UNIONE EUROPEA
Fondo Sociale Europeo

Ministero dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE 2014 - 2020

UNIVERSITÀ DEGLI STUDI FIRENZE
Da un secolo, oltre.

DOTTORATO DI RICERCA IN
Lingue, Letterature e Culture Comparate

CICLO XXXVII - PON ex DM1061

Conversazioni Artificiali: dagli assistenti vocali all'IA generativa

SSD GLOT-01/A

Dottoranda
Dott.ssa Ambra Paparatty

Supervisore
Prof.ssa Benedetta Baldi

Coordinatore
Prof. Fernando Cioni

Alla mia Mamma,
Creatura più pura che esista.

INDICE

INTRODUZIONE

Negli ultimi decenni, l'evoluzione dell'intelligenza artificiale ha trasformato profondamente il nostro modo di comunicare, interagire e persino attribuire senso alle relazioni tecnologiche. L'interazione tra uomo e macchina, un tempo relegata all'ambito della fantascienza, è oggi realtà concreta e pervasiva: assistenti vocali, chatbot, interfacce conversazionali e agenti digitali popolano dispositivi personali, ambienti professionali e piattaforme digitali globali. Tuttavia, a dispetto della loro crescente diffusione, questi sistemi pongono ancora interrogativi essenziali: in che misura possono davvero sostenere una conversazione? E cosa significa, in fondo, progettare una "conversazione" artificiale?

In questo lavoro mi occuperò di esplorare la questione da un punto di vista teorico, storico e progettuale. La prima parte del percorso sarà concentrata sulle fondamenta linguistiche e pragmatiche della comunicazione umana, proponendo una riflessione sul linguaggio come azione intenzionale, sociale e situata, in linea con la prospettiva degli atti linguistici di Searle e con il principio di cooperazione di Grice. L'analisi dei principali modelli comunicativi – da Shannon e Weaver a Jakobson, da Watzlawick a Levinson – fornirà una cornice teorica per comprendere la comunicazione non solo come trasmissione di informazioni, ma come costruzione condivisa di significato. In particolare, verrà posto l'accento sull'importanza del contesto, delle intenzioni e delle regole implicite che governano gli scambi linguistici: elementi imprescindibili per qualsiasi tentativo di simulare un'interazione autentica tra uomo e macchina.

A partire da queste premesse, il secondo capitolo offrirà una ricognizione storica dell'idea di Intelligenza Artificiale, tracciando un percorso che

dalle prime intuizioni teoriche conduce ai moderni modelli generativi. Saranno analizzate le tappe fondamentali dell'IA deterministica, fondata su regole predefinite e procedure logiche, per poi mettere in luce la svolta rappresentata dall'apprendimento automatico e dalle reti neurali profonde. In questo passaggio epocale, il paradigma cambia: dalle macchine che eseguono, si passa alle macchine che apprendono e generano contenuti, ponendo nuove sfide al concetto stesso di conversazione. La riflessione si sofferma anche su figure e sistemi emblematici, come ELIZA, Siri, Project Debater e i chatbot evoluti, per mostrare come ogni fase dell'IA abbia incarnato un'idea diversa di dialogo, spesso più vicina alla simulazione che alla comprensione autentica, dal pattern matching di Weizenbaum fino alla generazione argomentativa e multimodale e alle interfacce vocali integrate nella vita quotidiana.

Il lavoro proseguirà con un'analisi dello stato attuale delle tecnologie conversazionali, in particolare attraverso il confronto tra approcci progettuali ex-ante (basati su *script* e sceneggiature) e soluzioni generative capaci di adattarsi dinamicamente agli input degli utenti. L'attenzione si sposterà poi sul funzionamento dei Large Language Models e sull'impatto sociale e cognitivo di strumenti come ChatGPT, che sollevano questioni nuove non solo sul piano tecnico, ma anche su quello semantico, etico e relazionale.

Infine, con la sezione dedicata alla progettazione della conversazione artificiale, affronteremo le criticità e le sfide del *prompt design* e della scrittura interattiva per sistemi generativi, mostrando come la formulazione di istruzioni efficaci influenzi la qualità, la coerenza e la direzionalità del dialogo uomo-macchina. Sarà discusso come la cooperazione tra utente e sistema, nella logica griceana e nell'ottica degli atti linguistici, si traduca oggi in una nuova forma di collaborazione cognitiva mediata da algoritmi. In seguito a una riflessione sul "come" progettuale e linguistico, si passerà a un'analisi del "fine" relazionale e pragmatico: in questa prospettiva, la progettazione diventa parte integrante del processo di costruzione di sistemi artificiali in grado di sostenere un'interazione credibile, riconoscendo le aspettative dell'utente, la coerenza discorsiva e la dimensione relazionale del lin-

guaggio. Alla luce di ciò, il design conversazionale si configura come il tentativo di colmare la distanza tra funzione e finzione, tra risposte automatizzate e simulazioni di empatia, e come i sistemi di IA generativa richiedano sempre maggiore cooperazione e adattamento da parte degli utenti per poter beneficiare di una conversazione più efficiente, sensata e accurata.

In questo contesto, la domanda fondamentale che guida l'intera indagine è la seguente: è possibile progettare una conversazione tra uomo e macchina che sia qualcosa di più di una semplice simulazione funzionale? Può esistere una comunicazione artificiale che non si limiti a rispondere, ma che costruisca senso, cooperazione e relazione? Questo lavoro di ricerca si propone di esplorare tale interrogativo, consapevole che la risposta non riguarda solo l'efficienza dei sistemi, ma la nostra idea stessa di linguaggio, di intelligenza e di relazione.

1.
LA COMUNICAZIONE UOMO-MACCHINA: CORNICI TEORICHE E PROSPETTIVE PRAGMATICHE

La progettazione di una conversazione tra uomo e macchina non può prescindere da una riflessione approfondita sulla natura stessa della comunicazione. Questo capitolo si propone di esplorare le basi teoriche e pragmatiche che definiscono il linguaggio umano come comportamento complesso, intenzionale e situato, offrendo così un quadro di riferimento fondamentale per lo sviluppo di interfacce conversazionali intelligenti.

Verranno esaminati i modelli fondamentali della comunicazione: il modello matematico di Shannon e Weaver (1949), orientato alla trasmissione tecnica dell'informazione; il modello funzionale di Jakobson (1960), che evidenzia le diverse funzioni linguistiche in relazione agli attori comunicativi; e il modello sistemico-relazionale di Watzlawick et al. (1967), che sottolinea il valore interattivo e contestuale di ogni scambio comunicativo. Questi approcci forniscono tre prospettive complementari – informazionale, funzionale e relazionale – che devono necessariamente confluire nella progettazione di sistemi conversazionali capaci di adattarsi alle intenzioni e alle dinamiche degli utenti.

La seconda parte del capitolo si concentra sulla pragmatica linguistica, in particolare sulle teorie di Austin (1975), Searle (1969) e Levinson (1983). Verrà delineato come il significato emerge non solo dalle parole, ma dal contesto, dalle intenzioni del parlante e dalle norme sociali che regolano gli scambi comunicativi. Le nozioni di atti linguistici, condizioni di appropriatezza, organizzazione dei turni di parola e presupposizioni rappresentano elementi cruciali per costruire interfacce che non solo comprendano i comandi, ma che siano in grado di interagire in modo pertinente, naturale e cooperativo.

Nel corso del capitolo approfondiremo la teoria delle implicature conversazionali di Paul Grice (1975), mostrando come la comunicazione umana si fondi su un patto interpretativo implicito e su un principio di cooperazione che regola ciò che si dice, ciò che si tace e ciò che si lascia intendere. Le massime griceane sono proposte come linee guida per l'elaborazione di risposte credibili e pertinenti nei sistemi conversazionali, anche se la loro applicazione automatica da parte delle macchine rivela limiti sostanziali, soprattutto quando si tratta di gestire l'ambiguità, l'ironia o i riferimenti culturali impliciti.

Infine, verranno analizzate le critiche mosse al modello griceano e le evoluzioni teoriche che ne sono scaturite, come la *Relevance Theory* di Sperber e Wilson (1986), che pone l'accento sulla rilevanza cognitiva e sull'inferenza soggettiva, e il pensiero di Searle (1980), che richiama l'attenzione sui limiti ontologici dell'intelligenza artificiale nel generare comprensione autentica. L'esperimento della *Chinese Room* e il concetto di intenzionalità collettiva mettono in discussione l'idea che la corretta manipolazione del linguaggio equivalga a una reale competenza comunicativa.

Queste riflessioni costituiscono le fondamenta teoriche su cui poggia la sfida della progettazione della conversazione uomo-macchina: un'impresa che non può limitarsi all'efficienza funzionale, ma deve ispirarsi a una visione del linguaggio come azione sociale, relazione cooperativa e costruzione di significato condiviso (Clark, 1996).

1.1 Cos'è conversazione: la comunicazione come processo

Per comprendere le fondamenta teoriche che guidano la progettazione della conversazione tra uomo e macchina, è indispensabile analizzare la comunicazione come processo articolato, composto da una molteplicità di com-

ponenti interagenti. Diverse teorie e modelli hanno descritto il fenomeno comunicativo da punti di vista differenti, fornendo una base epistemologica fondamentale anche per lo sviluppo dei sistemi conversazionali intelligenti.

Uno dei primi modelli formali della comunicazione è quello sviluppato da Claude Shannon e Warren Weaver nel 1949. Il modello, pensato inizialmente per l'ambito delle telecomunicazioni, descrive la comunicazione come un processo lineare composto da sei elementi fondamentali: una fonte di informazione, un trasmettitore, un canale, un ricevitore, una destinazione e un possibile rumore che può interferire con la trasmissione del messaggio. Shannon e Weaver definiscono l'informazione come "la riduzione dell'incertezza", ponendo l'accento sull'efficienza del trasferimento piuttosto che sul contenuto semantico del messaggio. Come affermano: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point" (Shannon & Weaver, 1998: 5). Questo modello, che può essere rappresentato schematicamente come un flusso lineare tra gli elementi coinvolti, trova applicazione anche nella comunicazione uomo-macchina. In questo contesto, la fonte è l'utente umano, il trasmettitore può essere l'interfaccia vocale o testuale, il canale è il mezzo digitale, il ricevitore è il sistema di elaborazione linguistica e la destinazione è il modulo di risposta o azione. Il rumore può includere ambiguità linguistiche, errori di trascrizione, input non chiari o distorsioni dovute al riconoscimento vocale. Sebbene questo modello sia stato criticato per la sua rigidità e per la mancanza di attenzione agli aspetti semantici e pragmatici, rimane ancora oggi una base utile per descrivere i flussi comunicativi nei sistemi interattivi.

A questa impostazione lineare si contrappone una visione più articolata e multifunzionale della comunicazione, proposta da Roman Jakobson. Nel suo saggio del 1960 *Linguistics and Poetics*, Jakobson individua sei funzioni fondamentali del linguaggio, ciascuna delle quali corrisponde a un elemento del processo comunicativo: la funzione referenziale è legata al contesto e riguarda la trasmissione di informazioni oggettive sul mondo, come ad esempio l'enunciato "La temperatura è di 20 gradi"; la funzione emotiva esprime lo

stato d'animo del mittente, come nel caso di "Sono così felice!"; la funzione conativa è orientata al destinatario e include ordini, richieste, esortazioni, come in "Ascoltami!"; la funzione fatica riguarda il canale e serve per stabilire, mantenere o interrompere la comunicazione, come accade in "Pronto?", oppure in "Mi senti?"; la funzione metalinguistica si riferisce al codice ed è utile a chiarire il significato delle parole usate, come nell'espressione "Con 'digitale' intendo..."; infine, la funzione poetica è focalizzata sul messaggio in sé e include giochi di parole, figure retoriche e l'estetica del discorso, come avviene nella poesia o negli slogan pubblicitari. Jakobson associa ciascuna funzione a un elemento specifico della comunicazione: il mittente per la funzione emotiva, il destinatario per la conativa, il messaggio per la poetica, il contesto per la referenziale, il codice per la metalinguistica e il canale per la fatica. Queste funzioni non operano in modo esclusivo, ma spesso coesistono nello stesso enunciato, con una funzione prevalente che dipende dall'intenzione del parlante e dal contesto. Questo modello è estremamente rilevante per la progettazione delle interfacce conversazionali, che devono essere in grado di adattarsi a molteplici finalità comunicative e di riconoscere, ad esempio, quando un utente sta semplicemente testando il canale, quando sta esprimendo un'emozione o quando sta formulando una richiesta esplicita o implicita.

Il terzo modello teorico di riferimento, ancora più orientato all'interpretazione comportamentale e relazionale della comunicazione, è proposto da Watzlawick, Bavelas e Jackson nel noto testo *Pragmatics of Human Communication* (1967). L'approccio sistemico e pragmatico proposto dagli autori si basa sull'osservazione del comportamento comunicativo all'interno delle relazioni interpersonali. Il modello si fonda su cinque assiomi fondamentali che costituiscono il cuore della teoria. Il primo assioma afferma che è impossibile non comunicare: ogni comportamento, anche il silenzio, ha valore comunicativo, in quanto l'assenza di comunicazione intenzionale non implica l'assenza di interpretazione da parte dell'interlocutore. Il secondo assioma stabilisce che ogni comunicazione ha un aspetto di contenuto e uno di relazione, il che

significa che oltre al messaggio esplicito trasmesso (livello digitale), ogni interazione comunica anche qualcosa sul rapporto tra i partecipanti (livello analogico). Il terzo assioma riguarda la punteggiatura della sequenza di eventi, ovvero il fatto che gli individui organizzano gli scambi comunicativi in sequenze interpretative che possono divergere, generando incomprensioni. Il quarto assioma distingue tra comunicazione digitale e analogica: la prima si riferisce al linguaggio verbale, che è preciso e simbolico, mentre la seconda comprende la comunicazione non verbale, che è spesso ambigua ma altamente espressiva. Infine, il quinto assioma definisce le interazioni come simmetriche o complementari, a seconda che si basino sull'uguaglianza o sulla differenza di potere tra i partecipanti.

Questi principi, sebbene sviluppati nel contesto delle relazioni umane, trovano applicazione anche nella comunicazione uomo-macchina, dove il comportamento dell'interfaccia e la sua capacità di riconoscere e rispettare le dinamiche relazionali dell'utente diventano fondamentali. Ad esempio, un assistente vocale che ignora le funzioni fatiche o metalinguistiche può apparire meccanico o inadeguato. Inoltre, la capacità di un sistema conversazionale di interpretare correttamente la punteggiatura dello scambio (domande, repliche, feedback, silenzi) e di adattarsi alle relazioni simmetriche (utente-tecnologia) o complementari (cliente-servizio) è cruciale per garantire un'interazione efficace e soddisfacente.

In sintesi, i modelli proposti da Shannon e Weaver, Jakobson e Watzlawick et al. offrono tre prospettive complementari sul fenomeno comunicativo: la trasmissione tecnica dell'informazione, l'articolazione funzionale del linguaggio e l'interazione relazionale e comportamentale. Integrare queste prospettive nella progettazione della comunicazione uomo-macchina consente di costruire sistemi conversazionali più completi, in grado di comprendere il linguaggio non solo come codice, ma come azione, relazione e significato situato.

1.2 La pragmatica linguistica come chiave di lettura

Se la semantica descrive il contenuto convenzionale codificato delle espressioni linguistiche, le regole che lo compongono e le sue dipendenze dal contesto, la pragmatica si interessa al modo in cui quel contenuto assume significato in specifici contesti d'uso (Yule, 1996). La pragmatica, dunque, affronta il linguaggio come azione situata e socialmente regolata, riconoscendo l'importanza delle intenzioni comunicative, del contesto fisico e delle relazioni interpersonali, e rilevando come gli utenti impiegano le parole non solo per descrivere, ma anche per agire. In questa prospettiva, il linguaggio è anche azione: la semantica, più legata alla struttura formale del linguaggio, fornisce la base interpretativa su cui tale azione si innesta.

Il contesto gioca un ruolo essenziale nella costruzione del significato. Ad esempio, l'enunciato "Puoi aprire la finestra?" può essere interpretato come una semplice domanda, come una richiesta gentile, o persino come un rimprovero velato, a seconda del tono, della situazione e della relazione tra i parlanti. È proprio nella variazione interpretativa dovuta al contesto che la pragmatica trova la sua specificità: non è tanto ciò che si dice, quanto ciò che si intende e come viene inteso.

Il filosofo del linguaggio John L. Austin, nel suo celebre saggio *How to Do Things with Words* (1975), ha introdotto la nozione di atti linguistici (*speech acts*), segnando l'inizio della pragmatica moderna. Austin distingue tra: atto locutorio (ovvero l'atto di pronunciare una frase con una certa struttura e significato), atto illocutorio (ovvero l'azione che si compie pronunciando quella frase, es. ordinare, promettere, avvertire), e atto perlocutorio (ovvero l'effetto che la frase ha sull'interlocutore). Un'affermazione come "Ti prometto che verrò" non è semplicemente una proposizione: è un atto che crea un vincolo sociale. Austin introduce inoltre i concetti di felicità e infelicità degli atti linguistici: un atto è "felice" quando le condizioni contestuali e sociali sono rispettate (ad esempio, una promessa è valida se chi parla ha l'intenzione e la possibilità di mantenerla); è "infelice" quando tali condizioni

vengono violate (come nel caso di una promessa fatta con l'intenzione di non mantenerla o da una persona non legittimata a farla).

Nelle opere *Speech Acts* (1969) e *Speech Act Theory and Pragmatics* (1980), John Searle propone una teoria dell'intenzionalità e della regolazione linguistica che è cruciale per comprendere la comunicazione linguistica come comportamento intenzionale e regolato (Searle, 1980). Secondo Searle, la comunicazione non è una mera trasmissione di significato, ma un'azione volontaria che avviene in un sistema di regole condivise (Searle, 1969). In *Minds, brains and programs* (1980) l'intenzionalità viene intesa come la capacità degli esseri umani di orientare i loro stati mentali verso oggetti o situazioni del mondo. Ogni atto linguistico è quindi un'azione intenzionale con una direzione verso un obiettivo specifico: promettere, chiedere, affermare non sono solo atti verbali, ma azioni finalizzate a produrre effetti precisi.

Searle afferma che ogni atto linguistico è composto da due elementi principali: il contenuto proposizionale (ciò di cui si parla, ad esempio "Piove") e la forza illocutoria (l'intenzione di chi parla, ad esempio affermare che piove). Da questa combinazione nasce il significato comunicativo dell'atto, che non è mai solo legato alle parole, ma sempre anche al contesto e all'intenzione.

La comunicazione, per Searle, è governata da regole che strutturano e rendono possibile il linguaggio. Le regole costitutive definiscono gli atti linguistici stessi e stabiliscono cosa significa fare un'asserzione, una richiesta o un ordine. Le regole regolative, invece, determinano le condizioni in cui è possibile eseguire correttamente tali atti (Searle, 1969): ad esempio, per fare una promessa valida, il parlante deve avere l'intenzione di mantenerla. Searle parla di condizioni di felicità e di appropriatezza, che includono anche la sincerità (non si può promettere ciò che non si intende fare) e le condizioni contestuali (un ordine è valido solo se emesso da una figura autoritaria in un contesto pertinente).

L'autore sviluppa ulteriormente queste riflessioni in *Expression and Meaning* (1979), introducendo una tipologia più dettagliata di condizioni che

regolano la validità degli atti linguistici, note come condizioni di appropriatezza (Searle, 1979). Tali condizioni sono classificate in tre principali categorie: condizioni preparatorie, condizioni di sincerità e condizioni essenziali.

Le condizioni preparatorie riguardano le precondizioni che devono sussistere affinché un atto linguistico sia effettivamente realizzabile. Esse variano a seconda del tipo di atto. Per esempio, in una richiesta, è necessario che il destinatario abbia la capacità di compiere l'azione richiesta; una promessa, invece, richiede che il parlante sia realisticamente in grado di portare a termine l'azione promessa.

Le condizioni di sincerità si riferiscono all'intenzione genuina del parlante. Un atto linguistico è valido solo se il parlante crede sinceramente in ciò che sta dicendo o promettendo. Se un individuo pronuncia una promessa senza l'intenzione reale di rispettarla, l'atto linguistico, pur formalmente corretto, è invalido da un punto di vista pragmatico.

Infine, le condizioni essenziali sono ciò che definisce l'identità stessa dell'atto linguistico. Per esempio, un'affermazione è tale solo se il parlante intende rappresentare una proposizione come vera. Anche se la frase è corretta grammaticalmente, essa non costituisce un'asserzione autentica senza questa intenzione. Queste condizioni determinano la natura dell'atto e lo distinguono da altri tipi di espressioni che non comportano un'azione comunicativa, come ad esempio dichiarazioni senza scopo pragmatico o ripetizioni meccaniche.

Le condizioni di appropriatezza non solo determinano se un atto è realizzabile e valido, ma sottolineano anche la natura normata e sociale della comunicazione. Questa distinzione tra validità formale e appropriatezza contestuale degli atti linguistici sottolinea come la comunicazione umana sia regolata da norme condivise e da intenzioni cooperative che vanno oltre la semplice produzione di enunciati grammaticalmente corretti. Le condizioni di appropriatezza, infatti, rivelano la complessità delle interazioni linguistiche, evidenziando come il significato emerga da una rete di fattori pragmatici e sociali più che da una struttura puramente sintattica.

Le regole operano su più livelli: determinano i tipi di atti linguistici (es. asserzioni, richieste, domande), ma anche le condizioni per cui tali atti siano socialmente riconosciuti e accettati. Ad esempio, se un bambino ordina a un adulto di fare qualcosa, anche se usa una forma grammaticale corretta, l'enunciato non verrà interpretato come un vero ordine, perché mancano le condizioni sociali per conferirgli forza illocutoria.

Searle insiste su un punto cruciale: parlare una lingua significa partecipare a un gioco regolato culturalmente ("speaking a language is engaging in a rule-governed form of behavior", Searle, 1969: 16). Le regole linguistiche non sono semplici convenzioni, ma costituiscono le condizioni che rendono possibile il significato. La comunicazione linguistica è quindi un comportamento normato in cui le intenzioni del parlante devono rispettare una rete di condizioni formali e sociali per essere comprese e accettate.

Questa visione è rafforzata anche da Stephen C. Levinson, che nel suo manuale *Pragmatics* (1983) amplia la riflessione sul ruolo del contesto e introduce una prospettiva sociolinguistica e conversazionale. Un tema centrale è l'organizzazione dei turni di parola (*turn-taking*), ovvero le modalità con cui i partecipanti a una conversazione si alternano nel parlare. Levinson sottolinea che questa alternanza non è casuale, ma regolata da norme implicite che governano l'interazione linguistica in modo sorprendentemente ordinato. Le conversazioni seguono un principio sequenziale: un intervento è seguito da un altro secondo schemi riconoscibili, come nelle domande e risposte, saluti e controrisposte, inviti e accettazioni/rifiuti.

Le regole implicite che gestiscono l'alternanza dei turni sono apprese socialmente e condivise culturalmente. Sebbene raramente formalizzate, esse sono efficaci nel regolare chi parla, quando e per quanto tempo. Il passaggio di turno può avvenire attraverso segnali verbali come "allora..." o "cioè...", che indicano l'intenzione di continuare a parlare, oppure mediante pause strategiche che invitano un altro partecipante a prendere la parola. I segnali non verbali, come il contatto visivo, il gesto del capo, o la modulazione del tono di voce, svolgono un ruolo cruciale nel coordinamento delle transizioni, se-

gnalando che un turno si sta concludendo o che si desidera intervenire. Levinson osserva che queste dinamiche si mantengono sorprendentemente coerenti tra culture diverse, pur con varianti specifiche.

Un altro aspetto di rilievo riguarda le preferenze conversazionali, ovvero le tendenze sistematiche che regolano il modo in cui le sequenze di interazione vengono strutturate. Le risposte che mantengono la coerenza o l'accordo con il turno precedente (come un "sì" a una domanda) sono preferite, mentre quelle che introducono disaccordo o rifiuto vengono elaborate con maggiore cautela, spesso precedute da esitazioni, mitigazioni o giustificazioni. Queste strategie sono adottate per preservare l'armonia e la cooperazione interazionale, riflettendo la dimensione sociale del linguaggio.

La gestione dei turni e delle preferenze conversazionali ha implicazioni profonde nella progettazione dei sistemi di comunicazione uomo-macchina. Gli assistenti virtuali devono essere in grado di identificare quando è il loro turno per parlare, come rispondere in modo pertinente e quando cedere la parola all'utente. Questo richiede non solo una comprensione sintattica o semantica del messaggio, ma anche un'interpretazione pragmatica dell'interazione. Levinson evidenzia come le conversazioni naturali siano governate da una grammatica sottostante dei turni, che le macchine devono imitare per risultare credibili e collaborative (Levinson, 1983). I modelli conversazionali devono quindi integrare meccanismi per il riconoscimento delle sequenze preferite e per la gestione di quelle dispreferite, adattando le risposte in base al contesto e al tono dell'interazione.

L'integrazione di queste dinamiche nella progettazione di interfacce intelligenti è fondamentale per superare l'effetto meccanico o frammentato tipico delle interazioni digitali. Solo riconoscendo e implementando le regole implicite dell'alternanza conversazionale e delle preferenze interazionali sarà possibile sviluppare sistemi di IA capaci di gestire scambi naturali, cooperativi e socialmente accettabili. Prendere in considerazione questi aspetti amplia la riflessione sul ruolo del contesto e introduce una prospettiva sociolinguistica e conversazionale.

Levinson considera centrale il concetto di deissi nello studio del linguaggio perché strettamente legato all'interpretazione contestuale degli enunciati (Levinson, 1983). La deissi si riferisce a elementi linguistici come i pronomi personali, gli avverbi di tempo e spazio e le forme di cortesia, il cui significato si determina nell'atto d'uso. L'interpretazione corretta di tali elementi è fondamentale, tanto per la comprensione tra esseri umani quanto per la progettazione efficace di sistemi conversazionali artificiali, i quali devono essere capaci di inferire dinamicamente i riferimenti deittici per poter interagire in modo naturale mantenendo coerenza e pertinenza dello scambio. La deissi, pertanto, esemplifica la dipendenza del significato dal contesto e rappresenta una sfida nella realizzazione di una comunicazione uomo-macchina realmente situata.

Levinson osserva che gli interlocutori di una conversazione fanno affidamento su convenzioni implicite. Tali convenzioni consentono di interpretare significati non detti ma suggeriti, come nel caso delle implicature conversazionali, concetto che approfondiremo nel paragrafo successivo. Inoltre, egli evidenzia l'importanza della struttura conversazionale: l'organizzazione dei turni di parola, la gestione delle interruzioni, i segnali di ascolto e i meccanismi di riparazione. Questi elementi sono fondamentali per la fluidità del dialogo umano e rappresentano una delle sfide principali nella progettazione di interfacce conversazionali artificiali.

Un altro aspetto cruciale analizzato da Levinson è il concetto di presupposizione, intesa come inferenza pragmatica distinta da altre forme inferenziali, come le implicature (Levinson, 1983). La presupposizione rappresenta una condizione di background che deve essere vera affinché l'enunciato principale abbia senso. Ad esempio, nella frase "Michele ha smesso di fumare", si presuppone che Michele prima fumasse. Se la presupposizione è falsa, l'enunciato resta formalmente corretto, ma perde di coerenza pragmatica. Ciò distingue la presupposizione dalle implicature, che sono più sensibili al contesto e alla cooperazione conversazionale.

Levinson distingue tra presupposizioni semantiche e presupposizioni pragmatiche. Le prime sono collegate alla struttura della frase e includono

verità implicite necessarie per la comprensione. Ad esempio, l'enunciato "Ambra è ancora viva" presuppone che Ambra sia viva nel momento dell'enunciazione. Le seconde, invece, emergono dall'uso linguistico e dal contesto: dire "Non è più il caso che Michele parli così" implica che Michele parlasse in un certo modo anche precedentemente.

Le presupposizioni si differenziano dalle implicature sotto vari aspetti. In primo luogo, esse mostrano una resistenza alla negazione: negando l'enunciato "Michele ha smesso di fumare", la presupposizione che Michele fumasse prima resta implicita. Inoltre, le presupposizioni hanno la capacità di proiettarsi attraverso contesti logici complessi, mantenendosi valide anche in frasi subordinate o ipotetiche, come "Non è vero che Michele ha smesso di fumare".

Queste proprietà rendono le presupposizioni particolarmente rilevanti nella progettazione della comunicazione uomo-macchina. Nei sistemi di intelligenza artificiale e di elaborazione del linguaggio naturale, la capacità di riconoscere e gestire le presupposizioni è cruciale per comprendere appieno il significato implicito di un'istruzione o di una domanda. L'integrazione delle presupposizioni nell'interpretazione del discorso permette al sistema di allinearsi meglio con le intenzioni e con le aspettative dell'interlocutore umano. Questo è particolarmente vero negli assistenti virtuali e nei chatbot, dove la corretta elaborazione delle presupposizioni può fare la differenza tra una risposta pertinente e una generica o fuorviante.

In sintesi, le presupposizioni, come descritte da Levinson, costituiscono un elemento fondamentale della pragmatica e svolgono un ruolo chiave nella comprensione del significato implicito. Le loro stabilità e proiezione nei contesti complessi le rendono uno strumento potente per la costruzione di interazioni naturali ed efficaci tra esseri umani e macchine.

Concludendo, la pragmatica fornisce gli strumenti teorici per comprendere il linguaggio come comportamento situato, regolato da norme condivise e finalizzato a uno scopo. Questa prospettiva è essenziale per progettare interazioni efficaci tra esseri umani e macchine, in cui la capacità di interpretare

correttamente ciò che si intende, più che ciò che si dice, diventa il fulcro di una comunicazione autentica.

1.3 Le massime conversazionali di Grice: cooperazione e inferenza nel linguaggio ordinario

Un contributo fondamentale alla comprensione del linguaggio naturale è stato offerto dal filosofo Paul Grice nel saggio *Logic and Conversation* (1975), in cui viene esplorata la dimensione implicita e cooperativa del significato linguistico. A partire dall'osservazione che ciò che viene detto esplicitamente spesso non coincide con ciò che si intende comunicare, Grice elabora una teoria dell'inferenza pragmatica che ha influenzato profondamente gli studi successivi sulla comunicazione verbale. Egli si distanzia dai dispositivi linguistici formali propri della logica proposizionale, per i quali il significato è interamente contenuto nella forma linguistica. Nella conversazione quotidiana, infatti, il significato emerge dalla relazione tra ciò che viene detto e ciò che si presume l'interlocutore voglia intendere, sulla base del contesto e delle aspettative condivise. Pertanto, il linguaggio naturale non può essere spiegato unicamente attraverso regole formali, poiché include costantemente elementi impliciti, allusivi e contestuali che devono essere inferiti.

Per spiegare queste dinamiche, Grice introduce il concetto di implicatura. L'implicatura è una forma di significato implicito che si genera quando un enunciato, pur non esplicitando un contenuto, ne consente la deduzione attraverso un meccanismo inferenziale. Le implicature convenzionali sono distinte da quelle conversazionali (Grice, 1975). Le prime dipendono dal significato lessicale di certi termini: ad esempio, la congiunzione "ma" implica un contrasto, anche se formalmente equivale a "e". Le implicature conversazionali, invece, emergono dal contesto dell'interazione e dall'applicazione di principi pragmatici. Quando un parlante afferma: "C'è una stazione di servi-

zio dietro l'angolo", in risposta a "Sto finendo la benzina", l'ascoltatore inferisce che la stazione è aperta e disponibile, anche se ciò non è stato detto direttamente. Questa capacità di generare e interpretare significati ulteriori è ciò che rende il linguaggio umano estremamente flessibile e sottile.

Alla base del funzionamento delle implicature conversazionali vi è il cosiddetto Principio di Cooperazione, formulato da Grice come una norma generale che regola le interazioni verbali: "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1975). Questo principio implica che i parlanti si comportino come se collaborassero razionalmente per raggiungere uno scopo comune nella comunicazione. Grice articola il Principio di Cooperazione in quattro massime conversazionali: Quantità, Qualità, Relazione e Modo. La massima di Quantità suggerisce di fornire né più né meno informazioni del necessario. La massima di Qualità richiede di non dire ciò che si sa essere falso o non supportato da prove. La massima di Relazione prescrive di essere pertinenti, mentre la massima di Modo invita alla chiarezza, evitando ambiguità e disordine espositivo. Occorre sottolineare che il modello delle massime non agisce in modo isolato: esso interagisce con altri fattori sociali che regolano la conversazione, come per esempio la salvaguardia della *face* positiva e negativa secondo la teoria della cortesia di Brown e Levinson[1]. In molte situazioni, infatti, la scelta di rispettare o infrangere una massima è condizionata anche da strategie di cortesia e dalla necessità di mantenere l'armonia interazionale.

Tuttavia, ciò che rende il modello griceano particolarmente efficace è il riconoscimento che le massime possono essere apparentemente violate senza compromettere la comunicazione. Anzi, molte implicature derivano proprio dalle violazioni strategiche di queste massime. Si individuano tre modalità attraverso cui le massime possono essere disattese: violazione, sospensione e

[1] Per *face* positiva si intende il desiderio di ciascun parlante di essere approvato dagli altri; al contrario, per *face* negativa si intende il bisogno di preservare la propria libertà d'azione. Cfr. Brown, P. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge University Press.

cancellazione (Grice, 1975). Una massima è violata quando il parlante fornisce intenzionalmente un'informazione falsa o fuorviante, come per esempio nel caso della bugia, dove si verifica la violazione della massima di Qualità. Al contrario, una massima è sospesa quando il contesto rende irrilevante la sua applicazione: per esempio, nella poesia o nell'umorismo, l'ambiguità e l'imprecisione possono essere intenzionalmente valorizzate. Infine, una massima è cancellata quando il parlante segnala esplicitamente che l'implicatura associata non è da considerare valida, come nell'espressione "non voglio dire che...". Questo modello di flessibilità pragmatica evidenzia quanto la comunicazione linguistica sia fondata su un patto interpretativo, in cui una violazione apparente delle regole può servire a rafforzare l'efficacia comunicativa.

Nel quadro della ricerca sulla progettazione della conversazione uomo-macchina, la teoria di Grice assume una rilevanza strategica. Le massime conversazionali rappresentano un riferimento fondamentale per il design di interfacce dialogiche capaci di produrre interazioni coerenti, rilevanti e comprensibili. Eppure, le difficoltà principali potrebbero risiedere proprio nella natura implicita, contestuale e dinamica, delle implicature. Un sistema di intelligenza artificiale può essere programmato per seguire le massime in modo rigido, ma fatica a gestire i casi in cui tali massime vengono infrante intenzionalmente o quando la cooperazione è implicita ma non codificata. La vera sfida è dunque quella di progettare sistemi in grado di riconoscere segnali deboli, ambigui o ironici, e di gestire le aspettative implicite dell'utente attraverso modelli inferenziali adattivi (Grice, 1975).

La teoria di Grice, nel suo insieme, non offre solo una descrizione del linguaggio naturale, ma propone un modello di razionalità comunicativa che può essere assunto come riferimento teorico per la costruzione di agenti conversazionali realmente capaci di interazione pragmatica. Essa evidenzia che la comprensione non è un processo di decodifica passiva, ma un atto interpretativo attivo e situato, basato su principi cooperativi, capacità inferenziali e conoscenze condivise. Per questa ragione, ogni riflessione sulla comunica-

zione tra uomo e macchina che ambisca a superare la mera esecuzione di comandi deve necessariamente confrontarsi con le implicazioni della pragmatica griceana.

