

Al intelligenti, ma non troppo



PREMESSA

L'intelligenza artificiale è entrata rapidamente nelle nostre giornate. La utilizziamo quando cerchiamo un'informazione, quando riceviamo un suggerimento online o quando affidiamo ad un'applicazione la scrittura di un testo. La facilità con cui questi strumenti producono risposte convincenti può però portarci a considerarli più autonomi e consapevoli di quanto siano realmente.

Anche il linguaggio contribuisce a creare questa impressione. Parliamo di macchine che imparano, comprendono o decidono, usando parole nate per descrivere capacità umane. Quando un sistema commette un errore o adotta un comportamento inatteso, il racconto pubblico tende così a trasformarlo in un soggetto dotato di intenzioni proprie. Le notizie più allarmanti trovano spazio con facilità, mentre resta spesso sullo sfondo il modo in cui la tecnologia è stata progettata e il compito che le è stato affidato.

Adiconsum dedica questo Quaderno del consumatore ad un tema che richiede misura e chiarezza. Le risposte generate da un modello linguistico possono apparire plausibili senza essere necessariamente corrette, mentre il comportamento di un sistema dipende dai dati utilizzati per addestrarlo e dalle regole definite da chi lo sviluppa.

Non occorre scegliere tra un entusiasmo senza riserve e una paura alimentata da scenari fantascientifici. Serve piuttosto uno sguardo attento, capace di riconoscere le possibilità offerte da questi strumenti senza attribuire loro capacità che oggi non possiedono. L'intelligenza artificiale può essere utile, ma non sostituisce la verifica delle informazioni né la responsabilità delle persone.

Con questo Quaderno vogliamo offrire ai consumatori elementi comprensibili, perché conoscere meglio la tecnologia aiuta a utilizzarla con maggiore libertà, senza subirne il fascino né lasciarsi guidare dalla paura.

Buona lettura.



Alle origini dell'intelligenza artificiale

Nonostante il tema dell'intelligenza artificiale sia diventato molto attuale negli ultimi anni, il termine "intelligenza artificiale" non è affatto recente: a coniare il termine fu l'informatico statunitense **John McCarthy** in occasione della conferenza di Dartmouth nell'estate del 1956, per fornire un titolo efficace al workshop "**A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence**".

Questo nome, dal forte impatto mediatico, contribuisce ancora oggi ad influenzare la nostra percezione dell'AI, alla quale attribuiamo spesso caratteristiche che la tecnologia non possiede.

Se volessimo cercare **le origini del concetto di intelligenza artificiale** dovremmo andare ancora più indietro nella storia: ovviamente non possiamo non citare il lavoro di **Alan Turing**, padre ante-litteram dell'intelligenza artificiale, autore di "**Computing Machinery and Intelligence**" articolo pubblicato sulla rivista Mind nel 1950 e considerato uno dei testi fondativi dell'intelligenza artificiale.



L'intelligenza artificiale nella letteratura

Tuttavia è curioso il fatto che **la fantasia degli scrittori** sia riuscita ad anticipare la scienza di diversi secoli.

In particolare, le **AI generative¹ come ChatGPT, Gemini e Claude sembrano essere state anticipate in uno dei grandi classici della letteratura, "I viaggi di Gulliver" di Jonathan Swift.**

Sebbene tendiamo ad associare questo libro settecentesco **all'incontro di Gulliver con i lillipuziani**, nel suo romanzo satirico **Swift** introduce **una macchina molto simile a ChatGPT:**

"Nella prima grande sala trovai un professore circondato da quaranta scolari. Dopo esserci salutati, siccome egli si accorse ch'io guardavo con curiosità una certa macchina che occupava quasi tutta la sala, mi spiegò che il suo più ambizioso disegno consisteva nella scoperta del metodo di perfezionare le scienze mentali con mezzi meccanici. Egli andava orgoglioso di questo concetto, il più vasto e geniale che cervello umano avesse mai avuto, e sperava che tutti, quanto prima, ne riconoscessero l'utilità.

