

# DA TURING A CHATGPT: LA STORIA DEI CHATBOT CONVERSAZIONALI



## SOMMARIO

- Da Turing ai primi Chatbot
- Intelligenza Artificiale e Chatbot
- Natural Language Programming
- I moderni Chatbot conversazionali: prima e dopo ChatGPT
- ChatGPT e il cambio di paradigma
- Quale futuro per i chatbot "tradizionali"?
- Tutti i volti dei chatbot

---

*A computer program that can hold a conversation with a person, usually over the internet.*

**Oxford Dictionary**

*Non sarà l'intelligenza artificiale a rubarti il lavoro, ma colui/colei che utilizzerà l'intelligenza artificiale nel proprio lavoro.*

**Anonimo**

---

## DA TURING AI PRIMI CHATBOT

Probabilmente Alan Turing è conosciuto ai più, grazie al film *The imitation game* del 2014, come il matematico inglese che alla fine della Seconda guerra mondiale ha dato il proprio contributo alla decrittazione dei messaggi nazisti crittografati con la macchina Enigma.

Non solo, perché Turing è anche l'autore di un famoso articolo del 1950 dal titolo "*Computing Machinery and Intelligence*" in cui ipotizza la misurazione dell'intelligenza di una "macchina" attraverso la sua capacità di simulare una conversazione senza che il suo interlocutore umano se ne possa accorgere. Questa prova è passata alla storia come **Test di Turing**. Il test fu chiamato da Turing, nell'articolo del '50, *The Imitation Game*, da cui anche il titolo del film del 2014.

Il superamento del test, in estrema sintesi, prevede che a due entità, una umana ed un'altra software vengano poste delle domande da un "giudice" il quale, in base alle risposte, non sia in grado di stabilire se l'entità che ha fornito la risposta sia umana o software. Fino ad oggi il test di Turing ha resistito ma, come vedremo anche in questo documento, le cose stanno cambiando molto in fretta, o forse sono già cambiate.

Lo scopo di questo testo è quello di raccontare brevemente la storia dei Chabot, da Turing fino ad oggi così da conoscerne gli aspetti e comprendere quali potrebbero essere le sue evoluzioni future.



## INTELLIGENZA ARTIFICIALE E CHATBOT

Nel lavoro di Alan Turing non si parla ancora esplicitamente di **Intelligenza Artificiale** e **Chatbot**: i termini, infatti, sono stati coniati anni dopo. In particolare il termine Intelligenza Artificiale farà la sua comparsa per la prima volta in un workshop del '56 a cura di John McCarthy (Dartmouth College), Marvin Minsky (Harvard University), Nathaniel Rochester (IBM), Claude Shannon (Bell Telephone Laboratories) circa 2 anni dopo la morte del grande matematico inglese avvenuta nel 1954. Tuttavia, è evidente che nel suo lavoro e nel suo pensiero visionario, ci siano i primi segnali della rivoluzione che stiamo vivendo oggi.

Il termine chatbot, invece, è stato coniato solo agli inizi degli anni '90 da Michael Mauldin, creatore del chatbot Verbot, come unione dei termini chat e robot. Quello che è considerato il primo chatbot della storia è però **ELIZA**, sviluppato nel 1966 da Joseph Weizenbaum, Professore di Computer Science al Massachusetts Institute of Technology (MIT) che era stato profondamente influenzato dal lavoro di Alan Turing.

**ELIZA fu progettato per essere in grado di rispondere a domande aperte**, ed era in grado di riconoscere parole chiave o intere frasi all'interno dell'input dell'utente e rispondere usando queste parole chiave attraverso risposte preconfigurate, così da dare l'impressione all'utente di una conversazione "reale".

Come vedremo in questo documento, questo è lo stesso approccio che è oggi alla base dei moderni chatbot commerciali presenti sul mercato.

## NATURAL LANGUAGE PROGRAMMING

La tecnologia alla base di ELIZA, così come quella dei moderni chatbot presenti sul mercato, è conosciuta con l'acronimo **NLP, Natural Language Programming**. La NLP è una branca dell'intelligenza artificiale mirata allo **studio del linguaggio e alla possibilità di trattare quest'ultimo attraverso algoritmi e modelli automatizzati**.

Ricerche in questo ambito iniziano già dagli anni '40 sul finire della Seconda guerra mondiale, ma è solo negli anni successivi che, grazie alla disponibilità di computer e risorse di calcolo sempre più a buon mercato, questa branca riesce ad evolvere.