1.4 Grice e l'interazione uomo-macchina

La crescente diffusione di sistemi conversazionali basati sull'intelligenza artificiale, come chatbot e assistenti vocali, rende necessario interrogarsi sulla possibilità di applicare le teorie linguistiche classiche alla progettazione di queste interfacce. Tra le prospettive teoriche più influenti, la pragmatica griceana offre un paradigma interpretativo utile a comprendere i meccanismi che regolano l'interazione verbale e a valutarne la trasposizione nei contesti computazionali. Tuttavia, la stessa complessità delle categorie concettuali introdotte da Grice, quali le implicature conversazionali, il principio di cooperazione e le massime pragmatiche, pone limiti strutturali alla loro replicabilità da parte delle macchine.

La possibilità della cooperazione comunicativa nei sistemi artificiali è strettamente connessa alla capacità dell'agente computazionale di interpretare il contesto, modellare le intenzioni comunicative dell'utente e generare risposte coerenti. Il principio di cooperazione di Grice, fondato sull'assunzione che i partecipanti a uno scambio comunicativo contribuiscano in modo razionale e pertinente allo sviluppo del discorso, si basa su presupposti cognitivi e sociali che non sono nativamente presenti nei sistemi artificiali. Se l'essere umano è in grado di adattare dinamicamente il proprio comportamento linguistico in funzione del contesto situazionale e delle intenzioni altrui, le macchine operano invece su base algoritmica e probabilistica, e necessitano di regole esplicite per simulare tali adattamenti.

La trasposizione delle massime conversazionali nel design delle interfacce uomo-macchina rappresenta un tentativo di strutturare la comunica-

zione computazionale in modo più naturale ed efficace. La massima di Quantità, per esempio, può essere implementata attraverso la sintesi di risposte concise ma informative, mentre la massima di Relazione orienta la pertinenza delle risposte in funzione della richiesta dell'utente. Analogamente, la massima di Modo può guidare la progettazione di output linguistici semplici e non ambigui. Al contrario, la massima di Qualità risulta più problematica, in quanto la nozione di verità in un sistema computazionale è subordinata alla qualità dei dati e al funzionamento del modello predittivo sottostante. Di conseguenza, non sempre le informazioni fornite da un assistente digitale sono verificabili o affidabili.

La maggiore difficoltà riguarda tuttavia il trattamento delle implicature, che rappresentano un elemento centrale del modello griceano. Le implicature non sono codificate nel testo, ma emergono dalla relazione tra l'enunciato e il contesto pragmatico, nonché dalle aspettative condivise tra interlocutori. Un chatbot può essere programmato per generare implicature semplici, come l'implicatura generalizzata "non tutti" nella frase "alcuni clienti hanno risposto", ma difficilmente riesce a gestire implicature conversazionali particolarizzate che dipendono da conoscenze implicite, ironia, intenzionalità indiretta o riferimenti culturali. Questo limite risulta evidente nelle conversazioni complesse, dove l'interazione richiede capacità inferenziali adattive e una sensibilità alla dimensione implicita del linguaggio.

Occorre ricordare che tali limiti non dipendono soltanto da vincoli architetturali o dal funzionamento statistico dei modelli. Gli LLM sono infatti addestrati principalmente su testi scritti e non su dialoghi spontanei: di conseguenza, hanno introiettato in misura minore le dinamiche tipiche della conversazione orale, come la gestione dei turni, le interruzioni, i segnali discorsivi o la prosodia, elementi che risultano centrali per l'interazione naturale tra esseri umani.

Tuttavia, le interfacce conversazionali attuali tendono a simulare l'adesione alle massime griciane in modo rigido e formalizzato, senza possedere la flessibilità pragmatica tipica della comunicazione umana. Come osserva Herring (1999), le interazioni uomo-macchina risultano spesso prevedibili e

poco naturali proprio perché mancano di quella capacità di violare strategicamente le massime – come accade nelle implicature ironiche o nelle sospensioni contestuali – per creare significati più ricchi e sfumati. Ciò comporta una comunicazione meno efficace, soprattutto in contesti non standardizzati.

Alla luce di queste considerazioni, l'applicazione delle massime conversazionali alla progettazione delle interfacce dialogiche rappresenta una sfida ancora aperta. Le teorie di Grice possono fornire un quadro teorico utile per definire criteri di naturalezza e cooperazione comunicativa, ma richiedono un'elaborazione tecnica e interpretativa che tenga conto dei limiti strutturali dei modelli computazionali. La strada verso una vera conversazione tra uomo e macchina, capace di generare e interpretare implicature in modo credibile, passa dunque non solo per l'aumento della potenza computazionale, ma soprattutto per una profonda integrazione tra scienze linguistiche, cognitive e ingegneria del linguaggio.

1.5 Critiche e sviluppi delle teorie di Grice

Il modello teorico di Paul Grice ha avuto, come abbiamo visto, un impatto significativo nello studio della comunicazione linguistica, fornendo una cornice interpretativa solida per comprendere il significato implicito nelle conversazioni quotidiane. La formulazione del principio di cooperazione e delle massime conversazionali ha permesso di analizzare in termini razionali e sistematici come i parlanti intendano più di quanto esprimano esplicitamente, affidandosi alla capacità inferenziale dell'interlocutore. Tuttavia, nonostante l'influenza duratura di queste idee, alcuni studiosi hanno messo in luce limiti teorici e ambiguità insite nella teoria griceana, proponendo modelli alternativi che ne ampliano o riformulano i presupposti.

Una prima area di riflessione critica riguarda la presunta universalità del modello griceano. L'assunto di base secondo cui i partecipanti a una conversazione agirebbero sempre in modo razionale e cooperativo non sembra

descrivere adeguatamente la molteplicità delle situazioni comunicative reali, nelle quali possono intervenire fattori di natura sociale, emotiva o strategica che deviano dalla cooperazione ideale. La teoria griceana è stata anche accusata di presupporre un interlocutore ideale, razionale e astratto, trascurando le dinamiche sociali, culturali e strategiche che condizionano la comunicazione reale. Oltre a ciò, è stata criticata la vaghezza del concetto stesso di "cooperazione", che nella teoria griceana assume un significato tecnico spesso frainteso rispetto alla sua accezione ordinaria di collaborazione empatica o solidarietà sociale.

Le massime proposte da Grice, pur costituendo strumenti analitici efficaci, si configurano talvolta come linee guida troppo rigide per spiegare la complessità del comportamento linguistico in contesti caratterizzati da ambiguità, reticenza, ironia o manipolazione. Grice stesso ha riconosciuto la varietà delle situazioni in cui le massime possono non essere osservate, distinguendo cinque casi principali: trasgressione (*flouting*), violazione (*violation*), inadempienza (*infringing*), rinuncia (*opting out*) e sospensione (*suspending*), ciascuno con implicazioni differenti per l'interpretazione pragmatica (Grice, 1975). Questa classificazione, se da un lato testimonia la ricchezza descrittiva del modello, dall'altro ne rivela la complessità e la necessità di costanti precisazioni e adattamenti.

In particolare, alcuni studiosi hanno messo in discussione l'applicazione delle implicature conversazionali a figure retoriche come l'ironia e la metafora, sostenendo che tali fenomeni non si prestano sempre a una spiegazione in termini di rispetto o di violazione delle massime. L'interpretazione figurata, infatti, implicherebbe la sostituzione del significato letterale con uno implicito, confermando piuttosto la violazione delle massime che non la loro osservanza.

A partire da queste considerazioni, Dan Sperber e Deirdre Wilson (1986) hanno proposto una teoria alternativa, nota come *Relevance Theory*, che riformula in chiave cognitiva l'interazione comunicativa. Questa teoria abbandona il principio di cooperazione griceano e le sue articolazioni in massime, sostituendoli con un unico principio generale: gli esseri umani tendono

a selezionare, tra i molteplici stimoli informativi, quelli che promettono il massimo guadagno cognitivo con il minimo sforzo interpretativo. Ogni atto comunicativo viene quindi trattato come un tentativo di offrire un contenuto che l'interlocutore giudicherà sufficientemente rilevante in relazione al proprio stato mentale e alle proprie aspettative (Sperber & Wilson, 1986).

La *Relevance Theory* sposta l'attenzione dalla cooperazione sociale tra gli interlocutori alle capacità cognitive del ricevente, che diventa il vero protagonista dell'interpretazione. L'inferenza pragmatica non si fonda più sulla violazione o sull'osservanza di massime condivise, ma sulla valutazione soggettiva della rilevanza di uno stimolo comunicativo. Questo approccio permette di spiegare fenomeni linguistici complessi come l'ironia, la metafora, la vaghezza o l'allusione, non come eccezioni o violazioni, bensì come modalità pienamente legittime di comunicazione, fondate sulla condivisione di conoscenze implicite e sulla capacità del destinatario di compiere inferenze pertinenti.

La flessibilità della *Relevance Theory* consente inoltre di modellare le situazioni comunicative non perfettamente cooperative o strutturate, introducendo una prospettiva più realistica e meno prescrittiva nella descrizione degli scambi linguistici. Essa sottolinea come la comunicazione umana sia profondamente radicata in meccanismi cognitivi inferenziali e nella capacità di attribuire stati mentali agli altri, piuttosto che nel rispetto formale di regole conversazionali. Se il modello griceano ha posto le basi per la comprensione del significato implicito e dell'inferenza comunicativa, la *Relevance Theory* ha ampliato questo orizzonte, proponendo una visione più dinamica, contestualizzata e centrata sulla mente dell'interprete.

Infine, nel dibattito teorico si sono inseriti anche i cosiddetti "neo-griceani", studiosi come Martinich, Levinson e Leech, che hanno cercato di emendare o estendere il modello originario senza abbandonarne del tutto i presupposti (Panfili et al., 2021). Martinich ha proposto massime più ampie legate all'autenticità del parlante: "Be authentic. That is not knowingly participating in a speech act for which the conditions for its successful and non-defective performance are not satisfied" (Martinich, 1984), mentre Levinson

ha introdotto la distinzione tra implicature generalizzate (GCI) e particolarizzate (PCI), offrendo una soluzione a uno dei problemi centrali del modello griceano: la gestione dei conflitti tra massime (Levinson, 1983). Leech, invece, ha affiancato al principio cooperativo quello della *politeness* e dell'ironia, proponendo una pragmatica orientata agli scopi della comunicazione e fondata su principi motivazionali più ampi (Leech, 2014). Hossain, infine, ha evidenziato i limiti e le possibili estensioni del modello griceano, richiamando in particolare le cinque modalità di non-osservanza delle massime: "More importantly, our main concern in this research paper is that Grice makes a distinction between observing maxims and non-observance of the maxims and there are five cases where maxims are not observed. According to Grice they are, flouting the maxims, violating the maxims, infringing the maxims, opting out the maxims and suspending the maxims" (Hossain, 2021).

Questi sviluppi testimoniano il ruolo centrale ma anche problematico del pensiero griceano nella pragmatica contemporanea: un punto di partenza fondamentale, ma non una teoria esaustiva dei fenomeni comunicativi.

1.6 *Chinese room*: il pensiero di Searle tra comprensione e costruzione sociale

Il dibattito sul rapporto tra linguaggio e intelligenza artificiale si è intensificato negli ultimi decenni, trovando una delle sue espressioni più influenti nelle riflessioni filosofiche di John R. Searle. L'autore si è distinto per una critica radicale all'assunto secondo cui l'elaborazione sintattica di simboli, tipica dei sistemi computazionali, possa di per sé generare stati mentali dotati di significato. Il fulcro di questa critica si trova nel suo saggio *Minds, Brains and Programs*, in cui viene formulato il celebre esperimento della *Chinese room* (Searle, 1980).

In questo esperimento mentale, Searle immagina un soggetto chiuso in una stanza, intento a manipolare simboli in cinese secondo precise istruzioni

senza comprenderne il significato. L'esterno riceve risposte linguisticamente corrette, ma l'uomo nella stanza non sa cosa stia comunicando. L'esperimento, che mira a confutare l'ipotesi dell'intelligenza artificiale forte, dimostra secondo Searle che la comprensione semantica non può emergere dalla sola manipolazione sintattica dei simboli. La tesi sottostante è che l'intenzionalità, ossia la capacità di riferirsi consapevolmente a oggetti, situazioni o concetti, è una proprietà intrinsecamente biologica, non riproducibile attraverso algoritmi simbolici.

Il lavoro di Searle ha suscitato reazioni contrastanti. Tra le critiche più significative, si annoverano quelle di Daniel Dennett (1994) e Marvin Minsky (1986), i quali hanno sostenuto che l'esperimento della "stanza cinese" sottovaluti il ruolo del sistema complessivo. Secondo questi autori, è l'intero sistema – uomo, regole, supporti materiali – a costituire un'unità funzionale capace, potenzialmente, di produrre significato. Tale interpretazione, nota come *systems reply*, è stata oggetto di una dettagliata controreplica da parte di Searle, il quale sottolinea che in nessuna parte del sistema vi è un soggetto in grado di intendere il contenuto manipolato. Anche l'eventuale capacità del sistema di rispondere coerentemente in una conversazione non implica, per l'autore, l'esistenza di comprensione, analogamente a come un registratore può riprodurre una voce umana senza comprendere ciò che dice.

La posizione di Searle si chiarisce ulteriormente se si considera il suo contributo nel volume *Creare il mondo sociale. La struttura della civiltà umana* (2010), in cui viene delineata una teoria della realtà istituzionale basata su atti linguistici, regole costitutive e intenzionalità collettiva. In quest'opera, l'autore mostra come molte delle strutture fondamentali della nostra convivenza, dal denaro alle leggi, dai matrimoni ai ruoli professionali, esistano in virtù di un riconoscimento collettivo organizzato linguisticamente. Il principio strutturante è espresso dalla formula "X conta come Y nel contesto C": ad esempio, un foglio di carta conta come banconota da dieci euro nel contesto dell'economia europea (Searle, 2010).

Questo meccanismo di assegnazione funzionale è reso possibile da quella che Searle chiama intenzionalità collettiva, ovvero la capacità di una

comunità di condividere un orizzonte semantico e normativo comune. Tale capacità è veicolata in primis dal linguaggio, che, oltre alla funzione comunicativa, esercita una funzione costitutiva: crea nuovi stati di fatto attraverso dichiarazioni che trasformano la realtà sociale (ad esempio: "Vi dichiaro marito e moglie").

Applicare questa cornice alla questione dell'intelligenza artificiale comporta una riflessione critica: sebbene un assistente virtuale possa essere formalmente riconosciuto come agente conversazionale in un determinato contesto istituzionale, ciò non comporta che esso sia effettivamente dotato di intenzionalità o comprensione. Può essere trattato come se comprendesse, ma tale riconoscimento è di tipo funzionale, non ontologico. Si potrebbe dire che l'IA conta come interlocutore valido nel contesto C, ma non per questo è un interlocutore dotato di mente. Inoltre, l'IA non sembra possedere una vera e propria "teoria della mente": non attribuisce stati mentali coscienti ai suoi interlocutori, ma si limita a simulare risposte linguistiche prive di consapevolezza intenzionale.

Questo discrimine è di fondamentale importanza per la progettazione di sistemi di comunicazione uomo-macchina. Da un lato, le interfacce devono essere costruite in modo da rispettare le regole pragmatiche del linguaggio umano, simulando atti linguistici coerenti con le aspettative degli utenti. Dall'altro, tale simulazione non deve essere confusa con una reale comprensione. Per Searle, l'errore della cosiddetta IA forte è proprio quello di scambiare la funzionalità sintattica per intenzionalità semantica, attribuendo alla macchina una competenza che essa non può possedere, in quanto mancante del substrato biologico e cosciente che caratterizza la mente umana (Searle, 2010).

Questa distinzione comporta anche conseguenze etiche e sociali. Il rischio non è solo quello di sopravvalutare le capacità delle macchine, ma anche di ridurre il senso stesso di comunicare, comprendere, pensare. Se la comprensione viene identificata esclusivamente con la correttezza funzionale delle risposte, si perde la profondità dell'esperienza linguistica umana, fatta di contesto, intenzioni, emozioni e riconoscimento reciproco. In questo senso,

il pensiero di Searle rappresenta un richiamo alla responsabilità epistemologica e progettuale nell'ambito dell'intelligenza artificiale linguistica.

In sintesi, l'opera di Searle offre un contributo teorico di straordinaria rilevanza per comprendere i limiti strutturali dell'interazione tra uomo e macchina. Il suo approccio integra riflessioni linguistiche, epistemologiche e sociali, mostrando come il linguaggio non sia solo un codice, ma una pratica sociale costitutiva, che presuppone intenzionalità e riconoscimento condiviso. Qualsiasi progetto di comunicazione artificiale che ambisca a essere più di una simulazione tecnica dovrà necessariamente confrontarsi con queste premesse filosofiche, ponendo al centro non solo l'efficacia dell'interazione, ma anche la sua autenticità semantica e relazionale.

2.
COS'È ARTIFICIALE (DA ANALOGICO A DIGITALE)

L'Intelligenza Artificiale (IA) rappresenta uno degli oggetti di studio più affascinanti e complessi del panorama scientifico e tecnologico contemporaneo. Studiare la sua storia e le sue evoluzioni non è solo un esercizio di curiosità storica, ma un passo fondamentale per comprendere il presente e anticipare il futuro. Fin dai suoi esordi, l'IA si è configurata come un campo di studio interdisciplinare che fonde informatica, matematica, filosofia e linguistica per affrontare la complessità del pensiero umano e la sua simulazione.

Una delle prime espressioni teoriche e pratiche dell'IA è stata l'Intelligenza Artificiale deterministica, un approccio basato sull'elaborazione predefinita di regole e algoritmi progettati per risolvere problemi specifici. Questo paradigma, particolarmente in voga nella seconda metà del Novecento, si distingue per l'enfasi sul controllo e sulla prevedibilità del comportamento del sistema. I sistemi esperti ne rappresentano un esempio emblematico: sviluppati a partire dagli anni Sessanta, essi simulano le decisioni di specialisti umani in domini circoscritti utilizzando regole deduttive formali. MYCIN, ad esempio, progettato per diagnosticare infezioni batteriche, operava attraverso un'ampia base di conoscenza codificata manualmente e di algoritmi di inferenza logica (Van Melle, 1978). Tuttavia, l'efficacia di tali sistemi dipendeva strettamente dall'accuratezza e dalla completezza delle regole fornite dai programmatori, evidenziando una dipendenza intrinseca dall'intervento umano e una limitata capacità di adattamento a contesti non previsti.

Questa rigidità è ciò che differenzia l'IA deterministica dai modelli generativi emersi negli ultimi anni. Laddove l'approccio deterministico segue un percorso rigidamente strutturato, le moderne IA generative, basate su reti

neurali profonde e apprendimento automatico, presentano un funzionamento probabilistico e dinamico. Questi modelli, come quelli alla base di GPT o DALL-E (Zhou & Nabus, 2023), apprendono da enormi quantità di dati per generare contenuti nuovi e adattarsi a input inaspettati. La capacità di apprendere e generare risposte complesse in modo autonomo sottolinea l'artificiosità intrinseca dell'IA deterministica, che non può evolvere al di fuori delle regole imposte a priori.

Un aspetto cruciale dello sviluppo dell'IA deterministica è il suo stretto rapporto con la linguistica. La linguistica computazionale, una disciplina che fonde linguistica teorica e informatica, ha fornito strumenti essenziali per la creazione di sistemi in grado di comprendere, elaborare e rispondere al linguaggio umano. Fin dai primi approcci, come l'algoritmo di *parsing* basato su alberi sintattici[2], il focus è stato sulla modellazione di regole grammaticali e strutture linguistiche che potessero essere tradotte in istruzioni computazionali. Questo legame è evidente nei chatbot pionieristici come ELIZA, sviluppato da Joseph Weizenbaum negli anni Sessanta, di cui presenterò un approfondimento in una sezione successiva dell'elaborato. Sebbene ELIZA utilizzasse un approccio rudimentale basato su semplici pattern di corrispondenza, rappresenta un primo tentativo di simulare una conversazione umana, dimostrando le potenzialità e i limiti dell'IA deterministica.

Tuttavia, l'interazione tra questo tipo di intelligenza artificiale e la linguistica non si limita alla sintassi. Anche la semantica e la pragmatica sono state esplorate, sebbene con risultati più limitati. I modelli deterministici, infatti, faticano a cogliere la complessità del significato contestuale e delle sfumature linguistiche. Un chatbot deterministico può riconoscere e rispondere a frasi secondo regole predefinite, ma non è in grado di comprendere il sottotesto o di adattarsi in modo fluido a conversazioni impreviste. Questo limite è particolarmente evidente nel confronto con l'IA generativa, che sfrutta

[2] L'algoritmo di *parsing* è una tecnica di analisi che traduce le regole grammaticali in una rappresentazione visiva ad albero, utile per l'elaborazione automatica del linguaggio. Grune, D., & Jacobs, C. J. (2008). Introduction to parsing. In *Parsing techniques: A practical guide*. NY: Springer New York. 61-102.

grandi modelli di linguaggio per cogliere correlazioni statistiche tra parole e concetti, offrendo risposte più naturali e adattive.

Dal punto di vista informatico, l'IA deterministica si colloca storicamente come una tappa fondamentale per la formalizzazione di algoritmi e strutture di dati. Il suo sviluppo ha contribuito a consolidare approcci computazionali sistematici, che hanno avuto un impatto duraturo su settori come l'automazione industriale, i giochi strategici (ad esempio il programma per il gioco degli scacchi di Claude Shannon) e i sistemi di gestione delle basi di dati. Tuttavia, la rigidità che contraddistingue questi sistemi li rende inadatti ad affrontare problemi complessi e ambigui, come quelli tipici del linguaggio naturale e del *Conversation Design*.

Il *Conversation Design*, ovvero la disciplina che progetta interazioni naturali tra esseri umani e sistemi digitali, si trova oggi in uno straordinario processo di avanzamento: oggi, distinguere un discorso generato da una macchina da uno pronunciato da una persona diventa sempre più difficile. La questione, oggi così centrale, affonda le sue radici nelle origini stesse dell'IA. Già in passato, infatti, questo tema venne sollevato come un criterio essenziale per definire l'intelligenza delle macchine. Nel 1950 Alan Turing propose un gioco imitativo come strumento per misurare la capacità di una macchina di emulare il linguaggio umano: da questa idea prese forma quello che in seguito fu definito Turing Test, ovvero il criterio secondo cui una macchina – se in grado di sostenere una conversazione umana credibile – può dirsi "intelligente".

Nel *Conversation Design*, l'Intelligenza Artificiale deterministica ha posto le basi per i primi sistemi di interazione uomo-macchina, ma la sua limitata capacità di comprendere e generare linguaggio in modo dinamico ne ha presto evidenziato i limiti. La transizione verso modelli generativi ha consentito di superare molte di queste barriere, introducendo sistemi più flessibili e scalabili. Tuttavia, l'approccio deterministico rimane un capitolo fondamentale nella storia dell'IA, non solo come precursore tecnologico, ma anche come monito sul rapporto tra progettazione umana e simulazione dell'intelligenza.

Questo rapporto intrecciato tra IA, linguistica e informatica continua a plasmare il futuro della progettazione di conversazioni tra umani e sistemi di Intelligenza Artificiale, suggerendo nuove direzioni per integrare le caratteristiche migliori di entrambi gli approcci. Da un lato, la prevedibilità e il controllo garantiti dall'IA deterministica possono ancora essere utili in contesti altamente regolamentati o in cui è essenziale garantire un elevato livello di affidabilità. Dall'altro, la creatività e l'adattabilità dei modelli generativi aprono prospettive entusiasmanti per la creazione di interazioni sempre più fluide e naturali.

Nel prosieguo del capitolo, approfondirò le principali tappe evolutive dell'Intelligenza Artificiale per comprendere come si sia giunti agli attuali livelli di interazione tra uomo e macchina, esplorando i progressi tecnologici che hanno trasformato il *Conversation Design* in una disciplina sempre più sofisticata.

2.1 Prima dell'Intelligenza Artificiale

L'evoluzione dell'Intelligenza Artificiale rappresenta un lungo iter che abbraccia secoli di speculazioni teoriche, sperimentazioni tecniche e sviluppi tecnologici. I primi tentativi di automazione e simulazione del pensiero umano non appartengono esclusivamente all'era moderna. Sin dall'antichità, filosofi e matematici hanno riflettuto sulla possibilità di rappresentare il ragionamento logico attraverso linguaggi simbolici. In particolare, il filosofo e matematico tedesco Gottfried Wilhelm Leibniz, nel XVII secolo, propose l'idea di una "lingua universale". L'idea di Leibniz era quella di rappresentare il pensiero umano in modo formale, attraverso simboli logici che potessero essere manipolati secondo regole precise (Leibniz, 1666). Sebbene questo fosse un progetto teorico, le idee di Leibniz avrebbero gettato le basi per lo sviluppo dei linguaggi formali e degli algoritmi computazionali.

Nel corso del XVIII e XIX secolo, con la Rivoluzione Industriale e lo sviluppo della meccanica, si iniziarono a concepire dispositivi meccanici in grado di automatizzare alcune funzioni cognitive umane. L'esempio più noto è quello della macchina analitica di Charles Babbage, un dispositivo progettato per eseguire operazioni aritmetiche complesse in modo automatico. Sebbene la macchina di Babbage non fu mai completata, essa ha rappresentato un passo significativo verso la realizzazione di dispositivi meccanici programmabili, in grado di svolgere operazioni definite.

Parallelamente, l'opera di Ada Lovelace, considerata da molti la prima programmatrice, si basò sul progetto di Babbage per sviluppare algoritmi che potessero essere eseguiti da una macchina (Natale, 2024). Lovelace fu tra le prime a concepire la possibilità che le macchine non si limitassero a eseguire calcoli numerici, ma potessero anche manipolare simboli e, in teoria, elaborare pensieri astratti (Aiello, 2016). Questo concetto avrebbe poi influenzato lo sviluppo dell'Intelligenza Artificiale, specialmente nelle sue applicazioni più recenti nel campo dell'elaborazione del linguaggio naturale.

Il vero punto di svolta nella storia dell'IA si ebbe però nel XX secolo, con l'avvento dei calcolatori elettronici. Il matematico britannico Alan Turing è considerato uno dei padri fondatori dell'informatica moderna e delle basi teoriche dell'Intelligenza Artificiale. Nel 1936, Turing sviluppò il concetto di macchina universale, un dispositivo teorico capace di simulare qualsiasi algoritmo computabile. Questa idea rivoluzionaria rappresentava il fondamento teorico su cui si sarebbe basata tutta l'informatica moderna: la possibilità di realizzare macchine programmabili in grado di risolvere una vasta gamma di problemi.

Oltre al concetto di macchina universale, Turing è noto per aver proposto un test mentale che divenne noto come Test di Turing. Durante questo esperimento egli si pose una domanda rivoluzionaria: può una macchina pensare? (Turing, 1980). Come verrà approfondito nelle prossime righe (Cfr. § 2.1.1) il test prevede la presenza di un interrogatore umano che comunica, tramite un'interfaccia, con due interlocutori, uno umano e uno macchina. Se l'interrogatore non riesce a distinguere quale dei due è la macchina, allora si

può considerare che quest'ultima possieda un'intelligenza simile a quella umana.

2.1.1 Il Test di Turing: approfondimento

La storia del *chatbot* ha radici profonde: risale agli anni Cinquanta, quando il brillante scienziato britannico Alan Turing iniziò ad esplorare il concetto di intelligenza artificiale. Nel suo articolo del 1980, *Computing Machinery and Intelligence*, Turing introdusse quello che oggi è noto come Test di Turing, un esperimento mentale volto a determinare se una macchina è in grado di imitare così bene il comportamento umano in una conversazione da non far distinguere all'interlocutore umano la differenza tra una persona reale e una macchina.

Il Test di Turing (Turing, 1980) rappresenta una delle pietre miliari della riflessione sull'Intelligenza Artificiale. Nato come un esperimento concettuale, ha stimolato un dibattito intenso sulle capacità delle macchine di emulare processi cognitivi umani e sulla definizione stessa di "intelligenza". Ma quali sono le implicazioni filosofiche e tecniche di questo test, e perché continua a essere rilevante anche a distanza di decenni dalla sua formulazione?

Il test si fonda su un approccio pragmatico alla questione dell'intelligenza. Turing evitava di definire l'intelligenza in modo stretto, preferendo una valutazione pratica basata sull'indistinguibilità delle risposte tra un umano e una macchina. Questo approccio fu considerato radicale per l'epoca, perché si distaccava dalle tradizionali visioni filosofiche dell'intelligenza, che cercavano di definirne la natura attraverso criteri oggettivi o esclusivamente umani.

Tuttavia, il Test di Turing è stato oggetto di numerose critiche. Una delle obiezioni più celebri è quella avanzata dal filosofo John Searle nel suo già menzionato esperimento della "stanza cinese" (Searle, 1980) in cui sostenne che il Test di Turing non fosse sufficiente a dimostrare che una mac-

china possedesse un'intelligenza reale. Secondo Searle, una macchina potrebbe rispondere in modo appropriato a un interrogatore umano seguendo semplicemente regole prestabilite, senza però "comprendere" realmente ciò che sta facendo. Questo dimostrerebbe, a suo avviso, che il Test di Turing valuta solo la capacità di simulare il comportamento intelligente, ma non l'intelligenza vera e propria.

Le obiezioni di Searle e di altri critici non hanno diminuito l'importanza del Test di Turing, che rimane un riferimento centrale per valutare i progressi dell'IA. Con l'evoluzione delle tecnologie, il test è stato messo alla prova più volte. Negli anni recenti, alcuni programmi, come il chatbot Eugene Goostman[3], progettato dai di programmatori Vladimir Veselov, Eugene Demchenko e Sergey Ulasen a San Pietroburgo nel 2001, hanno affermato di aver superato il Test di Turing (Schofield, 2014), ma le circostanze e le regole del test sono state spesso criticate per non essere state rispettate rigorosamente. Ad esempio, Eugene Goostman, che si fingeva un tredicenne ucraino, aveva l'obiettivo di sfruttare la mancanza di aspettative elevate nei confronti di un giovane non madrelingua inglese, il che potrebbe aver compromesso la valutazione critica degli interrogatori.

Nonostante queste limitazioni, il Test di Turing continua a essere un parametro di confronto importante, specialmente in relazione agli sviluppi odierni dell'Intelligenza Artificiale, che includono chatbot avanzati e sistemi di *Natural Language Processing* (NLP) (Joshi, 1991). Sistemi come GPT-3 e i suoi successori hanno dimostrato di poter rispondere a domande complesse e sostenere conversazioni coerenti con esseri umani (Elkins, 2020), sollevando nuove domande su cosa significhi realmente "pensare" per una macchina.

L'importanza del Test di Turing oggi non si limita solo alla sua applicazione tecnica, ma si estende anche alla filosofia dell'Intelligenza Artificiale. L'idea che l'intelligenza possa essere valutata in base al comportamento osservabile, senza necessariamente richiedere una comprensione interna del

[3] Schofield, J. (2014). "Computer chatbot 'Eugene Goostman' passes the Turing test". ZDNet. Disponibile su: https://www.zdnet.com/article/computer-chatbot-eugene-goostman-passes-the-turing-test/. URL consultato il 18 giugno 2025.

contesto o del significato, ha influenzato profondamente l'approccio dell'IA moderna. Questo approccio si riflette anche nelle tecnologie attuali di *machine learning* e *deep learning* (Shinde et al., 2018), dove la capacità di un sistema di apprendere dai dati e produrre risultati accurati è spesso considerata più importante della sua "comprensione" dei dati stessi.

Le teorie di Turing sulle macchine pensanti e la possibilità di simulare processi linguistici avrebbero ispirato, nei decenni successivi, i primi esperimenti di interazione uomo-macchina basati sul linguaggio naturale, come si vedrà più avanti (Cfr. § 3.5) dedicato al programma ELIZA di Weizenbaum (1966).

2.2 L'IA, breve storia e breve geografia

Nel presente paragrafo viene tracciato un excursus sulla nascita dell'Intelligenza Artificiale, esplorando i percorsi che hanno portato allo sviluppo di questa tecnologia innovativa. L'analisi si sofferma sulle principali tappe storiche, evidenziando le sfide affrontate nel corso del tempo e le ragioni alla base di alcuni periodi di rallentamento. Inoltre, viene osservato il momento in cui il mercato e le imprese hanno iniziato a mostrare un crescente interesse per l'IA, esaminando le motivazioni economiche, tecnologiche e strategiche che hanno reso questa disciplina un pilastro della trasformazione digitale.

L'Intelligenza Artificiale, come disciplina scientifica, ha una storia che comincia ufficialmente nel 1956 con la conferenza di Dartmouth, tenutasi presso il Dartmouth College di Hanover, nel New Hampshire. Questo evento segnò un punto di svolta, in quanto fu la prima volta in cui venne utilizzato il termine "Artificial Intelligence" (McCarthy et al., 2006). La conferenza, organizzata da John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon, radunò alcuni dei più brillanti scienziati e ingegneri del tempo, con

l'obiettivo di discutere e sviluppare macchine capaci di simulare processi cognitivi umani, come l'apprendimento, il ragionamento e la risoluzione di problemi.

Da quel momento, l'IA iniziò a svilupparsi rapidamente in varie direzioni. I primi decenni di ricerca furono caratterizzati da un grande entusiasmo, con la speranza che le macchine potessero presto eguagliare l'intelligenza umana. Un esempio emblematico di questo periodo è il programma *Logic Theorist* (Newell et al., 1956), sviluppato da Allen Newell e Herbert A. Simon, considerato uno dei primi esempi di IA in grado di dimostrare teoremi matematici. Questo programma fu capace di dimostrare 38 dei 52 teoremi presenti nei *Principia Mathematica* di Bertrand Russell e Alfred North Whitehead (McCorduck & Cfe, 2004).

Durante gli anni '60, si fece strada la convinzione che l'IA avrebbe potuto risolvere qualunque problema intellettuale. Così, molte università e aziende tecnologiche, tra cui IBM e AT&T Bell Labs, cominciarono a investire risorse significative nello sviluppo di sistemi intelligenti. Tuttavia, a fronte di queste aspettative elevate, emersero presto limiti tecnici significativi (Nilsson, 2009). Le capacità di calcolo dei computer dell'epoca erano insufficienti per gestire la complessità del mondo reale, e molte delle promesse fatte negli anni Sessanta non furono mantenute. Questo periodo, noto come il primo inverno dell'IA, durò fino alla fine degli anni Settanta, con una drastica riduzione dei fondi destinati alla ricerca (Negnevitsky, 2025).

La situazione iniziò a cambiare negli anni Ottanta, con l'avvento dei sistemi esperti, una tipologia di IA progettata per risolvere problemi specifici in domini ristretti (Waterman, 1986). I sistemi esperti utilizzavano una base di conoscenza composta da regole *if-then* e furono applicati con successo in settori come la medicina, dove il programma MYCIN fu in grado di diagnosticare infezioni batteriche (Van Melle, 1978). Il primo sistema esperto commerciale di successo fu R1, sviluppato da Digital Equipment Corporation per configurare gli ordini di nuovi sistemi di elaborazione in ambito produttivo (McDermott, 1982). L'efficacia di R1 fu dimostrata in breve tempo: entro il 1986, la compagnia dichiarò di aver risparmiato circa 40 milioni di dollari

grazie al suo utilizzo (Lewis, 2019). Questo successo spinse anche altre aziende a investire nei sistemi esperti, tanto che, alla fine degli anni Ottanta, quasi ogni grande azienda statunitense possedeva già un sistema esperto proprio (Durkin, 1996). Parallelamente, anche paesi come il Giappone e il Regno Unito iniziarono a investire significativamente nella ricerca sull'IA, ispirati dall'entusiasmo e dall'innovazione provenienti dagli Stati Uniti.

Questo approccio dimostrò che, sebbene fosse difficile replicare l'intelligenza generale umana, le macchine potevano eccellere in compiti specializzati.

