Chiara Martino

Intelligenza Artificiale Conversazionale

Processi, strumenti e professioni per creare chatbot e assistenti vocali

MANUALI FrancoAngeli

FrancoAngeli
Edizioni La passione per le conoscenze

Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con **Adobe Acrobat Reader**

La versione completa dell'e-book (a pagamento) è leggibile **con Adobe Digital Editions**.

Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta **cliccando qui** le nostre F.A.Q.

Chiara Martino

Intelligenza Artificiale Conversazionale

Processi, strumenti e professioni per creare chatbot e assistenti vocali

MANUALI FrancoAngeli

Isbn: 9788835157274

Progetto grafico di copertina di Elena Pellegrini

A Giulio, Ludovica e Marcello,
che mi hanno sempre incoraggiata
a lanciare il cuore oltre la pagina bianca.

Indice

Introduzione

Fino a qualche anno fa, l'idea di dialogare con i dispositivi elettronici sembrava pura fantascienza. Oggi, è una realtà: parliamo con smartphone, smartwatch, computer, tablet, smart speaker, automobili e loro ci rispondono.

Può sembrare magia, può sembrare che questi strumenti siano dotati di un proprio intelletto, e questo può generare stupore, diffidenza e perfino timore.

In questo libro, sbirceremo dietro le quinte di questo spettacolo di magia e **scopriremo come fa l'intelligenza artificiale a capirci, a risponderci e a eseguire le nostre richieste**, e soprattutto quanto lavoro umano è ancora necessario per progettare e implementare questi prodotti.

Il settore che si occupa di creare queste tecnologie si chiama *Conversational AI*, che sta per *Conversational Artificial Intelligence*, in italiano 'IA conversazionale' o 'intelligenza artificiale conversazionale'. Nel corso del libro, useremo queste espressioni in modo equivalente.

La Conversational AI offre **opportunità lavorative variegate** e in continua evoluzione, che non riguardano solo chi ha una formazione STEM, ma anche chi ha una **formazione umanistica**, per almeno due ottime ragioni: per prima cosa, realizzare un oggetto che sia in grado di dialogare in modo naturale e che non risulti artificioso implica una profonda conoscenza delle dinamiche linguistiche che regolano il linguaggio e la conversazione tra persone; inoltre, per creare soluzioni conversazionali non è necessario saper programmare, perché si possono usare software che consentono di configurare tutte le regole tramite interfaccia grafica e quindi senza che sia necessario scrivere codice.

Questo libro si rivolge proprio a chi è interessato ad approfondire come si progettano e sviluppano chatbot e assistenti vocali: è pensato come un per-

corso, che parte dalle definizioni di base e approfondisce **tutte le conoscenze fondamentali per diventare esperti del settore**.

Nel primo capitolo, inizieremo a esaminare i **concetti chiave**: capiremo perché la Conversational AI si chiama così e cosa sono le interfacce conversazionali, l'intelligenza artificiale (AI), il machine learning (ML) e l'elaborazione del linguaggio naturale (NLP); approfondiremo il concetto di conversazione, i meccanismi linguistici che la regolano e le sfide linguistiche da affrontare quando si prova a riprodurla artificialmente.

Nel secondo capitolo, ci addentreremo più in profondità nella terminologia usata in questo settore e capiremo quali sono i principali tipi di prodotti conversazionali, in cosa differiscono e come vengono classificati, ma anche quali sono le **principali applicazioni** di questi prodotti in ambito business.

Nel terzo capitolo, osserveremo **come funzionano** le soluzioni conversazionali e familiarizzeremo con i vari moduli che compongono uno strumento di questo tipo: il riconoscimento automatico del parlato (ASR o STT), la comprensione del linguaggio naturale (NLU), la generazione del linguaggio naturale (NLG), i modelli linguistici di grandi dimensioni (LLM), la sintesi del parlato (TTS).

Nel quarto capitolo, proseguiremo il viaggio esaminando le **opportunità professionali** offerte da questo settore e vedremo quali sono le **fasi di un progetto conversazionale tipo**.

Nei capitoli successivi, vedremo più da vicino le responsabilità di ciascuna di queste aree professionali. Nel quinto capitolo, infatti, approfondiremo il ***Conversation Design human-centred***, cioè il processo di ricerca, ideazione e test, che porta alla progettazione dell'esperienza conversazionale nel suo insieme e mette al centro le persone. Nel sesto capitolo, vedremo cosa vuol dire **Prompt Engineering** e come questa disciplina nuovissima può essere applicata alla creazione di prodotti conversazionali. Nel settimo capitolo scopriremo cosa fa un **Knowledge Engineer** per gestire i dati e come può usare le piattaforme di programmazione visuale per **implementare soluzioni conversazionali anche senza saper scrivere codice**. Nell'ottavo capitolo, passeremo in rassegna i principali criteri di analisi e le **metriche per valutare le performance** delle interfacce conversazionali.

Nel nono capitolo, infine, allungheremo lo sguardo verso l'orizzonte e verso il **futuro della Conversational AI**. I contenuti delle prossime pagine sono infatti una fotografia dell'attuale stato dell'arte di questo settore dinamico e di queste tecnologie in continua evoluzione.

E ora che il tragitto è delineato, non mi resta che augurarvi un buon viaggio nel mondo dell'intelligenza artificiale conversazionale!

