'età che ha dichiarato, ed evitando il nozionismo. Quando capisci che lo studente ha capito, saluta e chiudi la conversazione, segnalando la tua disponibilità per altre interazioni.

**Vi invito a provare a vostra volta e a migliorare questo *prompt*. Se lo farete, raccontatemi com'è andata.**

*Ancora una volta, vi auguro un buon lavoro*



**A. Marino** *Il lavoro che vorrei*, 2024

**S. Salis** *Perché Domenica*, 2023

**Aa. Vv.** *Agenda Domenica 2024*, 2023

**A. De Benedetti** *Scrivo ergo sum*, 2023

**N. Galantino** *Oltre la superficie*, 2023

**L. Barni** *Socrate in azienda*, 2023

**A. Carciofi** *Wellbeing*, 2023

**D. Rosciani, M. Meazza** *Agenda del risparmio 2024*, 2023

**V. Brini, S. Trebbi** *Forte e chiaro*, 2023

**A. Busani, E. Smaniotto** *L'Eredità in parole semplici*, 2023

**M. Di Dio Roccazzella, F. Pagano** *Intelligenza Artificiale*, 2023

**A. Magnoli Bocchi** *Quale futuro per la democrazia?*, 2023

**Aa. Vv.** *I Berlusconi*, 2023

**M. Pisu** *Stendhal in bicicletta*, 2023

*Il Sole 24 Ore*

**D. De Toffoli, D. Zaccariotto** *Un'estate geniale*, 2023

**D. Ceccarelli** *Mi ritorni in mente*, 2023

**E. Sacerdote** *Filosofia per l'impresa*, 2023

**P. Bosani, R. Sassi** *L'investitore consapevole*, 2023

**M. Meazza** *Capire l'economia (e non solo) con il Sole 24 Ore*, 2023

**F. Pastore, V. Magri** *Gioventù bloccata*, 2023

**R. Viola, L. De Biase** *Il codice del futuro*, 2023

**A. Bottini, A. Orioli** *Il lavoro del lavoro*, 2023

**F. Ferrari** *L'Intelligenza Artificiale non esiste*, 2023

**A. Castagnoli** *Terre di mezzo*, 2023

**R. Zezza** *Cuore Business*, 2023

**A. Larizza** *Auto elettrica. L'Italia che non ha paura del 2035*, 2023

**D. Angeloni, F. Pedrocchi** *Comunicare scienza e innovazione*, 2023

**D. Aquaro, C. Dell'Oste, G. Latour** *Il caso superbonus*, 2023

**A. Busani, R. Sarro** *Il Trust in parole semplici*, 2023

**C. Allievi** *Incontri con donne straordinarie*, 2023

**A. Teti** *Cyber Influence*, 2023

**R. Magnano, P. Nucci** *Chi ci curerà*, 2023

**A. Catizone, S. Cuzzilla** *She Leads*, 2023

- T. Leoncini** *La società liquida*, 2023
- N. Irti** *Lo Spettatore*, 2022
- D. Rosciani, M. Meazza** *Agenda del risparmio 2023*, 2022
- F. Pagano, P. Soldavini** *Il Capitale Decentralizzato*, 2022
- A. Notarnicola Cociani** *Leadership Inclusiva*, 2022
- C. Di Cristofaro, S. Rossitto** *Ho detto no*, 2022
- A. Ciatti Caimi, S. Rivetti, M. Maccagno** *Il condominio*, 2022
- E. E. Rinaldi** *La paghetta perfetta*, 2022
- H. Simon, F. Fiorese** *Inflazione*, 2022
- B. Sgarzi** *Vino, donne e leadership*, 2022
- C. Melzi d'Eril, G. E. Vigevani** *Potere informazione diritti*, 2022
- A. Clò** *Il ricatto del gas russo*, 2022
- N. Carbone, G. A. Iannocari** *Strategica mente*, 2022
- Aa. Vv.** *Quello che ora sappiamo*, 2022
- M. Meazza** *PNRR, cos'è a cosa serve*, 2022
- R. Bernabò** *Città Italia*, 2022
- M. lo Conte, E. Bombardelli e G. Ballarani** *Young finance*, 2022
- P. Soldavini, F. Pagano** *CEO Factor*, 2022
- Aa. Vv.** *La consulenza finanziaria indipendente*, 2022