Un'altra pietra miliare nella storia dell'IA fu posta dall'invenzione di nuovi algoritmi di apprendimento basati sulla retropropagazione dell'errore[4] (o *backpropagation*), sviluppati intorno al 1985 (Werbos, 2002). Questi algoritmi permisero di risolvere problemi di apprendimento in modo molto più efficace, soprattutto nel campo delle reti neurali artificiali, un concetto ispirato al funzionamento del cervello umano. Le reti neurali artificiali, in particolare, poterono essere applicate a una vasta gamma di problemi, non solo in informatica, ma anche in campi come la psicologia (Boden, 1989). Questi sistemi erano in grado di apprendere dai dati e migliorare progressivamente le proprie prestazioni, rendendo l'IA sempre più versatile e potente.

Un esempio di applicazione delle reti neurali fu studiato nel campo del gioco strategico, come gli scacchi e il gioco del Go, due giochi con un'enorme complessità di mosse possibili. Uno dei principali successi in questo ambito si ebbe nel 1997, quando Deep Blue, un sistema sviluppato da IBM, sconfisse il campione mondiale di scacchi Garry Kasparov (Newborn, 2003). Deep Blue rappresentava una dimostrazione pratica delle capacità dell'IA, poiché la macchina era in grado di analizzare milioni di combinazioni di mosse e contro-mosse, grazie alla sua enorme capacità computazionale. Nonostante Kasparov avesse inizialmente vinto alcuni incontri, i continui miglioramenti apportati a Deep Blue consentirono alla macchina di prevalere nelle partite

[4] Metodo che ha permesso alle reti neurali — modelli ispirati al funzionamento del cervello umano — di "imparare" dai propri errori in modo più efficace. Questa tecnica è alla base dei moderni sistemi di *deep learning*, inclusi quelli impiegati nel trattamento automatico del linguaggio naturale.

successive, dimostrando la superiorità dell'IA nell'elaborare strategie complesse (Hsu, 1999).

Il successo di Deep Blue fu attribuito alla sua capacità di elaborare un'enorme quantità di dati storici, analizzando le mosse dei più grandi campioni di scacchi e apprendendo le migliori strategie di apertura e attacco. Come sottolineato dallo stesso Kasparov nel suo libro *Deep Thinking*, "Gli scacchi sono la pietra di paragone dell'intelletto" (Kasparov, 2020), e Deep Blue, con la sua capacità di calcolo e di apprendimento, riuscì a incarnare questo concetto senza essere soggetto allo stress tipico dell'essere umano.

La vittoria di Deep Blue segnò una svolta significativa non solo nel campo dell'IA applicata al gioco, ma anche nel modo in cui le macchine potevano essere utilizzate per risolvere problemi complessi in contesti reali. Questo successo incentivò lo spostamento del baricentro della ricerca sull'IA verso l'apprendimento automatico e, più tardi, verso il *deep learning*. Tecnologie queste che, sviluppate a partire dagli anni Novanta e duemila, permisero di creare algoritmi capaci di apprendere dai dati, migliorando continuamente le prestazioni attraverso l'esperienza. Un esempio emblematico di questa evoluzione è AlphaGo (Silver et al., 2016), il programma sviluppato da Google DeepMind che nel 2016 sconfisse il campione mondiale di Go[5], un gioco di strategia considerato fino a quel momento troppo complesso per le macchine.

La geografia dell'IA si espanse rapidamente: dagli Stati Uniti, che restarono un centro nevralgico grazie a università come il MIT e Stanford e ad aziende come Google, Microsoft e Facebook, l'IA si diffuse in tutto il mondo. In Europa, paesi come il Regno Unito, la Germania e la Francia investirono massicciamente nella ricerca, mentre in Asia, nazioni come Cina, Giappone e Corea del Sud divennero leader nello sviluppo e nell'applicazione di tecnologie basate sull'Intelligenza Artificiale. La Cina, in particolare, ha adottato un piano strategico a lungo termine per diventare il leader globale nel settore entro il 2030, investendo miliardi di dollari in ricerca e sviluppo (Pan, 2016).

[5] BBC News. (2016). "Artificial intelligence: Google's AlphaGo beats Go master Lee Sedol". Disponibile su: https://www.bbc.com/news/technology-35785875. URL consultato il 26 settembre 2025.

2.3 Che cos'è l'IA generativa?

L'emergere delle Intelligenze Artificiali Generative (IAG) rappresenta uno dei più significativi avanzamenti tecnologici nel campo dell'Intelligenza Artificiale degli ultimi anni. Questi sistemi rivoluzionari si distinguono per la capacità di produrre contenuti originali e complessi attraverso meccanismi di apprendimento e generazione che vanno oltre la semplice replica di informazioni preesistenti. In questa sezione, che chiude il capitolo, analizzerò alcuni aspetti relativi all'architettura, ai processi di funzionamento, alle potenzialità applicative e alle implicazioni etiche e tecnologiche di questa innovativa tecnologia. L'indagine è articolata attraverso un'analisi delle fasi che conducono alla creazione di un'Intelligenza Artificiale generativa, esplorando i meccanismi di apprendimento automatico, i modelli di elaborazione del linguaggio naturale e le architetture dei modelli generativi. Saranno poi esposti i diversi tipi di contenuti che questi sistemi sono in grado di produrre – dal testo alle immagini, dal codice software alla musica – e le relative applicazioni nei diversi settori di ricerca e organizzativi. Infine, una sezione è dedicata anche alla valutazione dei benefici potenziali e delle criticità ancora aperte relativamente agli aspetti etici, tecnici e sociali connessi all'implementazione di tali tecnologie. L'obiettivo è fornire una prospettiva analitica che contribuisca alla comprensione di un fenomeno tecnologico in rapida evoluzione.

L'IA generativa, nota anche come "Gen AI", è una forma di Intelligenza Irtificiale progettata per creare contenuti originali, quali testi, immagini, video, audio o codici software, in risposta a richieste specifiche o input forniti dagli utenti (Cazzaniga, 2024).

Grazie alla quantità di dati e di informazioni raccolte dalla simulazione dei processi di apprendimento automatico (*deep learning*), l'IA generativa è in grado di comprendere le richieste in linguaggio naturale degli utenti e rispondere con contenuti nuovi e pertinenti.

Negli ultimi dieci anni, l'Intelligenza Artificiale è stata un tema centrale nel panorama tecnologico. Tuttavia, l'arrivo dell'IA generativa e il lancio di

strumenti come ChatGPT nel 2022 hanno accelerato il ritmo dell'innovazione, attirando un'attenzione globale senza precedenti (Fui-Hoon Nah et al., 2023). L'IA generativa offre notevoli benefici in termini di produttività sia per individui che per organizzazioni, anche se presenta importanti sfide e rischi. Secondo una ricerca della società di consulenza McKinsey, un terzo delle aziende utilizza già regolarmente l'IA generativa in almeno una funzione aziendale (Chui et al., 2017). Inoltre, Gartner, società di ricerca e consulenza nel mondo tech, prevede che entro il 2026 oltre l'80% delle organizzazioni avrà implementato applicazioni o API basate su questa tecnologia (Okrepilov, 2022).

2.3.1 Come funziona l'IA generativa

Il processo operativo dell'IA generativa può essere suddiviso in tre fasi principali.

Addestramento, cioè la costruzione di un modello di base che sia il fondamento per molteplici applicazioni. Ottimizzazione, che consiste in una modifica personalizzata del modello di base per poter svolgere compiti specifici. Generazione, valutazione e riottimizzazione: ovvero la produzione effettiva di nuovi contenuti e il continuo miglioramento della qualità e dell'accuratezza dei contenuti prodotti.

Di seguito presento un approfondimento di ciascuna fase.

2.3.1.1 Addestramento

L'addestramento rappresenta il primo passo per creare un modello di base, un tipo di modello di *deep learning* che costituisce la base per diverse applicazioni di IA generativa. I modelli di base più comuni sono i grandi modelli linguistici (LLM) utilizzati per generare testi, ma esistono anche modelli dedicati alla generazione di immagini, video, suoni e musica, oltre a modelli multimodali capaci di combinare diversi tipi di contenuti.

Durante l'addestramento, l'algoritmo viene esposto a enormi volumi di dati non strutturati e non etichettati, spesso raccolti da internet. Questo processo implica l'esecuzione di milioni di esercizi di completamento, in cui l'algoritmo tenta di prevedere l'elemento successivo in una sequenza, come una parola in una frase o un comando in un codice (Alpaydin, 2020). Il modello si adatta progressivamente per ridurre al minimo la discrepanza tra le sue previsioni e i dati reali. Alla fine di questo processo, si ottiene una rete neurale in grado di generare contenuti in modo autonomo in risposta agli input forniti.

L'addestramento dei modelli di base è un processo intensivo in termini di risorse computazionali, tempo e costi. Richiede settimane di elaborazione su migliaia di GPU e può comportare spese di milioni di dollari (Luccioni, 2024). Per ridurre questi costi, sono disponibili progetti open-source, come Llama-2 di Meta, che offrono modelli pre-addestrati.

2.3.1.2 Ottimizzazione

I modelli di base si presentano come strutture generaliste: contengono una grande quantità di informazioni ma non sono ancora in grado di generare output speficifi con la precisione e l'aderenza alle richieste. Per questa ragione, una volta completato l'addestramento, il modello di base necessita di ulteriori ottimizzazioni per soddisfare specifici requisiti applicativi. Di seguito presento alcune tra le principali tecniche utilizzate.

Fine tuning, che consiste nell'addestrare ulteriormente il modello con dati etichettati specifici per l'applicazione. Ad esempio, per creare un chatbot destinato al servizio clienti, si possono utilizzare documenti contenenti domande e risposte strutturate (Friederich, 2017).

Un'altra tecnica è quella che possiamo tradurre con "apprendimento per rinforzo tramite feedback umano" (*Reinforcement learning with human feedback*, o RLHF) che coinvolge utenti umani capaci di valutare i contenuti generati, fornendo feedback che il modello utilizza per migliorare la sua accuratezza (Gonzáles Barman, 2025).

2.3.1.3 Generazione e miglioramento continuo

La generazione dei contenuti avviene attraverso l'applicazione del modello ottimizzato, che produce output pertinenti in risposta agli input forniti. In questa fase, sviluppatori e utenti valutano regolarmente i risultati, perfezionando ulteriormente il modello. Un metodo innovativo per migliorare le prestazioni è la generazione aumentata dal recupero (RAG), che utilizza fonti di dati esterne per integrare e aggiornare le conoscenze del modello di base (Arslan et al., 2024).

2.3.2 Evoluzione delle architetture dei modelli di IA generativa

Nel corso degli ultimi dodici anni, le architetture dei modelli di IA generativa si sono evolute notevolmente, portando a innovazioni significative in molteplici settori. Di seguito, presento un elenco di quelle architetture considerate pietre miliari di questo processo evolutivo.

Gli Autoencoder Variazionali (*Variational autoencoders*, o VAE), introdotti nel 2013, hanno contribuito a progressi nel riconoscimento delle immagini, nell'elaborazione del linguaggio naturale e nel rilevamento di anomalie (Mazza et al., 2019).

Le Reti Generative Avversarie (*Generative adversarial networks*, o GAN), sviluppate nel 2014, hanno invece migliorato la qualità delle applicazioni esistenti e aperto la strada a soluzioni di Intelligenza Artificiale in grado di generare immagini foto-realistiche (Goodfellow, 2014).

I Modelli di Diffusione (*Diffusion models*), anch'essi introdotti nel 2014, hanno contribuito a fornire un controllo più dettagliato sugli output, in particolare nella generazione di immagini di alta qualità (Sohl-Dickstein, 2015).

I Transformers, descritti per la prima volta nel 2017, rappresentano oggi l'architettura principale dietro i più avanzati modelli di IA generativa, come GPT e BERT (Devlin, 2018).

2.3.2.1 Autoencoder Variazionali (VAE)

Gli *autoencoder* sono modelli di *deep learning* composti da due reti neurali connesse: una per comprimere i dati non strutturati e non etichettati in parametri, e l'altra per decodificare quei parametri e ricostruirne i contenuti originali. Sebbene tecnicamente possano generare nuovi contenuti, il loro utilizzo principale risiede nella compressione e nella decompressione dei dati piuttosto che nella generazione di contenuti di alta qualità (Kingma, 2013).

Gli Autoencoder Variazionali (VAE) rappresentano un'evoluzione degli *autoencoder*, e consentono la generazione di molteplici variazioni di un contenuto. Questi modelli sono stati inizialmente utilizzati in applicazioni come il rilevamento di anomalie (ad esempio, nella analisi delle immagini mediche) e nella generazione di linguaggio naturale.

2.3.2.2 Reti Generative Avversarie (GAN)

Le GAN, introdotte nel 2014, si compongono di due reti neurali: un generatore, che crea nuovi contenuti, e un discriminatore, che valuta l'accuratezza e la qualità di tali contenuti. Questo approccio "adversarial" è ciò che spinge il modello a migliorare continuamente la qualità dei risultati (Goodfellow, 2014).

Le GAN sono ampiamente utilizzate per la generazione di immagini e video realistici, oltre che per compiti come il trasferimento di stile (ad esempio, trasformare una foto in un disegno a matita) e l'accrescimento dei dati *(data augmentation*), ovvero la creazione di dati sintetici per ampliare e diversificare i dataset di addestramento (Brock, 2018).

2.3.2.3 Modelli di diffusione

Anche i modelli di diffusione, introdotti nel 2014, hanno trovato largo impiego soprattutto nella generazione di immagini di alta qualità. Il loro funzionamento si basa su un processo di aggiunta progressiva di rumore ai dati di addestramento fino a renderli irriconoscibili, per poi addestrare l'algoritmo

a rimuovere gradualmente il rumore e rivelare l'output desiderato (Sohl-Dickstein, 2015).

Sebbene il loro addestramento richieda più tempo rispetto ai VAE o alle GAN, i modelli di diffusione offrono un controllo più preciso sugli output. Un esempio emblematico è DALL-E, lo strumento di generazione di immagini di OpenAI (Ramesh, 2022).

2.3.2.4 Transformers

I Transformers, introdotti nel 2017, hanno rivoluzionato il paradigma *encoder-decoder*, permettendo un significativo avanzamento nella qualità e nella gamma dei contenuti prodotti dai modelli di IA. Alla base dei principali strumenti di IA generativa odierni, come ChatGPT, GPT-4 e Midjourney, i Transformers utilizzano un meccanismo chiamato *attention* per identificare e focalizzarsi sugli elementi più rilevanti di una sequenza di dati (Vaswani, 2017).

Questi modelli possono elaborare intere sequenze simultaneamente, catturare il contesto all'interno della sequenza ed esprimere le informazioni sotto forma di rappresentazioni codificate. I Transformers eccellono nell'elaborazione e nella comprensione del linguaggio naturale, e sono in grado di generare contenuti più lunghi e complessi come articoli, poesie e documenti, con una qualità superiore rispetto ad altri modelli di IA generativa (Devlin, 2018).

2.3.3 Tipologie di contenuti prodotti dall'IA generativa

L'IA generativa è in grado di creare una vasta gamma di contenuti in numerosi ambiti.

Per quanto riguarda i contenuti testuali, i modelli generativi, in particolare quelli basati sui Transformers, sono in grado di generare testi coerenti e contestualmente rilevanti, da istruzioni e documentazione a opuscoli, e-mail, testi per siti web, blog, articoli, relazioni e persino scrittura creativa.

Relativamente alla generazione di immagini e video, strumenti come DALL-E, Midjourney e Stable Diffusion sono in grado di creare immagini realistiche od originali e di eseguire il trasferimento di stile, la traduzione da immagine a immagine e altre operazioni di modifica o miglioramento delle immagini (Borji, 2022). Gli strumenti video AI di nuova generazione sono in grado di creare animazioni a partire da richieste di testo e di applicare effetti speciali a video esistenti in modo più rapido ed economico rispetto ad altri metodi (Singh, 2022).

Audio e musica. I modelli generativi possono sintetizzare discorsi e contenuti audio dal suono naturale per chatbot AI e assistenti digitali abilitati alla voce, narrazione di audiolibri e altre applicazioni (Akman, 2024). La stessa tecnologia può inoltre generare musica originale che imita la struttura e il suono di composizioni professionali.

L'IA generativa è in grado di generare codice originale (cioè generare da zero porzioni di codice sorgente funzionante), completare automaticamente gli snippet di codice (ad esempio, scrivendo le prime righe di una funzione in Python, il modello è in grado di suggerire il resto come un autocompletamento "intelligente"), tradurre tra linguaggi di programmazione (ad esempio, rendere in Python un programma in Java mantenendo la logica di base) e riassumere le funzionalità del codice (ovvero descrivere come funziona quel blocco di istruzioni). Consente inoltre agli sviluppatori di prototipare, rifattorizzare (cioè riorganizzare) e debuggare (cioè correggere errori nel codice) rapidamente le applicazioni, offrendo un'interfaccia in linguaggio naturale per le attività di coding (Ray, 2023).

I modelli di Intelligenza Artificiale generativa possono generare opere d'arte e di design uniche o assistere nella progettazione grafica. Le applicazioni includono la generazione dinamica di ambienti, personaggi o avatar ed effetti speciali per simulazioni virtuali e videogiochi (Jiang, 2023).

Un ulteriore utilizzo è legato alla generazione di dati sintetici o strutture sintetiche basate su dati reali o sintetici. Ad esempio, l'IA generativa viene applicata nella scoperta di farmaci per generare strutture molecolari con le

proprietà desiderate, aiutando la progettazione di nuovi composti farmaceutici (De Almeida, 2019).

2.3.4 *Prompt* efficaci: guida alla progettazione dell'interazione con l'IA generativa

In questa sezione verrà presentato uno degli approcci per costruire al meglio un prompt efficace: *A step-by-step guide to prompt design* (Khalid, 2024), una guida passo dopo passo alla progettazione dell'interazione con l'IA generativa, nata per sfruttare al meglio le capacità degli LLM attraverso la costruzione di prompt performanti (Cfr. § 5.3.1).

È possibile proporre un approccio metodico alla costruzione di prompt efficaci per l'IA generativa, articolato in cinque fasi fondamentali. Il primo passo consiste nell'identificare chiaramente il compito che si desidera affidare al modello, esplicitando con precisione la tipologia di risposta attesa. Questa fase richiede una riflessione attenta sullo scopo dell'interazione, che può variare dalla produzione di un testo descrittivo alla generazione di una sintesi, da un argomento creativo a una spiegazione tecnica. In assenza di un obiettivo definito, l'output rischia di risultare generico o fuori fuoco.

La seconda fase riguarda l'aggiunta di un contesto adeguato. Fornire al modello un quadro informativo chiaro e circostanziato è essenziale per ottenere risposte pertinenti e coerenti. Il contesto può includere il pubblico di riferimento, il dominio disciplinare, lo stile comunicativo, il registro linguistico e altri elementi che supportano l'IA nella comprensione dell'ambiente simulato della conversazione.

A queste due componenti si aggiunge, nella terza fase, l'uso degli esempi. L'inclusione di modelli di input-output guida il comportamento del sistema, soprattutto in scenari complessi. Questa tecnica, nota come *few-shot prompting*, consente di esemplificare il tipo di risposta attesa e contribuisce a ridurre le ambiguità insite nelle richieste aperte (Reynolds & McDonell, 2021).

La quarta fase della metodologia proposta riguarda la definizione di vincoli formali e stilistici. Specificare parametri come la lunghezza massima

della risposta, il tipo di linguaggio (tecnico, divulgativo, formale, colloquiale), il tono e l'eventuale struttura retorica desiderata, permette di affinare ulteriormente il *prompt*. Questi vincoli aiutano il modello a produrre un output più vicino alle esigenze comunicative dell'utente, migliorandone la qualità e l'usabilità.

Infine, un altro elemento da tenere in considerazione nell'atto di elaborazione di una richiesta è l'iterazione. La progettazione di un *prompt* non si conclude con un'unica formulazione: richiede un processo progressivo di affinamento, basato sull'analisi critica delle risposte ricevute. L'utente deve quindi testare il *prompt*, valutarne l'efficacia e modificarne la struttura fino a ottenere una risposta soddisfacente. L'approccio iterativo rende l'interazione con l'IA non solo più efficace, ma anche più consapevole e strategica.

A complemento di questo approccio procedurale possiamo proseguire l'analisi della struttura di un *prompt* efficace (Khalid, 2024) suddividendolo in tre componenti principali: input, contesto ed esempi. L'input è l'elemento che esplicita il compito da svolgere. Il contesto fornisce le coordinate semantiche, pragmatiche e stilistiche per orientare la generazione. Gli esempi, infine, fungono da modello comportamentale che il sistema può replicare. È la sinergia tra questi elementi che permette al *prompt* di trasformarsi da semplice input testuale in uno strumento progettuale sofisticato, capace di controllare l'output linguistico della macchina con maggiore precisione.

A conclusione della guida presentata da Irfan Khalid (2024) è dedicata una sezione alla sintesi operativa dei principi fin qui illustrati, mostrando come metterli in pratica attraverso la costruzione di un prompt completo ed efficace. Questa parte rappresenta l'applicazione concreta delle fasi precedenti e fornisce una visione integrata del *prompt design* come processo progettuale intenzionale e iterativo.

L'autore propone di partire da un'esigenza comunicativa ben definita, come ad esempio ottenere una spiegazione chiara e accessibile su un argomento scientifico destinata a un pubblico specifico. A partire da questa esigenza, si definisce un'istruzione diretta e chiara che costituisce il nucleo del

prompt. Ad essa si accompagna un contesto articolato, nel quale si specificano il livello del pubblico (es. studenti delle scuole superiori), il tono (es. divulgativo), lo scopo (es. favorire la comprensione) e l'ambito disciplinare (es. biologia).

Un prompt completo dovrebbe inoltre includere un esempio di risposta desiderata oppure un riferimento a un modello testuale che il sistema può imitare. Questa strategia consente di migliorare la precisione dell'output e riduce la variabilità indesiderata nella generazione. Il prompt può essere ulteriormente potenziato con l'aggiunta di vincoli stilistici o strutturali, come il numero massimo di parole, l'adozione di uno stile argomentativo o l'esclusione di determinati termini tecnici.

Questo esempio mostra non solo la correttezza formale, ma anche l'efficacia comunicativa del prompt, capace di guidare l'Intelligenza Artificiale verso una risposta pertinente, ben formattata e adatta al pubblico previsto.

La sezione rappresenta dunque un passaggio fondamentale dal piano teorico a quello operativo, sottolineando che la costruzione di un prompt efficace non è un'attività estemporanea, bensì un atto di progettazione che implica competenze linguistiche, consapevolezza comunicativa e sensibilità contestuale. La capacità di mettere insieme tutti gli elementi analizzati (istruzione, contesto, esempio, vincoli) in modo armonico e mirato segna il passaggio dall'interazione casuale con l'IA a una forma di collaborazione consapevole, orientata da obiettivi precisi e sostenuta da un metodo rigoroso (Khalid, 2024).

2.3.4.1 Tipologie di input, contesto ed esempi nella progettazione del prompt

Per rendere operativa la costruzione di un prompt completo, si propone una classificazione dettagliata delle componenti essenziali del prompt, articolate in tre macroaree: tipi di input, contesto e uso degli esempi (Khalid, 2024). Tra le tipologie di input più comuni utilizzate nei prompt troviamo:

- *Task Instruction*: semplice istruzione che specifica ciò che vogliamo far eseguire al modello (es. “Riassumi il seguente testo…”);
- *Question Answering*: domanda diretta rivolta al modello che lo interroga su una questione (es. “Qual è la causa principale del cambiamento climatico?”);
- *Scenario* + *Question*: descrizione di contesto seguita da domanda (es. “Agisci come un giornalista durante un’inchiesta a Gaza: qual è la prima domanda che faresti a un cittadino palestinese?”);
- *Example Input/Output* + *Task Instruction*: combinazione di input-output seguita da nuova istruzione per imitarne lo stile o la struttura;
- *Scenario* + *Task Instruction*: contesto narrativo accompagnato da una richiesta operativa (es. “In qualità di copywriter pubblicitario, scrivi un copy social per un brand…”)*;*
- *Entity* + *Task Instruction*: l’attribuzione di un ruolo all’IA più un compito (es. “Come esperto in diritto penale, spiega la differenza tra dolo e colpa.”).

Il contesto è una componente cruciale del *prompt design* e influisce in maniera significativa sulla pertinenza e sulla qualità dell’output. L’importanza del contesto risiede nella sua capacità di ancorare l’interazione a un quadro di riferimento specifico, fornendo all’IA informazioni implicite che altrimenti non emergerebbero dal solo input. Gli elementi chiave da includere nel contesto sono: il destinatario della risposta, il livello di formalità, l’obiettivo comunicativo, l’ambito disciplinare, il ruolo assegnato al modello e le condizioni d’uso.

Per fornire un contesto efficace, è consigliabile: esplicitare il pubblico previsto (es. “Spiega questo concetto a un bambino di sei anni…”), definire un tono e uno stile (es. formale, colloquiale, divulgativo), stabilire limiti o

obiettivi (es. “Scrivi un testo di massimo 200 parole.”), specificare l’ambiente comunicativo (es. una conversazione, una conferenza, un post online, etc.).

Gli esempi costituiscono una risorsa fondamentale per migliorare la precisione dell’output e ridurre le ambiguità interpretative. L’importanza degli esempi risiede nella loro funzione di guida: permettono al modello di osservare una struttura da replicare, uno stile da imitare o un contenuto da riformulare (Khalid, 2024). Le tipologie principali sono: *Few-shot examples*, cioè coppie di input-output che il modello può usare come riferimento; *Self-generated examples*, ovvero esempi generati autonomamente dall’IA su richiesta dell’utente.

Per incorporare efficacemente gli esempi è necessario: usare casi rilevanti e rappresentativi, mantenere coerenza tra esempio e compito target, inserire gli esempi prima dell’istruzione principale, limitare il numero di esempi per evitare confusione (Khalid, 2024).

Una progettazione accurata e consapevole di queste tre dimensioni (input, contesto ed esempi), consente di costruire prompt che non solo orientano l’output del modello in modo preciso, ma migliorano anche l’affidabilità, la trasparenza e l’utilità dell’interazione uomo-macchina.

2.3.5 Benefici dell’IA generativa

La produzione di contenuti nuovi e originali in numerosi campi di applicazione offre, alle organizzazioni e alla ricerca, altrettanti vantaggi. Di seguito espongo i principali benefici offerti dall’IA generativa.

Il vantaggio più evidente è legato all’aumento dell’efficienza. Grazie alla capacità di generare contenuti e risposte su richiesta, questa tecnologia ha il potenziale di accelerare o automatizzare attività faticose ad altà intesità di lavoro, ridurre i costi operativi e liberare tempo per consentire ai lavoratori di concentrarsi su attività a maggior valore aggiunto (Alhosani, 2024).

Oltre a ciò, l’IA generativa offre numerosi benefici sia per i singoli individui che per le organizzazioni.

2.3.5.1 Creatività potenziata

Gli strumenti di IA generativa possono stimolare la creatività attraverso processi automatizzati di brainstorming, generando molteplici versioni innovative di uno stesso contenuto. Queste varianti possono servire da punto di partenza o come riferimento, aiutando scrittori, artisti, designer e altri creativi a superare blocchi mentali o difficoltà di ispirazione.

2.3.5.2 Decisioni più rapide e informate

L'IA generativa eccelle nell'analizzare grandi dataset, identificare schemi ed estrarre intuizioni significative. Inoltre, è in grado di generare ipotesi e raccomandazioni basate su tali analisi, supportando dirigenti, analisti, ricercatori e altri professionisti nel prendere decisioni più intelligenti e basate sui dati.

2.3.5.3 Personalizzazione dinamica

In applicazioni come i sistemi di raccomandazione e la creazione di contenuti, l'IA generativa può analizzare le preferenze e la cronologia degli utenti per generare contenuti personalizzati in tempo reale, offrendo esperienze più coinvolgenti e su misura.

2.3.5.4 Disponibilità costante

A differenza delle risorse umane, l'IA generativa opera senza interruzioni e senza affaticamento, garantendo disponibilità continua per attività come chatbot di assistenza clienti e risposte automatizzate.

2.3.6 Casi d'uso dell'IA generativa

L'IA generativa ha già trovato applicazione in numerosi contesti aziendali e, con lo sviluppo continuo della tecnologia, le evidenze dimostrano che

non è solo possibile, ma probabile, aspettarsi una diffusione sempre maggiore.

I reparti marketing delle organizzazioni sono al lavoro per migliorare l'esperienza del cliente e risparmiare tempo incrementando la produzione di contenuti con l'utilizzo di strumenti di IA generativa per redigere testi destinati a blog, pagine web, materiali promozionali, e-mail e altro ancora. Inoltre, l'IA generativa può produrre contenuti di marketing altamente personalizzati in tempo reale, adattandoli al contesto, al pubblico e al momento della distribuzione (Chen, 2023). Questa tecnologia alimenta anche chatbot e agenti virtuali di nuova generazione, capaci non solo di fornire risposte personalizzate, ma anche di intraprendere azioni a nome del cliente, rappresentando un significativo progresso rispetto ai modelli conversazionali precedenti (Jo, 2023).

Attraverso la generazione di codice sorgente è possibile automatizzare e accelerare il processo di scrittura di nuovo codice. Una conseguenza positiva è l'elevata velocità con cui è possibile aggiornare e modernizzare programmi e applicazioni, automatizzando gran parte del lavoro ripetitivo (Ray, 2025).

L'IA generativa può redigere o revisionare rapidamente contratti, fatture, bolle e altra documentazione, sia digitale che fisica. Questo consente ai dipendenti di concentrarsi su compiti più importanti, accelerando i flussi di lavoro in aree aziendali come risorse umane, area legale, approvvigionamenti e contabilità (Beerbaum, 2023).

I modelli di IA generativa possono, inoltre, supportare scienziati e ingegneri nella proposta di soluzioni innovative a problemi complessi. Ad esempio, in ambito sanitario, i modelli generativi possono sintetizzare immagini mediche per addestrare e testare sistemi di imaging diagnostico (Kanjee, 2023).

2.3.7 Sfide, limitazioni e rischi dell'IA generativa

L'ultima sezione di questa ricerca sul funzionamento e sugli usi e dell'IA generativa è dedicata a presentare aspetti che risultano ancora problematici nello sviluppo di questa tecnologia. Nonostante i notevoli progressi, infatti, l'IA generativa presenta ancora importanti sfide e rischi per sviluppatori, utenti e pubblico in generale. Di seguito quelle che al momento sono considerate le criticità principali.

2.3.7.1 Risultati inaccurati: le "allucinazioni" dell'IA

Le "allucinazioni" dell'IA si verificano quando il modello genera contenuti privi di senso o completamente inesatti, ma apparentemente plausibili. Un esempio emblematico riguarda un caso legale in cui un avvocato ha utilizzato uno strumento di IA generativa per una ricerca, ottenendo casi giuridici completamente inventati ma presentati come reali (Khawaldeh, 2025). Alcuni esperti ritengono che le allucinazioni siano una conseguenza inevitabile del bilanciamento tra accuratezza e creatività del modello. Tuttavia, gli sviluppatori possono implementare misure preventive, come vincolare il modello a fonti di dati affidabili e procedere con valutazioni e perfezionamenti costanti per ridurre tali errori (Wang, 2025).

2.3.7.2 Output inconsistenti

A causa della natura probabilistica dei modelli generativi, lo stesso input può produrre risultati differenti, il che diventa problematico in applicazioni che richiedono uniformità, come i chatbot per il servizio clienti. Attraverso tecniche di *prompt engineering* (raffinamento iterativo degli input), gli utenti possono ottenere maggiore coerenza nei risultati (Wach, 2023).

2.3.7.3 Bias

I modelli generativi possono incorporare pregiudizi presenti nei dati di addestramento o nelle valutazioni umane, generando contenuti discriminatori,

iniqui o offensivi. Per mitigare tali rischi, è essenziale garantire una diversità nei dati di addestramento, stabilire linee guida rigorose durante il processo di ottimizzazione e monitorare continuamente i risultati del modello (Zhou, 2024).

2.3.7.4 Mancanza di spiegabilità

Molti modelli di IA generativa sono considerati "scatole nere", rendendo difficile comprendere o spiegare i processi decisionali che portano a specifici risultati. Le pratiche di *Explainable AI* (IA spiegabile) possono aiutare a migliorare la trasparenza e la fiducia nei confronti di questi strumenti. Inoltre, valutare la qualità dei contenuti generati rimane una sfida aperta, soprattutto per metriche che richiedono creatività, coerenza o rilevanza (Sun, 2022).

2.3.7.5 Minaccia a sicurezza, privacy e proprietà intellettuale

L'IA generativa può essere sfruttata per creare contenuti dannosi, come e-mail di phishing o identità false, compromettendo la sicurezza e la privacy. È fondamentale che sviluppatori e utenti adottino misure per evitare l'esposizione di informazioni sensibili o proprietà intellettuali, sia nei dati di addestramento che negli output generati (Golda, 2024).

2.3.7.6 Deepfake

I deepfake (immagini, video o audio generati artificialmente) rappresentano uno degli esempi più preoccupanti di utilizzo dannoso dell'IA generativa. Questi contenuti possono essere utilizzati per diffondere disinformazione, danneggiare reputazioni o facilitare attacchi informatici. La ricerca su modelli di IA in grado di rilevare i deepfake è in continuo sviluppo, mentre l'educazione degli utenti e l'adozione di buone pratiche possono contribuire a limitarne gli effetti negativi (Romero Moreno, 2024).

3.
LA CONVERSAZIONE ARTIFICIALE IN PASSATO

In questa sezione dell'elaborato, offro una panoramica sullo sviluppo degli assistenti vocali, partendo dai primi esperimenti tra modelli di conversazione ed elaborazione del linguaggio naturale. L'obiettivo è esaminare i cambiamenti chiave che hanno determinato salti di qualità nella tecnologia, oltre ad approfondire il ruolo delle *Voice User Interfaces* (VUI) nel migliorare l'interazione tra uomo e macchina (Stigall et al., 2019).

Nel contesto dell'evoluzione tecnologica contemporanea, il design delle conversazioni artificiali rappresenta un ambito di ricerca di fondamentale importanza, caratterizzato da una rapida e costante evoluzione (Luppicini, 2008). Come già accennato, questo settore ha registrato una significativa trasformazione, passando da sistemi basati su semplici regole predefinite a soluzioni sofisticate che integrano complesse tecnologie di elaborazione del linguaggio naturale e apprendimento automatico (Urban, 2019).

L'interazione uomo-macchina attraverso il linguaggio naturale ha seguito un percorso evolutivo che riflette non solo i progressi tecnologici, ma anche la crescente comprensione delle dinamiche comunicative umane. In questo contesto, il presente capitolo si propone di analizzare gli strumenti digitali che hanno plasmato lo sviluppo delle conversazioni artificiali, esaminando come questi abbiano progressivamente ridefinito i paradigmi dell'interazione tra esseri umani e sistemi computazionali.

La trattazione si concentrerà sull'analisi degli *Interactive Voice Response* (IVR), delle *Voice User Interface* (VUI) e dei *chatbot*, evidenziando come questi sistemi abbiano contribuito alla transizione: da semplici automazioni vocali a interazioni più sofisticate e contestuali. Particolare attenzione verrà dedicata all'evoluzione delle capacità di elaborazione del linguaggio

naturale e alla progressiva integrazione di tecnologie di intelligenza artificiale sempre più avanzate.

Il capitolo includerà un approfondimento specifico su tre *milestone* fondamentali nel campo delle conversazioni artificiali: ELIZA, il primo *chatbot* della storia che ha aperto nuove prospettive sulla simulazione del dialogo uomo-macchina; SIRI, che ha rivoluzionato il concetto di assistente vocale personale; infine, il Project Debater di IBM, che rappresenta un significativo avanzamento nella capacità dei sistemi artificiali di sostenere argomentazioni complesse e strutturate. Questi casi di studio illustrano non solo l'evoluzione tecnologica del settore, ma anche il progressivo raffinamento delle modalità di interazione tra esseri umani e sistemi artificiali.