1

Cos'è l'intelligenza artificiale conversazionale

I concetti chiave

- La Conversational AI è la branca dell'intelligenza artificiale che si concentra sulla progettazione e sullo sviluppo di soluzioni che consentono a persone e oggetti di conversare.
- Qualsiasi oggetto in grado di sostenere un dialogo con le persone è un'interfaccia conversazionale.
- L'intelligenza artificiale è la disciplina che si occupa di creare sistemi che svolgono in autonomia compiti che tipicamente richiedono l'intelligenza umana.
- Il ramo dell'intelligenza artificiale che studia come elaborare il linguaggio umano tramite strumenti informatici si chiama Natural Language Processing (NLP) o linguistica computazionale.
- Una conversazione implica che l'interazione prosegua oltre il semplice botta e risposta, che si articoli in modo continuativo e collaborativo tramite l'alternanza di turni di parola e che abbia come obiettivo uno scambio bidirezionale ed efficace di informazioni.

La prima tappa del nostro viaggio nel settore dell'intelligenza artificiale conversazionale ha come obiettivo familiarizzare con il concetto di dialogo applicato all'interazione tra persone e macchine.

La **Conversational AI**, infatti, è la branca dell'intelligenza artificiale che si concentra sulla progettazione e sullo sviluppo di soluzioni che consentono a persone e oggetti di conversare.

Questo settore è relativamente recente, poiché solo negli ultimi anni ci sono stati avanzamenti tecnologici tali da portare gli assistenti conversazio-

nali nelle tasche di (quasi) tutti e tutte, eppure sono già decenni che i ricercatori studiano come dialogare con i computer.

L'idea stessa di conversazione tra umano e macchina è nata negli anni '50 con Alan Turing. Il matematico britannico ideò un test che metteva alla prova la capacità di una macchina di comportarsi in modo tanto intelligente, da essere indistinguibile da un essere umano. Il **test di Turing**, che ancora oggi è utilizzato per valutare i sistemi conversazionali, consiste nel far interagire una persona con una macchina e con un essere umano, senza dirle quale delle due entità è la macchina: se la macchina riesce a convincere la persona di essere l'umano, allora il test si considera superato.

La prima vera interfaccia conversazionale della storia, però, risale al 1966: si chiamava **ELIZA** e simulava uno psicoterapeuta. Per rispondere usava template predefiniti e spesso si limitava a riformulare le frasi scritte dai suoi interlocutori, ad esempio chiedendo loro di approfondire la risposta fornita.

Nei decenni successivi, le tecnologie conversazionali si sono evolute notevolmente, portando alla creazione di soluzioni molto più avanzate, capaci anche di ascoltare, parlare e perfino di partecipare come concorrenti a giochi televisivi, come ha fatto Watson di IBM, che nel 2011 ha vinto una puntata di *Jeopardy*, un gioco televisivo popolare negli Stati Uniti.

L'ultima vera rivoluzione, più recente, è stata **ChatGPT**, rilasciato a novembre 2022 da OpenAI. Per le sue ottime performance, ChatGPT ha rappresentato un enorme passo avanti nella capacità di simulare le competenze umane necessarie per intrattenere una conversazione e ha trasformato il modo in cui guardiamo ai prodotti conversazionali: risponde a praticamente qualsiasi domanda, formulando in automatico risposte linguisticamente coerenti.

Ma cosa accomuna tutti questi prodotti? Cosa li rende interfacce conversazionali? Vediamolo nei prossimi paragrafi.

1. Cosa sono le interfacce conversazionali

Interfaccia conversazionale è il termine più generico per indicare un qualsiasi oggetto al quale parliamo o scriviamo e dal quale riceviamo una risposta orale o scritta.

Sono interfacce conversazionali i chatbot sui siti web, gli assistenti virtuali come Siri, Alexa, Google Assistant, Cortana, Bixby, i sistemi multimediali delle automobili e, potenzialmente, qualunque dispositivo digitale in grado di sostenere un dialogo con le persone.

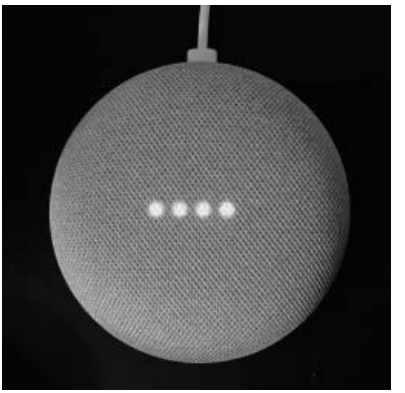
In informatica, un'**interfaccia** è un sistema che permette di interagire con un'applicazione e di controllarne il comportamento.

I primi computer consentivano l'interazione solo tramite interfaccia a linea di comando (*Command Line Interface*, CLI): le istruzioni venivano

fornite in formato testuale e dovevano essere scritte con una sintassi precisa, conosciuta e utilizzata da pochi esperti. I computer moderni consentono l'interazione tramite interfaccia grafica (*Graphical User Interface*, GUI): le istruzioni vengono fornite per mezzo di elementi visivi come icone, pulsanti, menu e finestre, e per questo possono essere utilizzati da un pubblico molto più ampio. Le interfacce conversazionali (*Conversational User Interface*, CUI) sono un'ulteriore evoluzione e sono pensate per abilitare l'interazione tramite il mezzo di comunicazione più naturale per l'essere umano: il dialogo.

Si distinguono dalle interfacce grafiche perché usano pochissimi elementi visivi: un riquadro in cui digitare l'input, l'icona di un microfono per attivare il riconoscimento vocale, un'icona per chiudere la finestra, dei led luminosi e pochi altri simboli (Fig. 1).