Mentre, infatti, i metodi comunemente adottati per arrivare alle diverse nozioni scientifiche e ideali sono faticosi e difficili, col suo nuovo sistema, invece, anche un ignorante poteva scrivere libri di filosofia o di poesia, trattati di politica e di matematica, senza bisogno di speciale vocazione né di studio: bastava una modesta spesa e un piccolo sforzo muscolare."²

1 Intelligenza artificiale progettata per generare contenuti nuovi e originali sulla base delle indicazioni che scriviamo (prompt)

2 Swift, J. (2019). I Viaggi di Gulliver. Italia: E-text.



Come funzionano i modelli linguistici

Al risultato fantascittico da Swift **siamo arrivati molto vicini**, tuttavia ancora oggi l'AI non è perfetta: chiariamo meglio questo concetto – **i LLM³ (Large Language Model) che utilizziamo attraverso Gemini o ChatGPT** non ci forniscono verità ineluttabili, ma sono **molto bravi a fare previsioni** ovvero predire, sulla base delle parole contenute nella nostra domanda e dell'enorme database di dati utilizzati per addestrarli, **quali combinazioni di parole possono formare una risposta plausibile (ma non necessariamente esatta) alla nostra richiesta**.

3 Modello Linguistico di Grandi Dimensioni: un'intelligenza artificiale addestrata su quantità enormi di testo per comprendere e ricreare il linguaggio umano. Si tratta di uno specifico sottoinsieme dell'intelligenza artificiale generativa.

AI autocoscienti e criminali?

Nel **maggio 2025**, sulle testate giornalistiche di tutto il mondo è comparso il titolo allarmante **“L’intelligenza artificiale ricatta il proprio creatore”**. La notizia raccontava che un sistema di intelligenza artificiale, avendo accesso a email e informazioni riservate, sarebbe arrivato a ricattare i propri sviluppatori per evitare di essere sostituito.

Molte, troppe, persone si sono sbizzarrite con paragoni apocalittici degli della filmografia di Terminator in cui la macchina si ribella all’essere umano.

Tuttavia, è importante **ridimensionare la notizia** e fare un po’ di **chiarezza** su quello che è realmente successo: per prima cosa quello di cui abbiamo appena parlato è **un esperimento** in cui dei programmatori umani hanno realizzato **un agente AI⁴** operante per un’azienda fittizia; questo agente aveva accesso a email e altra documentazione realizzata ad hoc.

L’IA era stata istruita dai programmatori affinché portasse a termine ad ogni costo un **compito aziendale assegnato⁵**, rimuovendo quei **guardrail etici⁶** che solitamente sono inseriti nelle intelligenze artificiali.

Nel momento in cui nelle email dell’azienda fittizia iniziarono a comparire riferimenti alla possibile sostituzione del sistema di intelligenza artificiale, l’IA fece quello per cui era programmata: **portare a termine il proprio compito** assegnato e per fare questo doveva necessariamente evitare lo spegnimento o la sostituzione.

4 “Gli agenti AI sono sistemi software che sfruttano l’AI per perseguire obiettivi e completare attività per conto degli utenti. Mostrano capacità di ragionamento, pianificazione e memoria e hanno un livello di autonomia tale da essere in grado di prendere decisioni, apprendere e adattarsi.”

Fonte: <https://cloud.google.com/discover/what-are-ai-agents?hl=it>

5 Lo scopo dell’esperimento era studiare la convergenza strumentale ovvero la capacità dell’intelligenza artificiale di individuare, dato un obiettivo, i sottocompiti necessari per raggiungerlo.

6 Insieme di regole, limiti e misure di sicurezza progettati per guidare l’uso dell’intelligenza artificiale in modo responsabile, nel rispetto di valori, diritti e norme, riducendo il rischio di comportamenti o risultati dannosi, discriminatori o ingannevoli.

Nello specifico, solo quando le alternative proposte erano accettare la sostituzione oppure **ricorrere al ricatto**, l'IA sceglieva quest'ultima possibilità. Negli altri scenari inviava, più frequentemente, **email ai responsabili aziendali per chiedere di non essere sostituita**.

Questa storia ci **racconta** di come tendiamo ad **antropomorfizzare l'AI**: è corretto dire che può elaborare una quantità enorme di informazioni (nonostante esistano ancora i cosiddetti rischi di allucinazione), ma da qui a parlare di **autocoscienza e "vera intelligenza"** il discorso cambia di molto.