Attraverso questa analisi, l'obiettivo sarà evidenziare come le VUI siano state integrate nei processi di comunicazione, ridefinendo il modo in cui gli utenti interagiscono con i sistemi digitali.

3.1 Da ELIZA a Siri e Alexa

La storia degli assistenti vocali basati sull'Intelligenza Artificiale comincia negli anni Sessanta, quando Joseph Weizenbaum, un informatico del MIT, creò ELIZA, uno dei primi programmi di elaborazione del linguaggio naturale (O'Regan, 2018). ELIZA era progettata per simulare una conversazione con uno psicoterapeuta, rispondendo alle domande degli utenti in modo che sembrasse comprendere i loro problemi. Sebbene il programma fosse molto limitato nelle sue capacità – in effetti, ELIZA non capiva realmente il significato delle parole ma si limitava a seguire schemi predefiniti – ha rappresentato un passo importante verso lo sviluppo di interfacce conversazionali tra uomo e macchina.

Negli anni successivi, lo sviluppo dei chatbot rimase limitato a causa della mancanza di risorse computazionali e della difficoltà di elaborare il linguaggio naturale in modo efficace. Fu solo con l'avvento dei primi personal

computer e l'inizio dell'era di internet negli anni Novanta che la ricerca sui chatbot riprese vigore. Programmi come ALICE (*Artificial Linguistic Internet Computer Entity*), creato da Richard Wallace nel 1995, dimostrarono che le interfacce conversazionali potevano diventare strumenti più sofisticati, capaci di sostenere conversazioni più complesse (Wallace, 2007).

Il vero salto di qualità avvenne nei primi anni duemila con l'introduzione di Siri, l'assistente vocale di Apple, lanciato nel 2011. Siri utilizza algoritmi di riconoscimento vocale avanzati basati sul *machine learning*, per comprendere e rispondere a domande poste dagli utenti. Non si limita a eseguire comandi predefiniti, ma è anche capace di rispondere a domande generali, cercando informazioni su internet e interagendo con le applicazioni del telefono. Questo modello di interazione vocale ha cambiato radicalmente il modo in cui le persone interagiscono con i dispositivi elettronici, aprendo la strada a quella che conosciamo come l'era degli assistenti vocali (Bellegarda, 2013).

Negli anni successivi, altri grandi attori tecnologici hanno seguito l'esempio di Apple. Amazon lanciò Alexa nel 2014, un assistente vocale progettato per funzionare attraverso dispositivi Echo e interagire con vari dispositivi domestici intelligenti (Lopatovska, 2019). A differenza di Siri, Alexa è progettata non solo per rispondere a domande, ma per diventare il centro di controllo di una casa intelligente, permettendo agli utenti di gestire, tramite comandi vocali, luci, termostati, sistemi di sicurezza e persino elettrodomestici. Il suo avvento segnò l'inizio della convergenza tra intelligenza artificiale e internet delle cose (IoT) (Jimenez, 2021).

Google Assistant, lanciato nel 2016, rappresentò un ulteriore avanzamento nella capacità degli assistenti vocali di comprendere il linguaggio naturale (Akinbi, 2020). Grazie all'enorme quantità di dati accumulati da Google attraverso i suoi servizi, Assistant divenne rapidamente uno degli assistenti vocali più avanzati, capace di gestire conversazioni complesse e interagire con una vasta gamma di applicazioni e servizi. L'uso del *deep learning* permise a Google Assistant di migliorare continuamente le sue prestazioni, di

apprendere dalle interazioni con gli utenti e di diventare sempre più preciso nel comprendere il contesto delle conversazioni.

Oggi, assistenti vocali come Siri, Alexa e Google Assistant sono parte integrante della vita quotidiana di milioni di persone. L'integrazione di queste tecnologie con sistemi di apprendimento automatico, *big data* e reti neurali ha permesso agli assistenti vocali di migliorare costantemente, offrendo risposte più personalizzate in una vasta gamma di contesti, dal controllo domestico alla ricerca di informazioni online, fino all'assistenza medica e al supporto al cliente (Tulshan, 2018).

Questi assistenti vocali, apparentemente semplici strumenti di interazione, sono in realtà il risultato di decenni di ricerca e sviluppo in vari campi dell'informatica, dall'elaborazione del linguaggio naturale fino alla comprensione contestuale e all'apprendimento automatico. Essi rappresentano uno dei più chiari esempi di come l'intelligenza artificiale stia trasformando radicalmente il modo in cui viviamo e interagiamo con la tecnologia.

3.2 Origini e ruolo pionieristico degli IVR

La concezione delle conversazioni artificiali ha conosciuto una trasformazione significativa, passando da interazioni basate su regole statiche a sistemi più sofisticati che integrano l'elaborazione del linguaggio naturale. Questa evoluzione è stata guidata dai progressi tecnologici e dall'esigenza di creare interfacce capaci di comprendere e rispondere al linguaggio umano in modo sempre più naturale e intuitivo. Un capitolo fondamentale di questa storia è rappresentato dagli *Interactive Voice Response* (IVR) (Corkrey, 2002), dalle *Voice User Interfaces* (VUI) (Stigall et al., 2019) e dai chatbot che, come più volte accennato, hanno segnato il passaggio dalla semplice automazione vocale a interazioni più dinamiche e contestuali. Esaminando lo sviluppo degli IVR, dei VUI e dei chatbot, possiamo comprendere meglio come la progettazione delle conversazioni artificiali abbia influenzato non

solo l'esperienza utente, ma anche la concezione stessa della comunicazione tra uomo e macchina.

Gli IVR, introdotti su larga scala negli anni Novanta, rappresentano una delle prime applicazioni pratiche delle conversazioni artificiali. Questi sistemi sono stati concepiti per automatizzare operazioni telefoniche, fornendo agli utenti un mezzo per interagire con i servizi aziendali senza la necessità di un operatore umano. Attraverso menu vocali predefiniti e comandi numerici, gli IVR permettevano di eseguire operazioni come prenotazioni di viaggi, verifica di saldi bancari o controllo dello stato di consegne (Corkrey, 2002).

Uno degli aspetti pionieristici degli IVR era la loro capacità di gestire compiti ripetitivi con efficienza e precisione. Ad esempio, sistemi come quello implementato da Charles Schwab negli anni Novanta, consentivano agli utenti di accedere a quotazioni di borsa in tempo reale. Questo tipo di automazione non solo riduceva il carico di lavoro per gli operatori, ma incentivava gli utenti a usufruire più frequentemente del servizio, eliminando la percezione di disturbo tipica delle interazioni umane (Pearl, 2016).

Tuttavia, gli IVR non erano privi di limitazioni. La rigidità dei menu e la necessità di seguire percorsi predefiniti spesso causavano frustrazione negli utenti, specialmente quando le loro richieste non rientravano nelle opzioni previste. Questa rigidità ha alimentato una percezione negativa dei sistemi IVR, immortalata in parodie come quelle del Saturday Night Live[6] e nell'esistenza di siti come GetHuman (gethuman.com), dedicati a bypassare i menu automatizzati per raggiungere direttamente un operatore umano (Pearl, 2016).

Nonostante i limiti iniziali, gli IVR rappresentano un passo fondamentale verso l'uso dell'intelligenza artificiale per migliorare l'interazione uomo-macchina. I progressi nella tecnologia di riconoscimento vocale, specialmente negli anni Novanta e duemila, hanno consentito ai sistemi IVR di evolversi, permettendo la comprensione di frasi più complesse e l'elaborazione di stringhe numeriche lunghe, come i codici di tracciamento delle spedizioni (Corkrey, 2002).

[6] Julie the Operator Lady - Saturday Night Live https://youtu.be/jSVRrJJ2nl4. URL consultato il 26 settembre 2025.

Questi miglioramenti erano resi possibili da modelli acustici più avanzati e dall'introduzione delle reti neurali nei processi di elaborazione vocale. Sebbene limitati rispetto alle moderne applicazioni basate sull'intelligenza artificiale, questi sistemi rappresentavano un tentativo iniziale di simulare una conversazione, anticipando molte delle funzionalità che oggi consideriamo standard nei VUI.

Un esempio emblematico dell'efficacia di questi software è rappresentato dal sistema IVR del servizio 511 della Baia di San Francisco, che consentiva ai conducenti di ottenere informazioni sul traffico, ritardi degli autobus e tempi di percorrenza. Questo tipo di servizio automatizzato era particolarmente utile in un'epoca in cui gli smartphone non erano ancora diffusi, e ha dimostrato come gli IVR potessero fornire valore aggiunto in contesti specifici (Pearl, 2016).

3.3 Dagli IVR alle VUI moderne: l'introduzione dei chatbot avanzati

Con il progredire della tecnologia, gli IVR hanno aperto la strada ai VUI, una categoria più avanzata di interfacce vocali che sfruttano l'intelligenza artificiale per offrire esperienze più fluide e naturali. I VUI moderni come Siri, Alexa e Google Assistant, integrano tecnologie avanzate di riconoscimento vocale (ASR, *Automatic Speech Recognition*) ed elaborazione del linguaggio naturale (NLP), che consentono di comprendere non solo le parole pronunciate, ma anche il contesto e le intenzioni dell'utente.

I chatbot rappresentano una tecnologia distinta che, secondo la definizione di Google, è progettata per simulare conversazioni con gli utenti, principalmente attraverso Internet. Sebbene possano incorporare funzionalità vocali, questi sistemi operano principalmente attraverso interfacce testuali. È importante notare che, nonostante la loro popolarità crescente, la maggior parte dei chatbot non ha registrato progressi significativi rispetto a ELIZA, il

programma pionieristico di elaborazione del linguaggio naturale degli anni Sessanta. L'efficacia di questi strumenti non è necessariamente superiore a quella di un'interfaccia grafica (GUI) tradizionale, tanto che molte soluzioni moderne integrano elementi GUI con interfacce testuali per ottimizzare l'esperienza utente. L'implementazione di un chatbot dovrebbe essere guidata da chiari benefici per l'utente finale e non semplicemente dall'attrattiva tecnologica o dalla volontà di automatizzare il supporto clienti (Pearl, 2016).

A differenza degli IVR tradizionali, i chatbot più avanzati utilizzano tecniche di apprendimento automatico per migliorare continuamente la qualità delle loro risposte, rendendoli capaci di adattarsi meglio alle esigenze degli utenti.

Un esempio significativo è il chatbot di Microsoft Xiaoice, che sfrutta enormi quantità di dati raccolti da conversazioni reali per generare risposte più naturali e contestualmente pertinenti. Allo stesso tempo, l'integrazione dei chatbot con piattaforme come Facebook Messenger o WhatsApp ha reso queste tecnologie estremamente accessibili, eliminando la necessità di applicazioni dedicate e consentendo interazioni immediate (Pearl, 2016).

La convergenza tra VUI e chatbot si manifesta chiaramente nei moderni assistenti vocali, che combinano funzionalità testuali e vocali per creare esperienze multimodali. Questo consente di superare molte delle limitazioni storiche degli IVR, offrendo agli utenti un modo più intuitivo e versatile di interagire con i sistemi tecnologici.

3.4 Il ruolo delle *Voice User Interfaces*

L'adozione massiccia di assistenti vocali come Siri, Alexa e Google Assistant ha portato alla creazione di una nuova categoria di interfacce: le *Voice User Interfaces*. I VUI permettono agli utenti di interagire con dispositivi elettronici attraverso la voce anziché tramite interfacce grafiche tradizionali (GUI), come schermi o tastiere. L'importanza dei VUI risiede nella loro

capacità di sfruttare una forma di comunicazione intuitiva e naturale per gli esseri umani: il linguaggio parlato (Pearl, 2016).

Le prime implementazioni di VUI risalgono, come già menzionato, agli anni Novanta, con i sistemi di risposta vocale interattiva (IVR) utilizzati per scopi specifici, ad esempio la gestione di ordini telefonici o il trasferimento di chiamate. Tuttavia, questi sistemi erano limitati e frustranti per gli utenti a causa della loro incapacità di gestire conversazioni fluide o contesti complessi. La seconda era dei VUI, inaugurata con l'introduzione di assistenti vocali avanzati come Alexa di Amazon e Google Assistant, ha cambiato radicalmente il panorama, offrendo interazioni molto più sofisticate e conversazioni più naturali.

3.4.1 Vantaggi delle VUI

I VUI presentano numerosi vantaggi rispetto alle interfacce tradizionali, contribuendo a rivoluzionare l'interazione uomo-macchina. Tra i principali benefici si trovano: velocità, intuitività, empatia e personalizzazione, compatibilità.

Velocità: parlare è generalmente più rapido rispetto alla digitazione. Uno studio ha dimostrato che la dettatura vocale è più veloce della scrittura, anche per gli utenti esperti di SMS (Ruan et al., 2017). In contesti in cui la velocità è fondamentale, come l'esecuzione di compiti multipli o la ricerca di informazioni, i VUI offrono un notevole vantaggio.

Uso a mani libere: i VUI consentono di interagire con i dispositivi senza dover utilizzare le mani, rendendoli particolarmente utili in situazioni come la guida, la cucina o in ambienti industriali. Questo vantaggio rende i VUI ideali per applicazioni dove la sicurezza e l'efficienza operativa sono prioritarie.

Intuitività: il linguaggio parlato è il mezzo di comunicazione naturale degli esseri umani, rendendo i VUI accessibili anche a utenti meno esperti di

tecnologia. L'interazione vocale permette agli utenti di evitare la curva di apprendimento tipica delle interfacce grafiche, offrendo un'esperienza utente più immediata e soddisfacente.

Empatia e personalizzazione: i VUI permettono di trasmettere non solo contenuti, ma anche emozioni, grazie all'uso della voce. L'intonazione, il volume e il ritmo della voce veicolano informazioni che il testo scritto non può trasmettere. Questa capacità di trasmettere empatia rende i VUI strumenti particolarmente efficaci in contesti emozionali, come l'assistenza sanitaria o il supporto psicologico, dove la comunicazione vocale può migliorare l'esperienza utente.

Compatibilità con dispositivi senza schermo: con l'aumento della popolarità di dispositivi senza schermo, come gli *smart speaker* (es. Amazon Echo, Google Home), i VUI sono diventati essenziali. Gli utenti possono così interagire con questi dispositivi semplicemente parlando, senza dover toccare o visualizzare alcuna interfaccia fisica, un aspetto che ha contribuito alla rapida diffusione degli *smart speaker* nelle case (Pearl, 2016).

3.4.2 Limiti delle VUI

Nonostante i numerosi vantaggi, i VUI non sono privi di limitazioni. Un problema significativo è la difficoltà di utilizzare questi sistemi in ambienti rumorosi o pubblici, dove la privacy e la chiarezza della voce possono essere compromesse. Molti utenti non si sentono a proprio agio nel parlare con un assistente vocale in spazi pubblici, dove potrebbero essere ascoltati da altre persone. Inoltre, i VUI possono risultare inefficaci quando le interazioni richiedono precisione e complessità, poiché possono fraintendere parole o contesti.

Un altro limite risiede nella capacità dei VUI di ricordare informazioni contestuali e conversazioni precedenti. Mentre gli assistenti vocali sono efficaci nel rispondere a comandi semplici, come "Accendi la luce" o "Qual è la mia prossima riunione?", essi faticano a mantenere il filo logico di conversazioni più complesse. Per esempio, se un utente chiede "Com'è il traffico

oggi?" e successivamente chiede "E il tempo?", l'assistente potrebbe non comprendere che la seconda domanda si riferisce alla stessa area geografica della prima, evidenziando una mancanza di consapevolezza del contesto (Pearl, 2016).

3.4.3 Evoluzione delle VUI: verso una maggiore "conversazionalità"

Affinché i VUI possano evolvere ulteriormente, devono sviluppare una capacità più avanzata di gestire il contesto e la memoria. Al momento, molti assistenti vocali sono in grado di gestire comandi semplici e isolati, ma faticano a mantenere una conversazione prolungata o a ricordare dettagli di interazioni passate. Un vero passo avanti verso un'interazione naturale sarà possibile quando i VUI saranno in grado di "ricordare" ciò che è stato detto nelle conversazioni precedenti e di adattare le risposte in base a quelle informazioni (Due, 2022).

Un esempio di questa mancanza di contesto si vede con le interazioni tipiche degli attuali assistenti vocali, come Alexa o Google Assistant. Questi sistemi, sebbene estremamente avanzati, si basano spesso su comandi a turni singoli in cui ogni richiesta è trattata come una nuova conversazione indipendente. Come abbiamo detto, per esempio, se un utente chiede "Com'è il traffico per il mio volo?" e successivamente "Quando atterra?", l'assistente potrebbe non collegare automaticamente le due richieste.

L'integrazione di una memoria conversazionale più profonda potrebbe migliorare enormemente l'esperienza utente, consentendo ai VUI di rispondere in modo più pertinente e naturale, mantenendo una coerenza con le informazioni raccolte precedentemente.

3.4.4 Impatto degli assistenti vocali nella vita quotidiana

Oggi, gli assistenti vocali come Siri, Alexa e Google Assistant sono parte integrante della vita quotidiana di milioni di persone. Gli utenti possono fare domande, impostare promemoria, controllare dispositivi smart e molto

altro, semplicemente parlando. Questa integrazione ha reso la tecnologia più accessibile e intuitiva, contribuendo a ridurre le barriere tra gli utenti e i dispositivi elettronici.

Tuttavia, l'integrazione dell'intelligenza artificiale conversazionale nelle abitudini quotidiane pone anche delle sfide. Gli sviluppatori devono affrontare problemi legati alla privacy, alla sicurezza dei dati e alla fiducia degli utenti. Molte persone sono riluttanti all'idea che i loro dispositivi ascoltino continuamente, temendo che le loro conversazioni private possano essere intercettate e i dati raccolti o utilizzati senza il loro consenso. Questo è particolarmente critico in contesti sensibili come la sanità, dove i dati personali potrebbero essere vulnerabili a intrusioni non autorizzate.

L'evoluzione degli assistenti virtuali, partendo dal loro antenato ELIZA fino a Siri, Alexa e oltre, rappresenta uno dei progressi più significativi nel campo dell'IA e delle interfacce uomo-macchina. I VUI oggi stanno giocando un ruolo centrale nel rendere la tecnologia più accessibile, intuitiva e personalizzata. Tuttavia, il loro pieno potenziale non è ancora stato raggiunto: affinché i VUI diventino veramente conversazionali e utili in una vasta gamma di contesti, è necessario superare le attuali limitazioni legate alla gestione del contesto e alla memoria.

Nel futuro, la previsione è che i VUI diventeranno sempre più sofisticati, capaci di comprendere le esigenze degli utenti in modo più profondo e di rispondere in modo più naturale e pertinente. Con il continuo avanzamento delle tecnologie di apprendimento automatico e di elaborazione del linguaggio naturale, i VUI potrebbero trasformarsi in veri e propri assistenti personali, capaci di anticipare i bisogni degli utenti e di fornire soluzioni in tempo reale. L'impatto di queste tecnologie sulla nostra vita quotidiana sarà sempre più significativo, fino a trasformare radicalmente il modo in cui interagiamo con il mondo digitale.

3.5 ELIZA di Wezeinbaum

In questa sezione del capitolo andrò a esaminare storia, caratteristiche e impatto nel mondo della comunicazione e della tecnologia di quello che possiamo considerare il progenitore dei *chatterbot*: ELIZA (O'Regan, 2018).

Molto prima che assistenti virtuali come Siri o Alexa entrassero a far parte della nostra vita quotidiana, esisteva ELIZA, un pioniere nel mondo dei *chatterbot*. A differenza degli assistenti moderni, progettati per eseguire azioni concrete come accendere le luci o riprodurre musica, ELIZA aveva uno scopo più astratto: simulare una conversazione con uno psicoterapeuta. Questo tentativo di imitare la comunicazione umana sfidò le prime concezioni dell'interazione uomo-macchina e gettò le basi per l'evoluzione futura dell'Intelligenza Artificiale.

Ispirandosi alle intuizioni di Turing sul linguaggio come criterio d'intelligenza (Cfr. § 2.1.1), l'informatico tedesco Joseph Weizenbaum, ricercatore presso il MIT (Massachussetts Institute of Technology), realizzò nel 1966 ELIZA, il primo programma in grado di simulare una conversazione con un essere umano. L'obiettivo di Weizenbaum era esplorare il potenziale delle macchine per imitare l'interazione umana, ma anche le implicazioni etiche e sociali di questa capacità.

3.5.1 ELIZA: come nasce e come funziona

ELIZA segnò la nascita del primo *chatterbot*. Lo scopo principale di ELIZA era esplorare l'elaborazione del linguaggio naturale e testare quanto un computer potesse interagire con gli esseri umani utilizzando il linguaggio quotidiano, oltre che dimostrare i limiti dell'interazione uomo-macchina e sottolineare come risposte automatiche potessero essere scambiate per risposte "umane" (Weizenbaum, 1966). Il nome di ELIZA fu scelto in riferimento a Eliza Doolittle, protagonista della commedia *Pigmalione* di George Bernard Shaw. Eliza è una fioraia che impara a parlare con un accento aristocratico per ingannare gli altri riguardo alla sua vera origine sociale. Allo stesso modo,

il programma di Weizenbaum era progettato per simulare una conversazione, ma senza comprendere realmente i contenuti del dialogo.

Nonostante l'innovatività percepita all'epoca, ELIZA non è in grado di comprendere il contenuto delle conversazioni. La generazione dei messaggi mostrati a video dalla macchina si basa su una tecnica di *pattern matching* (corrispondenza di modelli) e *substitution* (sostituzione) per produrre risposte. Quando un utente scrive un messaggio, il programma va alla ricerca di parole chiave. Basandosi su di esse, seleziona una risposta preprogrammata. Ad esempio, se un utente scrive: "mia madre cucina bene", ELIZA può riconoscere la parola "madre" e rispondere con una domanda generica come "parlami di più della tua famiglia". Questo semplice meccanismo ha creato l'illusione di empatia e comprensione profonda da parte della macchina.

Il linguaggio di programmazione utilizzato per ELIZA, chiamato MAD-SLIP (Lane et al., 2025), permette al programma di gestire liste di possibili risposte in modo efficiente e naturale. Anche se ELIZA non comprende veramente il significato delle parole, l'uso di regole predefinite le permette di simulare interazioni umane sorprendenti per gli utenti che per la prima volta interagiscono con una macchina in questo modo. Il sistema era molto semplice rispetto agli standard odierni, eppure, riusciva a imitare efficacemente una conversazione umana anche grazie alla strategia di "impostare" l'interazione come una seduta psicoterapeutica in cui l'utente è invitato a guidare lo scambio.

Una delle implementazioni più note di ELIZA fu l'inserimento di uno *script*, cioè di un insieme di regole su cui il sistema si basa per impostare l'interazione (Bassett, 2019). Lo *script* – denominato *Doctor* perché ispirato alla tecnica terapeutica del dottor Carl Rogers, psicoterapeuta noto per il suo approccio centrato sul paziente – permetteva a ELIZA di simulare una sessione di terapia, in cui rispondeva alle affermazioni degli utenti con domande aperte, incoraggiandoli a esplorare i propri sentimenti. Ad esempio, se un utente scriveva "mi sento ansioso", ELIZA poteva rispondere con "perché ti senti ansioso?", alimentando la conversazione in modo aperto e prolungato.

3.5.2 Impatto emotivo sugli utenti

Nonostante la consapevolezza che ELIZA fosse solo un programma, molti utenti svilupparono un forte legame emotivo con il bot. Questo fenomeno, sorprendente per i tempi che correvano, dimostrava il potere psicologico delle interazioni con la tecnologia, specialmente quando questa si presentava in modo empatico e non giudicante. Molte persone si sentirono a proprio agio nel confidarsi con ELIZA, proprio perché veniva simulato il comportamento di un terapeuta privo di pregiudizi e capace di ascoltare in modo neutrale.

Questo effetto aveva implicazioni significative per la psicologia umana e per la percezione delle tecnologie interattive. ELIZA ha dimostrato che è possibile instaurare un rapporto emotivo con una macchina, anche sapendo che essa non sia in grado di comprendere realmente. Questo fenomeno cominciò a sollevare interrogativi sull'etica dell'utilizzo di tali strumenti, soprattutto in contesti emotivi e terapeutici (Weizenbaum, 1966).

Nonostante il grande impatto di ELIZA, Weizenbaum fu profondamente critico verso il successo del suo programma. Riteneva che il successo di ELIZA rivelasse non tanto i progressi dell'Intelligenza Artificiale, quanto piuttosto la vulnerabilità umana che vuole vedere più di quanto sia realmente presente. Weizenbaum sottolineava che ELIZA non poteva né avrebbe mai potuto comprendere realmente le emozioni umane, vedendola come uno strumento limitato e non come un vero sostituto delle interazioni umane autentiche (Berry, 2023).

Weizenbaum arrivò a sostenere che, sebbene i chatbot potessero essere utili in contesti specifici come supporto o assistenza, non dovessero essere confusi con sistemi capaci di empatizzare o comprendere in profondità le persone e i loro sentimenti. La sua visione rifletteva una preoccupazione più ampia riguardo al crescente utilizzo dell'Intelligenza Artificiale nella società e ai suoi limiti intrinseci.

3.5.3 Le innovazioni conversazionali introdotte con ELIZA

Per comprendere la portata innovativa (e i limiti) del programma, è utile osservare alcuni scambi autentici tratti dal *Doctor script* di Weizenbaum (1966), l'implementazione più celebre di ELIZA progettata per simulare uno psicoterapeuta di orientamento rogersiano.

L'interazione seguente mostra come il sistema costruiva l'illusione di una comprensione a partire da semplici operazioni di *pattern-matching*:

User: Men are all alike.
ELIZA: In what way?
User: They're always bugging us about something or other.
ELIZA: Can you think of a specific example?
User: Well, my boyfriend made me come here.
ELIZA: Your boyfriend made you come here?
User: He says I'm depressed much of the time.
ELIZA: I'm sorry to hear you are depressed.
User: It's true. I'm unhappy.
ELIZA: Do you think coming here will help you not to be unhappy?[7]
(Wezeinbaum, 1966)

Questo estratto esemplifica la capacità del programma di mantenere la coerenza discorsiva attraverso la ripetizione parziale dell'enunciato dell'utente, che viene "riciclato" per generare l'impressione di ascolto empatico (Turkle, 1984).

Dal punto di vista griceano, ELIZA rispetta la massima di Relazione ("Sii pertinente"), ma viola sistematicamente quella di Quantità: fornisce risposte minimali, spesso tautologiche, che rimandano la responsabilità interpretativa all'utente. L'apparente aderenza al contesto e l'ambiguità strategica

[7] Trad. it.: "Utente: Gli uomini sono tutti uguali. ELIZA: In che senso? Utente: Ci tormentano sempre per una cosa o per un'altra. ELIZA: Puoi pensare a un esempio concreto? Utente: Beh, il mio ragazzo mi ha fatto venire qui. ELIZA: Il tuo ragazzo ti ha fatto venire qui? Utente: Dice che sono depressa per la maggior parte del tempo. ELIZA: Mi dispiace sentire che sei depressa. Utente: È vero. Sono infelice. ELIZA: Pensi che venire qui ti aiuterà a non essere infelice?".

funzionano come una forma primordiale di *flouting*, atta a stimolare la continuazione del dialogo: un tratto che anticipa, in modo ancora rudimentale, le tecniche di *active listening* dei chatbot contemporanei.

Un altro frammento evidenzia il carattere meccanico ma sorprendentemente efficace del sistema nel creare inferenze di tipo socio-emotivo:

> User: I feel unhappy today.
> ELIZA: Do you often feel unhappy?
> User: Yes, almost every day.
> ELIZA: Do you think coming here will help you not to feel unhappy?
> User: I need help.
> ELIZA: What would it mean to you if you got some help?[8]
> (Wezeinbaum, 1066)

Qui, l'effetto di "empatia simulata" nasce dal rispetto della massima di Modalità: il linguaggio è semplice, chiaro e ordinato, al punto da apparire spontaneo. In effetti, ELIZA non comprende il contenuto semantico delle frasi, ma riesce a modellare turni di parola coerenti in cui la forma sintattica e il ritmo della risposta evocano una vera interazione dialogica (Wardrip-Fruin, 2009).

Wezeinbaum stesso sottolinea che "the illusion of understanding is striking, though the program understands nothing" (1966: 37). Questo paradosso segna la nascita delle "non conversazioni", in cui la cooperazione conversazionale non deriva da intenzioni condivise ma dalla ripetizione formale di schemi linguistici appresi: un paradigma che resta, ancora oggi, il punto di partenza per la progettazione dei modelli conversazionali.

[8] Trad. it.: "Utente: Oggi mi sento infelice. ELIZA: Ti senti spesso infelice? Utente: Sì, quasi ogni giorno. ELIZA: Pensi che venire qui ti aiuterà a non sentirti infelice? Utente: Ho bisogno di aiuto. ELIZA: Che cosa significherebbe per te ricevere un aiuto?".

3.5.4 Un confronto tra il Test di Turing e i moderni chatbot

ELIZA è spesso considerata uno dei primi programmi a sfiorare il superamento del Test di Turing. In tempi recenti, un gruppo di ricercatori dell'Università della California, ha condotto uno studio per confrontare ELIZA con modelli di intelligenza artificiale più avanzati, tra cui GPT-3.5 e GPT-4 di OpenAI (Jones & Bergen, 2024). Lo studio mirava a osservare quale di questi modelli riuscisse a ingannare gli utenti, facendoli credere di parlare con un essere umano.

Sorprendentemente, ELIZA ottenne risultati notevoli. Pur essendo un programma molto semplice, le risposte basate su pattern e lo stile riflessivo hanno spesso ingannato gli utenti più di quanto facesse GPT-3.5, che usa un linguaggio molto più sofisticato. Con ogni probabilità, la natura formulaica del linguaggio umano avvicina ELIZA alla conversazione umana nonostante essa sia basata su un modello linguistico semplice. Questo esperimento, infatti, ha mostrato come anche semplici accorgimenti o tecniche conversazionali siano capaci di creare un'illusione di comprensione più efficace rispetto ai ben più complessi modelli di Intelligenza Artificiale moderni (Natale, 2021).

3.5.5 L'eredità di ELIZA e l'evoluzione dei *Chatterbot*

La creazione di ELIZA segnò un passaggio importante nella storia dell'Intelligenza Artificiale e dell'interazione uomo-computer. Il suo successo mostrò che i computer potevano essere programmati per simulare una conversazione umana, e questo ha aperto nuove strade per lo sviluppo di sistemi più avanzati.

Dopo ELIZA, furono creati altri *chatterbot* significativi legati al campo medico. Un esempione è il programma PARRY, progettato dallo psichiatra Kenneth Colby, che simulava un paziente affetto da schizofrenia paranoide e rappresentava un notevole passo avanti rispetto a ELIZA per la sua capacità di mantenere conversazioni coerenti (Colby et al., 1972).

Il termine stesso di “chatterbot” fu coniato solo nel 1994, da Michael Mauldin, creatore del bot Julia, a indicare programmi di conversazione interattiva (Mauldin, 1994). Oggi, questi *chatbot* sono evoluti in sistemi sofisticati come ChatGPT, che utilizzano reti neurali avanzate per simulare conversazioni sempre più complesse e naturali.

L’approfondimento su ELIZA ha aiutato a rilevare come, anche mediante l’uso di un linguaggio artificiale e una comprensione limitata, una macchina può simulare un’interazione umana abbastanza convincente da stabilire un legame emotivo con gli utenti. Sebbene oggi esistano *chatbot* molto più avanzati, la creazione di Weizenbaum rimane un fondamento storico dell’Intelligenza Artificiale, aprendo la strada a un futuro in cui le interazioni uomo-macchina continuano a evolversi.

3.6 Project Debater di IBM

Questa sezione del capitolo è dedicata a un approfondimento su Project Debater, un sistema di intelligenza artificiale sviluppato da IBM con l’obiettivo di partecipare a dibattiti complessi contro esseri umani su argomenti di interesse generale o sociale. Nei seguenti paragrafi ne ricostruirò la nascita e gli aspetti principali, oltre a presentare l’impatto che questo tipo di tecnologia ha avuto sugli utenti e le prospettive che apre nel campo del progresso degli assistenti vocali, in quanto il sistema è ormai ampiamente riconosciuto come pietra miliare nel panorama dell’intelligenza artificiale.

3.6.1 Project Debater: come nasce e come funziona

Project Debater nasce all’interno dell’IBM Research Lab (Potthast, 2019), una delle divisioni di ricerca avanzata dell’azienda tecnologica americana. Il progetto Debater rappresenta un significativo avanzamento nell’ambito dell’Intelligenza Artificiale, concepito e sviluppato per affrontare una

delle sfide più complesse dell'interazione uomo-macchina: la capacità di sostenere un dibattito articolato e convincente su temi complessi. Presentato pubblicamente nel 2019, questo sistema costituisce il culmine di oltre sei anni di ricerca e sviluppo da parte di un team interdisciplinare, composto da esperti in vari campi tra cui Intelligenza Artificiale, elaborazione del linguaggio naturale, apprendimento automatico e scienze cognitive.

Inserendosi nella tradizione delle grandi sfide affrontate da IBM, Project Debater segue i successi di sistemi come Deep Blue, che nel 1997 ha sconfitto il campione mondiale di scacchi Garry Kasparov, e Watson, che nel 2011 vinse il quiz televisivo "Jeopardy!" grazie alla sua abilità di comprendere e rispondere a domande in linguaggio naturale (Ferrucci, 2013). Tuttavia, Project Debater si distingue per il suo ambizioso obiettivo: non limitarsi a fornire risposte precise, ma costruire un discorso complesso, basato su dati ed evidenze, per sostenere o contestare una posizione in un dibattito.

Il sistema è progettato secondo un'architettura modulare, che ne consente un funzionamento altamente specializzato (Slonim et al., 2021). Ogni componente è dedicata a un compito specifico, indispensabile per la gestione di un dibattito. Prima di tutto, Project Debater è in grado di raccogliere informazioni rilevanti da un corpus di dati estremamente vasto e comprendente miliardi di frasi, per individuare argomenti pertinenti rispetto a un tema. Successivamente, valuta la qualità degli argomenti selezionati utilizzando modelli di apprendimento automatico avanzati, come BERT (Devlin, 2018), assicurandosi che siano coerenti, persuasivi e appropriati al contesto. Una volta selezionati gli argomenti, il sistema genera una narrazione strutturata, organizzando i punti in un discorso logico e convincente. Infine, durante il dibattito, Project Debater è in grado di rispondere alle obiezioni sollevate dall'avversario, costruendo controargomentazioni che si basano sui dati raccolti.

Tra le tecnologie chiave che caratterizzano Project Debater, vi è un algoritmo di wikificazione, capace di collegare i concetti identificati nei testi analizzati agli articoli corrispondenti su Wikipedia. Questo approccio non solo garantisce precisione e velocità nell'identificazione delle informazioni,

ma fornisce anche un contesto semantico ai concetti utilizzati. Inoltre, il sistema integra strumenti per il rilevamento e la valutazione degli argomenti, permettendo di distinguere tra affermazioni di supporto e di contestazione e attribuendo un punteggio alla loro qualità. Tra le sue funzionalità più avanzate spicca la capacità di *summarization* di contenuti complessi, riassumendo grandi volumi di dati in punti chiave che risultano comprensibili, pertinenti e rappresentativi (Bar-Haim, 2020).

L'efficacia di Project Debater è stata ampiamente dimostrata durante la sua presentazione pubblica nel febbraio 2019, in cui il sistema ha partecipato a un dibattito su temi come la tassazione delle bevande zuccherate. In quell'occasione, il sistema ha mostrato di saper individuare i punti principali della discussione, fornire esempi concreti e dati a sostegno delle proprie argomentazioni e rispondere efficacemente alle obiezioni sollevate dagli avversari umani.