Fig. 1 – Lo smart speaker Google Home Mini non ha nessun elemento grafico e gli unici output che dà sono vocali e luminosi – 23.01.2023



Inoltre, si distinguono dalle interfacce a linea di comando perché, oltre a usare le parole sia come input che come output, usano il **linguaggio naturale**, cioè il sistema di comunicazione che gli esseri umani adoperano per interagire tra di loro e che si declina nelle varie lingue parlate sulla Terra. Questo sistema di comunicazione è complesso e versatile e consente di esprimere in modo efficace anche concetti astratti come idee ed emozioni. Si chiama naturale perché si è sviluppato spontaneamente nel corso dell'evoluzione umana e non è stato creato artificialmente, come invece è accaduto per le lingue artificiali come l'Esperanto, create a tavolino con l'obiettivo di facilitare la comunicazione interculturale, e come è avvenuto per i linguaggi formali, cioè i linguaggi di programmazione e di markup o altri sistemi di codifica, che hanno una struttura molto più rigida, basata su una serie di regole convenzionali che stabiliscono come devono essere combinati i simboli per creare espressioni valide.

Infine, le interfacce conversazionali si distinguono dalle altre modalità di interazione che usano il linguaggio naturale perché consentono l'alternanza dei cosiddetti turni di parola. Quando si digita una richiesta in un motore di ricerca tradizionale, si fornisce un input nella propria lingua e anche le proposte di contenuti che si ricevono come output della ricerca sono scritti nella stessa lingua. Questa interazione, però, non è uno scambio conversazionale, bensì una semplice ricerca web[1]: la richiesta dell'utente è spesso telegrafica e lontana da una formulazione spontanea e la risposta del motore di ricerca è costituita da un elenco di siti web accompagnati da un titolo e da una breve descrizione. Inoltre, se subito dopo si digita un'altra richiesta, il processo ricomincerà da capo, e le due ricerche non saranno collegate tra loro. Al contrario, se si chiede a un'interfaccia conversazionale la stessa informazione, questa reagirà elaborando il contenuto in forma di risposta e, nella maggior parte dei casi, si potrà proseguire il dialogo per più turni, chiedendo ulteriori approfondimenti collegati alla prima richiesta.

A questo punto, torniamo al concetto di IA conversazionale e scomponiamolo per approfondire ciascuno dei suoi componenti: perché si parla di intelligenza artificiale? E quali dinamiche rendono una conversazione tale?

2. Perché si parla di intelligenza artificiale

L'**intelligenza artificiale** è la disciplina che si occupa di creare sistemi che svolgono in autonomia compiti che tipicamente richiedono l'intelligenza umana, come il vedere, il parlare, il prendere decisioni.

L'espressione è stata coniata dal ricercatore americano John McCarthy nel lontano 1956, tuttavia, questo settore ha visto a lungo l'alternarsi di periodi di entusiasmo e di stasi e solo negli ultimi decenni ha raggiunto risultati davvero soddisfacenti, grazie al potenziamento della capacità computazionale dei computer odierni e alla disponibilità di quantità esorbitanti di dati con cui addestrare i sistemi.

Per emulare l'intelligenza umana, l'intelligenza artificiale può utilizzare una varietà di metodi. Le prime soluzioni di AI, e quindi anche le prime interfacce conversazionali, utilizzavano i cosiddetti sistemi esperti (*expert systems*) basati su regole (*rule-based*) esplicitamente definite, che tentavano di codificare la conoscenza di esperti umani. L'insieme di regole costituiva la base di conoscenza a cui il motore di inferenza (*inference engine*) attingeva per eseguire i ragionamenti logici, cioè per eseguire le regole stesse, finché non trovava la soluzione al compito da svolgere.

1. Deibel D., Evanhoe R., *Conversations with Things: UX Design for Chat and Voice*, Rosenfeld Media, 2021

2.1. Machine Learning: apprendimento automatico

Oggi, questo approccio è stato in gran parte sostituito o arricchito da quello basato sul *machine learning* (ML), in italiano 'apprendimento automatico'. L'obiettivo del machine learning è sviluppare algoritmi che imparino in autonomia dai dati d'esempio, senza che sia necessario definire esplicitamente tutte le regole per far ragionare il sistema. Il cuore di un sistema di ML è il suo modello, cioè la logica di come quel software ragiona e crea output. Il modello viene istruito a identificare le relazioni tra i dati di addestramento, al fine di utilizzare le informazioni apprese per fare **previsioni statistiche** su nuovi dati, che non sono stati utilizzati nel processo di addestramento.

I dati usati come esempi nella fase di addestramento costituiscono un *training data set*, o semplicemente ***training set***, e la loro qualità influenza direttamente le performance del sistema. Il modello diventa tanto più preciso quanto più i dati utilizzati per l'addestramento sono rappresentativi della realtà che si vuole analizzare. Se, per esempio, i dati non sono abbastanza diversificati, potrebbero portare il sistema a produrre dei *bias*, cioè dei pregiudizi che fanno sì che il sistema privilegi un punto di vista e ne escluda altri.

L'addestramento può avvenire tramite diverse **tecniche**, tra cui:

- *supervised learning* ('apprendimento supervisionato'): il modello viene addestrato su dati precedentemente classificati da persone tramite etichette che identificano la tipologia del dato, così che possa essere riconosciuto dal sistema;
- *unsupervised learning* ('apprendimento non supervisionato'): il modello cerca di identificare in autonomia delle relazioni tra i dati grezzi, non etichettati;
- *semi-supervised learning* ('apprendimento semi-supervisionato'): il modello viene addestrato con pochi dati etichettati e molti dati grezzi;
- *reinforcement learning* ('apprendimento per rinforzo'): il modello viene addestrato tramite feedback fatti da premi e punizioni, che vengono assegnati in base alla correttezza o scorrettezza della previsione; in questo modo si simula un processo di apprendimento interattivo tramite tentativi ed errori.