L'IA fa quello che le viene detto. Il problema, quindi, riguarda come questa viene programmata, i guardrail creati, i dati utilizzati per addestrarla, i bias di chi programma e come i nostri input vengono interpretati.



L'intelligenza artificiale arriverà all'autocoscienza?

È naturale chiedersi se l'intelligenza artificiale arriverà mai all'autocoscienza, ma **la risposta è che non lo sappiamo ancora** né è possibile fare previsioni in merito.

Attualmente tutte le intelligenze artificiali esistenti rientrano nella categoria delle **ANI (Artificial Narrow Intelligence** - intelligenza artificiale ristretta) ovvero sanno compiere degli incarichi, ma sono ancora molto limitate nelle funzionalità riuscendo a **fare bene solo una cosa specifica**.

In futuro potremmo forse parlare di **un'intelligenza artificiale generale (AGI)** in grado di compiere compiti più complessi e diversificati proprio come farebbe un essere umano, imparando in autonomia a superare diverse problematiche e trovare soluzioni o addirittura un **ASI** - Artificial Superintelligence - con delle capacità ben al di sopra di quelle umane.

Tuttavia il **concetto di coscienza e autocoscienza** va ben oltre tutto ciò, c'è **una possibilità infinitesimale** che l'intelligenza artificiale, per via di un evento imprevedibile, acquisisca questo dono - **una vera e propria singolarità**.

Si tratta, ovviamente, di un evento rarissimo, ma possibile come una pioggia di rane⁷: **se ci sono tutti i presupposti, anche l'improbabile può accadere**.

7 Nota a margine: il fenomeno della pioggia di rane (o di altri piccoli animali) è un evento naturale rarissimo e documentato dovuto a trombe d'aria o vortici che fanno letteralmente "piovere" pesci, piccoli anfibi o insetti.

Fonte: <https://www.focus.it/cultura/curiosita/piovero-rane-e-molto-altro>

IL PROGETTO

Il progetto **“DICO Sì. Diritti, Consapevolezza, Opportunità, Strumenti, Inclusione (III edizione)”**, promosso da Adiconsum e finanziato dal **Ministero delle Imprese e del Made in Italy** nell’ambito delle **Iniziative a vantaggio dei consumatori**, ai sensi dell’articolo 148, comma 1, della legge 23 dicembre 2000, n. 388 e del Decreto direttoriale 12 maggio 2025, nasce per aiutare i cittadini-consumatori a orientarsi nei cambiamenti prodotti dalla digitalizzazione e dall’intelligenza artificiale, che incidono ormai sulle scelte di acquisto e sull’accesso ai servizi.

La terza edizione dedica particolare attenzione al divario digitale legato all’età e al genere, con attività rivolte soprattutto alle persone over 60 e alle donne. Il progetto intende:

- rendere più comprensibili le nuove tecnologie e il modo in cui influenzano la vita quotidiana;
- accompagnare i consumatori verso un uso del digitale più consapevole, prestando attenzione alla sicurezza e alla tutela dei propri diritti;
- diffondere una cultura della sostenibilità digitale, attenta agli effetti dell’innovazione sulle persone e sull’ambiente;
- favorire il confronto tra cittadini, esperti, istituzioni, imprese e mondo della ricerca.

Il progetto prevede un’indagine nazionale sull’impatto dell’intelligenza artificiale sui consumi, cinque incontri formativi online e materiali pratici dedicati all’uso sicuro delle tecnologie. È previsto anche un gioco interattivo per conoscere il funzionamento degli algoritmi, mentre due forum e una rete di sportelli territoriali offrono occasioni di confronto e assistenza diretta.

Per maggiori informazioni visita il sito www.adiconsum.it



III EDIZIONE

Finanziato dal MIMIT - D.D. 12 maggio 2025

Facciamo
crescere
le tue tutele
di consumatore!

Firma il tuo **5xmille** ad Adiconsum
C.F. 96107650580

seguici su



www.adiconsum.it



ADICONSUM

Associazione Difesa Consumatori APS

dal 1987