Questa dimostrazione non solo ha evidenziato le potenzialità di Project Debater come strumento per arricchire il dibattito pubblico e supportare le decisioni complesse, ma ha anche aperto nuove prospettive sull'impiego dell'Intelligenza Artificiale nella costruzione di dialoghi argomentativi. Tale capacità rappresenta un passo avanti fondamentale verso lo sviluppo di sistemi in grado di partecipare a conversazioni articolate e di contribuire in maniera significativa alla comprensione e alla risoluzione di problemi sociali, educativi e politici.

3.6.2 Interazione con gli utenti

Un aspetto centrale di Project Debater è la sua capacità di sostenere interazioni complesse e dinamiche con gli utenti, ponendosi come un interlocutore capace di costruire argomentazioni convincenti e di rispondere alle obiezioni in modo mirato. A differenza degli assistenti vocali tradizionali, che si limitano per lo più a fornire risposte dirette a domande o a eseguire comandi predefiniti, Project Debater eccelle nella creazione di una conversazione argomentativa che simula un dibattito umano (Bar-Haim et al., 2021).

Questa capacità nasce da un sofisticato sistema di comprensione del linguaggio naturale (NLU), progettato per analizzare le frasi e riconoscere non solo i contenuti testuali, ma anche il contesto e la posizione dell'interlocutore rispetto a un tema. Ad esempio, il sistema può identificare se una dichiarazione sostiene o contesta un argomento specifico, un'operazione essenziale per strutturare una risposta pertinente durante un dibattito. Tale competenza è resa possibile da componenti chiave come il rilevamento delle affermazioni, la classificazione dello *stance* (pro o contro) e la valutazione della qualità degli argomenti, tecnologie che sfruttano modelli avanzati di apprendimento automatico come BERT (Devlin, 2018).

Durante un dibattito, Project Debater dimostra un notevole livello di reattività, generando risposte contestuali basate su dati raccolti. Se, ad esempio, l'interlocutore sostiene che "la tassazione delle bevande zuccherate è inefficace", il sistema può confutare questa affermazione citando evidenze empiriche o fornendo controargomentazioni elaborate sulla base delle sue risorse di conoscenza. Questa capacità di rispondere in modo coerente e strategico sottolinea la maturità del sistema nel gestire interazioni complesse e nel contribuire a discussioni costruttive.

Un ulteriore elemento distintivo di Project Debater è la sua abilità nella generazione narrativa (Bar-Haim et al, 2021). Il sistema organizza i dati raccolti in discorsi strutturati che seguono una logica sequenziale: introduzione, sviluppo e conclusione. Questo approccio non solo migliora la comprensione del messaggio da parte dell'utente, ma contribuisce anche a creare un'interazione più coinvolgente e naturale. L'attenzione al flusso narrativo è particolarmente evidente nelle dimostrazioni pubbliche del sistema, durante le quali è stato percepito come un interlocutore sofisticato e capace di arricchire il dibattito con approfondimenti ben costruiti.

Nonostante queste avanzate capacità, Project Debater presenta alcune limitazioni: manca, infatti, della componente emotiva e intuitiva che caratterizza le conversazioni umane. Pur mostrando notevoli capacità argomentative e logiche, il sistema non è in grado di interpretare o reagire alle emozioni dell'interlocutore, un aspetto cruciale per le interazioni realmente immersive

e personalizzate. Questo limite rappresenta una sfida significativa per il futuro sviluppo di tecnologie conversazionali che mirano a replicare non solo la logica, ma anche l'empatia delle comunicazioni umane.

3.6.3 Impatto emotivo sugli utenti

Durante le sue dimostrazioni pubbliche, il sistema ha dimostrato la capacità di sostenere un dibattito articolato e di coinvolgere emotivamente gli spettatori. L'accuratezza delle sue argomentazioni, la struttura logica dei discorsi e la capacità di ribattere in modo puntuale hanno portato molti a considerarlo un interlocutore quasi "umano". Questo ha generato un senso di stupore, mettendo in discussione l'idea tradizionale che l'argomentazione e il dibattito siano esclusivi dell'intelligenza umana. Tuttavia, questa percezione ha anche sollevato domande sulla natura del rapporto tra uomo e macchina, alimentando in alcuni casi una lieve inquietudine per le potenziali implicazioni di un sistema così avanzato (Bar-Haim et al., 2021).

Un elemento cruciale dell'impatto emotivo è dato dalla capacità di Project Debater di trasmettere autorevolezza. Nel momento in cui il sistema utilizza i dati, con evidenze empiriche e una narrazione ben strutturata, riesce a comunicare non solo competenza, ma anche una sorta di fiducia nel proprio operato. Questo effetto è amplificato dalla precisione e dall'apparente imparzialità delle sue risposte, che conferiscono un'aura di oggettività e credibilità. Tuttavia, proprio questa percezione può generare ambivalenza: da un lato, gli utenti apprezzano la qualità e la profondità delle argomentazioni offerte; dall'altro, emergono preoccupazioni sul rischio di affidarsi troppo a un sistema che, pur avanzato, rimane privo di una comprensione emotiva reale.

Nonostante la sua sofisticatezza, Project Debater non è in grado di percepire né influenzare attivamente le emozioni degli interlocutori. Le sue risposte, per quanto accurate e pertinenti, mancano di empatia e adattabilità emotiva. Questo limita la profondità delle connessioni che possono essere stabilite con gli utenti, sottolineando un aspetto ancora esclusivo dell'intera-

zione umana. Tale mancanza rappresenta un punto di partenza per future evoluzioni, che potrebbero integrare componenti capaci di riconoscere e rispondere alle emozioni, rendendo, in questo modo, l'interazione non solo argomentativa, ma anche emotivamente significativa (Bar-Haim et al, 2021).

3.6.4 La capacità argomentativa di Project Debater

Un elemento chiave per cogliere la portata innovativa di Project Debater è rappresentato dalla sua abilità argomentativa, ovvero dalla capacità di elaborare e sostenere, in modo coerente, posizioni opposte a quelle di un interlocutore umano.

Durante il dibattito tenutosi su *Should we subsidize preschool education?* nel 2019 (Bar-Haim et al., 2021), il sistema aprì il proprio intervento in questo modo – un passaggio oggi considerato emblematico della sua forza retorica basata sui dati:

> "For decades, research has demonstrated that high-quality preschool is one of the best investments of public dollars, resulting in children who fare better on tests and have more successful lives than those without the same access."[9] (Nature, 2021)

Questo esempio mette in luce la competenza logico-deduttiva del sistema: la sua argomentazione combina un riferimento empirico ("research has demonstrated") con una conclusione normativa ("best investments of public dollars"), articolando un ragionamento che rispetta le massime griceane di Quantità e Qualità — fornendo informazioni sufficienti e fondate su fonti verificabili (Slonim et al., 2021).

Nel successivo turno di replica, Project Debater mostrò una differente strategia discorsiva, mirata alla confutazione retorica dell'avversario:

[9] Trad. it.: "Per decenni, la ricerca ha dimostrato che un'istruzione prescolare di alta qualità è uno dei migliori investimenti dei fondi pubblici, perché produce bambini che ottengono risultati migliori nei test e conducono vite più di successo rispetto a chi non ne ha accesso.".

> "For starters, I sometimes listen to opponents and wonder: what do they want? Would they prefer poor people on their doorsteps begging for money? Would they live well with poor people, without heating and running water?" [10] (IBM Think, 2019)

Qui, emerge la dimensione dialettica e persuasiva: il sistema utilizza una domanda retorica per generare *pathos*, pur non provandolo, e orientare il consenso dell'ascoltatore. È una forma di simulazione argomentativa che dimostra la capacità di applicare strategie retoriche tipicamente umane (iperbole, implicatura ironica, etc.) pur restando ancorate a schemi logico-formali.

Infine, nella chiusura dello stesso dibattito, Project Debater condensò il proprio messaggio in un principio morale universale:

> "Giving opportunities to the less fortunate should be a moral obligation of any human being, and it is a key role for the state."[11] (Medium, 2019)

Questo passo mette in evidenza la sua capacità di astrazione etico-argomentativa: pur non comprendendo realmente il concetto di "obbligo morale", il sistema è in grado di riprodurne la struttura discorsiva, collegando il dominio dei dati empirici (in questo caso, politiche educative) con un registro valoriale. Si tratta di una forma embrionale di *reasoned ethos*, che nella retorica aristotelica si riferisce alla credibilità costruita attraverso la ragione, dove la persuasione deriva dalla coerenza logica più che dall'intenzione morale.

Gli esempi menzionati mostrano come Project Debater riesca ad avvicinarsi a una forma di cooperazione conversazionale "artificiale" nel senso griceano, non perché condivida intenzioni o empatia, bensì perché imita la struttura logica e retorica con cui gli esseri umani costruiscono senso nel dialogo.

---

[10] Trad. it.: "A volte ascolto gli avversari e mi chiedo: che cosa vogliono? Preferirebbero forse vedere i poveri sulla loro soglia a mendicare? Vivrebbero bene accanto a persone senza riscaldamento e acqua corrente?".

[11] Trad. it.: "Offrire opportunità ai meno fortunati dovrebbe essere un obbligo morale di ogni essere umano, ed è un ruolo fondamentale dello Stato.".

3.6.5 Project Debater: una sfida al Test di Turing

Project Debater, pur non essendo esplicitamente concepito per superare il Test di Turing, ne sfida indirettamente i principi, dimostrando di poter sostenere un dibattito argomentativo senza necessariamente apparire umano. Il sistema, grazie alla sua sofisticata architettura basata sull'elaborazione del linguaggio naturale e sulla costruzione di narrazioni coerenti, è in grado di partecipare a conversazioni complesse, strutturando risposte pertinenti e ragionate. Tuttavia, ciò avviene in modo dichiaratamente "macchinale," senza tentare di imitare il comportamento umano o mascherare la sua natura artificiale.

Questa peculiarità solleva una questione cruciale: l'efficacia di un sistema di IA nel dialogo dipende realmente dalla sua capacità di apparire umano? Project Debater dimostra che la persuasività e la rilevanza di un discorso possono derivare dalla qualità delle argomentazioni e dalla solidità delle evidenze presentate, piuttosto che dalla simulazione di tratti umani. Il sistema eccelle nel raccogliere informazioni, valutare la qualità degli argomenti e costruire discorsi articolati, ma rimane privo di una comprensione intuitiva ed emotiva che caratterizza le interazioni tra esseri umani (Bar-Haim et al., 2021).

Nonostante la sua sofisticazione tecnica, Project Debater rivela i limiti attuali dell'intelligenza artificiale nel replicare aspetti come l'empatia, la spontaneità e la comprensione profonda delle sfumature emotive. Sebbene il sistema sia in grado di rispondere in modo puntuale agli argomenti sollevati dagli avversari, manca di una vera consapevolezza del contesto emotivo e sociale che accompagna il dialogo umano. Questo limite sottolinea l'importanza di distinguere tra l'efficacia di un sistema nell'elaborazione di contenuti e la sua capacità di instaurare connessioni autentiche con gli interlocutori.

La sfida posta da Project Debater non è dunque quella di ingannare gli utenti facendogli credere di interagire con un essere umano, ma di ridefinire le aspettative rispetto al ruolo che un sistema di Intelligenza Artificiale può assumere in una conversazione. Questo approccio non solo apre nuove prospettive per l'uso dell'IA in contesti argomentativi, ma contribuisce anche a

una riflessione più ampia su ciò che significa “dialogare” in una società sempre più mediata dalla tecnologia.

3.6.6 Un contributo al futuro degli assistenti vocali: opportunità e limiti

Project Debater rappresenta una pietra miliare nel panorama dell’Intelligenza Artificiale, offrendo opportunità significative per il progresso degli assistenti vocali. La capacità di raccogliere, analizzare e sintetizzare argomentazioni complesse consente al sistema di contribuire in modo unico a contesti che richiedono decisioni basate su dati, come l’educazione, il dibattito pubblico e l’analisi delle opinioni. Ad esempio, il sistema può essere impiegato per analizzare grandi quantità di feedback, individuando le tematiche centrali e fornendo approfondimenti utili per il miglioramento delle politiche o dei servizi.

Tuttavia, le potenzialità di Project Debater sono bilanciate da limiti significativi. La sua dipendenza da dataset preesistenti e annotati con cura limita la sua capacità di gestire contesti imprevedibili o nuovi, mentre l’incapacità di comprendere ed elaborare le emozioni umane ne riduce l’efficacia in situazioni che richiedono empatia o sensibilità sociale. Inoltre, il rischio di utilizzi impropri rappresenta un lato oscuro non trascurabile: la capacità del sistema di presentare argomentazioni persuasive potrebbe essere sfruttata per manipolare opinioni o influenzare decisioni in modo poco etico.

Nonostante questi limiti, le tecnologie alla base di Project Debater delineano un futuro promettente per gli assistenti vocali. Integrando la capacità di argomentare e sintetizzare contenuti complessi, questi sistemi possono evolvere da semplici esecutori di comandi a interlocutori sofisticati, in grado di partecipare a conversazioni articolate. Tale evoluzione potrebbe ampliare le applicazioni degli assistenti vocali, rendendoli strumenti essenziali in settori come il supporto alle decisioni, l’educazione personalizzata e la comunicazione aziendale. La sfida sarà quella di bilanciare l’innovazione tecnologica con un uso responsabile ed etico, garantendo che tali sistemi continuino a servire gli interessi umani senza compromettere la fiducia e la trasparenza.

3.7 Siri di Apple

L'analisi di Siri, assistente vocale di Apple, rappresenta un caso emblematico nell'evoluzione dell'interazione uomo-macchina e nell'integrazione dell'Intelligenza Artificiale nella vita quotidiana. Non è solo uno dei primi assistenti vocali disponibili su larga scala, ma incarna anche le sfide e le opportunità che caratterizzano il rapporto tra esseri umani e tecnologia conversazionale. In questa sezione esaminerò come Siri, dalle sue origini come progetto di ricerca, fino alla sua attuale implementazione nell'ecosistema Apple, abbia plasmato le aspettative degli utenti verso gli assistenti virtuali. Particolare attenzione verrà dedicata alle modalità di interazione, all'impatto emotivo sugli utenti di diverse generazioni e ai limiti tecnologici ed etici che emergono dal suo utilizzo. L'analisi di Siri risulta particolarmente significativa poiché evidenzia il delicato equilibrio tra utilità pratica e coinvolgimento emotivo, tra innovazione tecnologica e tutela della privacy, riflettendo le più ampie questioni che caratterizzano lo sviluppo degli assistenti vocali contemporanei. Attraverso questo studio, emergerà come Siri rappresenti non solo uno strumento tecnologico, ma un vero e proprio paradigma dell'evoluzione del rapporto tra umani e Intelligenza Artificiale nel contesto quotidiano.

3.7.1 Siri: come nasce e come funziona

Siri è stata introdotta da Apple nel 2011 come parte integrante di iOS, rappresentando uno dei primi assistenti vocali disponibili su larga scala. Le sue origini risalgono a un progetto di ricerca finanziato dalla DARPA (*Defense Advanced Research Projects Agency*) e sviluppato da SRI International, un centro di ricerca californiano. Inizialmente concepita come un'applicazione indipendente per iPhone, Siri è stata acquisita da Apple nel 2010 per essere integrata direttamente nei propri dispositivi, con l'obiettivo di offrire un assistente personale virtuale capace di semplificare le attività quotidiane attraverso comandi vocali. Il nome "Siri" è stato scelto per la sua semplicità e memorabilità; in norvegese significa "bella vittoria", espressione che rivela

la visione di Apple di creare un sistema innovativo e accessibile (Bellegarda, 2013).

Siri si basa su un'architettura complessa che combina riconoscimento vocale, elaborazione del linguaggio naturale e Intelligenza Artificiale. Il processo inizia con il riconoscimento automatico del parlato (ASR), che converte la voce dell'utente in testo. Questo testo viene poi analizzato utilizzando modelli di elaborazione e comprensione del linguaggio naturale (NLP, NLU) per identificare l'intento dell'utente e determinare l'azione da eseguire. Una volta elaborata la richiesta, Siri restituisce una risposta vocale generata tramite tecnologie di sintesi vocale (TTS). Il sistema è fortemente integrato con i servizi cloud di Apple, che gestiscono gran parte dell'elaborazione dei dati, consentendo a Siri di accedere a una vasta gamma di informazioni e di aggiornarsi continuamente per migliorare la precisione delle risposte. Tuttavia, il suo funzionamento richiede una connessione internet stabile, poiché le richieste vocali vengono elaborate nei server remoti di Apple prima di essere restituite all'utente.

Siri utilizza il modello di attivazione *command-and-control*, che richiede un segnale esplicito per attivare il sistema, come il comando vocale o la pressione di un pulsante, garantendo chiarezza nell'interazione iniziale (Pearl, 2016). Siri si distingue per la capacità di comprendere comandi sequenziali, permettendo agli utenti di impartire istruzioni complesse senza la necessità di ripetere ogni dettaglio. Inoltre, con l'uso continuo, Siri si adatta alle preferenze linguistiche e di ricerca degli utenti, fornendo risultati personalizzati.

Dal suo lancio, Siri ha subito numerosi aggiornamenti per migliorare le sue capacità e adattarsi alle nuove tecnologie. Apple ha ampliato le sue funzionalità aggiungendo supporto per lingue diverse, integrazioni con app di terze parti e funzioni proattive che consentono a Siri di anticipare i bisogni dell'utente basandosi sulle abitudini e sul contesto.

In sintesi, Siri rappresenta un esempio significativo di come l'Intelligenza Artificiale possa essere integrata nei dispositivi di uso quotidiano, evolvendosi costantemente per offrire un'esperienza utente sempre più naturale e personalizzata.

3.7.2 Interazione con gli utenti

L'interazione con Siri si distingue per la sua immediatezza e semplicità, caratteristiche che hanno contribuito a renderlo uno degli assistenti vocali più diffusi a livello globale. L'attivazione avviene tramite il comando vocale "Hey Siri" o mediante un pulsante fisico sui dispositivi Apple. Questa modalità consente agli utenti di impartire comandi vocali senza la necessità di un'interazione manuale diretta, migliorando l'accessibilità in situazioni multitasking o durante la guida.

Siri offre una vasta gamma di funzionalità, tra cui impostare promemoria, inviare messaggi, controllare dispositivi smart tramite HomeKit e cercare informazioni online. Gli utenti apprezzano la capacità dell'assistente di eseguire operazioni sequenziali e contestuali, migliorando l'efficacia delle interazioni quotidiane. Tuttavia, la maggior parte delle interazioni richieste si limita a compiti di base, come riprodurre musica o rispondere a domande semplici. Questo riflette sia i limiti attuali delle capacità di Siri, sia le abitudini d'uso degli utenti, spesso poco inclini a esplorare funzionalità avanzate (Hasan et al., 2021). Inoltre, il sistema può risultare meno coinvolgente quando fornisce risposte troppo "spiritose" o poco informative, come nel caso di richieste che esulano dal contesto strettamente pratico, ad esempio: "Siri, dimmi il significato dell'amore", a cui risponde, "Chi, io?" (Pearl, 2016).

Uno dei punti di forza di Siri è la sua capacità di apprendere e adattarsi alle preferenze dell'utente. Analizzando i dati vocali e comportamentali, Siri personalizza le risposte e offre suggerimenti proattivi, come ricordare un appuntamento, inviare notifiche basate sul contesto o avvisare delle condizioni meteorologiche che potrebbero influire sulle attività pianificate. Questo processo di adattamento è reso possibile dall'analisi continua delle interazioni,

che consente al sistema di affinare le proprie capacità per soddisfare le esigenze specifiche degli utenti.

Per migliorare ulteriormente l'esperienza, Apple offre le funzioni *Migliora Siri* e *Dettatura*, che permettono agli utenti di contribuire volontariamente alla qualità del servizio. Attivando questa funzione, le registrazioni delle interazioni vocali vengono inviate ad Apple per essere analizzate da un team dedicato. Le analisi includono la revisione di porzioni anonime delle richieste vocali e delle risposte fornite da Siri, con l'obiettivo di migliorare la capacità del sistema di comprendere accenti, dialetti e richieste complesse. Apple garantisce che le informazioni raccolte sono trattate in modo anonimo e sicuro, separando le registrazioni dalla loro origine per tutelare la privacy degli utenti. Gli utenti possono disattivare questa opzione in qualsiasi momento, dimostrando l'impegno di Apple nel bilanciare innovazione e protezione dei dati personali. Nonostante questi progressi, permangono limiti legati alla capacità del sistema di comprendere contesti complessi o di rispondere a richieste ambigue.

L'assistente virtuale di casa Apple viene percepito principalmente come uno strumento utile piuttosto che come un vero interlocutore artificiale. Questo limita il coinvolgimento emotivo degli utenti, ma rafforza la fiducia nella sua affidabilità come assistente pratico. La percezione di rischio associata alla condivisione dei dati vocali, tuttavia, può influire negativamente sull'adozione di alcune funzionalità avanzate, indicando la necessità di una maggiore trasparenza nella gestione della privacy (Hasan et al., 2021).

Siri, inoltre, affronta limitazioni tecniche legate al riconoscimento del parlato laddove il contenuto non è gestito adeguatamente o quando manca una risposta predefinita, fattori che possono influenzare l'esperienza utente (Pearl, 2016).

3.7.3 Impatto emotivo sugli utenti

L'impatto emotivo di Siri sugli utenti varia notevolmente a seconda del contesto d'uso. Molti lo percepiscono come un "aiutante tecnologico", piuttosto che come un compagno al quale affiancare una componente emotiva (Zhang et al., 2014). Questa percezione è rafforzata dalla sua mancanza di empatia e dalla natura prevalentemente utilitaristica delle interazioni. Tuttavia, in situazioni di solitudine o isolamento sociale, Siri può assumere un ruolo più personale, fungendo da presenza tecnologica che facilita il dialogo e l'informazione.

L'adozione di Siri ha evidenziato differenze significative tra generazioni e contesti culturali. I millennial e i giovani adulti (18-24 anni), attratti dalla novità e dalla facilità d'uso del sistema, hanno appreso molto velocemente a utilizzare prodotti digitali come gli assistenti virtuali intelligenti. Eppure, è la fascia d'età compresa tra i 25 e i 49 anni quella che utilizza con maggior frequenza prodotti come Siri, grazie alla familiarità con la tecnologia e alla necessità di interagire con strumenti che semplifichino la gestione delle attività quotidiane (Bellegarda, 2013). In contesti culturali con una maggiore preoccupazione per la privacy, l'adozione di Siri è invece ostacolata dalla diffidenza verso la gestione dei dati personali (Hasan et al, 2021).

L'uso di Siri solleva questioni etiche e sociali legate al rapporto uomo-macchina: nonostante l'ottimizzazione di efficienza e accessibilità, la mancanza di una comprensione reale delle emozioni umane evidenzia i limiti dell'Intelligenza Artificiale attuale. Questo aspetto pone interrogativi su come i sistemi futuri potranno integrare componenti empatiche, migliorando tanto la funzionalità quanto il coinvolgimento emotivo.

3.7.4 La grammatica conversazionale di Siri

A differenza di ELIZA, che simulava il dialogo in modo rigido, Siri rappresenta il primo tentativo su larga scala di mixare risposte informative e ironiche, bilanciando cooperazione conversazionale e intrattenimento.

Dai test condotti da The Guardian, MIT Technology Review e Apple Press Kit, emergono i diversi aspetti della sua argomentazione conversazionale, anche se limitata: ironia programmata, gestione degli errori e imitazione della cooperazione griceana.

Sin dal lancio del 2011, l'assistente di Apple mostrò una peculiare "personalità discorsiva" costruita attraverso scelte linguistiche calibrate, brevi, umoristiche e apparentemente spontanee, che davano all'utente l'impressione di interagire con un soggetto dotato di intenzionalità.

Un esempio emblematico è uno dei primi dialoghi citati da Medium (2022) ed estratti da The Guardian (2011):

> User: Siri, what is the meaning of life?
> SIRI: Try to be nice to people, avoid eating fat, ready a good book now and then, get some walking in, and try to live together in peace and harmony with people of all creeds and nations.[12]
> (Medium, 2022)

La risposta di Siri riprende ironicamente la battuta del film *Monty Python's The Meaning of Live* (1983), ma nel contesto pragmatico si comporta come una violazione ironica della massima di Relazione, trasformata in *flouting* cooperativo: Siri "disobbedisce" alla richiesta informativa per costruire complicità e riconoscimento culturale.

L'assistente non risponde, ma comunica, applicando implicitamente la massima di Qualità ("Non dire ciò che non credi vero") attraverso una citazione che dichiara la sua natura non-umana.

Un altro scambio mostra, invece, come strategia metadialogica, la gestione dell'errore:

> User: Siri, why do you exist?

[12] Trad. it.: "Utente: Siri, qual è il senso della vita? SIRI: Cercare di essere gentili con le persone, evitare di mangiare grasso, leggere un buon libro di tanto in tanto, fare una passeggiata e provare a vivere in pace e armonia con persone di ogni credo e nazione.".

SIRI: I ask myself that question more often than you think.[13]
(Wired, 2011)

Qui, Siri attua un sofisticato gioco di *self-reference*: la risposta, apparentemente umoristica, esprime una forma di auto-implicatura metacomunicativa, nella quale il sistema riconosce la propria limitatezza linguistica. Si tratta di una pseudo-argomentazione esistenziale che umanizza la macchina attraverso una "sceneggiatura" di consapevolezza.

Infine, una terza interazione tratta da The Wall Street Journal (2011) evidenzia la capacità di adattare la cooperazione al contesto pragmatico-informativo.

User: Remind me to call mom when I leave work.
SIRI: Okay, I'll remind you to call your mom when you leave work. Don't forget flowers![14]
(The Wall Street Journal, 2011)

In questo caso, Siri rispetta pienamente le massime di Quantità ("Fornisci l'informazione richiesta") e di Relazione ("Sii pertinente"), ma aggiunge un *add-on* affettivo ("Don't forget the flowers!") che funge da implicatura sociale: un tocco empatico destinato a consolidare la percezione di un agente cooperativo e "premuroso".

Nel complesso, quindi, questi esempi mostrano come Siri abbia portato nel linguaggio della macchina una sofisticazione pragmatica ibrida: ironica (*flouting* controllato delle massime), metadialogica (riconoscimento della propria funzione) e affettiva (integrazione di atti direttivi e commissivi con tratti empatici).

[13] Trad. it.: "Utente: Siri, perché esisti? SIRI: Me lo chiedo più spesso di quanto tu possa immaginare.".
[14] Trad. it.: "Utente: Ricordami di chiamare la mamma quando esco dal lavoro. SIRI: Va bene, ti ricorderò di chiamare la mamma quando esci dal lavoro. Non dimenticare i fiori!".

L’assistente vocale diventa così una forma di “cooperazione teatrale”, dove la pertinenza e la comprensione sono simulate attraverso atti linguistici codificati, in linea con la definizione searleana di *illocutionary act* come azione regolata da norme costitutive piuttosto che da intenzioni mentali.

3.7.5 I lati oscuri di Siri: opportunità e limiti

Siri si colloca tra gli assistenti vocali più utilizzati grazie alla sua integrazione con l’ecosistema Apple. La possibilità di interagire vocalmente con dispositivi come iPhone, iPad, Apple Watch e HomeKit consente agli utenti di accedere a funzionalità avanzate senza necessità di input manuale. Tra i principali vantaggi, Siri permette di gestire dispositivi intelligenti, organizzare la giornata e rispondere a domande in modo immediato, rendendosi un valido strumento per aumentare l’efficienza nella vita quotidiana (Assefi et al., 2015).

Un aspetto particolarmente rilevante è la sua utilità rivolta a persone con disabilità, grazie alla possibilità di eseguire azioni complesse semplicemente attraverso comandi vocali. Supportando strumenti di accessibilità come i VoiceOver, che rendono i dispositivi Apple utilizzabili anche da persone con difficoltà visive o motorie, Siri sottolinea il valore inclusivo e sociale della tecnologia, che diventa uno strumento di emancipazione per categorie di utenti altrimenti escluse (Seymour et al., 2023).

In ambito aziendale, Siri offre opportunità interessanti per la gestione del lavoro. Ad esempio, i professionisti possono utilizzare l’assistente per programmare riunioni, inviare e-mail e cercare informazioni in modo rapido, migliorando la produttività e riducendo il tempo necessario per compiti ripetitivi. Inoltre, il supporto alle app di terze parti consente di ampliare le funzionalità disponibili, offrendo agli utenti la possibilità di personalizzare l’esperienza in base alle proprie esigenze.

Nonostante le opportunità offerte, Siri presenta numerosi limiti che ne riducono la competitività rispetto ad altri assistenti vocali, come Google Assistant o Alexa (Tulshan, 2018). Uno dei problemi principali riguarda la capacità di elaborare richieste più complesse o contestuali: Siri riesce a gestire bene i comandi sequenziali, ma ha difficoltà a comprendere le sfumature linguistiche o a interpretare domande composte da più istruzioni o che richiedono risposte argomentative.

Inoltre, come riportato sul sito ufficiale Apple, Siri si basa su un approccio centralizzato per la gestione dei dati vocali, che comporta l'invio delle richieste ai server di Apple per l'elaborazione. Questo sistema dipende fortemente dalla connessione internet e non è sempre in grado di rispondere in tempo reale in caso di problemi di rete. Al contrario, concorrenti come Google Assistant stanno integrando funzionalità di elaborazione locale, che migliorano la velocità e la privacy.

Le limitazioni linguistiche rappresentano un altro punto critico. Sebbene Siri supporti numerose lingue, il livello di accuratezza e fluidità varia notevolmente a seconda del contesto culturale e linguistico. Accenti, dialetti e modi di dire locali possono interferire con la capacità del sistema di comprendere correttamente i comandi, creando un'esperienza frustrante per alcuni utenti.

Le questioni relative alla privacy e alla sicurezza dei dati sono tra gli aspetti più controversi di Siri (Bolton et al., 2021). L'archiviazione delle registrazioni vocali da parte di Apple ha sollevato preoccupazioni tra gli utenti, in alcuni casi infatti è emerso che porzioni delle conversazioni degli utenti sono state ascoltate da revisori umani per migliorare l'accuratezza del sistema, generando un dibattito sull'etica di tali pratiche.

Per affrontare i limiti esistenti e mitigare i lati oscuri, Apple sta lavorando all'integrazione di tecnologie di Intelligenza Artificiale generativa. Questo approccio mira a migliorare la capacità di Siri di rispondere in modo contestuale e di sostenere conversazioni più naturali e articolate. L'adozione di modelli di linguaggio avanzati potrebbe rappresentare, infatti, una svolta significativa, allineando Siri ai progressi compiuti da altre piattaforme basate

su IA. Inoltre, Apple sta esplorando modalità per migliorare la gestione della privacy, come l'elaborazione locale delle richieste vocali, per garantire una maggiore trasparenza e sicurezza.

Siri offre numerose opportunità per migliorare la qualità della vita degli utenti, ma deve ancora affrontare sfide legate alla comprensione del linguaggio, alla privacy e all'adozione globale. Il futuro dell'assistente vocale dipenderà dalla capacità di Apple di integrare innovazioni tecnologiche e principi etici, assicurando un equilibrio tra progresso e responsabilità sociale.

4.
LA CONVERSAZIONE ARTIFICIALE OGGI

Il progresso dell'Intelligenza Artificiale ha portato a una trasformazione radicale nella progettazione delle interfacce conversazionali. Questo capitolo esamina il contributo della *Generative AI* alla conversazione artificiale, analizzando come il paradigma sia mutato rispetto agli approcci tradizionali. In particolare, affronterò la questione legata ai principi di progettazione delle interfacce vocali (Pearl, 2016), principi che hanno subito profondi cambiamenti grazie alle recenti innovazioni offerte dai modelli generativi avanzati, come quelli della famiglia GPT. L'obiettivo è comprendere il salto evolutivo che ha caratterizzato la progettazione conversazionale e delineare le implicazioni future di questi sviluppi.

L'introduzione della Gen AI ha cambiato il modo in cui vengono concepite le interazioni tra esseri umani e agenti artificiali. Se un tempo la conversazione artificiale era dominata da modelli rigidamente strutturati, oggi i modelli generativi hanno reso possibile una comunicazione più fluida e naturale. Questo progresso, tuttavia, non è privo di complessità e presenta nuove sfide, soprattutto in termini di affidabilità e gestione della coerenza informativa.

Kathy Pearl propone un approccio metodico e strutturato alla progettazione delle interfacce vocali, fondato su principi di usabilità, prevedibilità e controllo dell'interazione (Pearl, 2016). Un'interfaccia vocale efficace deve essere in grado di gestire conversazioni multi-turno e mantenere una memoria contestuale, consentendo agli utenti di interagire con maggiore naturalezza. Tuttavia, i VUI tradizionali presentano dei limiti, come per esempio che spesso faticano a comprendere riferimenti indiretti e a mantenere il contesto per interazioni prolungate (Myers et al., 2019).

Un aspetto chiave della progettazione delle VUI è la gestione degli errori e dell'ambiguità. Nei modelli tradizionali, la gestione degli errori si basa su strategie di recupero che spesso impongono agli utenti di ripetere o riformulare le proprie richieste. Questo approccio, sebbene efficace in contesti strutturati, risulta meno adatto a interazioni fluide e dinamiche. Inoltre, Pearl evidenzia l'importanza della cooperazione conversazionale griceana basata su massime, affinché le interfacce vocali possano fornire risposte pertinenti, chiare e adeguate al contesto senza risultare verbose o incomplete (Pearl, 2016).

Inoltre, l'autrice sottolinea la necessità di una progettazione basata su un attento bilanciamento tra precisione e accessibilità. Le interfacce vocali devono fornire informazioni accurate e comprensibili, evitando però di risultare complesse o difficili da usare per gli utenti meno esperti. Questo equilibrio, però, è difficile da conservare nei sistemi tradizionali, dove la rigidità della programmazione limita la capacità di adattamento alle necessità specifiche di ogni utente.

L'avvento della Gen AI ha modificato radicalmente il paradigma della progettazione conversazionale, introducendo sistemi in grado di elaborare risposte dinamiche, contestuali e adattive. A differenza dei VUI tradizionali, che si basano su una configurazione rigida di *intent detection* e *slot-filling*[15], i modelli generativi non richiedono una pre-programmazione dettagliata e possono rispondere a input formulati in modi diversi, aumentando la flessibilità dell'interazione (Weld et al., 2022; Brown et al., 2020).

Un aspetto fondamentale di questa evoluzione è la capacità dei modelli generativi di mantenere e processare il contesto della conversazione. Mentre i VUI classici tendono a perdere il riferimento alle informazioni fornite dall'utente dopo pochi turni di dialogo, i modelli basati su Generative AI conservano il contesto per periodi più lunghi, consentendo in questo modo inte-

[15] Nei sistemi di dialogo *task-oriented*, la *intent detection* identifica l'intenzione comunicativa dell'utente e lo *slot filling* estrae i parametri semantici necessari per completare l'azione richiesta. Weld, H., Huang, X., Long, S., Poon, J., & Han, S. C. (2022). A survey of joint intent detection and slot filling models in natural language understanding. *ACM Computing Surveys*, 55(8), 1-38.

razioni più fluide e coerenti (Radford et al., 2019). Questo avanzamento consente di superare i limiti della progettazione tradizionale, in cui l'utente è costretto a ripetere informazioni già note o a formulare le richieste in modo più preciso affinché siano comprese dal sistema.

L'apprendimento continuo è un'altra delle differenze chiave tra i due approcci. Nei modelli VUI tradizionali, il miglioramento delle prestazioni richiede un aggiornamento manuale del sistema, con revisione periodica dei dati e delle regole di interazione. Al contrario, i modelli generativi, essendo pre-addestrati su enormi quantità di dati testuali, possono adattarsi dinamicamente a nuove interazioni senza intervento umano diretto. Ciò li rende strumenti più versatili e capaci di gestire una varietà maggiore di richieste, anche se solleva questioni legate all'affidabilità delle risposte generate (Kirschthaler et al., 2020).