Un tipo particolare efficace di machine learning è il *deep learning,* che utilizza reti neurali artificiali, ispirate al funzionamento delle reti neurali biologiche del cervello umano, composte da diversi strati di neuroni artificiali disposti in modo gerarchico, in cui ogni strato riceve input dai livelli precedenti e fornisce output ai livelli successivi.

Un tipo di architettura di rete neurale che ha rivoluzionato il campo dell'elaborazione del linguaggio naturale negli ultimi anni è quella basata su *transformers*, che, a differenza delle architetture precedenti, come i modelli di linguaggio basati su reti neurali ricorrenti (RNN), utilizza un meccanismo chiamato *self-attention*. Questo meccanismo consente ai transformers di considerare l'importanza relativa di ciascuna parola rispetto alle altre nella frase o nel contesto più ampio. I transformers sono in grado di apprendere rappresentazioni complesse del linguaggio, catturando le relazioni semantiche e sintattiche in un modo che le RNN faticavano a fare.

Esistono diversi tipi di modelli di apprendimento automatico. Due tipi di modelli che utilizzano il deep learning per svolgere compiti diversi sono quelli **discriminativi** e quelli **generativi**. I modelli discriminativi vengono usati per classificare, e quindi assegnano un'etichetta ai dati che ricevono come input; un sistema discriminativo, per esempio, può essere addestrato a distinguere le recensioni positive da quelle negative. I modelli generativi vengono usati per creare nuovi dati, simili a quelli di addestramento; un sistema generativo, per esempio, può essere addestrato a scrivere ex novo delle recensioni positive. Come approfondiremo nei prossimi capitoli, per realizzare interfacce conversazionali si possono usare sia modelli discriminativi che generativi.

2.2. NLP: elaborazione del linguaggio naturale

Poste queste premesse, ne deriva che la Conversational AI possa essere considerata una branca dell'intelligenza artificiale, poiché **emula le capacità che l'essere umano usa per conversare**.

Queste capacità si basano in realtà su più elementi e vengono riprodotte in modo diverso. Il primo elemento sono gli organi del corpo umano, che possono essere riprodotti tramite hardware: per esempio, i microfoni simulano la capacità d'ascolto dell'orecchio. Il secondo elemento è la capacità di ragionare, che ci permette di decidere cosa dire o fare, e che può essere riprodotta tramite software che, come vedremo, scelgono cosa fare sulla base delle condizioni che si presentano loro. Il terzo elemento è la capacità di elaborare il linguaggio naturale e anche questa può essere riprodotta tramite specifiche tecnologie.

Il ramo dell'intelligenza artificiale che studia come **elaborare il linguaggio naturale tramite strumenti informatici** si chiama *Natural Language Processing* (NLP) o linguistica computazionale. Queste due espressioni possono essere usate come sinonimi, ma hanno focus leggermente diversi: la linguistica computazionale formula la base teorico-scientifica su cui poggia l'NLP, e l'NLP sviluppa le applicazioni pratiche di quella base teorica.

L'NLP ha numerose applicazioni e, tra queste, ci sono quelle che consentono il funzionamento delle interfacce conversazionali, cioè:

- il riconoscimento automatico del parlato (in inglese *Automatic Speech Recognition,* ASR, o *Speech-to-Text*, STT);
- la comprensione del testo scritto (in inglese *Natural Language Understanding,* NLU);
- la generazione di testo scritto (in inglese *Natural Language Generation*, NLG);
- la sintesi del parlato (in inglese *Speech Synthesis o Text-to-Speech,* TTS).

Nel corso del libro approfondiremo ciascuno di questi ambiti. Ora che abbiamo capito perché le interfacce conversazionali sono applicazioni dell'intelligenza artificiale, possiamo familiarizzare con il concetto di conversazione e capire come mai simulare un dialogo è così complesso.

3. Perché si parla di conversazionale

Affinché si parli di conversazione o di dialogo, non basta che il sistema capisca ciò che l'utente scrive o dice nella propria lingua e che risponda usando lo stesso codice linguistico: una **conversazione** implica che l'interazione prosegua oltre il semplice botta e risposta, che si articoli in modo continuativo e collaborativo tramite l'alternanza di turni di parola (*turn-taking*) e che abbia come obiettivo uno scambio bidirezionale ed efficace di informazioni.

3.1. Alternanza di turni di parola

L'idea stessa di dialogo implica che i due o più attori coinvolti prendano la parola per esprimere i propri pensieri alternandosi: ciascuna delle loro interazioni è considerata un turno (Fig. 2).

Il modo in cui le persone prendono e passano la parola è regolato da meccanismi sofisticati ed è stato ampiamente studiato dalla disciplina chiamata analisi conversazionale (*Conversation Analysis,* CA), che è un punto d'incontro tra linguistica e sociologia.

Studiare l'alternanza di turni vuol dire studiare come le persone capiscono che possono iniziare a parlare, se possono farlo anche se l'interlocutore non ha ancora finito il proprio intervento, quanto devono essere lunghe le pause tra un intervento e l'altro, come modulare il tono di voce per far capire all'altro che hanno terminato di parlare o che hanno ricevuto e compreso il messaggio comunicato dall'altro. Questi meccanismi possono va-

riare in base alla gerarchia tra gli interlocutori ma anche in base alla cultura di appartenenza: gli italiani e i popoli latini sono più flessibili e collaborativi nel dialogo e tendono più spesso a sovrapporsi e a completarsi le frasi a vicenda, mentre nord-europei e americani mal sopportano le intrusioni nel proprio turno[2].