Applicazioni di NLP sono ad esempio:

- **NER (Named Entity Recognition)**: la capacità di riconoscere all'interno di un testo entità come nomi di aziende, nomi di persona, luoghi e indirizzi, date.
- **POS (Part of Speech Recognition)**: la possibilità di riconoscere parti sintattiche di un testo come verbi, avverbi, aggettivi, articoli, nomi comuni.
- **Topic extraction e Topic modeling**: la capacità estrarre dal testo concetti e/o parole chiave che in qualche modo identificano l'argomento di cui si parla nel testo stesso. Un utilizzo pratico di quest'ultima funzionalità è quella di classificare automaticamente dei testi/documenti analizzandone il contenuto.

È evidente, quindi, come l'uso combinato di queste ed altre funzionalità ancora più sofisticate, insieme a potenza di calcolo a basso costo ed ingenti quantità di dati oggi disponibili, possa portare ad ottenere risultati mai visti prima e far diventare realtà il sogno di Alan Turing: quello cioè di **avere una macchina le cui risposte siano indistinguibili da quelle di un essere umano**.

## I MODERNI CHATBOT CONVERSAZIONALI: PRIMA E DOPO CHATGPT

**ChatGPT**, nell'universo dei chatbot conversazionali, rappresenta uno spartiacque: c'è stato un prima e ci sarà un dopo ChatGPT, proprio com'è già capitato altre volte nella storia della tecnologia (pensiamo al prima e dopo l'iPhone). ChatGPT ha portato infatti il mercato e la competizione tra chatbot ad un nuovo livello orientando anche una **nuova customer experience** a cui, presto o tardi, tutti dovranno adeguarsi.

Chi ha provato ad utilizzare ChatGPT rispetto ai classici chatbot che si trovano, ad esempio, sul sito web della banca online, del proprio fornitore di telefonia, del supermercato avrà notato un livello di fluidità della "conversazione" e di "comprensione" delle richieste anche in diverse lingue, che, complice anche la scelta di scrivere la risposta come in un Word Editor, danno la sensazione all'utente che dall'altra parte ci sia davvero un'entità intelligente che sta scrivendo.

Una cosa che invece accompagna, almeno fino a oggi, l'esperienza con i chatbot tradizionali, rispetto a ChatGPT, è un certo senso di frustrazione, quando si rischia di entrare nel loop per cui il chatbot non "capisce" esattamente il nostro intento e risponde in maniera apparentemente casuale o errata, portandoci, il più delle volte, ad abbandonare la conversazione o a chiedere di parlare con un operatore umano, vanificando, di fatto, l'obiettivo del chatbot stesso nell'ambito del customer support: quello di migliorare i tempi di risposta scaricando l'operatore da attività più banali ed automatizzabili.

Ma perché questo succede?

Il problema non è nella tecnologia e nei modelli in sé che, come abbiamo visto parlando di NLP, possono essere molto potenti e sofisticati, ma è essenzialmente nel modo in cui queste piattaforme sono state progettate e che, come vedremo, non è molto dissimile dall'idea originale con cui è stato realizzato il primo chatbot ELIZA negli anni '60.

Realizzare oggi un chatbot mediante le piattaforme presenti sul mercato significa essenzialmente disegnare un flusso conversazionale, più o meno complesso, attraverso una interfaccia code-less che permetta al chatbot di far evolvere la conversazione secondo degli step predefiniti.

**L'applicazione di tecniche di NLP avviene nel momento in cui si deve addestrare il chatbot a "capire" quello che l'utente umano sta chiedendo: questo approccio è definito tecnicamente intent based.**

Un intent è essenzialmente "il motivo" per cui un utente contatta il chatbot e che, quest'ultimo, dovrebbe rilevare e gestire.

Si tratta a tutti gli effetti di una macchina a stati finiti, dove ciascun intent può essere assimilato ad uno stato della macchina e la risposta dell'utente è l'evento che determina la transizione da uno stato all'altro.

Facciamo un esempio.

Ipotizziamo che una catena di alberghi renda disponibile sul proprio sito web un chatbot o meglio un concierge digitale, la cui finalità sia quella di dare informazioni agli utenti ma anche quella di permettere la prenotazione di soggiorni.

Ipotizziamo che l'utente contatti il chatbot scrivendo:

*"Buongiorno, vorrei riservare una stanza con check-in il giorno X e check-out il giorno Y"*

In questo caso il chatbot dovrà fare in linea di principio almeno queste 5 operazioni:

1. Rilevare l'intent ovvero sia "PRENOTAZIONE STANZA" rilevando anche che, nella frase scritta in linguaggio naturale dall'utente, quest'ultimo ha già inserito le date di arrivo e partenza
2. Fare ulteriori domande all'utente, ad esempio relativamente al numero di occupanti ed alla tipologia di stanza
3. Controllare la disponibilità di quanto richiesto informando l'utente o della mancanza di disponibilità o del prezzo del soggiorno
4. Riservare fisicamente la stanza, qualora l'utente decida di procedere, attraverso il sistema di prenotazione dell'albergo, gestendone il pagamento
5. Dare riscontro all'utente, ad esempio via e-mail, dell'avvenuta operazione.