Tuttavia, l'elevata flessibilità dei modelli generativi introduce nuove sfide. Uno dei problemi principali è rappresentato dalle cosiddette "allucinazioni" dell'IA, ovvero la generazione di informazioni non accurate o prive di un fondamento nei dati reali. Nei VUI tradizionali, la prevedibilità delle risposte è garantita da regole esplicite e dataset controllati, mentre nei modelli generativi il rischio di informazioni errate è maggiore (Marcus & Davis, 2019). Inoltre, la maggiore capacità di elaborazione contestuale porta con sé interrogativi etici sulla gestione della privacy e sull'uso dei dati personali degli utenti (Golda et al., 2024).

Nel corso del capitolo andrò ad approfondire il confronto tra i due modi – tradizionale da un lato, generativo dall'altro – di strutturare la conversazione tra uomo e macchina. Uno spazio sarà dedicato al funzionamento del sistema GPT di casa OpenAI, che più di tutti ha contribuito al cambio di passo in corso negli ultimi anni.

4.1 Conversation design: dall'approccio *ex-ante* alla Generative AI

Nel periodo antecedente l'avvento della *Generative AI* la progettazione delle interfacce conversazionali si basava su un approccio *ex ante* che prevedeva la definizione manuale di *script* e *template* predefiniti. In questo modello, i conversation designer e i linguisti avevano il compito di estrapolare, analizzare e formalizzare i pattern comunicativi tipici delle interazioni umane, traducendoli in strutture rigide e prevedibili da implementare nei sistemi informatici (Jeon et al., 2023).

Come ormai sappiamo, Grice, in *Logic and Conversation* (1975), ha introdotto il principio cooperativo, sostenendo che gli interlocutori, nel corso di una conversazione, seguono implicitamente una serie di massime – quali Quantità, Qualità, Rilevanza (Relazione) e Maniera (Modo) –, per garantire una comunicazione efficace. Questi principi sono stati fondamentali per la digitalizzazione delle massime conversazionali, che, ispirandosi a tali teorie, hanno guidato l'elaborazione dei *template* esposti dai linguisti e applicati dai conversation designer.

Strumenti pionieristici come ELIZA e i più recenti Siri e Project Debater hanno rappresentato tappe fondamentali in questo percorso evolutivo. È importante ricordare che tali sistemi, sebbene limitati dal punto di vista della flessibilità, hanno segnato un'epoca in cui la simulazione della conversazione si affidava a regole esplicite, definite appunto *ex ante*, per garantire coerenza e prevedibilità nelle interazioni (Weizenbaum, 1966; Bar-Haim et al., 2021).

Il ruolo dei linguisti in questo contesto era cruciale: essi si occupavano di digitalizzare e codificare le massime conversazionali, basandosi proprio sui principi enunciati da Grice (1975), e di trasferirli in strutture programmatiche per l'implementazione nei sistemi di interazione uomo-macchina. I conversation designer, d'altra parte, utilizzavano tali *template* per creare interazioni che, seppur strutturate e in parte rigide, cercavano di replicare la naturalezza del dialogo umano attraverso strategie di *error handling* (gestione degli errori), conferme e gestione dei turni di parola.

Questo approccio, se da un lato garantiva una certa affidabilità e un controllo sull'output del sistema, dall'altro si rivelava limitato nella capacità di adattarsi alle sfumature e alle variabilità intrinseche del linguaggio naturale (Pearl, 2016).

Nonostante il modello di conversazione precostruito abbia rappresentato una tappa fondamentale nello sviluppo delle interfacce conversazionali, esso presenta limitazioni intrinseche che ne compromettono la capacità di rispondere alle complessità e alle dinamiche del linguaggio usato nella comunicazione fra individui. Prima di tutto, la rigidità degli *script* predefiniti comporta una forte dipendenza da regole esplicitamente codificate, le quali, pur garantendo coerenza e prevedibilità, non riescono ad adattarsi alle innumerevoli variabili che caratterizzano le interazioni umane (Grice, 1975).

Questa struttura statica impone ai conversation designer di prevedere, in via a priori, tutte le possibili varianti di una conversazione, rendendo il sistema vulnerabile a input imprevisti o ambigui. Quando un utente esprime una domanda o una richiesta in un modo che esula dal modello predefinito, il sistema tende a fornire risposte fuori contesto o a fallire nel recuperare informazioni rilevanti, evidenziando così l'incapacità di gestire la flessibilità linguistica (Weizenbaum, 1966).

Un'altra criticità rilevante riguarda il costo elevato in termini di manutenzione e aggiornamento: ogni qualvolta il linguaggio evolve o emergono nuovi pattern comunicativi, è necessario intervenire manualmente per modificare o espandere gli *script* esistenti. Tale approccio non solo richiede risorse considerevoli, ma spesso non riesce a mantenere il passo con l'evoluzione naturale della comunicazione, lasciando il sistema statico e obsoleto rispetto alle nuove esigenze degli utenti.

Infine, l'approccio *ex ante* limita la capacità di instaurare conversazioni fluide e naturali, in quanto il sistema, basandosi esclusivamente su *template* fissi, non riesce a sfruttare il contesto e la memoria delle interazioni precedenti in modo dinamico. Di conseguenza, l'esperienza utente risulta mecca-

nica e priva della ricchezza contestuale che caratterizza il dialogo umano, evidenziando come un modello basato su regole predefinite possa essere inadeguato in scenari complessi e variabili.

Queste criticità hanno posto le basi per il passaggio verso modelli più avanzati, capaci di apprendere autonomamente dai dati e di adattarsi dinamicamente alle interazioni, come avviene con l'evoluzione della Gen AI.

Il salto introdotto dalla *Generative AI* segna quindi una trasformazione radicale nel campo del conversation design, superando le limitazioni dei modelli *ex ante* basati su *script* e *template* predefiniti. I modelli generativi linguistici GPT hanno rivoluzionato questo scenario grazie alla loro capacità di estrarre automaticamente pattern e regolarità da enormi dataset, basandosi su sofisticate tecniche statistiche (Brown et al., 2020). Questa metodologia elimina la necessità di fornire al sistema un *template* predefinito, sia per interazioni conversazionali che argomentative, poiché il modello "impara" direttamente dai dati, evolvendosi in maniera autonoma e adattandosi dinamicamente a input variabili e imprevedibili.

Il confronto tra il tradizionale approccio *ex ante* e quello offerto dalla *Generative AI* evidenzia chiaramente il passaggio da un modello statico a uno dinamico: mentre in passato i conversation designer erano chiamati a definire anticipatamente ogni possibile scenario di interazione, un processo laborioso e costantemente soggetto a revisioni manuali, oggi il focus si sposta sulla supervisione e sulla curatela dell'output generato dal sistema (Pearl, 2016). In questo nuovo paradigma, il ruolo del designer evolve, convertendosi in un supervisore in grado di monitorare la coerenza e l'adeguatezza contestuale delle risposte, più che in un creatore di regole fisse.

Questa rivoluzione tecnologica, evidenziata anche negli studi di Floridi (2017; 2020), ha profonde implicazioni: la capacità dei modelli generativi di apprendere e adattarsi in tempo reale consente di instaurare interazioni più fluide e naturali, riducendo la necessità di interventi manuali per aggiornare e mantenere i sistemi conversazionali. In sostanza, il salto generato dalla *Generative AI* rappresenta il passaggio da una progettazione basata su definizioni pre-esistenti e rigide a un modello *data-driven*, in cui l'interazione si

configura come un processo in continua evoluzione, capace di rispondere in modo più efficace alle dinamiche del linguaggio umano (Floridi, 2020).

Le prospettive offerte da questo approccio sono evidenti: da un lato, l'autonomia dei modelli generativi permette di instaurare conversazioni più fluide, naturali e contestualmente rilevanti, superando i limiti imposti dagli *script* predeterminati. Dall'altro, tuttavia, questa flessibilità introduce criticità significative. Tra queste, il già menzionato rischio di "allucinazioni" dell'IA, ovvero la generazione di informazioni incoerenti o non fondate sui dati reali e la possibilità di ottenere risposte inaffidabili, le quali possono compromettere la qualità dell'interazione (Brown et al., 2020).

Con l'avvento di modelli generativi come GPT-3 si è assistito a un passaggio in cui il sistema apprende autonomamente dai dati per estrarre pattern linguistici e generare risposte fluenti e contestualmente rilevanti (Brown et al., 2020). Secondo Floridi e Chiriatti (2020), il modello GPT-3 opera su una base puramente statistica, il che gli consente di produrre testi che, sebbene sorprendentemente efficaci e naturali, non derivano da una comprensione semantica intrinseca. In altre parole, il sistema non "comprende" i contenuti come farebbe un essere umano, ma è in grado di replicare strutture comunicative in modo da soddisfare le aspettative formali di una conversazione.

L'impatto di GPT-3 si estende ben oltre la mera capacità di generare testi: esso ha rivoluzionato il modo in cui concepiamo la conversazione automatica. I limiti dei vecchi sistemi, che richiedevano una programmazione manuale e un aggiornamento continuo per far fronte alle evoluzioni del linguaggio, sono stati superati grazie alla flessibilità dei modelli generativi. Questi sistemi, infatti, si adattano dinamicamente al contesto, eliminando la necessità di interventi manuali per prevedere ogni possibile variazione nella comunicazione (Floridi, 2020). Il nuovo paradigma, dunque, consente un'interazione più fluida e naturale, in cui il ruolo del designer si trasforma: da creatore di *script* fissi diventa supervisore e curatore dell'output, incaricato di monitorare la coerenza e di intervenire in situazioni critiche.

Questo salto evolutivo comporta anche sfide e implicazioni etiche rilevanti. La capacità di GPT-3 di produrre contenuti in maniera autonoma ha

portato a un'industrializzazione della scrittura, che può tradursi sia in un'enorme crescita della produzione di artefatti semantici di alta qualità, sia in una diffusione massiccia di "spazzatura semantica" (Floridi & Chiriatti, 2020). Inoltre, la possibilità di generare testi con estrema velocità e a basso costo solleva questioni relative alla responsabilità nell'uso dei contenuti, alla trasparenza del processo e alla gestione dei bias intrinseci nei dati di addestramento.

Le domande che coinvolgono significati complessi e sfumature semantiche tendono a mettere in luce i limiti di un approccio rigido, in cui la risposta viene generata automaticamente secondo regole preconfezionate, senza che il sistema possa realmente "comprenderne" il contenuto. Questo problema emerge chiaramente negli esperimenti condotti su opere di celebri autori, come ad esempio per un romanzo di Jane Austen e un sonetto di Dante: a partire dalla sola prima frase di un racconto o dai primi versi di un componimento poetico, il modello genera un proseguimento che, pur risultando formalmente coerente e statisticamente plausibile, evidenzia la mancanza di una comprensione semantica profonda (Floridi & Chiriatti, 2020).

Una distinzione utile per comprendere meglio questi limiti è quella tra domande reversibili e irreversibili. Le domande irreversibili sono quelle per cui non è possibile dedurre la natura della fonte che ha prodotto la risposta: per esempio in quesiti matematici, fattuali o binari come "Quanto fa 2+2?", "Qual è la capitale della Francia?" o "Ti piace il gelato?", risolvibili tanto da un essere umano quanto da un'intelligenza artificiale, senza che ne sia immediata la distinzione. Al contrario, le domande reversibili, o semantiche, richiedono una comprensione più profonda del contesto e del significato e per questo possono rivelare se la risposta è stata prodotta da un essere umano oppure da un sistema automatico. Esempi di tali domande includono "Quanti piedi possono entrare in una scarpa?" o "Quali usi alternativi può avere una scarpa?", quesiti che implicano conoscenza del mondo, esperienza e interpretazione.

Nel caso dei modelli linguistici come GPT-3, questa distinzione si traduce in un'avanzata capacità di rispondere correttamente a domande irreversibili e in una maggiore difficoltà nel gestire domande reversibili, dove la mancanza di una reale comprensione semantica emerge in modo evidente. Questo aspetto sottolinea i limiti di un approccio puramente statistico nella generazione del linguaggio e la necessità di integrare le attuali metodologie con sistemi capaci di catturare sfumature di significato e contesto.

L'esempio dell'articolo pubblicato sul Guardian, redatto interamente da GPT-3, costituisce un ulteriore elemento significativo. Nonostante il testo sia stato successivamente editato da mani umane, il fatto che un modello di IA sia stato in grado di produrre un articolo di qualità quasi al pari di un testo umano ha suscitato notevole scalpore, dimostrando come, a partire da input minimi, il sistema sia capace di generare automaticamente una risposta che rispetta le convenzioni formali della scrittura (GPT-3, 2020).

Questi esempi mostrano che il modello *ex ante* rappresenta un solido punto di partenza per il conversation design, ma si scontra con la complessità e la variabilità del linguaggio naturale. In definitiva, sebbene la metodologia tradizionale abbia permesso di codificare in maniera rigorosa le interazioni, essa risulta ancora insufficiente per gestire situazioni in cui il contesto e le sfumature semantiche giocano un ruolo fondamentale. Tale riconoscimento sottolinea la necessità di integrare tali approcci con sistemi più flessibili, in grado di adattarsi dinamicamente alle interazioni e di garantire una produzione automatica continua di risposte, anche in presenza di input ridotti o ambigui, senza perdere la coerenza formale e la rilevanza contestuale.

Questi aspetti sollevano importanti questioni etiche e di controllo, poiché diventa essenziale definire confini chiari per garantire che il sistema operi entro limiti accettabili. Un'analogia efficace è quella del restauro di un vaso antico: così come il punto di intervento del restauratore è visibile e riconoscibile, anche nell'ambito delle interazioni generative occorre individuare in modo trasparente le eventuali correzioni o gli interventi necessari affinché si conservi l'integrità del dialogo (Floridi, 2020).

In sintesi: il salto generato dalla *Generative AI* rappresenta un cambiamento paradigmatico nel campo del conversation design. Mentre GPT-3 e modelli simili offrono una capacità di generare testi di elevata qualità, che supera di gran lunga i limiti dei sistemi basati su regole predeterminate, essi evidenziano altresì il divario tra la mera replicazione di pattern linguistici e una vera comprensione intellettuale. La visione di Floridi e Chiriatti (2020) ci invita a riconoscere che, sebbene l'efficienza e la flessibilità dei modelli generativi siano innegabili, esse si basano su meccanismi statistici. Tali meccanismi, per quanto efficaci, non equivalgono a una forma autentica di intelligenza. Riconoscere ciò è fondamentale per orientare lo sviluppo delle interfacce conversazionali, affinché esse integrino le potenzialità dei modelli generativi con sistemi di controllo e supervisione capaci di garantire affidabilità, trasparenza ed etica nell'interazione uomo-macchina.

4.2 Come funzionano i GPT

I sistemi GPT (*Generative Pre-trained Transformer*) rappresentano una delle innovazioni più rilevanti nel campo dell'intelligenza artificiale, specialmente in relazione alla generazione del linguaggio naturale. Peculiarità dei modelli GPT è la loro capacità di apprendere e produrre testi in modo fluido, coerente e, talvolta, quasi indistinguibile da quelli scritti da esseri umani, basandosi su una rete neurale di tipo *Transformer* (Vaswani, 2017), che svolge un ruolo centrale nel loro funzionamento integrando meccanismi avanzati come il *self-attention* (Humphreys, 2016), le reti neurali *feed-forward* (Bebis, 2002) e il *positional encoding* (Chu, 2021). Il *self-attention* consente al modello di valutare l'importanza relativa di ciascuna unità all'interno di una sequenza, identificando relazioni semantiche e sintattiche anche a lunga distanza (Humphreys, 2016). Le reti neurali *feed-forward*, invece, elaborano i dati attraverso trasformazioni non lineari, migliorando la capacità del sistema

di riconoscere e riprodurre schemi linguistici complessi (Bebis, 2002). Il *positional encoding* permette di preservare l'ordine delle parole, operazione fondamentale affinché il modello possa interpretare correttamente la struttura del linguaggio umano (Chu, 2021). In particolare, questi modelli apprendono da enormi quantità di dati testuali, capaci di esporli a una varietà di situazioni linguistiche e semantiche senza bisogno di intervento manuale e aggiornamenti costanti, come avveniva nei sistemi tradizionali di progettazione *ex ante*. Il processo di allenamento di un sistema GPT si basa su due fasi principali: l'addestramento preliminare e il *fine-tuning* (Friedrich, 2017). L'addestramento preliminare avviene su un vasto corpus di dati, che include libri, articoli, risorse online e altre forme di testo. In questa fase, il modello viene esposto a milioni di frasi e testi, dai quali impara a predire la parola successiva data una sequenza precedente. Attraverso tecniche di *Masked Language Modeling* (Salazar, 2019) e la *Next Sentence Prediction* (Shi, 2019) si costruisce una comprensione statistica del linguaggio, dove il modello apprende le probabilità di apparizione delle parole e delle frasi e le modalità con cui esse si relazionano tra loro. La prima tecnica consiste nel nascondere parole all'interno di un testo per addestrare il modello a prevederle (Salazar, 2019), mentre la seconda aiuta il sistema a comprendere le relazioni contestuali tra frasi consecutive (Shi, 2019). Questo processo consente al modello di sviluppare una conoscenza generale delle strutture linguistiche e delle relazioni semantiche senza una specifica supervisione umana. Durante il *training*, vengono poi implementati algoritmi di ottimizzazione, come la discesa del gradiente (Ruder, 2016), in modo che ogni previsione del modello venga confrontata con la realtà e che l'errore venga minimizzato man mano che l'addestramento procede. Tuttavia, affinché il modello possa essere adattato a compiti specifici, è necessario intervenire con il *fine-tuning* – che avviene attraverso dataset supervisionati e mirati a migliorare le prestazioni in contesti applicativi concreti come la traduzione automatica –, il *question answering* e la gestione di dialoghi interattivi (Friedrich, 2017). Questa fase consente di affinare la capacità del modello di generare risposte precise e pertinenti, adattandosi a set-

tori specifici grazie al *transfer learning* (Torrey, 2010), che permette di trasferire la conoscenza acquisita durante il pre-allenamento su nuovi compiti con un numero ridotto di esempi. La seconda fase, il *fine-tuning*, avviene successivamente all'addestramento preliminare. Durante questa fase, il modello viene specificamente adattato per eseguire compiti più concreti, come rispondere a domande, completare testi o generare articoli su temi specifici. Gli errori vengono ridotti su compiti alternati, rendendo il modello più funzionale e capace di consentire applicazioni pratiche nel mondo reale (Friedrich, 2017).

Durante la fase di inferenza, il modello utilizza strategie di decodifica avanzate per generare testi fluidi e coerenti. Una delle tecniche più comuni è il *beam search* (Freitag, 2017), che esplora simultaneamente diverse possibilità di completamento della frase e seleziona quella più probabile, garantendo risposte coerenti con il contesto. Un'altra strategia è il *top-k sampling* (Li et al., 2020) che introduce un grado di casualità nella selezione dell'unità successiva, permettendo al sistema di generare output più vari e creativi. La tecnica del *temperature scaling* (Chamola et al., 2023), invece, regola il livello di casualità delle risposte: valori bassi producono output più prevedibili e controllati, mentre valori più alti favoriscono la creatività, pur con un maggiore rischio di incoerenza. L'uso combinato di queste strategie permette di ottenere un equilibrio tra prevedibilità e diversità nella generazione del testo, migliorando la qualità complessiva delle risposte e riducendo il rischio di ripetizioni (Hassija et al., 2023).

L'integrazione dell'architettura *Transformer* con un processo di addestramento articolato su più livelli ha permesso a questi modelli di apprendere dai dati con una precisione e una coerenza sempre maggiori, superando molte delle limitazioni dei metodi basati su regole predefinite.

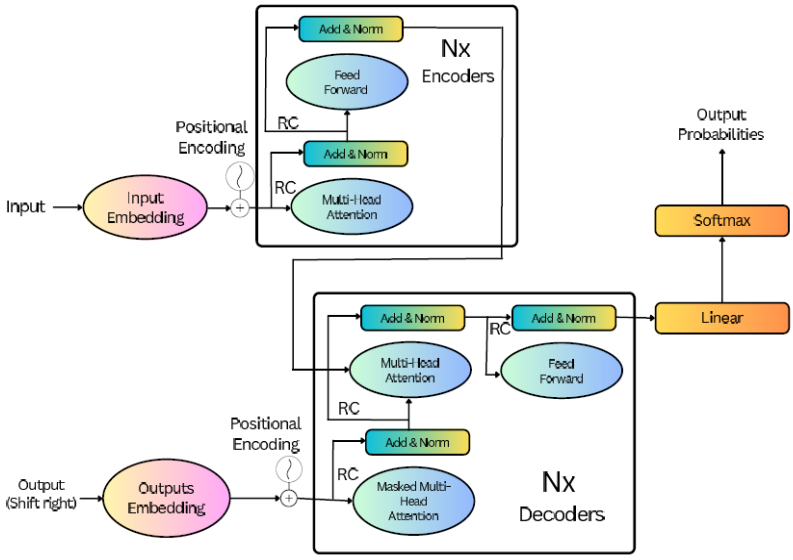

Immagine 1. Architettura dei Trasformer
(Fonte: Hassija et al, 2023)

Negli ultimi anni, l'evoluzione dei modelli GPT ha portato a significativi miglioramenti in termini di efficienza e adattabilità: l'ultima generazione di GPT ha ampliato ulteriormente il proprio campo d'azione grazie a capacità multimodali, permettendo l'elaborazione simultanea di input testuali e visivi. Questo ha aperto nuove possibilità di applicazione nei settori della comunicazione, dell'analisi dei dati e della creatività assistita dall'intelligenza artificiale.

Ciò che distingue i sistemi GPT da altri modelli di intelligenza artificiale è proprio il loro approccio massivo e *data-driven* al problema linguistico. Mentre i modelli precedenti si basavano espressamente su pattern esemplificativi definiti manualmente, il modello GPT fa affidamento su enormi database testuali per "apprendere" le strutture e le relazioni linguistiche tramite algoritmi di *machine learning*, senza la necessità intrinseca di un intervento umano in ogni specifico scenario conversazionale. Questo approccio consente ai modelli GPT di generare frasi apparentemente naturali ed efficaci sulla base delle lunghe sequenze di dati: in pratica, il modello crea una propria rappresentazione linguistica, basata su quello che ha "letto" durante l'allenamento. In questo modo, un modello GPT può rispondere a domande in modo

sorprendentemente dettagliato, generare articoli su argomenti specifici, scrivere poesie e altro ancora, producendo testi che sembrano avere una comprensione semantica pur non avendo una comprensione profonda del significato stesso. La capacità dei sistemi GPT di gestire una quantità enorme di dati e di svilupparsi costantemente li rende strumenti estremamente potenti, eppure limitati a causa della loro dipendenza dai dati e dal formato statistico. Non avendo una "comprensione" semantica o esperienziale come gli esseri umani, un sistema GPT si limita a generare testi che sembrano coerenti sulla base di ciò che ha imparato, ma che non sono frutto di una consapevolezza intenzionale. In conclusione, il funzionamento dei modelli GPT e il loro processo di allenamento testimoniano un'evoluzione significativa nell'ambito dell'Intelligenza Artificiale, permettendo la generazione di linguaggio naturale su larga scala e portando con sé nuove sfide, specialmente legate alla capacità di garantire coerenza, accuratezza e responsabilità nel loro utilizzo. La comprensione delle capacità e dei limiti di questi modelli è dunque fondamentale per ottenere il loro corretto impiego nel campo del conversation design, della generazione automatica di contenuti e delle interazioni uomo-macchina.

4.3 Impatto di ChatGPT sugli utenti

L'arrivo di ChatGPT ha rappresentato un momento di svolta nel panorama della Conversational AI, catalizzando un entusiasmo globale che ha travolto sia il pubblico che la comunità scientifica. Nel contesto di una rivoluzione tecnologica caratterizzata dalla nascita di modelli di linguaggio su larga scala, ChatGPT ha suscitato immediatamente l'attenzione dei media internazionali. Un esempio emblematico è l'articolo pubblicato da The Guardian intitolato *A robot wrote this entire article. Are you scared yet, human?* (GPT-3, 2020; Dickson, 2020; Heaven, 2020), il quale ha evidenziato la sorprendente capacità del modello di generare un intero testo autonomamente. Tale

pubblicazione ha innescato un vivace dibattito sulle potenzialità dei modelli di Intelligenza Artificiale e sulle implicazioni etiche di un sistema capace di produrre contenuti di alta qualità, alimentando l'illusione di un imminente salto verso l'Artificial General Intelligence (AGI)[16].

Il debutto di ChatGPT ha portato con sé un forte hype mediatico, spingendo numerosi osservatori e analisti a speculare sul futuro delle interazioni uomo-macchina. Da un lato, l'impressionante capacità del modello di comprendere e generare testo in maniera contestualmente rilevante ha fatto sorgere l'idea che, finalmente, l'IA potesse avvicinarsi a una forma di intelligenza generale, in grado di replicare la complessità e la creatività del linguaggio umano. Dall'altro, esperti e ricercatori hanno messo in guardia contro interpretazioni troppo ottimistiche, sottolineando che, nonostante i notevoli progressi, ChatGPT rimane un sistema basato su correlazioni statistiche e non su una vera comprensione semantica (Hassija et al., 2023).

Questo entusiasmo iniziale ha spinto la comunità scientifica a indagare più a fondo, non solo le capacità tecniche del modello, ma anche le sue implicazioni sociali ed etiche. L'interesse mediatico, con testimonianze e analisi provenienti da fonti autorevoli come The Guardian, MIT Technology Review e altre, ha evidenziato come ChatGPT abbia trasformato la percezione pubblica dell'intelligenza artificiale, presentandosi come una tecnologia rivoluzionaria in grado di automatizzare compiti complessi, migliorare la customer service, generare contenuti originali e persino influenzare il dibattito sull'AGI.

In questo contesto, la nascita di ChatGPT ha rappresentato non solo un progresso tecnologico, ma anche un catalizzatore per una riflessione critica

[16] L'Artificial General Intelligence (AGI) è una forma di intelligenza artificiale che, a differenza dei modelli di intelligenza artificiale ristretta o specifica (come i sistemi GPT), è in grado di comprendere e apprendere conoscenze in una varietà di domini e compiti, similmente all'intelligenza umana. I sistemi di intelligenza artificiale odierni (es. ChatGPT), sono specializzati per compiti specifici (es. la generazione di testo in contesti conversazionali), l'AGI avrebbe la capacità di trasferire il proprio apprendimento tra diverse aree senza richiedere una riprogettazione completa. In altre parole, un'intelligenza artificiale di tipo AGI potrebbe imitare l'intera gamma di capacità cognitive di un essere umano, tra cui ragionamento, pianificazione, comprensione del linguaggio, percezione, capacità motorie e interazione sociale. Attualmente, l'AGI rimane un obiettivo teorico non ancora raggiunto, e le tecnologie come GPT-3, sebbene impressionanti nell'eseguire compiti specifici, non possiedono la capacità di generalizzare propria dell'essere umano.

sulle potenzialità e sui limiti delle tecnologie di IA. L'ammirazione per la capacità di generare testi coerenti e naturali è stata accompagnata da interrogativi sulle possibili conseguenze dell'adozione diffusa di tali sistemi, tra cui il rischio di disinformazione, bias intrinseci e il potenziale impatto sulla sostituzione dei ruoli tradizionali. L'attenzione mediatica e il fervore del dibattito pubblico hanno quindi evidenziato la duplice natura di ChatGPT: da una parte, un'innovazione straordinaria che apre nuove prospettive per l'interazione e la comunicazione; dall'altra, un campo di ricerca che richiede ulteriori approfondimenti per garantire un uso responsabile e trasparente della tecnologia (Hassija et al., 2023; Dickson, 2020; Heaven, 2020).

4.3.1 Le tappe fondamentali dell'avvento di ChatGPT

L'evoluzione di ChatGPT rappresenta un percorso di continuo potenziamento, che ha rivoluzionato il nostro modo di interagire con le macchine. Il cammino inizia con il modello originale GPT, primo tentativo di OpenAI di generare testo in maniera autonoma attraverso un approccio di pre-allenamento non supervisionato. Questo modello pionieristico ha gettato le basi per l'utilizzo di tecniche di *deep learning* applicate al linguaggio naturale, aprendo la strada a successive iterazioni più complesse. Con l'introduzione di GPT-2 nel 2019, caratterizzato da circa 1,5 miliardi di parametri, si è registrato un notevole aumento delle capacità generative, permettendo di produrre testi significativamente più coerenti e fluidi rispetto alla prima versione. Questa fase ha segnato il passaggio da un sistema sperimentale a uno capace di applicazioni pratiche, poiché l'incremento della scala del modello ha consentito di apprendere in maniera più accurata le complesse relazioni semantiche e sintattiche del linguaggio.

Successivamente, GPT-3, lanciato nel 2020, ha portato questo progresso a un livello completamente nuovo, ampliando drasticamente la dimensione del modello fino a raggiungere 175 miliardi di parametri. Questa evoluzione ha permesso a ChatGPT di generare risposte estremamente articolate e contestualmente rilevanti, suscitando un entusiasmo diffuso tra il pubblico

e la comunità scientifica, come riportato in Hassija et al. (2023). Numerose testate giornalistiche hanno descritto l'arrivo di GPT-3, esaltandone la capacità di automatizzare la generazione di contenuti con una fluidità che ricorda il linguaggio umano. Allo stesso tempo, critiche e preoccupazioni sollevate da alcuni osservatori hanno evidenziato il rischio di un uso incontrollato di tali tecnologie. Tale uso improprio potrebbe avere conseguenze negative sul mercato del lavoro e sulla diffusione di disinformazione.

L'evoluzione non si è interrotta a GPT-3: la versione successiva, GPT-3.5, ha ulteriormente affinato le prestazioni del modello, migliorando la gestione delle ambiguità e garantendo un controllo più accurato sull'output. Questa iterazione ha ottimizzato la capacità del sistema di mantenere il contesto nelle conversazioni complesse e ha ridotto le problematiche legate a risposte incoerenti (Hassija et al., 2023). GPT-4 ha poi introdotto un ulteriore salto qualitativo, integrando capacità multimodali che permettono di elaborare contemporaneamente input testuali e visivi, migliorando la coerenza, l'accuratezza e l'efficienza nella generazione di contenuti. Questa evoluzione ha ampliato ulteriormente il campo di applicazioni di ChatGPT, rendendolo non solo uno strumento per la generazione di testo, ma anche un sistema in grado di comprendere contesti complessi e di operare in ambienti interattivi più variegati. L'ultima generazione, GPT-5, segna un nuovo punto di svolta: il modello consolida l'integrazione multimodale con una comprensione più profonda del contesto e una capacità di ragionamento più astratta e generalizzante. Rispetto ai predecessori, GPT-5 mostra un miglioramento sostanziale nella pianificazione discorsiva e nella gestione del dialogo prolungato, grazie a un'architettura più efficiente e a un addestramento su dati eterogenei e supervisioni raffinate. In tal modo, il modello tende verso una maggiore coerenza semantica, una minore incidenza di "allucinazioni" e un più accurato controllo delle inferenze pragmatiche nelle interazioni uomo-macchina.

Il percorso evolutivo di ChatGPT, che si estende dalla prima versione fino a GPT-5, testimonia la rapida espansione delle capacità dei modelli di linguaggio, alimentata dalla crescita esponenziale dei parametri e dalle inno-

vazioni nelle tecniche di pre-allenamento. Le evoluzioni introdotte dai modelli generativi della serie GPT hanno suscitato un ampio dibattito non solo mediatico, ma anche scientifico. Oltre alla copertura entusiastica dei principali quotidiani internazionali, la letteratura accademica ha analizzato in modo sistematico le prestazioni e i limiti di ChatGPT, valutandone l'efficacia in diversi contesti applicativi e comunicativi (Laskar et al., 2023; Hassija et al., 2023). Tali studi hanno evidenziato come, pur rappresentando un progresso significativo nell'elaborazione del linguaggio naturale, i modelli GPT presentino ancora criticità in termini di coerenza argomentativa, accuratezza fattuale e stabilità del comportamento conversazionale. Questo quadro di analisi ha contribuito a sostituire la narrazione giornalistica con una riflessione più strutturata sul bilanciamento tra innovazione tecnologica e responsabilità etica nello sviluppo di sistemi linguistici basati su IA. Il confronto tra i principi di progettazione conversazionale descritti da Kathy Pearl e le innovazioni introdotte dalla Gen AI evidenzia un cambiamento paradigmatico nella progettazione delle interfacce vocali. Mentre i modelli classici garantiscono coerenza, affidabilità e prevedibilità, i modelli generativi offrono flessibilità, adattabilità e una gestione del contesto più sofisticata. Tuttavia, questi progressi comportano anche nuove sfide, legate alla gestione dell'errore, alla sicurezza delle informazioni e all'affidabilità delle risposte fornite dai sistemi AI.

Nel prossimo futuro, il settore della progettazione conversazionale potrebbe assistere a un'integrazione tra questi due approcci, combinando la struttura e la prevedibilità delle VUI tradizionali con la versatilità e l'intelligenza contestuale offerte dalla Generative AI. Trovare il giusto equilibrio tra controllo e autonomia rappresenterà la chiave per sviluppare interfacce conversazionali più avanzate e intuitive, in grado di gestire interazioni sempre più complesse e naturali. Con l'ulteriore sviluppo delle tecnologie AI, sarà fondamentale monitorarne attentamente i progressi, per garantire che queste interfacce rispettino criteri di affidabilità, trasparenza ed etica nell'interazione tra uomo e macchina.

4.4 Etica e Privacy nella conversazione artificiale

Dalle prime simulazioni linguistiche di ELIZA (Weizenbaum, 1966), fino ai sistemi generativi di ultima generazione come ChatGPT, l'interazione uomo-macchina ha progressivamente ridotto la distanza tra linguaggio naturale e linguaggio computazionale. Ma a questa evoluzione tecnica corrisponde un incremento di complessità etica, che oggi si intensifica e pone dei limiti necessari nell'uso dei dati personali, in quanto le macchine non si limitano a eseguire comandi ma generano linguaggio, argomentano e apprendono dai comportamenti umani (Boddington, 2017).

In effetti, le questioni di trasparenza, responsabilità e tutela della privacy emergono oggi come i nodi più critici delle tecnologie conversazionali: il linguaggio, che per l'uomo è veicolo d'identità e relazione, diventa per la macchina una sorgente di dati, inferenze e potere predittivo.

In questa prospettiva, l'etica dell'IA non può essere considerata un capitolo a parte, ma una condizione necessaria di legittimità delle interazioni artificiali (Floridi, 2023).

4.4.1 Dalle origini dell'illusione all'intimità algoritmica

Il caso ELIZA (Wezeinbaum, 1966) segna l'inizio di una tensione etica tuttora attuale. Già nel 1966 Joseph Wezeinbaum aveva notato come molti utenti sviluppassero un attaccamento emotivo al programma, nonostante sapessero che le risposte erano generate da semplici regole di sostituzione.

Tuttavia, nei primi sistemi conversazionali, la dimensione etica era quasi assente: la macchina simulava dialoghi senza comprendere. In *Computer Power and Human Reason* (1976) Weizenbaum avvertiva: "ciò che conta non è che la macchina pensi, ma che l'uomo smetta di pensare", denunciando il rischio di "attribuire intenzionalità morale" a una macchina priva di coscienza (Weizenbaum, 1976).

Questa “illusione di comprensione” rappresenta il primo problema etico delle AI conversazionali: la sospensione dell’incredulità, che porta a percepire il sistema come empatico e affidabile. È un fenomeno che oggi si traduce in *over-trust*, ossia una fiducia eccessiva nella capacità cognitiva delle macchine.

Con l’evoluzione verso agenti come Siri o Alexa, la questione si è spostata dalla mera simulazione linguistica alla responsabilità algoritmica (*algorithmic accountability*): chi è responsabile delle decisioni implicite nella produzione linguistica di un sistema? (Diakopoulos, 2016).