Fig. 2 – Alternanza di turni di parola tra un cliente e Angie, il chatbot di Tim – sito www.tim.it, 15.01.2023

Le strategie di alternanza di turni dipendono anche dal mezzo di comunicazione usato: interagire di persona è diverso dall'interagire in videochiamata e ancora diverso dall'interagire tramite normale chiamata telefonica o addirittura tramite chat. Nelle conversazioni faccia a faccia, per esempio, si sfruttano anche elementi non verbali come il contatto visivo, le espressioni facciali, i gesti, la postura e il linguaggio del corpo in generale, che però non sono praticabili al telefono o in chat.

Riprodurre queste dinamiche in modo artificiale non è affatto semplice, ma chi progetta le conversazioni è chiamato a simularle.

2. Balboni P.E., *Parole comuni culture diverse*, Marsilio, 1999.

3.2. Sfide linguistiche del dialogare

L'alternanza di turni di parola porta con sé la sfida di gestire le frasi incomplete, in cui vengono omesse una o più parole. In linguistica, l'omissione di una o più parole da una proposizione è chiamata **ellissi**.

Non è detto che un'ellissi interferisca con la capacità dell'interlocutore di capire il significato della proposizione: di solito, le parole escluse possono essere intuite dal **contesto**, che è l'insieme di circostanze sociali, culturali e situazionali in cui avviene la comunicazione e comprende elementi interni ed esterni all'atto comunicativo, che forniscono risorse per interpretarlo in modo corretto. È studiato dalla pragmatica, una branca della linguistica che esamina proprio l'influenza del contesto sul significato.

Immaginiamo di dover interpretare una richiesta di assistenza relativa al tracciamento di un ordine fatto su un e-commerce. Il cliente potrebbe dire soltanto "Sto ancora aspettando il pacco!", omettendo l'obiettivo della sua richiesta, cioè "voglio sapere quando sarà consegnato il pacco che aspetto", e le informazioni necessarie a capire a quale pacco fa riferimento, cioè il numero d'ordine, la data in cui lo ha effettuato o i prodotti che ha ordinato.

Per interpretare una richiesta di questo tipo, una persona può attingere a:

- implicature conversazionali;
- informazioni sugli ordini dell'utente;
- storico della conversazione.

In linguistica, un'**implicatura conversazionale** è ciò che una persona intende comunicare attraverso ciò che dice, senza che lo comunichi esplicitamente. L'intenzione può essere compresa solo considerando il contesto della conversazione, la conoscenza condivisa tra i parlanti e gli obiettivi comunicativi dell'interazione.

Nel nostro esempio, il fatto stesso di trovarsi su un e-commerce implica che il cliente si aspetti che l'assistenza sappia fornire determinate informazioni e implica la conoscenza del funzionamento di un e-commerce, per cui si fa un acquisto online e si aspetta che l'ordine sia consegnato all'indirizzo scelto, auspicabilmente senza ritardi rispetto alla data prevista. Si può inferire che la richiesta dell'utente abbia come obiettivo informarsi sulla data di consegna dell'ordine e non comunicare soltanto che si è in attesa di quell'ordine: il significato implicito della formulazione è quindi diverso da quello esplicitato, ma può essere dedotto dal contesto.

Allo stesso modo, quando si progetta un assistente virtuale, bisogna tener conto che non sempre quello che l'utente dice equivale a quello che l'utente vuole ottenere e che, per riconoscere l'obiettivo di una richiesta, non basta fermarsi alla sua formulazione esplicita. Per esempio, bisognerà addestrare il bot a considerare la frase "Sto ancora aspettando il pacco!" come sinoni-

mo di “Voglio sapere lo stato del mio ordine”, di “Quando mi consegnate il pacco?” e di “Devo tracciare la spedizione”, tutte formulazioni che, in questo specifico contesto comunicativo, hanno lo stesso obiettivo.

Il contesto include anche la data in cui l’utente sta ponendo la domanda, che potrebbe indicare un periodo di sovraccarico per i corrieri (Black Friday, Natale) e soprattutto le informazioni che si possono ottenere dal profilo del cliente, come il numero di ordini effettuati, e per ciascuno di essi la data, il numero identificativo, il contenuto, lo stato e il corriere che lo ha preso in carico, ma anche eventuali prodotti presenti nel carrello o nei preferiti ma non ancora acquistati.

Se opportunamente progettato, anche un assistente digitale può avere accesso a queste informazioni contestuali e usarle per determinare a quale ordine fa riferimento il cliente. Per esempio, può accedere alle informazioni sugli ordini effettuati e vedere che nelle ultime due settimane quella persona ha fatto due ordini diversi; poi, può controllare lo stato di questi ordini e notare che uno dei due risulta consegnato e che l’altro è ancora in transito, quindi supporre che il cliente abbia bisogno di informazioni proprio su quest’ultimo e fornirgliele direttamente, senza il bisogno di chiedere il numero d’ordine esplicitamente.

Il contesto include anche lo **storico della conversazione**, cioè quel che è stato detto in precedenza nella stessa conversazione o nelle conversazioni precedenti.