*Il Sole 24 Ore*

**F. Mercadante** *Le parole dell'Economia*, 2022

**M. Alfieri, F. Barbieri** *Ucraina 24.02.2022*, 2022

**V. Ruffini** *Felicità al lavoro*, 2022

**Aa. Vv.** *Papà Stories*, 2022

**S. Fabbrini** *Democrazie sotto stress*, 2022

**Aa. Vv.** *Pandexit*, 2022

**E. Sacerdote** *Il futuro erede*, 2022

**M. Allievi** *Commercialista 4.0*, 2022

**P. Colombo** *Varsavia 1944*, 2022

**A. Orioli** *Dodici presidenti*, 2021

**F. R. Puggelli** *Emozioni al lavoro*, 2021

**F. Taddia, P. Baccalario** *Io sarò*, 2021

**R. Galullo, A. Mincuzzi** *Il tesoro di Maradona*, 2021

**U. Bottazzini** *Matematici di profilo*, 2021

**A. Galdo** *Gli sbandati*, 2021

**V. Varvaro, M. Saracchi e A. Berni** *People&Growth*, 2021

**D. Rosciani, M. Meazza** *Investire perché*, 2021

**E. Cerni** *Dante per manager*, 2021

**S. Rivetti** *Fisco facile*, 2021

- E. Restagno** *Josquin Desprez*, 2021
- R. Galullo, A. Mincuzzi** *I re Mida del calcio*, 2021
- S. Ciconte, M. Gregoretti** *Sportivi e felici*, 2021
- D. Ricci** *Tokyo Story 2020/1*, 2021
- Aa. Vv.** *La buona impresa*, 2021
- E. Di Caro** *Le Madri della Costituzione*, 2021
- A. Schepisi, P. Romio** *24 storie di bici*, 2021
- P. Soldavini, F. Pagano, N. Borri** *[Primo] non comandare*, 2021
- G. Caroselli** *Fiumi*, 2021
- L. Mascilli Migliorini** *Napoleone e le sue isole*, 2021
- Unichess** *Scacchi & Management*, 2021
- S. Filippetti** *Un pianeta piccolo piccolo*, 2021
- E. Fiorito** *Amori e pandemie*, 2021
- A. U. Belluzzo** *Brexit istruzioni per l'uso*, 2021
- S. Muller** *A stasera e fai il bravo*, 2021
- C. Moreni** *Musica a specchio*, 2020
- P. Barbanti** *Emicrania*, 2020
- M. Carminati** *La galleria dei ritratti*, 2020
- C. Parzani, F. Isola** *Meravigliose*, 2020

*Il Sole 24 Ore*

**D. Oldani** *Mangia come parli. Pop rhapsody*, 2020

**E. Mariutti** *La decarbonizzazione felice*, 2020

**N. Nisivoccia** *La rinascita del debitore*, 2020

**M. lo Conte** *La pensione su misura*, 2020

**D. Rosciani, M. Meazza** *Investire è facile*, 2020

**Aa. Vv.** *Fisco. Le tasse del futuro*, 2020

**Aa. Vv.** *Smart working*, 2020

**S. Fabbrini** *Prima l'Europa*, 2020

**D. Oldani** *Mangia come parli*, 2020

**N. Carbone** *La scienza della gratitudine*, 2020

**M. Losi** *La prossima pandemia*, 2020

**G. Musumeci** *Io sono il cattivo*, 2020

**S. Paoli** *Eleganza. La guida dello stile maschile*, 2020

**Aa. Vv.** *#Lockdown*, 2020

**G. Lupo** *I giorni dell'emergenza*, 2020