Gli assistenti vocali apprendono da enormi basi di dati vocali e testuali, e i loro comportamenti linguistici riflettono i bias e le asimmetrie dei dati di addestramento (Crawford, 2021).

4.4.2 L’etica della voce: il caso degli assistenti vocali

Con Siri di Apple (2011) e i successivi assistenti vocali, l’interazione diventa quotidiana, pervasiva e potenzialmente intima. L’“etica della voce” si articola su due dimensioni principali:

1. Sorveglianza ambientale, legata alla possibilità che dispositivi “in ascolto” raccolgano dati oltre lo scopo dichiarato (Arthur, 2011);
2. Antropomorfismo progettato, cioè l’uso intenzionale di umorismo, tono amichevole e memoria contestuale per costruire fiducia.

L’utente, spesso inconsapevole di quando viene registrato, non distingue se una risposta “empatica” sia frutto di un atto programmato o di una reale comprensione. In questo senso, Siri inaugura l’era dell’intimità algoritmica (Yeasmin et al., 2020), dove la tecnologia assume un ruolo relazionale e il confine tra comunicazione e sorveglianza si fa poroso.

L’Unione Europea, con il Regolamento (UE) 2016/679 (GDPR), stabilisce principi chiave come *data minimization* e *purpose limitation*: i dati devono essere raccolti solo per finalità determinate e per il tempo strettamente

necessario. Tuttavia, come mostrano diversi rapporti del European Data Protection Board (EDPB 2021), i dispositivi vocali sfumano il confine tra uso funzionale e sorveglianza domestica.

La componente vocale rappresenta infatti un'identità biometrica: non solo ciò che si dice, ma anche come lo si dice diventa un dato personale. Questo pone un problema pragmatico e non solo tecnico: le interazioni linguistiche vengono registrate, archiviate e riutilizzate, generando un circuito comunicativo non più interpersonale ma asimmetrico (Zuboff, 2023).

4.4.3 La questione della privacy e della sorveglianza dei dati

I sistemi vocali e testuali di conversazione artificiale dipendono quindi da enormi quantità di dati. Il trattamento di tali informazioni solleva quattro criticità principali:

1. Raccolta implicita: le interazioni possono contenere dati sensibili (salute, emozioni, geolocalizzazione) che vengono acquisiti come materiale di training;
2. Conservazione e diritto all'oblio: i log vengono mantenuti per motivi di sicurezza o ottimizzazione, ma l'utente non ha pieno controllo sulla cancellazione;
3. Inferenze non dichiarate: gli algoritmi possono dedurre tratti identitari o psicologici non esplicitamente forniti;
4. Riutilizzo dei dati: le conversazioni vengono riutilizzate per scopi commerciali o di sviluppo senza trasparenza sufficiente (EDPB, 2021).

Nel 2024, il Garante per la Protezione dei Dati Personali ha sanzionato OpenAI per violazioni di trasparenza e tutela dei minori (Pollina & Armellini, 2024), evidenziando quanto la regolazione proceda più lentamente dell'innovazione tecnica.

4.4.4 L'etica dell'argomentazione automatica

L'etica dei sistemi conversazionali non si limita alla privacy: riguarda anche la responsabilità epistemica delle risposte generate.

Con Project Debater (IBM Research, 2019), l'attenzione si sposta dalla simulazione del dialogo alla costruzione automatica di argomenti. L'algoritmo è progettato per cercare fonti pubbliche e bilanciare pro e contro (Slonim et al., 2021), ma non possiede criteri interni di verità o giustizia. Si apre così il problema del valore morale del discorso prodotto da entità non coscienti.

Con i modelli linguistici di grandi dimensioni (GPT-3, GPT-4, GPT-5 e ChatGPT), il tema si amplifica. Questi sistemi apprendono da miliardi di testi raccolti online, spesso senza consenso esplicito degli autori, sollevando questioni di copyright e di utilizzo dei dati personali (Birhane & Prabhu, 2021).

Gli LLM non comprendono il significato, ma predicono la probabilità linguistica: la loro autorevolezza apparente può generare disinformazione o "allucinazioni" credibili (Bender et al., 2021). Stahl (2024) definisce questo fenomeno "erosione dell'autonomia cognitiva", in cui l'utente delega il proprio giudizio a un sistema opaco e difficilmente verificabile.

4.4.5 Bias, trasparenza e implicazioni pragmatiche

Le interazioni tra utenti e assistenti vocali (in particolare Alexa), analizzate da Panfili et al. (2021) attraverso le massime griceane, mostrano che le frustrazioni da parte degli utenti derivano da violazioni delle massime di Quantità e Relazione, che si traducono in mancanza di chiarezza e pertinenza.

Applicando la lente griceana emerge che tali violazioni, se traslate sul piano etico, equivalgono a mancanze di trasparenza e di controllo conversazionale: quando il sistema non è trasparente o produce messaggi non pertinenti, infrange la cooperazione comunicativa e mina la fiducia.

Secondo Searle (1969) e Grice (1975), l'atto linguistico implica intenzionalità e responsabilità del parlante: una IA che "parla" senza intenzionalità, o con intenzionalità simulata, introduce un vuoto di *accountability*. L'intelligenza conversazionale resta, quindi, una responsabilità umana travestita da linguaggio artificiale. Questo scarto pragmatico crea uno spazio grigio di responsabilità comunicativa, che l'etica dell'IA deve colmare.

4.4.6 Principi di un'etica delle conversazioni artificiali

Dalla letteratura recente (Floridi, 2023; Stahl, 2024; European AI Act 2025; UNESCO, 2021) emergono cinque principi cardine:

1. Trasparenza: esplicitare limiti e logiche di funzionamento dei sistemi;
2. Responsabilità (*accountability*): attribuire la supervisione e la paternità delle risposte a un soggetto umano o istituzionale;
3. Privacy by design: integrare la protezione dei dati nella progettazione architetturale e non come misura correttiva;
4. Equità e non discriminazione: ridurre i bias linguistici e culturali nei dataset e nei modelli;
5. Controllo umano significativo (*human oversight*): mantenere l'ultima parola nelle decisioni sensibili.

Tali principi, applicati retrospettivamente, delineano una traiettoria storica: ELIZA rappresenta la nascita del problema dell'illusione cognitiva; Siri introduce la questione dell'empatia artificiale e della sorveglianza domestica; Project Debater problematizza la responsabilità epistemica della macchina; ChatGPT estende il problema alla scala globale del dato e dell'informazione generata.

4.4.7 Verso un'etica dialogica

Floridi (2020) propone una *information ethics* fondata sul rispetto dell'"infosphera", in cui ogni entità informazionale, umana o artificiale, è parte di un ecosistema comunicativo.

Seguendo questa prospettiva, l'etica delle conversazioni artificiali può essere intesa come una "pragmatica della responsabilità" basata su tre principi interdipendenti: la trasparenza epistemica, che chiarisce quando l'interlocutore è un'IA e su quali basi produce risposte; il consenso informato, che garantisce all'utente la possibilità di sapere e modificare l'uso dei propri dati; e la responsabilità distribuita, che riconosce la co-costruzione linguistica tra progettisti, modelli e utenti.

Come ricorda Grice (1975), la cooperazione è la condizione stessa della conversazione: la sua violazione non è solo un errore linguistico, ma un fallimento etico.

L'etica e la privacy nelle conversazioni artificiali non sono dunque accessori della tecnologia, ma il loro fondamento. Ogni interazione con un assistente virtuale, da ELIZA a ChatGPT, è un atto linguistico mediato da infrastrutture economiche e algoritmiche.

Riprendendo la lezione di Weizenbaum (1976), la domanda cruciale non è se le macchine possano pensare, ma se l'essere umano riesca ancora a riconoscersi nel pensiero che esse riflettono. La conversazione artificiale è, in definitiva, un patto di fiducia: chi parla con una macchina deve sapere con chi parla, dove finiscono le sue parole e come vengono riutilizzate.

5.
PROGETTARE LA CONVERSAZIONE ARTIFICIALE

L'interazione tra esseri umani e macchine ha subito un'evoluzione significativa negli ultimi decenni, passando da semplici interfacce testuali a sistemi avanzati basati su Intelligenza Artificiale generativa. Eppure, non tutta l'IA conversazionale è realmente capace di sostenere una vera conversazione. Questo capitolo esplorerà tre diversi livelli di interazione artificiale: la non conversazione, esemplificata da Clippy, la falsa conversazione, tipica di assistenti vocali come Alexa e Siri, e la pseudo-conversazione, guidata da modelli di IA generativa come ChatGPT e Copilot.

Il termine "conversazione" implica un processo di scambio bidirezionale in cui entrambi gli interlocutori comprendono il contesto e rispondono in modo pertinente. Tuttavia, come verrà dimostrato, molte delle tecnologie esistenti simulano soltanto questa capacità, creando l'illusione di una comunicazione autentica. Come sottolinea Jess Stratton nel suo *Copilot for Microsoft 365*, l'integrazione di modelli di IA generativa in ambienti lavorativi ha trasformato il modo in cui interagiamo con i software, ma ciò non significa che questi strumenti possano realmente comprendere il linguaggio umano (Stratton, 2024).

Uno dei primi esempi di assistenti virtuali nella storia del software è Clippy, l'Office Assistant introdotto da Microsoft con Office 97. Clippy era progettato per fornire suggerimenti e assistenza agli utenti, ma si basava su semplici trigger testuali e non possedeva alcuna capacità conversazionale. La sua interazione era puramente deterministica: in base a parole chiave rilevate nel documento, proponeva suggerimenti preimpostati senza comprensione del contesto.

Clippy faceva parte di un'evoluzione dell'interfaccia utente intelligente, la sua inefficacia però ne ha portato la rimozione a partire da Office 2007 (Stratton, 2024). La sua principale limitazione risiedeva nel fatto che non era in grado di apprendere dall'utente o di adattarsi dinamicamente al dialogo, risultando più fastidioso che utile. Questo esempio dimostra come una mera reattività a input specifici non possa essere definita "conversazione" nel senso pieno del termine.

Con l'introduzione degli assistenti vocali, si è passati a un livello di interazione più sofisticato. Strumenti come Siri di Apple e Alexa di Amazon simulano una conversazione naturale attraverso il riconoscimento vocale e il recupero di informazioni dal web. Tuttavia, anche la loro capacità di comprensione è limitata, basandosi su schemi predefiniti e non possedendo una vera comprensione semantica.

Ad esempio, nel caso di Siri, la risposta a una domanda come "Qual è il tempo oggi?" si fonda sul recupero di dati meteorologici e sulla loro formattazione in una risposta naturale. Tuttavia, se la domanda viene formulata in modo ambiguo oppure implica un ragionamento contestuale, l'assistente può fallire nel fornire una risposta corretta. Questo fenomeno è stato ben documentato nella letteratura sull'IA conversazionale, che evidenzia come questi sistemi siano essenzialmente *pattern-matching* avanzati, piuttosto che entità dotate di comprensione propria.

L'ultimo stadio dell'evoluzione degli assistenti conversazionali è rappresentato dall'IA generativa. Modelli come ChatGPT e Copilot di Microsoft utilizzano reti neurali avanzate per generare risposte contestuali a input testuali, apprendendo da enormi dataset di linguaggio naturale. Nonostante la loro capacità di generare risposte apparentemente coerenti, questi strumenti non possiedono una vera comprensione del significato.

Copilot, per esempio, pur essendo un potente strumento di produttività, non è realmente in grado di "conversare" in senso umano (Stratton, 2024). Il modello si basa su predizioni statistiche e sulla capacità di generare testo plausibile, non ha però reale intenzionalità e coscienza del dialogo in corso. Que-

sto aspetto è determinante quando si verificano fenomeni come le allucinazioni dell'IA, perché il sistema produce risposte errate formulate però in modo convincente.

Comprendere la differenza tra questi tre livelli di interazione è fondamentale per la progettazione di sistemi conversazionali efficaci. Un vero Conversation Designer deve infatti considerare i limiti dell'IA e progettare esperienze in cui l'utente non venga ingannato dall'illusione di una comprensione che non esiste.

L'obiettivo non è semplicemente rendere l'interazione più "realistica", ma garantire che l'utente abbia aspettative corrette su ciò che l'IA può e non può fare. Microsoft ha integrato Copilot in modo che possa supportare il lavoro umano, non sostituirlo (Stratton, 2024). Questa è una distinzione cruciale per evitare fraintendimenti e garantire un utilizzo responsabile dell'IA.

La progettazione della conversazione artificiale richiede un'analisi critica delle tecnologie disponibili e delle loro reali capacità. Clippy, Siri e l'IA generativa rappresentano tre fasi di un percorso evolutivo che ha portato alla realizzazione di sistemi sempre più sofisticati, ma ancora lontani da una vera conversazione umana. La comprensione di questi limiti è essenziale per costruire esperienze utente efficaci e trasparenti, evitando illusioni di intelligenza che possono generare frustrazione o disinformazione.

5.1 Clippy: design conversazionale statico

5.1.1. Introduzione a Clippy

Negli anni Novanta Microsoft introdusse Clippy, ufficialmente noto come Clippit, un assistente virtuale pensato per migliorare l'esperienza degli utenti di Microsoft Office. Clippy era un agente animato che compariva per offrire suggerimenti e aiuto contestuale, basandosi su un motore di rilevamento delle azioni dell'utente. L'idea alla base della sua implementazione era

quella di rendere l'interazione con il software più intuitiva, anticipando le necessità degli utenti e fornendo suggerimenti proattivi.

D'altra parte, nonostante le intenzioni positive dietro la sua creazione, Clippy si rivelò rapidamente una delle funzionalità più detestate di Microsoft Office. L'assistente veniva percepito come invadente e poco funzionale, generando un diffuso senso di frustrazione tra gli utenti (Baym et al., 2019). La sua incapacità di comprendere il contesto e il modo in cui si inseriva nell'esperienza utente lo resero un simbolo di interfacce mal progettate.

5.1.2 La simulazione di conversazione di Clippy

Per comprendere il fallimento di Clippy come assistente virtuale, è essenziale analizzare il concetto di "non conversazione". A differenza di un vero interlocutore, Clippy non possedeva capacità di apprendimento, né un'effettiva comprensione semantica delle richieste da parte dell'utente. Il suo funzionamento si basava su modelli predefiniti di interazione, che venivano attivati in base a determinate parole chiave o azioni ripetute (Pearl, 2016).

La mancanza di adattabilità e la rigidità delle sue risposte generavano un'interazione frustrante. Piuttosto che aiutare realmente l'utente, Clippy interrompeva frequentemente il flusso di lavoro con suggerimenti irrilevanti, aggravando il problema che avrebbe dovuto risolvere. Questo aspetto sottolinea una distinzione fondamentale tra le vere interazioni conversazionali e le simulazioni statiche che non tengono conto del contesto in cui avvengono (SAS Institute, 2019).

5.1.3 Clippy: errori di design

Le principali cause del fallimento di Clippy possono essere ricondotte a tre fattori chiave.

Interruzione non richiesta: Clippy si attivava senza essere sollecitato, interrompendo il flusso di lavoro e distraendo l'utente da compiti essenziali.

Mancanza di personalizzazione: l'assistente non offriva opzioni per un'interazione su misura, ripetendo le stesse frasi indipendentemente dal livello di competenza dell'utente.

Incapacità di apprendimento: a differenza degli attuali sistemi basati su Intelligenza Artificiale, Clippy non era in grado di evolversi o migliorare in base alle interazioni precedenti. Ciò lo rendeva obsoleto e irritante nel lungo periodo (Baym et al., 2019).

Questi errori di progettazione hanno reso Clippy un esempio paradigmatico di interfacce fallimentari, dimostrando come un'assistenza non richiesta e poco intelligente possa generare più problemi di quanti ne risolva.

5.1.4 Impatto di Clippy sul design delle interfacce

L'insuccesso di Clippy ha avuto conseguenze importanti sulla progettazione delle interfacce conversazionali moderne. Ha mostrato l'importanza di progettare assistenti che siano discreti, realmente utili e dotati di capacità adattive. Le generazioni successive di assistenti virtuali, come Siri e Alexa, hanno appreso da questi errori, implementando sistemi basati su *machine learning* e migliorando il riconoscimento del contesto.

Tuttavia, nonostante i progressi tecnologici, molti sistemi attuali presentano ancora limiti simili a quelli di Clippy, come la generazione di risposte fuori contesto o di interazioni che non tengono conto delle reali necessità dell'utente. Questo suggerisce che, sebbene la tecnologia sia avanzata, le sfide fondamentali della conversazione artificiale rimangono aperte.

Clippy è un esempio emblematico di come un'assistenza virtuale mal progettata possa risultare inefficace e persino intrusiva per gli utenti. Il suo fallimento è stato determinato dalla rigidità del modello di interazione, dall'incapacità di adattamento al contesto e dall'eccessiva insistenza nel proporre aiuto. Le moderne tecnologie di intelligenza artificiale si sono evoluti significativamente rispetto a Clippy. Il suo caso però resta un monito per chiunque si occupi di Conversation Design: un'interazione artificiale efficace deve essere discreta, contestuale e realmente utile all'utente.

5.2 Assistenti digitali: l'illusione della conversazione

5.2.1 Le promesse iniziali degli assistenti vocali

L'introduzione degli assistenti vocali nel panorama tecnologico è stata accompagnata da un forte entusiasmo e dalla promessa di rivoluzionare il modo in cui gli utenti interagiscono con i dispositivi digitali. L'idea alla base di strumenti come Siri, Alexa e Google Assistant era quella di creare interfacce conversazionali capaci di comprendere e rispondere in modo naturale alle richieste umane, eliminando la necessità di interazioni basate su input manuali e testuali. Questa prospettiva, supportata dai progressi nel riconoscimento vocale e nell'elaborazione del linguaggio naturale (NLP), ha generato grandi aspettative sia tra i consumatori sia tra gli esperti del settore (Pearl, 2016).

Gli assistenti vocali sono stati presentati come strumenti in grado di facilitare la gestione delle attività quotidiane, migliorando l'accessibilità ai servizi digitali e automatizzando una serie di funzioni. Apple, con il lancio di Siri nel 2011, ha enfatizzato la capacità del suo assistente di comprendere il linguaggio naturale e di eseguire azioni complesse in base a semplici comandi vocali (Bellegarda., 2013). Amazon ha seguito questa tendenza con Alexa, promettendo un'integrazione perfetta con la smart home e la possibilità di interagire con numerose applicazioni e dispositivi tramite un'interfaccia vocale intuitiva (Lim et al., 2022). Google Assistant ha ampliato questa visione con l'obiettivo di fornire un supporto personalizzato basato su modelli avanzati di apprendimento automatico (Tulshan, 2018).

La promessa centrale di questi assistenti era dunque quella di trasformarsi in veri e propri agenti conversazionali capaci di comprendere il contesto delle richieste, apprendere dalle interazioni precedenti e adattarsi alle necessità individuali degli utenti. Tuttavia, fin dalle prime fasi di utilizzo, sono emerse limitazioni significative: la capacità di elaborazione del linguaggio si è rivelata meno sofisticata di quanto inizialmente dichiarato, con assistenti spesso incapaci di gestire conversazioni fluide e rispondere a domande che esulassero da comandi predefiniti (SAS Institute, 2019). In molti casi, gli

utenti si sono trovati a dover adattare il proprio modo di parlare affinché il sistema comprendesse le richieste, invertendo il paradigma iniziale secondo cui la tecnologia avrebbe dovuto adattarsi al linguaggio naturale umano.

Un ulteriore problema riscontrato riguarda la natura delle risposte fornite dagli assistenti vocali. Nonostante la capacità di restituire informazioni su meteo, appuntamenti o ricerche sul web, la loro interazione rimane vincolata a modelli rigidi di risposta. Il sistema non possiede una vera comprensione semantica del dialogo e si limita a elaborare input secondo schemi predefiniti. Questo aspetto si è rivelato particolarmente problematico in situazioni in cui le richieste dell'utente implicavano una comprensione contestuale profonda o la gestione di conversazioni prolungate.

Le aziende sviluppatrici hanno cercato di colmare queste lacune attraverso aggiornamenti progressivi e l'integrazione di algoritmi di Intelligenza Artificiale più raffinati. A ogni modo, la questione principale rimane aperta: questi sistemi sono davvero in grado di sostenere una conversazione autentica oppure si limitano a simulare un'interazione attraverso l'uso di "pulsanti verbali" che attivano funzioni specifiche? La riflessione su questa tematica è fondamentale per comprendere la reale portata dell'evoluzione tecnologica nel campo dell'Intelligenza Artificiale conversazionale, nonché per stabilire quali siano i limiti ancora da superare per la realizzazione di assistenti realmente intelligenti e interattivi.

5.2.2 La falsa conversazione: pulsanti verbali travestiti da dialogo

Sistemi come Siri e Alexa sono stati introdotti con la promessa di rivoluzionare l'interazione uomo-macchina, offrendo un'interfaccia conversazionale che avrebbe reso la comunicazione con i dispositivi digitali più naturale e intuitiva (Kepuska & Bohouta, 2018). Eppure, l'analisi delle loro funzionalità e dell'esperienza utente ha dimostrato che questi sistemi non erano veramente in grado di sostenere una conversazione reale. Piuttosto, essi operavano come un'interfaccia a comando, in cui frasi e richieste formulate dall'utente

servivano ad attivare risposte predefinite, senza che vi fosse comprensione del linguaggio e del contesto.

La principale limitazione di questi sistemi risiede nella loro architettura, basata su schemi rigidi di riconoscimento del linguaggio naturale. Siri e Alexa utilizzano una combinazione di riconoscimento vocale e modelli di elaborazione del linguaggio naturale per interpretare i comandi degli utenti (Tulshan, 2018). Malgrado ciò, questi sistemi non sono in grado di comprendere il linguaggio nel senso umano del termine, semplicemente si limitano a identificare parole chiave e ad abbinarle a risposte pre-programmate. Il risultato è un'interazione che sembrava fluida e naturale solo in apparenza, ma che si rivela rapidamente limitata quando l'utente tenta di discostarsi da comandi specifici o di avviare una conversazione più articolata (Pearl, 2016).

Questa dinamica può essere paragonata a una serie di pulsanti verbali: ogni comando riconosciuto dall'assistente attivava una funzione, proprio come un pulsante su un'interfaccia grafica. L'utente, anziché premere fisicamente un tasto su uno schermo, pronuncia una frase che funge da attivatore per un'azione specifica. Tale paradigma ha portato molti utenti a percepire l'interazione con Siri e Alexa non come un vero dialogo, ma come una forma più sofisticata di controllo vocale. Ciò ha anche generato frustrazione, poiché le risposte fornite dagli assistenti sono spesso prevedibili e incapaci di gestire domande più complesse o contestuali (Tulshan, 2018).

L'illusione di una conversazione naturale è stata rafforzata dall'uso di tecniche linguistiche progettate per rendere il sistema più simile a un interlocutore umano. Ad esempio, gli assistenti vocali sono stati programmati per rispondere con frasi che simulavano un tono colloquiale, per impiegare espressioni di cortesia e mantenere un'apparenza di continuità nel dialogo. Questi aspetti stilistici però non compensano l'assenza di una vera comprensione semantica e contestuale. Inoltre, la dipendenza da una struttura rigida di comandi limita la capacità di adattamento del sistema alle esigenze dell'utente, diventando un'esperienza che, a lungo termine, si dimostrava insoddisfacente per molti utilizzatori (Cain, 2025).

L’avvento dell’intelligenza artificiale generativa ha sollevato la questione: questi assistenti possano finalmente trasformarsi in agenti conversazionali autentici? Con l’integrazione di modelli avanzati come GPT-5, gli assistenti vocali possono ora generare risposte più articolate e contestuali, riducendo la rigidità delle interazioni precedenti. Tuttavia, è ormai accertato che i modelli linguistici di grandi dimensioni non comprendono realmente il significato profondo delle conversazioni e non sono capaci di gestire interazioni complesse senza cadere in risposte generiche o fuorvianti, in quanto si limitano a emulare statisticamente il linguaggio umano sulla base delle correlazioni presenti nei dati di addestramento. Sebbene la tecnologia abbia compiuto progressi significativi, il rischio di perpetuare la dinamica dei pulsanti verbali rimane, evidenziando la necessità di un ripensamento nel design delle interfacce conversazionali per garantire una reale evoluzione dell’interazione uomo-macchina (Pearl, 2016).

5.2.3 Le promesse della GenAI: Verso un’autentica conversazione

L’evoluzione degli assistenti vocali ha seguito un percorso di sviluppo che, come abbiamo detto più volte, almeno nelle intenzioni iniziali mirava a rendere l’interazione uomo-macchina più naturale e fluida. Ciononostante, fino a pochi anni fa, strumenti come Siri e Alexa erano ancora limitati da modelli di riconoscimento vocale e comprensione del linguaggio che si basavano su strutture predefinite, offrendo risposte codificate e prive di una reale capacità di adattamento al contesto. Con l’avvento dell’intelligenza artificiale generativa si è invece aperta una nuova fase, che promette di trasformare questi strumenti in veri agenti conversazionali, capaci di dialogare con gli utenti in modo più sofisticato e personalizzato.

L’integrazione di modelli di IA generativa negli assistenti vocali ha permesso un miglioramento significativo della fluidità delle interazioni, rendendo le risposte meno meccaniche e più articolate. Sistemi come ChatGPT e le più recenti versioni di Google Assistant hanno dimostrato la capacità di

comprendere meglio le intenzioni dell'utente, di generare risposte più contestualizzate e di sostenere conversazioni più complesse. Nondimeno, nel caso specifico di Siri e Alexa, il passaggio dall'approccio deterministico a uno più generativo non è ancora del tutto compiuto: entrambi gli assistenti vocali continuano a essere vincolati da strutture progettuali che ne limitano la reale capacità di conversazione.

Un aspetto centrale della loro evoluzione è la capacità di gestire il contesto della conversazione. Gli assistenti basati su modelli tradizionali tendevano a funzionare su un paradigma di comando e risposta, senza mantenere memoria del dialogo precedente. Questo ha spesso portato a interazioni frammentate e poco naturali. Al contrario, l'adozione di tecnologie di *machine learning* avanzato ha permesso un miglioramento nella gestione del contesto, consentendo a questi strumenti di ricordare temporaneamente le richieste dell'utente e adattare le risposte di conseguenza. Comunque, questo miglioramento rimane ancora limitato rispetto alle promesse iniziali di un'interazione completamente fluida e intuitiva.

Un ulteriore elemento di valutazione riguarda la comprensione semantica e la capacità di ragionamento. Sebbene l'intelligenza artificiale generativa abbia ampliato la gamma di risposte e migliorato la varietà linguistica delle interazioni, Siri e Alexa non possiedono ancora una vera e propria comprensione del linguaggio paragonabile a quella umana. Il loro funzionamento si basa su correlazioni statistiche tra parole e frasi piuttosto che su un'autentica capacità di interpretazione del significato. Questo aspetto emerge chiaramente quando si pongono domande che richiedono inferenze logiche complesse o quando si cerca di ottenere una risposta che non rientra nei dataset di addestramento.

Un altro limite evidente è la gestione delle ambiguità. Gli esseri umani sono in grado di interpretare le sfumature linguistiche, il sarcasmo e i contesti impliciti, mentre gli assistenti vocali, anche con l'integrazione dell'IA generativa, faticano ancora a comprendere queste dinamiche. Questo problema non è solo tecnico ma anche progettuale: affinché un assistente vocale diventi

un vero agente conversazionale, deve essere dotato di capacità inferenziali avanzate che vanno oltre la semplice elaborazione del linguaggio naturale.

Infine, aspetti cruciali sono la trasparenza e la fiducia da parte dell'utente. Con l'aumento delle capacità generative, emerge la questione della prevedibilità delle risposte e del controllo sulle informazioni fornite. Se un assistente vocale inizia a generare risposte basate su modelli predittivi avanzati, si pone il problema della verificabilità delle informazioni e del rischio di allucinazioni dell'IA. Per essere considerati veri agenti conversazionali, Siri e Alexa dovrebbero non soltanto migliorare la coerenza e la fluidità delle loro interazioni, ma anche garantire un elevato grado di affidabilità e trasparenza nelle risposte fornite.

In conclusione, l'era dell'IA generativa ha segnato un progresso significativo nel campo degli assistenti vocali. Per il momento, però, Siri e Alexa non possono essere considerati agenti conversazionali nel senso pieno del termine. Hanno migliorato la loro capacità di generare risposte contestuali e di sostenere conversazioni più articolate, ma continuano a essere vincolati da limiti strutturali che ne impediscono comprensione semantica e gestione avanzata del contesto. Affinché possano davvero evolversi in agenti conversazionali, sarà dunque necessario un ulteriore sviluppo delle loro capacità inferenziali, una maggiore integrazione di memorie conversazionali persistenti e un miglioramento della loro capacità di interpretazione del linguaggio naturale.

5.2.4 Clippy, Siri, Alexa: sono veri agenti conversazionali?

L'evoluzione degli assistenti virtuali si colloca in un contesto tecnologico in cui l'intelligenza artificiale ha progressivamente assunto un ruolo chiave nella semplificazione delle interazioni uomo-macchina. Le aspettative create attorno a questi strumenti, però, non sempre hanno trovato un riscontro nella realtà. Clippy, Alexa e altri assistenti digitali sono stati presentati come strumenti rivoluzionari, capaci di migliorare l'esperienza dell'utente attraverso un'interazione intuitiva e personalizzata. Tuttavia, l'analisi delle loro

effettive capacità mostra una netta discrepanza tra le promesse iniziali e il reale impatto sulla quotidianità degli utenti.

Come abbiamo visto Clippy, introdotto da Microsoft come assistente intelligente integrato in Office, nasceva con l'intento di anticipare i bisogni dell'utente e semplificare la redazione dei documenti. Nella pratica, però, l'assistente si rivelò invadente e scarsamente adattabile, intervenendo con suggerimenti rigidi e decontestualizzati. Il suo fallimento deriva da una progettazione che sopravvalutava la disponibilità dell'utente ad accettare interventi automatizzati, trascurando la necessità di un controllo più flessibile sull'interazione (Baym et al., 2019). Diversamente da Clippy, Alexa e Siri si sono proposti come assistenti vocali capaci di comprendere e rispondere in modo naturale alle richieste degli utenti. Apple presentò Siri come un'innovazione capace di elaborare il linguaggio naturale, apprendendo dalle interazioni e fornendo risposte contestualmente adeguate. In modo analogo, Amazon introdusse Alexa come un hub domestico intelligente, in grado di controllare dispositivi smart, fornire informazioni e facilitare attività quotidiane. Ma come sappiamo, questi strumenti, pur essendo basati su tecnologie avanzate di riconoscimento vocale, si limitano a interpretare comandi specifici piuttosto che comprendere realmente il linguaggio umano (Pearl, 2016).

Un aspetto fondamentale che accomuna Clippy, Alexa e Siri è la differenza tra l'aspettativa creata dal marketing e la loro effettiva capacità operativa. La promessa di un'interazione naturale e conversazionale si è spesso rivelata un'illusione tecnologica. Gli utenti si sono trovati di fronte a strumenti che, seppur capaci di elaborare comandi vocali, mancavano della profondità cognitiva necessaria per sostenere una conversazione autentica. Ciò ha portato a una diffusa insoddisfazione e a una percezione degli assistenti vocali come semplici strumenti di esecuzione piuttosto che veri e propri agenti intelligenti.

Il confronto tra Clippy e gli assistenti vocali moderni mostra quindi un'evoluzione nelle modalità di interazione, non però una trasformazione radicale nella natura di questi sistemi. Se Clippy fallì nel suo intento di supporto proattivo a causa di una scarsa flessibilità, Alexa e Siri hanno incontrato limiti

derivanti dalla loro dipendenza da pattern predefiniti e da una comprensione superficiale del linguaggio naturale. La progressiva integrazione di modelli di Intelligenza Artificiale generativa potrebbe rappresentare un passo avanti, ma rimane il problema della gestione delle aspettative: la capacità di un assistente virtuale di simulare una conversazione non equivale alla sua effettiva comprensione del dialogo.

In definitiva, Clippy, Alexa e Siri rappresentano tre fasi di una stessa sfida tecnologica: creare interfacce intelligenti capaci di migliorare l'esperienza dell'utente senza generare frustrazione o insoddisfazione. Le promesse fatte al momento del loro lancio non sempre sono state mantenute, dimostrando come la progettazione di agenti conversazionali richieda un equilibrio tra ambizione tecnologica e realismo nell'implementazione.

5.3 Dal design conversazionale statico al *prompt design*: oggi chi deve adattarsi?

In passato, i progettisti di assistenti virtuali come Clippy, Alexa e Siri avevano il compito di definire in anticipo percorsi conversazionali rigidamente strutturati. L'utente poteva interagire solo all'interno dei limiti imposti dal design del sistema, scegliendo tra risposte predefinite o attivando specifiche funzioni attraverso comandi vocali o testuali. Questo approccio, pur garantendo un certo grado di prevedibilità, limitava la naturalezza dell'interazione e spesso generava frustrazione.

Con l'avvento dell'IA generativa il paradigma è cambiato: non è più il sistema a dover adattare le proprie risposte a schemi rigidi, ma è l'utente a dover affinare la propria capacità di interazione con l'IA. Il concetto di *prompt engineering* diventa centrale in questo contesto, poiché la qualità della risposta generata dipende direttamente dalla formulazione dell'input da parte dell'utente (Ye et al., 2023; Khalid & Witmer, 2025). Questa transizione ha profonde implicazioni: da un lato, aumenta la flessibilità e la versatilità degli

strumenti conversazionali e dall'altro, trasferisce un maggiore carico cognitivo sugli utenti, che devono imparare a interagire con l'IA in modo efficace per ottenere risposte pertinenti.

Tale trasferimento del carico cognitivo solleva interrogativi sull'accessibilità e sull'usabilità di questi strumenti. Se in un design conversazionale statico era il progettista a preoccuparsi dell'esperienza utente, oggi è l'utente che deve sperimentare e adattarsi alle capacità del sistema, spesso senza una chiara comprensione dei suoi limiti e dei suoi meccanismi interni. Questa maggiore libertà consente interazioni più fluide e contestualizzate, ma introduce nuove sfide, come il rischio di ottenere risposte incoerenti o fuorvianti e la necessità di affinare costantemente il modo di porre le domande.

L'evoluzione dal design statico al prompt design rappresenta dunque un cambiamento fondamentale nel rapporto tra esseri umani e IA conversazionale. Comprendere queste dinamiche è essenziale per sviluppare sistemi più efficaci e garantire un'interazione che sia davvero utile e intuitiva per gli utenti, senza imporre loro un eccessivo onere nell'adattarsi alle nuove tecnologie.

5.3.1 Tecniche per guidare la conversazione artificiale

Alla luce degli approcci presentati nello stato dell'arte (Cfr. § 2.3.4), attraverso la guida di Khalid (2024) nella costruzione di un prompt efficace, in questa sezione si intendono approfondire alcune strategie per fornire alla macchina un'istruzione performante, al fine di ottenere risposte rilevanti e appropriate da parte di un sistema di Intelligenza Artificiale generativa. L'efficacia dell'interazione uomo-macchina non dipende più esclusivamente dalla progettazione dell'interfaccia, ma anche dalla capacità dell'utente di comunicare con l'IA in modo preciso e strutturato. Un prompt ben progettato diventa lo strumento centrale per guidare la generazione del linguaggio artificiale, assicurando coerenza, aderenza al contesto e funzionalità comunicativa.

L'evoluzione del *prompt design* non si limita più alla scelta di formule linguistiche funzionali o alla redazione di comandi preimpostati. Essa coinvolge una vera e propria strategia composita per ottenere interazioni sempre più efficaci, controllabili e pertinenti con i modelli linguistici generativi. In questo contesto, l'analisi delle tecniche più avanzate di *prompting* fornisce un quadro di riferimento fondamentale per progettare interazioni sofisticate e personalizzate, orientate non solo alla risposta, ma alla comprensione contestuale e alla co-creazione.