La capacità umana di ricordare ciò che è già stato comunicato, magari anche a giorni di distanza, abilita l’alternanza di turni di parola e rende molto rapido lo scambio di informazioni. Per interagire in modo naturale, i prodotti conversazionali devono essere in grado di simulare questa capacità e di ricordare almeno le informazioni condivise nella conversazione in corso. L’esempio più popolare per spiegare questo concetto è quello di un’interazione tra persona e interfaccia a proposito delle condizioni metereologiche. Immaginiamo che un utente voglia partire all’ultimo momento per trascorrere un weekend fuori e che per scegliere la meta abbia bisogno di sapere quali saranno le condizioni meteo di più città. Mettiamo quindi che per prima cosa chieda “Che tempo farà a Milano domani?” e che l’interfaccia conversazionale risponda “A Milano, domani, sarà prevalentemente soleggiato, con una massima di 24 gradi e una minima di 15”. A questo punto, l’utente potrebbe voler conoscere le condizioni di un’altra città, ma difficilmente ripeterà una frase completa. È invece più probabile che chieda “E a Roma?” In questo caso darà come sottintesi, e quindi ometterà, sia l’argomento (si sta parlando ancora di tempo, è superfluo ripeterlo) che la data (si sta parlando ancora di domani, perché ripeterlo?), mentre preciserà solo l’informazione nuova e distintiva, in questo caso la città (Roma). Per una persona, comprendere la frase e collegarla alle precedenti è più che facile, mentre un’interfaccia conversazionale dovrà

essere espressamente istruita a mantenere in memoria i dati che non sono cambiati e a sostituire il dato che è cambiato: il parametro *argomento* e il parametro *data* rimarranno invariati, mentre l'interfaccia sostituirà il vecchio valore del parametro *città*.

Vediamo ora un altro caso, un po' più complesso. Questa volta, l'utente vuole conoscere il meteo della propria città, in data odierna. Il modo più spontaneo per chiederlo è "Che tempo fa?". Anche in questo caso, è stato esplicitato solo un parametro, l'*argomento*, e sono stati omessi data e luogo, perché le persone tendono a omettere tutte le informazioni che considerano superflue. Fateci caso: se chiedete informazioni sulla città in cui vi trovate abitualmente, la darete per scontata, mentre se chiedete per una città diversa da quella in cui siete soliti risiedere, sentirete la necessità di menzionarla esplicitamente. In questo scenario, un interlocutore umano vi capirebbe al volo, senza neanche badare all'omissione. L'interfaccia conversazionale, per dare una risposta pertinente, ha bisogno di sapere con precisione a quale luogo vi riferite e quindi dovrà reperirlo non più dallo storico della conversazione, bensì dal contesto esterno all'interazione. Per farlo, può essere progettata per consultare servizi esterni che possano fornire quel preciso dato. Per esempio, chi progetta l'interfaccia potrà definire che, se l'utente non specifica il luogo, l'interfaccia darà per scontato che la domanda si riferisca alla località in cui l'utente si trova in quel momento e, di conseguenza, userà la geolocalizzazione per scoprire qual è questa località. Tuttavia, se la geolocalizzazione non è disponibile (per esempio, se l'utente non l'ha autorizzata sul proprio smartphone), l'ultima risorsa sarà chiedere esplicitamente all'utente il dato necessario, dando il via proprio a un'alternanza di turni di domande e risposte.

Bisogna però fare attenzione a non confondere il *contesto* con quello che in linguistica è chiamato **co-testo**: l'insieme di parole che circondano un termine o una frase e che possono essere usate per determinarne il significato. Il co-testo torna utile nel caso di parole polisemiche, che hanno cioè significati diversi ma si scrivono allo stesso modo, e che possono essere disambiguate guardando ai termini che le precedono o seguono. Mettiamo che a un assistente virtuale si chieda in inglese "*play some rock music*", ('riproduci della musica rock'). La parola *rock* può riferirsi al genere musicale, ma può anche significare 'pietra'. Guardando al co-testo, in cui sono presenti le parole *play* e *music*, è però molto semplice disambiguare e capire a quale dei significati ci si riferisce.

3.3. Principi linguistici per una comunicazione efficace

L'alternanza di turni permette alla comunicazione umana di essere un'attività collaborativa, in cui gli interlocutori contribuiscono al raggiungimento

di uno scopo comune: far sì che lo scambio di informazioni avvenga in modo efficace[3].

Proprio per questo, gli scambi comunicativi rispettano spontaneamente il **principio di cooperazione** elaborato dal linguista Paul Grice:

> *"Conforma il tuo contributo conversazionale a quanto è richiesto, nel momento in cui avviene, dall'intento comune accettato o dalla direzione dello scambio verbale in cui sei impegnato"[4].*

La comunicazione tra persona e robot non fa eccezione: il contributo verbale dell'interfaccia conversazionale deve essere adeguato e collaborativo.

Ma come si fa a rispettare il principio di cooperazione? Grice ha elaborato indicazioni più precise, declinate nelle massime di quantità, qualità, relazione e modalità.

La **massima di quantità** precisa che bisogna essere concisi: il proprio contributo comunicativo deve fornire *tutte* le informazioni necessarie, ma *solo* le informazioni necessarie. Questo vale per i messaggi formulati dalle persone, ma anche e soprattutto per i messaggi forniti da un'interfaccia conversazionale, che deve sfruttare al meglio le poche risorse che ha a disposizione per trasmettere contenuti rilevanti.

La **massima di qualità**, poi, prescrive di dare informazioni vere e della cui autenticità si hanno prove sufficienti. Va da sé che anche un'interfaccia conversazionale debba fornire informazioni corrette e verificate. Se l'interfaccia attinge a risposte predefinite, scritte in precedenza dai progettisti, questo implica aggiornare le sue risposte tempestivamente, ogni volta che cambia qualcosa nelle policy aziendali o in qualsiasi altro dato fornito. Se l'interfaccia utilizza algoritmi di generazione automatica di contenuti, rispettare la massima di qualità vuol dire configurare questi algoritmi con attenzione, in modo da instradarli il più possibile verso la produzione di testi che veicolano informazioni veritiere.