## I MODERNI CHATBOT CONVERSAZIONALI: PRIMA E DOPO CHATGPT

Qualora l'utente, invece, si fosse limitato a scrivere: *"Buongiorno, vorrei riservare una stanza"*, ecco che il chatbot dovrebbe, nel formulare ulteriori domande, chiedere anche le date di arrivo e partenza, in quanto non espressamente dichiarate dall'utente.

Da questo semplice esempio si capisce quali possano essere **le criticità e le esigenze** nel progettare un chatbot intent based per un uso aperto al grande pubblico di internet:

1. Bisognerebbe progettare a priori tutti i possibili flussi della conversazione per fare in modo che questa abbia una certa fluidità
2. Bisognerebbe addestrare il chatbot con il maggior numero possibile di modi di dire in cui un utente potrebbe formulare la stessa richiesta in linguaggio naturale, per evitare di entrare nel loop di *"non ho capito forse intendevi dire..."*

## CHATGPT E IL CAMBIO DI PARADIGMA

Ma vediamo da vicino di cosa si tratta.

ChatGPT è a tutti gli effetti un "generatore" di testo: infatti il GPT che compare nel suo nome è l'acronimo di **Generative Pre-trained Transformer**.

Si tratta infatti di un **generatore di testo pre-addestrato con una mole impressionante di dati in diverse lingue**, in cui l'**input dell'utente, definito in gergo prompt, è la richiesta e, quella che sembra essere la risposta, altro non è che il completamento più probabile della richiesta**, dando l'impressione all'utente di aver ricevuto una risposta o ancora meglio di aver sostenuto una conversazione.

Facciamo un esempio: una delle applicazioni di cui si parla maggiormente è la capacità di questi sistemi di "generare" codice software. In realtà il chatbot non sta generando davvero codice, ma sta "semplicemente" completando la richiesta dell'utente con quello che è il completamento più probabile.

Per raccontarlo attraverso un'angolazione diversa, è una versione potenziata di StackOverflow (quello che fino ad oggi rimane il sito web punto di riferimento per gli sviluppatori a livello globale) con il vantaggio di non dover spulciare intere pagine web e thread per trovare quello che si cerca.

Sotto il "cofano" di ChatGPT, c'è GPT-3 che è la versione sviluppata da OpenAI (la società che ha rilasciato ChatGPT) di una applicazione di NLP conosciuta come LLMs, acronimo di **Large Language Model**. GPT-3 è addestrato con enormi quantità di dati testuali al fine di individuare relazioni tra parole e concetti all'interno dei testi. Questi modelli sono esplosi negli ultimi anni grazie alla potenza computazionale sempre più a buon mercato, e alla disponibilità enorme di dati da usare come training set.

Senza voler scendere in dettagli troppo tecnici, alla base dei modelli LLMs c'è una rete neurale che è l'applicazione di una particolare branca del Machine Learning, anche nota come Deep Learning.

In breve: **"addestrare" una rete neurale significa risolvere delle equazioni matematiche abbastanza complesse al fine di calcolare quelli che si definiscono iper-parametri della rete**. In GPT-3, il motore alla base di ChatGPT, sembra ci siano oltre 100 billion di parametri, mentre il training set di partenza sembra sia di oltre 500 GB di dati provenienti da diverse sorgenti come: libri, articoli, pagine web, documenti vari, in diverse lingue.



## QUALE FUTURO PER I CHATBOT “TRADIZIONALI”?

Davvero quindi, dopo quanto abbiamo appena raccontato, ChatGPT può segnare la fine dei chatbot conversazionali per come li abbiamo conosciuti fino ad oggi e descritti in questo articolo?

La risposta è sì e allo stesso tempo no.

Per inquadrare il problema, dobbiamo operare una distinzione tra:

- **Chatbot B2C**, quelli generalisti destinati al grande pubblico della rete (il chatbot pubblicato sul sito web), a cui un’infinità di utenti potrebbe chiedere potenzialmente di tutto ed in qualsiasi modo. In questo scenario progettare un chatbot intent based con performance anche solo vicine a quelle di ChatGPT è praticamente impossibile. Oggi, infatti, molti dei chatbot pubblicati su internet da società anche importanti, ha performance conversazionali al limite, cosa che ha contribuito a generare, in questi anni, un certo scetticismo verso questo tipo di tecnologia.
- Discorso diverso per i **chatbot B2B**, cioè quei chatbot che un’azienda potrebbe decidere di rilasciare per un uso “interno”, ad esempio per migliorare il proprio service desk aziendale, efficientando il servizio e scaricando gli operatori da attività ripetitive e a scarso valore aggiunto.