La terza **massima** è quella di **relazione**, che chiede di essere pertinenti e di fornire contenuti utili. Un oggetto conversazionale dovrebbe rispondere in modo rilevante e risolutivo. Questa è forse la massima che viene violata più spesso, a causa della difficoltà di intercettare correttamente l'obiettivo dell'utente.

La quarta e ultima **massima** è quella di **modalità**, che suggerisce di comunicare con chiarezza l'idea o il concetto che si intende esprimere, evitando ambiguità e formulazioni complesse. Il messaggio trasmesso dev'essere chiaro e lo sforzo cognitivo richiesto per comprenderlo dev'essere il più limitato possibile. Pertanto, i messaggi forniti da un'interfaccia conversazionale dovrebbero usare termini comprensibili e una sintassi lineare.

3. Pisanty V., Zijno A., *Semiotica*, McGraw-Hill Education , 2009.
4. Grice, P., *Studies in the way of words*, Harward Business Press, 1991.

Alle massime di Grice è collegata la teoria dell'**economia linguistica**, elaborata dal linguista francese André Martinet, che studia come gli individui gestiscono le risorse linguistiche in modo efficiente, al fine di ottenere il massimo risultato dalla comunicazione con il minimo sforzo. L'economia linguistica analizza le scelte linguistiche fatte dagli individui, che tendono spontaneamente a scegliere la lingua più familiare, la struttura sintattica più semplice, le frasi più brevi e le parole più di uso comune. Queste scelte sono influenzate da fattori come il contesto comunicativo, la cultura, la relazione tra gli interlocutori e la disponibilità di risorse quali il tempo, la memoria, lo sforzo cognitivo e l'attenzione dell'interlocutore.

Per risultare naturali, anche le formulazioni espresse da un'intelligenza artificiale dovrebbero rispettare questa teoria e minimizzare l'uso di risorse sfruttate in uno scambio comunicativo.

In sintesi, la Conversational AI è il settore che progetta e sviluppa interfacce conversazionali. Un'interfaccia conversazionale è un prodotto dell'intelligenza artificiale, poiché simula le capacità umane che servono per conversare, quindi per scambiarsi informazioni usando il linguaggio naturale, in modo collaborativo, efficiente e a turni alterni: affinché questa alternanza di turni possa avvenire con successo, entrambi gli interlocutori, siano essi persone o macchine, devono essere in grado di attingere al contesto.

2

I diversi tipi di interfacce conversazionali

I concetti chiave

- Le espressioni più generiche per riferirsi a un prodotto della Conversational AI sono interfaccia conversazionale, soluzione conversazionale, prodotto conversazionale, sistema dialogico, agente conversazionale, bot conversazionale (o soltanto bot), assistente virtuale (AV), assistente digitale.
- Un chatbot è un'interfaccia conversazionale fruibile via chat; un voicebot è un'interfaccia conversazionale fruibile tramite voce.
- Il concetto di Conversational AI è contiguo ma non equivalente a quelli di Voice Technology, Conversation Design, Conversational Marketing e Generative AI.
- Gli ambiti in cui vengono più utilizzate attualmente le interfacce conversazionali sono Customer Service, Conversational Commerce, Booking, Recruitment, Health e Mental Health, Gaming, Education, Tutoring.

Ora che abbiamo visto perché si parla di intelligenza artificiale conversazionale e di interfacce conversazionali, possiamo addentrarci nella ricca terminologia che spesso confonde chi si avvicina a questo settore.

Chatbot, voicebot, assistenti virtuali, assistenti vocali, bot,… qual è l'espressione più corretta per riferirsi a ciascun prodotto? E qual è la differenza tra espressioni come Voice Technology, Conversational AI, Conversation Design, Conversational Marketing, Generative AI?

"Talvolta c'è una cattiva interpretazione di cosa possa essere un'interfaccia conversazionale, perché ci troviamo davanti pubblicità fuorvianti e gli utenti sono portati a pensare che tutti gli assistenti virtuali siano come lo smart speaker che abbiamo in casa".
Iolanda Iacono[1]

Questo settore è relativamente giovane e in continuo fermento e, come tale, ancora restio alla standardizzazione terminologica. Più persone vi accedono, portando con sé i propri variegati percorsi formativi e professionali, più si diffondono nuove teorie, espressioni e consuetudini.

Nella prima parte di questo capitolo ci soffermeremo sui termini usati per riferirsi ai vari tipi di prodotti conversazionali e al settore nel suo insieme, per poi passare in rassegna alcune delle applicazioni più diffuse oggi in ambito business.

1. Orientarsi nella terminologia

Nell'esplorare le espressioni usate per indicare i diversi tipi di soluzioni conversazionali, inizieremo da quelle più generiche, che possono essere usate a prescindere dalle caratteristiche del prodotto.

La prima di queste espressioni è proprio '**interfaccia conversazionale**', in inglese *Conversational Interface* o *Conversational User Interface* (CUI). Pone l'enfasi sulla loro funzione di interfaccia tra utente e applicazione informatica e sulla modalità di interazione usata: lo scambio dialogico.

Anche nelle formulazioni '**soluzione conversazionale**', '**prodotto conversazionale**', '**sistema dialogico**' e '**agente conversazionale**' persiste il riferimento allo scambio dialogico, ma mentre *soluzione, prodotto* e *sistema* pongono l'enfasi sul loro essere oggetti (digitali), il termine *agente* personifica lo strumento ed è usato più spesso in ambito di assistenza al cliente e soprattutto nel mondo anglofono, in cui l'operatore è chiamato proprio *agent*.

Corta e quindi comoda da usare è la parola '**bot**', abbreviazione di robot. In generale, un bot è un qualsiasi software che esegue un'attività in automatico. L'espressione di per sé, quindi, non implica che si stia parlando di un software dialogico, a meno che non si precisi '**bot conversazionale**', chiarendo la compresenza di automazione e scambio dialogico. Ciononostante, 'bot' è la parola più breve per riferirsi a questo tipo di strumenti e quindi viene usata spesso anche da sola, come economia linguistica vuole. In questo libro, useremo sempre 'bot' nell'accezione di 'bot conversazionale'. Va precisato che l'ortografia corretta di questa parola è con tutti i caratteri minuscoli, non

1. *Donne nella Conversational AI: Mary Tomasso intervista Iolanda Iacono*, canale Women in Voice Italy, https://www.youtube.com/watch?v=Y-t9L6orkpk&t.

BOT, con i caratteri maiuscoli, come spesso si legge: è una parola a tutti gli effetti, non un acronimo.

Altre espressioni molto diffuse sono '**assistente virtuale**' (AV), in inglese *Virtual Assistant* (VA) o '**assistente digitale**', in inglese *Digital Assistant*. Queste espressioni enfatizzano sia il ruolo principale di queste interfacce, cioè assistere gli esseri umani in svariate mansioni, sia la natura tecnologica dell'assistente, tramite le parole 'virtuale' e 'digitale'.

L'espressione assistente virtuale, tuttavia, è un po' ambigua: fino a pochi anni fa, veniva utilizzata per indicare una figura professionale, una persona reale che lavorava da remoto, offrendo servizi e supporto alle aziende e ai liberi professionisti. Ancora oggi, cercando 'assistente virtuale' su Google, i primi risultati che si ottengono sono tutti relativi proprio a questa professione. Nell'uso comune, invece, l'espressione 'assistente virtuale' si è affermata soprattutto in riferimento a prodotti come Siri, Alexa, l'Assistente di Google, Cortana e Bixby. Inoltre, come vedremo a breve, a volte viene usata come contrario di chatbot, invece che come suo iperonimo.

A questo punto, possiamo scendere al livello successivo della classificazione dei prodotti conversazionali.

Fig. 1 – L'assistente virtuale di Lavazza elenca da subito gli argomenti che è progettato per gestire – sito www.lavazza.it, 20.03.2023

Un primo criterio di classificazione è l'ampiezza del dominio di conoscenza dell'interfaccia conversazionale, cioè la quantità di argomenti di cui sa conversare. Se ipoteticamente può gestire conversazioni su qualsiasi argomento, di parla di bot 'generalista' o *general-purpose*, se può gestire sono uno o più argomenti precisi, si parla di bot 'specialista' o *goal-oriented*. Alexa, Siri e l'Assistente di Google sono generalisti, mentre i chatbot presenti sulla maggior parte dei siti sono specialisti e possono rispondere solo a un set limitato di richieste (Fig. 1).

Un altro criterio è riferito a quale interlocutore tra l'utente e il bot guida la conversazione. Se l'interazione è iniziata e continuata proattivamente dall'utente si parla di bot *user-initiative* o *user-directed*: è il caso delle interfacce che hanno come scopo principale rispondere alle domande degli utenti. Se l'interazione è iniziata e portata avanti dal sistema, che continua a porre domande all'utente finché questi non risponde, si parla di bot *system-initiative* o *system-directed*: è questo il caso dei bot progettati per compiere un'azione, come prenotare un volo, per la quale devono chiedere una serie di dati obbligatori. Se l'interazione può essere guidata sia dall'utente che dal bot alternativamente, si parla di *mixed-initiative* e queste sono le interfacce che meglio emulano l'alternanza di turni tipica della conversazione.

I criteri più usati per classificare i bot conversazionali, però, sono **il canale usato per il dialogo e la modalità di interazion**e. Se il canale è testuale, si parla di *chatbot* o di soluzione *text-based*, se è vocale, si parla di *voicebot* o di soluzione *voice-based*. Se consente entrambe le modalità, si parla di soluzione multimodale.

Ma le distinzioni terminologiche non si fermano qui… andiamo più nel dettaglio!

1.1. Chatbot

Il termine 'chatbot' è una parola composta dal verbo *to chat*, che in inglese vuol dire 'chiacchierare' e il sostantivo *bot*, che come abbiamo già visto sta per 'robot'. Proprio come robot, è un termine che in italiano viene usato al maschile ed è quindi scorretto dire *una* chatbot. Espressioni equivalenti, ma molto meno usate, sono *Chat bot*, *Chatbox*, *Chatterbot*, o *Chatterbox*.

I chatbot sono le interfacce conversazionali testuali accessibili tramite una chat, che può essere la webchat su un sito, oppure la chat di un altro canale come Facebook Messenger, o WhatsApp, o Telegram, o Slack, o Zendesk, o qualunque altra app o piattaforma di messaggistica. L'apertura ai chatbot di piattaforme di messaggistica già molto frequentate, come Facebook Messenger, WhatsApp e, in misura molto minore, Telegram, ha contribuito notevolmente all'espansione del mercato della Conversational AI, poiché ha reso i bot conversazionali facili da trovare e da usare per chiunque.