In questo caso la scelta migliore, se non obbligata, è proprio quella dei chatbot intent based perché:

1. Gli utenti sono al più i dipendenti dell’azienda
2. Il perimetro è circoscritto e ben definito (es. il service desk)
3. Il vocabolario degli utenti è molto specifico e legato all’ambito di business e tecnologico dell’azienda di appartenenza
4. La quantità di dati aziendali per “addestrare” il chatbot è così esile, che pensare di realizzare un sistema che richiede milioni di dati di training è praticamente impossibile
5. La lingua è la stessa per tutti gli utenti
6. Maggior controllo della privacy

In questo secondo scenario, non solo i chatbot “tradizionali” e ChatGPT non sono in concorrenza ma anzi, il secondo potrebbe essere di aiuto per progettare i primi.

Progettare un chatbot intent based, abbiamo visto, consiste essenzialmente nell’individuare gli intent, disegnare i flussi conversazionali, e addestrare il chatbot con diverse varianti di frasi con cui un utente potrebbe richiedere quell’intent come “PRENOTAZIONE STANZA” nell’esempio dell’albergo.

ChatGPT, in ultima istanza, è un generatore di testo a partire da un prompt. Ecco quindi “la quadratura del cerchio”: **ChatGPT diventa il sistema con cui generare, in automatico, decine se non centinaia di frasi con cui addestrare un chatbot tradizionale a riconoscere i diversi intent dell’utente.**

Questo approccio ha diversi vantaggi tra cui:

1. Evitare agli sviluppatori una quantità di lavoro manuale e noioso come quello di “inventarsi” frasi con cui addestrare il chatbot.
2. Rendere più veloce ed automatizzabile lo sviluppo ed il test dello stesso chatbot.
3. Aumentare, in ultima analisi, la qualità del prodotto finale.



## TUTTI I VOLTI DEI CHATBOT

Un'ultima considerazione riguarda i dati.

Abbiamo visto come, addestrare modelli generativi tipo GPT-3, richieda una quantità enorme di dati di training, ma questi dati da dove arrivano?

È gioco forza che vengano presi da fonti diverse e soprattutto dal web: questo ha creato e creerà **inevitabilmente problemi legati al diritto d'autore, alla privacy e non ultimo problemi etici.**

Quest'ultimo aspetto, in particolare, è legato alla qualità e alla polarizzazione degli stessi dati: se addestriamo un modello con dati sbilanciati, non uniformi o non correttamente etichettati e filtrati, si corre il rischio reale di avere dei sistemi che genereranno risposte discriminatorie, razziste, incitanti all'odio.

Infine, un altro aspetto da considerare è legato all'**attività manuale che serve per creare ed etichettare dati di training di qualità.** Si tratta infatti, tipicamente, di un'attività manuale a basso valore aggiunto che spesso viene appaltata in outsourcing in condizioni di lavoro non sempre ideali, come ha rivelato un'inchiesta del Time proprio relativamente a ChatGPT.

Insomma, tra nuovi equilibri e nuove tecnologie, **i chatbot promettono di essere i protagonisti del prossimo futuro.**

## RIFERIMENTI

1. <https://phil415.pbworks.com/f/TuringComputing.pdf>
2. <https://analyticsindiamag.com/turing-test-key-contribution-field-artificial-intelligence/>
3. <https://www.forbes.com/sites/gilpress/2016/12/30/a-very-short-history-of-artificial-intelligence-ai/?sh=5e2f2eea6fba>
4. <https://analyticsindiamag.com/story-eliza-first-chatbot-developed-1966/>
5. <https://www.interlogica.it/en/insight-en/chatbot-history/>
6. [https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview\\_history.html](https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html)
7. <https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286>
8. <https://zapier.com/blog/how-does-chatgpt-work/>
9. <https://www.nposistemi.it/chatbot-cosa-sono-e-come-progettarli/>
10. Kate Crawford, "Né intelligente né artificiale. Il lato oscuro dell'IA", ed. il Mulino, 2021
11. <https://time.com/6247678/openai-chatgpt-kenya-workers/>
12. [https://it.wikipedia.org/wiki/Automa\\_a\\_stati\\_finiti](https://it.wikipedia.org/wiki/Automa_a_stati_finiti)