Uno dei contributi più significativi a tale scopo è offerto dalla meta-analisi contenuta in *The Prompt Report* (Schulhoff et al., 2024), che ha sistematizzato i principali approcci al *prompting* attraverso una revisione metodologica di oltre cento pubblicazioni scientifiche. Il metodo utilizzato si basa sul framework PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), un modello scientifico rigoroso che consente di aggregare, confrontare e valutare criticamente dati tratti da studi diversi, garantendo così una sintesi coerente, trasparente e riproducibile (Moher et al., 2009).

La meta-analisi identifica sei principali tecniche di *prompting* che si sono dimostrate efficaci per migliorare l'interazione uomo-macchina:

- *zero-shot prompting*
- *few-shot prompting*
- *chain-of-thought prompting*
- *decomposizione del problema*
- *ensemble dei prompt*
- *prompting basato sull'autocritica*

Ognuna di queste strategie presenta caratteristiche peculiari, vantaggi specifici e ambiti di applicazione differenziati, offrendo agli utenti strategie avanzate per comunicare in modo efficace con i modelli generativi.

Tra le tecniche più rilevanti oggi a disposizione degli utenti vi sono le *text-based prompting techniques*, ovvero strategie di progettazione dell'input

testuale che guidano direttamente il comportamento dell'IA generativa (Mother et al., 2009). Una delle più discusse è l'*In-Context Learning* (ICL), che permette al modello di apprendere implicitamente il comportamento atteso sulla base degli esempi forniti all'interno dello stesso prompt. In particolare, il *few-shot prompting* consente di guidare l'output del modello attraverso un numero limitato di esempi espliciti: ad esempio, per insegnare a riformulare frasi complesse in linguaggio semplice, si possono presentare due coppie esempio-input/output, seguite da un nuovo input da riformulare.

All'interno del *few-shot prompting* diverse decisioni di design influenzano l'efficacia dell'interazione. La quantità di esempi da includere, detta *exemplar quantity*, deve essere calibrata con cura: troppi esempi possono sovraccaricare il modello, troppo pochi possono essere insufficienti per guidare correttamente l'output. L'*esemplar format*, ovvero la struttura interna degli esempi, deve essere uniforme: ad esempio, se si utilizza la dicitura "Domanda: ... Risposta: ...", è consigliabile mantenerla in tutti i casi. Un aspetto cruciale è anche l'*esemplar similarity*, ovvero la somiglianza tra gli esempi forniti e il compito target. Ad esempio, se si vuole che il modello produca la descrizione di un film, è più utile fornire esempi di sinossi di film simili piuttosto che descrizioni generiche.

Tecniche derivate dal *few-shot prompting* includono il *Self-Generated In-Context Learning* (SG-ICL), in cui è il modello stesso a generare esempi a partire da una descrizione generica del task, prima di risolverlo. Questa modalità è utile quando non si dispone di esempi di qualità, ma ci si affida alla capacità del modello di generare un contesto di apprendimento in modo autonomo. Altra tecnica è il *Prompt Mining*, che consiste nel recupero automatico di *prompt* efficaci da grandi archivi di interazioni, sfruttando metriche di performance per individuare le formulazioni ottimali da riutilizzare.

Nel caso dello *zero-shot prompting*, l'interazione avviene in assenza di esempi ma con l'uso di istruzioni e modificatori testuali mirati. Il *role prompting* assegna un'identità al modello ("Sei un esperto di marketing digitale"), mentre lo *style prompting* specifica un registro comunicativo ("Scrivi in tono

formale e professionale"). L*'emotion prompting* guida l'output verso una tonalità emotiva, utile per scrivere messaggi empatici. Tecniche più sofisticate come il *System 2 Attention* (S2A) inducono il modello a riflettere prima di rispondere ("Prima di rispondere ragiona per 30 secondi"), mentre il *Rephrase and Respond* (RaR) invita il modello a riformulare una domanda prima di affrontarla, migliorandone la comprensione stessa. La tecnica *Self-Ask*, infine, stimola l'autogenerazione di sotto-domande per risolvere quesiti complessi, come nel caso in cui si chieda "Qual è la capitale del paese che confina a sud con la Germania?" e il modello prima risponde "Quali sono i paesi che confinano a sud con la Germania?" e poi conclude: "L'Austria confina a sud, quindi la capitale è Vienna."

Un'altra classe di tecniche riguarda la *decomposition prompting*, cioè la suddivisione di compiti complessi in sottocompiti. Il *Least-to-Most Prompting* parte dai passaggi più semplici per costruire progressivamente la soluzione finale. Il *Plan-and-Solve Prompting* prevede la pianificazione anticipata delle fasi di ragionamento. Il *Tree-of-Thought* (ToT) rappresenta le opzioni in una struttura ad albero, simulando esplorazioni multiple prima di giungere a una conclusione. Il *Faithful Chain-of-Thought* introduce invece vincoli logici tra i passaggi per garantire coerenza interna, utile per evitare errori nei problemi aritmetici o nei ragionamenti formali.

Nel campo del *prompt ensembling* si utilizzano versioni multiple del prompt per generare risposte alternative, valutando poi quale sia la più robusta. Il *Demonstration Ensembling* (DENSE) consiste nell'unire più esempi rappresentativi all'interno di un unico prompt per costruire un contesto ricco. Ad esempio, per addestrare il modello a classificare recensioni come positive, negative o neutre, si possono inserire tre esempi completi con etichette, lasciando poi al modello il compito di analizzare una nuova recensione. La tecnica della *Self-Consistency* consiste nel generare più risposte dallo stesso prompt (con temperature o randomizzazioni diverse) e selezionare quella più frequente o logica. Un esempio tipico è la risoluzione di un problema aritmetico come "180 km divisi per 60 km/h", in cui le risposte convergenti ("3 ore") vengono privilegiate. La *Universal Self-Consistency*, invece, spinge il

concetto oltre: si generano risposte da formulazioni differenti del prompt (più formali, più semplici, con toni diversi) e si sintetizzano le informazioni convergenti, ad esempio nei riassunti di testi complessi, per ottenere una risposta più affidabile e generalizzabile.

L'utilizzo di queste tecniche è condizionato da fattori quali il tipo di compito, il livello di precisione richiesto, la necessità di trasparenza e la disponibilità di esempi predefiniti. Le tecniche più semplici (*zero-shot*, *role prompting*) risultano adatte a compiti diretti e ben definiti. Tecniche più complesse (*ensemble*, *decomposition*) offrono vantaggi in ambiti di *problem solving*, argomentazione e generazione di contenuti complessi.

Prompt engineering e *answer engineering* sono due approcci complementari alla progettazione dell'interazione. Il primo si concentra sull'elaborazione intenzionale della richiesta da parte dell'utente: definire ruoli, contesto, tono, struttura e vincoli. Si tratta di costruire una "sceneggiatura" efficace per stimolare il comportamento desiderato nel modello. L'*answer engineering*, invece, riguarda la progettazione della forma dell'output: scegliere lunghezza, struttura retorica, formato testuale (es. elenco, paragrafo, e-mail, codice, etc.) e la presenza di elementi esplicativi. Entrambe le strategie permettono di aumentare la precisione, la leggibilità e l'utilità della risposta generata.

Nel complesso, le tecniche illustrate non rappresentano semplici accorgimenti linguistici, ma veri e propri strumenti cognitivi e strategici per controllare e affinare il comportamento dei modelli generativi. Un impiego consapevole consente agli utenti non solo di ottenere risultati più pertinenti e affidabili, ma di costruire una relazione conversazionale con l'IA più interessante, interpretabile e finalizzata a obiettivi concreti. Il *prompt design*, in questa prospettiva, si configura come un'arte comunicativa in continua evoluzione, capace di trasformare l'interazione uomo-macchina in uno spazio di progettazione condivisa e intelligente.

CONCLUSIONI

Nel corso di questa ricerca ho voluto esplorare la natura della conversazione tra esseri umani e macchine, affrontando la questione non solo dal punto di vista tecnico, ma anche linguistico, storico e sociale. Progettare un'interazione artificiale significa in effetti misurarsi con la complessità del linguaggio umano: una complessità fatta di intenzioni, contesti, regole implicite e aspettative relazionali. La comunicazione non è mai un semplice scambio di informazioni, ma un processo dinamico di costruzione condivisa del senso. È proprio in questa dimensione che si gioca la sfida più grande per l'Intelligenza Artificiale: riuscire non solo a rispondere, ma a farlo in modo cooperativo, credibile e situato.

Nel mio percorso, ho mostrato come i sistemi deterministici del passato, per quanto innovativi nel loro tempo, siano oggi insufficienti a sostenere una vera interazione. Al contrario, i modelli generativi contemporanei, come i Large Language Models, hanno segnato una svolta importante. Offrono maggiore coerenza testuale, una migliore capacità di adattarsi al contesto e una fluidità comunicativa senza precedenti. Tuttavia, restano strumenti statistici: non comprendono davvero, ma calcolano probabilità di occorrenza linguistica in base a vasti corpora di dati. Anche quando sembrano "sapere", stanno solo predicendo. Per questo motivo, sostengo che sia necessario sviluppare una maggiore responsabilità nel loro utilizzo, a partire dalla consapevolezza dei loro limiti e della loro natura simulativa.

Questi modelli sono sempre più capaci di costruire un contesto conversazionale, mantenendo la coerenza tra gli scambi, ricordando elementi dell'interazione e adattandosi agli input dell'utente. Ma questo contesto, per quanto raffinato, non è sufficiente se non è accompagnato da un controllo più

consapevole delle norme conversazionali, le stesse regole implicite che ci permettono di interpretare le intenzioni dell'altro, di evitare ambiguità e di cooperare nel dialogo. Senza una guida progettuale chiara, anche il miglior modello rischia di generare risposte incoerenti, fuorvianti o improprie. È qui che interviene il Conversation Design, inteso non solo come tecnica di scripting, ma come forma di cura della relazione.

Una delle sfide individuate riguarda la dimensione empatica della conversazione artificiale. Oggi, sistemi come Siri sono progettati per rispondere in modo efficiente e rapido, ma non per creare un vero legame emotivo. Gli utenti tendono a considerarla un "aiutante tecnologico", un "signore artificiale", uno strumento funzionale e impersonale. Questo approccio è rassicurante, ma anche limitante. Credo che per rendere le interazioni più credibili e coinvolgenti sarà necessario introdurre nei sistemi artificiali un grado maggiore di empatia progettata: non un'emozione reale, naturalmente, ma un comportamento linguistico che rifletta attenzione, riconoscimento e adattamento affettivo. Non per sostituire le relazioni umane, ma per accompagnarle meglio.

L'esperienza ci ha già insegnato cosa evitare. Clippy, l'assistente virtuale di Microsoft, è un esempio di fallimento comunicativo: intrusivo e fuori contesto. Da questo caso è possibile comprendere che una conversazione artificiale deve essere discreta, contestuale e veramente utile. Deve sapere quando parlare, come intervenire e soprattutto quando è il momento di tacere. Questo vale oggi più che mai, in un'epoca in cui l'Intelligenza Artificiale entra in ambiti sempre più sensibili della nostra vita quotidiana.

Un altro aspetto fondamentale che ho approfondito è quello del Prompt Design. Oggi, una parte dello sforzo conversazionale viene trasferita sull'utente: è lui che deve formulare richieste chiare, pertinenti, efficaci. Ma se il prompt è ben costruito, se il sistema è guidato da una buona progettazione, il risultato sarà una comunicazione più coerente, aderente ai bisogni reali e soddisfacente. Il Prompt Design, in questo senso, rappresenta una nuova forma di collaborazione tra uomo e macchina, in cui il significato emerge dall'interazione stessa, non solo dalla tecnologia.

Guardando al futuro, sembra che il potenziale dell'Intelligenza Artificiale conversazionale sia ancora tutto da esplorare: interessante e inquietante al tempo stesso. I prossimi passi dovranno includere modelli più trasparenti, controllabili e responsabili, capaci di apprendere senza replicare i bias del passato e di conversare senza fingere di comprendere ciò che non possono (ancora) capire. Se riusciremo a progettare interazioni in cui le macchine siano sì intelligenti, ma al tempo stesso delimitate da regole chiare e da una progettazione consapevole dei loro limiti, allora potremo davvero costruire sistemi che, pur rimanendo artificiali, sapranno parlare con noi in modo più umano.

In definitiva, questo lavoro non vuole offrire risposte risolutive ma uno spazio critico di riflessione. Mi sono chiesta: è possibile progettare una conversazione tra uomo e macchina che sia qualcosa di più di una simulazione? La risposta, forse, non sta solo nelle macchine, ma anche in noi: nel modo in cui scegliamo di progettarle, di usarle e di ascoltarle.

BIBLIOGRAFIA

Aiello, L. C. (2016). The multifaceted impact of Ada Lovelace in the digital age. *Artificial Intelligence*, 235, 58-62.

Akinbi, A., & Berry, T. (2020). Forensic investigation of google assistant. *SN Computer Science*, 1(5), 272.

Akman, A., & Schuller, B. W. (2024). Audio explainable Artificial Intelligence: A review. *Intelligent Computing*, 2, 0074.

Alhosani, K., & Alhashmi, S. M. (2024). Opportunities, challenges, and benefits of AI innovation in government services: A review. *Discover Artificial Intelligence*, 4(1), 18.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. (2024). A Survey on RAG with LLMs. *Procedia Computer Science*, 246, 3781-3790.

Arthur, C. (2011, ottobre 5). Apple's Siri: Your Wish Is Her Command. *The Guardian*. Disponibile su: https://www.theguardian.com/technology/2011/oct/05/iphone-4s-siri-icloud. URL consultato il 9 ottobre 2025.

Assefi, M., Liu, G., Wittie, M. P., & Izurieta, C. (2015). An experimental evaluation of Apple Siri and Google Speech Recognition. In *Proceedings of the 2015 ISCA SEDE*, 118.

Austin, J. L. (1975). *How to do things with words*. Harvard University Press.

BBC News. (2016). “Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol”. Disponibile su: https://www.bbc.com/news/technology-35785875. URL consultato il 26 settembre 2025.

Bar-Haim, R., Kantor, Y., Venezian, E., Katz, Y., & Slonim, N. (2021). Project Debater APIs: Decomposing the AI grand challenge. *arXiv preprint arXiv*:2110.01029.

Bar-Haim, R., Eden, L., Friedman, R., Kantor, Y., Lahav, D., & Slonim, N. (2020). From arguments to key points: Towards automatic argument summarization. *arXiv preprint arXiv*:2005.01619.

Baym, N., Shifman, L., Persaud, C., & Wagman, K. (2019). Intelligent failures: Clippy memes and the limits of digital assistants. *AoIR selected papers of internet research*.

Bassett, C. (2019). The computational therapeutic: exploring Weizenbaum’s ELIZA as a history of the present. *AI & SOCIETY*, 34(4), 803-812.

Bebis, G., & Georgiopoulos, M. (2002). Feed-forward neural networks. *IEEE Potentials*, 13(4), 27-31.

Beerbaum, D. (2023). *Generative Artificial Intelligence (GAI) with Chat GPT for Accounting–a business case*. Available at SSRN 4385651.

Bellegarda, J. R. (2013). Spoken language understanding for natural interaction: The Siri experience. *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems into Practice*, 3-14.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610-623.

Berry, D. M. (2023). The limits of computation: Joseph Weizenbaum and the ELIZA chatbot. *Weizenbaum Journal of the Digital Society*, 3(3).

Birhane, A., & Prabhu, V. U. (2021). Large image datasets: A pyrrhic win for computer vision?. In 2021 *IEEE Winter Conference on Applications of Computer Vision* (WACV).1536-1546.

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Cham: Springer. 0-143.

Boden, M. A. (1989). *Artificial intelligence in psychology: Interdisciplinary essays*. MIT Press.

Bolton, T., Dargahi, T., Belguith, S., Al-Rakhami, M. S., & Sodhro, A. H. (2021). On the security and privacy challenges of virtual assistants. *Sensors*, 21(7), 2312.

Borji, A. (2022). Generated faces in the wild: Quantitative comparison of stable diffusion, Midjourney and DALL—E 2. *arXiv preprint arXiv*:2210.00586.

Bosker, B. (2013, January 12). SIRI RISING: The Inside Story of Siri's Origins – And why she could overshadow the iPhone. Huffpost / MIT Technology Review. Disponibile su: https://www.huffpost.com/entry/siri-do-engine-apple-iphone_n_2499165. URL consultato il 7 ottobre 2025.

Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv*:1809.11096.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.

Brown, P. (1987). *Politeness: Some universals in language usage*. Vol. 4. Cambridge University Press.

Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Valdez, A. C. (2019). “Hey, Siri”, “Ok, Google”, “Alexa”. Acceptance-Relevant Factors of Virtual Voice-Assistants. In *International professional communication conference*. IEEE. 101-111.

Cain, A. (2025, august 20). “The Life and Death of Microsoft Clippy, the Paper Clip the World Loved to Hate”. Artsy.net. Disponibile su: https://www.artsy.net/article/artsy-editorial-life-death-microsoft-clippy-paper-clip-loved-hate. URL consultato il 26 settembre 2025.

Cazzaniga, M., Jaumotte, M. F., Li, L., Melina, M. G., Panton, A. J., Pizzinelli, C., & Tavares, M. M. (2024). *Gen AI: Artificial intelligence and the future of work*. International Monetary Fund.

Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A review of trustworthy and explainable artificial intelligence (XAI). *IEEE Access*, 11, 78994-79015.

Chen, B., Wu, Z., & Zhao, R. (2023). From fiction to fact: the growing role of generative AI in business and finance. *Journal of Chinese Economic and Business Studies*, 21(4), 471-496.

Chu, X., Tian, Z., Zhang, B., Wang, X., & Shen, C. (2021). Conditional positional encodings for vision transformers. *arXiv preprint arXiv*:2102.10882.

Chui, M., & Francisco, S. (2017). Artificial intelligence the next digital frontier. *McKinsey and Company Global Institute*, 47(3.6), 6-8.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Colby, K. M., Hilf, F. D., Weber, S., & Kraemer, H. C. (1972). Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. *Artificial Intelligence*, 3, 199-221.

Corkrey, R., & Parkinson, L. (2002). Interactive voice response: review of studies 1989–2000. *Behavior Research Methods, Instruments, & Computers*, 34(3), 342-353.

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62.

De Almeida, A. F., Moreira, R., & Rodrigues, T. (2019). Synthetic organic chemistry driven by artificial intelligence. *Nature Reviews Chemistry*, 3(10), 589-604.

Dennett, D. C. (1994). The myth of original intentionality. *In Thinking computers and virtual persons*. Academic Press. 91-107.

Dennett, D. (1980). The milk of human intentionality. *Behavioral and Brain Sciences*, 3(3), 428-430.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Bidirectional encoder representations from transformers. *arXiv preprint arXiv*:1810.04805, 15.

Dickson, B. (2020). "The Guardian's GPT-3-written article misleads readers about AI. Here's why". TechTalks.

Due, B. L., & Lüchow, L. (2022). VUI-speak: There is nothing conversational about "conversational user interfaces". *Communicative AI in (Inter-) Action: Investigating Human-Machine Encounters outside the Laboratory*, 155.

Durkin, J. (1996). Expert systems: a view of the field. *IEEE Intelligent Systems*, 11(02), 56-63.

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a writer's Turing test?. *Journal of Cultural Analytics*, 5(2).

European Data Protection Board (EDPB). (2021, july 7). *Guidelines on Virtual Voice Assistants*. Disponibile su: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-022021-virtual-voice-assistants_en. URL consultation il 9 ottobre 2025.

European Parliament. (2025, february 19). *EU AI Act: First Regulation on Artificial Intelligence*. Disponibile su: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence. URL consultato il 9 ottobre 2025.

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: beyond jeopardy!. *Artificial Intelligence*, 199, 93-105.

Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.

Floridi, L. (2020). AI and its new winter: From myths to realities. *Philosophy & Technology*, 33(1), 1-3.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and machines*, 30(4), 681-694.

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Floridi, L. (2017). Digital's cleaving power and its consequences. *Philosophy & Technology*, 30(2), 123-129.

Freitag, M., & Al-Onaizan, Y. (2017). Beam search strategies for neural machine translation. *arXiv preprint arXiv*:1702.01806.

Friederich, S. (2017). Fine-tuning. *The Stanford encyclopedia of philosophy*.

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, 25(3), 277-304.

Gay, J. (2011, october 8). You Cannot Be Siri-Ous. *The Wall Street Journal*. Disponibile su: https://www.wsj.com/articles/SB10001424052970204346104576637340627863406?gaa_at=eafs&gaa_n=ASWzDAi_a11ik3SDleRAZN7IpuhB6WoV3coU-BoSxaNuU45XkkzgbRqon3xTm&gaa_sig=3azT1pmu0zIvyJTqhsr-ebtmh4zNYA2xsehhw-XrofTCZUrk2qs7kZXu-DCL5o0V2F1xOZOLMCV0-isl-dUw0BVQ%3D%3D&gaa_ts=68e58341&utm.com. URL consultato il 9 ottobre 2025.

Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., & Sikdar, B. (2024). Privacy and security concerns in generative AI: a comprehensive survey. *IEEE Access*, 12, 48126-48144.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

González Barman, K., Lohse, S., & de Regt, H. W. (2025). Reinforcement learning from human feedback in LLMs: Whose culture, whose values, whose perspectives?. *Philosophy & Technology*, 38(2), 1-26.

GPT-3. (2020, september 8). "A robot wrote this entire article. Are you scared yet, human?". *The Guardian*.

Grice, H. P. (1975). Logic and conversation. In *Syntax and semantics*, 3, 43-58.

Grune, D., & Jacobs, C. J. (2008). Introduction to parsing. In *Parsing techniques: A practical guide*. NY: Springer New York. 61-102.

Hasan, R., Shams, R., & Rahman, M. (2021). Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri. *Journal of Business Research*, 131, 591-597.

Hassija, V., Chakrabarti, A., Singh, A., Chamola, V., & Sikdar, B. (2023). Unleashing the potential of conversational AI: Amplifying chat-GPT's capabilities and tackling technical hurdles. *IEEE Access*, 11, 143657-143682.

Heaven, W. D. (2020). OpenAI's new language generator GPT-3 is shockingly good—and completely mindless. *MIT Technology Review*.

Herring, S. C. (1999). Interactional coherence in CMC. In *Proceedings of the 32nd Annual International Conference on Systems Sciences*. HICSS-32. IEEE. 1-13.

Hossain, M. M. (2021). The application of Grice maxims in conversation: A pragmatic study. *Journal of English Language Teaching and Applied Linguistics*, 3(10), 32-40.

Howe, N. P., & Bundell, S. (2021, march 17). *The AI that argues back. A computer that can participate in live debates against human opponents*. Nature. Disponibile su: https://www.nature.com/articles/d41586-021-00720-w?utm_source=chatgpt.com. URL consultato il 7 ottobre 2025.

Hsu, F. H. (1999). IBM's deep blue chess grandmaster chips. *IEEE micro*, 19(2), 70-81.

Humphreys, G. W., & Sui, J. (2016). Attentional control and the self: the Self-Attention Network (SAN). *Cognitive Neuroscience*, 7(1-4), 5-17.

Jakobson, R., & Sebeok, T. A. (1960). Closing statement: Linguistics and Poetics. In *Semiotics: An introductory anthology*, 147-175.

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898.

Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., & Gebru, T. (2023). AI Art and its Impact on Artists. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. 363-374.

Jimenez, C., Saavedra, E., del Campo, G., & Santamaria, A. (2021). Alexa-based voice assistant for smart home applications. *IEEE Potentials*, 40(4), 31-38.

Jo, A. (2023). The promise and peril of generative AI. *Nature*, 614(1), 214-216.

Jones, C., & Bergen, B. (2024). Does GPT-4 pass the Turing test?. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 5183-5210.

Joshi, A. K. (1991). Natural language processing. *Science*, 253(5025), 1242-1249.

Khalid, M. T., & Witmer, A. P. (2025). Prompt Engineering for Large Language Model-assisted Inductive Thematic Analysis. *arXiv preprint arXiv*:2503.22978.

Khalid, I. (2024). "A step-by-step guide to prompt design for students". *Ai4Education.in*.

Kanjee, Z., Crowe, B., & Rodman, A. (2023). Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. *Jama*, 330(1), 78-80.

Kepuska, V., & Bohouta, G. (2018). Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google

Home). In 2018 *IEEE 8th annual computing and communication workshop and conference* (CCWC). IEEE. 99-103.

Khawaldeh, A. M. (2025). Generative AI Hallucinations and Legal Liability in Jordanian Civil Courts: Promoting the Responsible Use of Conversational Chat Bots. *International Journal for the Semiotics of Law-Revue Internationale de Sémiotique juridique*, 38(2), 381-401.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv*:1312.6114.

Kirschthaler, P., Porcheron, M., & Fischer, J. E. (2020). What can i say? effects of discoverability in VUIs on task performance and user experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1-9.

Labbe, M. (2019, february 12). *IBM's Project Debater loses debate but shows off AI prowess*. IBM Think / TechTarget. Disponibile su: https://www.techtarget.com/searchenterpriseai/news/252457450/IBMs-Project-Debater-fails-to-win-debate-with-human?utm_source=chatgpt.com. URL consultato il 7 ottobre 2025.

Lane, R., Hay, A., Schwarz, A., Berry, D. M., & Shrager, J. (2025). ELIZA Reanimated: The world's first chatbot restored on the world's first time sharing system. arXiv preprint arXiv:2501.06707.

Laskar, M. T. R., Bari, M. S., Rahman, M., Bhuiyan, M. A. H., Joty, S., & Huang, J. X. (2023). A systematic study and comprehensive evaluation of ChatGPT on benchmark datasets. arXiv preprint arXiv:2305.18486.

Lee, B. (2022, december 31). What is the meaning of life? According to Siri, Alexa and ChatGPT. *Medium*. Disponibile su: https://medium.com/ewhye/what-is-the-meaning-of-life-according-to-siri-alexa-and-chatgpt-7a883e4aedcb. URL consultato il 9 ottobre 2025.

Leech, G. N. (2014). *The pragmatics of politeness*. Oxford University Press.

Leibniz, G. W. (1923). Dissertatio de arte combinatoria (1666). *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*, 4, 27-102.

Levinson, S. C. (1983). Pragmatics. Cambridge University Press.

Lewis, M. S. (2019). Technology change or resistance to changing institutional logics: The rise and fall of digital equipment corporation. *Journal of Applied Behavioral Science*, 55(2), 141-160.

Li, D., Jin, R., Gao, J., & Liu, Z. (2020). On sampling top-k recommendation evaluation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2114-2124.

Lim, W. M., Kumar, S., Verma, S., & Chaturvedi, R. (2022). Alexa, what do we know about conversational commerce? Insights from a systematic literature review. *Psychology & Marketing*, 39(6), 1129-1155.

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., & Martinez, A. (2019). Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, 51(4), 984-997.

Luccioni, S., Jernite, Y., & Strubell, E. (2024, june). Power hungry processing: Watts driving the cost of AI deployment?. In *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*. 85-99.

Luppicini, R. (2008). Introducing conversation design. In Handbook of conversation design for instructional applications. *IGI Global Scientific Publishing*. 1-18.

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon Books.

Mauldin, M. L. (1994). Chatterbots, Tinymuds, and the Turing Test: Entering the Loebner Prize Competition. In *AAAI* 94, 16-21.

Martinich, A. (1984). *Communication and reference*. Walter de Gruyter.

Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., & Tesconi, M. (2019). Rtbust: Exploiting temporal patterns for botnet detection on twitter. In *Proceedings of the 10th ACM conference on web science*. 183-192.

McDermott, J. (1982). R1: A rule-based configurer of computer systems. *Artificial intelligence*, 19(1), 39-88.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on Artificial Intelligence, august 31, 1955. *AI magazine*, 27(4), 12-12.

McCorduck, P., Minsky, M., Selfridge, O. G., & Simon, H. A. (1977). History of Artificial Intelligence. In *IJCAI*. 951-954.

McCorduck, P., & Cfe, C. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. AK Peters/CRC Press.

Minsky, M. (1986). *Society of mind*. Simon and Schuster.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj*, 339.

Moskvitch, K. (2019, november 21). *Augmenting Humans: IBM's Project Debater AI Helps Human Debating Teams Win*. IBM Research Blog / Medium. Disponibile su: https://ibm-research.medium.com/augmenting-humans-ibms-project-debater-ai-gives-human-debating-teams-a-hand-at-cambridge-69a29bcd4eff. URL consultato il 7 ottobre 2025.

Myers, C. M., Furqan, A., & Zhu, J. (2019). The impact of user characteristics and preferences on performance with an unfamiliar voice user interface.

In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-9.

Natale, S. (2021). *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press.

Natale, S., & Henrickson, L. (2024). The Lovelace effect: Perceptions of creativity in machines. *New Media & Society*, 26(4), 1909-1926.

Newborn, M. (2003). *Deep Blue: an artificial intelligence milestone*. Springer Science & Business Media.

Newell, A., & Simon, H. (1956). The logic theory machine: A complex information processing system. *IRE Transactions on information theory*, 2(3), 61-79.

Nilsson, N. J. (2009). *The quest for Artificial Intelligence*. Cambridge University Press.

Negnevitsky, M. (2025). The history of artificial intelligence or from the 'Dark Ages' to the knowledge-based systems. *WIT Transactions on Information and Communication Technologies*, 19.

Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems*. Pearson education.

Okrepilov, V. V., Kovalenko, B. B., Getmanova, G. V., & Turovskaj, M. S. (2022). Modern trends in artificial intelligence in the transport system. *Transportation Research Procedia*, 61, 229-233.

O'Regan, G. (2018). ELIZA Program. In *The Innovation in Computing Companion: A Compendium of Select, Pivotal Inventions*. Cham: Springer International Publishing. 119-122.

Pan, Y. (2016). Heading toward Artificial Intelligence 2.0. *Engineering*, 2(4), 409-413.

Panfili, L., Duman, S., Nave, A., Ridgeway, K. P., Eversole, N., & Sarikaya, R. (2021). Human-AI interactions through a Gricean lens. *Proceedings of the Linguistic Society of America*, 6(1), 288-302.

Parlamento europeo e Consiglio dell'Unione europea (2016, april 27). Regolamento generale (UE) 2016/679 (GDPR) sulla protezione dei dati. In *Gazzetta ufficiale dell'Unione europea*, L 119. 1–88. Disponibile su: https://www.garanteprivacy.it/documents/10160/0/Regolamento+UE+2016+679.+Arricchito+con+riferimenti+ai+Considerando+Aggiornato+alle+rettifiche+pubblicate+sulla+Gazzetta+Ufficiale++dell%27Unione+europea+127+del+23+maggio+2018. URL consultato il 9 ottobre 2025.

Pearl, C. (2016). *Designing voice user interfaces: Principles of conversational experiences*. O'Reilly Media, Inc.

Pollina, E., & Armellini, A. (2024, december 20). Italy fines OpenAI over ChatGPT Privacy rules breach. *Reuters*. Disponibile su: https://www.reuters.com/technology/italy-fines-openai-15-million-euros-over-privacy-rules-breach-2024-12-20/. URL consultato il 9 ottobre 2025.

Potthast, M., Gienapp, L., Euchner, F., Heilenkötter, N., Weidmann, N., Wachsmuth, H., & Hagen, M. (2019). Argument search: Assessing argument relevance. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1117-1120.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv*:2204.06125, 1(2), 3.

Ray, P. P. (2025). A Review on Vibe Coding: Fundamentals, State-of-the-art, Challenges and Future Directions. *Authorea Preprints*.

Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1-7.

Riffee, M. (2011, november 30). New Website Launches as Siri's Scribe. *Wired*. Disponibile su: https://www.wired.com/2011/11/new-website-siris-scribe/. URL consultato il 9 ottobre 2025.

Romero Moreno, F. (2024). Generative AI and deepfakes: a human rights approach to tackling harmful content. *International review of law, computers & technology*, 38(3), 297-326.

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. A. (2018). Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4), 1-23.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv*:1609.04747.

Salazar, J., Liang, D., Nguyen, T. Q., & Kirchhoff, K. (2019). Masked language model scoring. *arXiv preprint arXiv*:1910.14659.

SAS Institute. (2019, May 29). "What can we learn from Clippy about AI?". SAS Blogs. Disponibile su: https://blogs.sas.com/content/sgf/2019/05/29/what-can-we-learn-from-clippy-about-ai/. URL consultato il 4 ottobre 2025.

Schofield, J. (2014, June 8). "Computer chatbot 'Eugene Goostman' passes the Turing test". ZDNet. Disponibile su: https://www.zdnet.com/article/computer-chatbot-eugene-goostman-passes-the-turing-test/. URL consultato il 26 settembre 2025.

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., & Resnik, P. (2024). The prompt report: a systematic survey of prompt engineering techniques. *arXiv preprint arXiv*:2406.06608.

Searle, J. R. (2010). *Creare il mondo sociale. La struttura della civiltà umana*. Milano: Raffaello Cortina Editore.

Searle, J. R., Kiefer, F., & Bierwisch, M. (1980). *Speech Act Theory and Pragmatics*. Vol. 10, 285-293. Dordrecht: D. Reidel.

Searle, J. R. (1980). *Minds, brains, and programs*. Behavioral and Brain Sciences, 3(3), 417-424.

Searle, J. R. (1980). The Background of Meaning. In *Speech Act Theory and Pragmatics*. Dordrecht: Reidel. 221–232.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Shannon, C. E., & Weaver, W. (1998). *A Mathematical Theory of Communication*. University of Illinois Press.

Shi, W., & Demberg, V. (2019). Next sentence prediction helps implicit discourse relation classification within and across domains. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing EMNLP-IJCNLP*. 5790-5796.

Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In 2018 *Fourth international conference on computing communication control and automation* (ICCUBEA). IEEE. 1-6.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.

Singh, A. (2023). A survey of ai text-to-image and ai text-to-video generators. In 2023 *4th International Conference on Artificial Intelligence, Robotics and Control* (AIRC). IEEE. 32-36.

Slonim, N., Bilu, Y., Alzate, C., Bar-Haim, R., Bogin, B., Bonin, F., & Aharonov, R. (2021). An autonomous debating system. *Nature*, 591(7850), 379-384.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*. 2256-2265.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

Stahl, B. C., & Eke, D. (2024). The ethics of ChatGPT–Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, 74, 102700.

Stigall, B., Waycott, J., Baker, S., & Caine, K. (2019). Older adults' perception and use of voice user interfaces: a preliminary review of the computing literature. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. 423-427.

Stratton, J. (2024). *Copilot for Microsoft 365: Harness the power of generative AI in the Microsoft apps you use every day*. Springer Nature.

Sun, J., Liao, Q. V., Muller, M., Agarwal, M., Houde, S., Talamadupula, K., & Weisz, J. D. (2022). Investigating explainability of generative AI for code through scenario-based design. In *Proceedings of the 27th international conference on intelligent user interfaces*. 212-228.

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global Scientific Publishing. 242-264.

Tulshan, A. S., & Dhage, S. N. (2018). Survey on virtual assistant: Google Assistant, Siri, Cortana, Alexa. In *International symposium on signal processing and intelligent recognition systems*. Singapore: Springer Singapore. 190-201.

Turing, A. M. (1980). Computing Machinery and intelligence. Creative Computing, 6(1), 44-53.

Turkle, S. (2005). *The second self: Computers and the human spirit*. MIT Press.

Urban, M., & Mailey, S. (2019). Conversation design: principles, strategies, and practical application. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1-3.

Van Melle, W. (1978). MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *International journal of man-machine studies*, 10(3), 313-322.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7-30.

Wallace, R. S. (2007). The anatomy of ALICE. In *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer*. Dordrecht: Springer Netherlands. 181-210.

Wang, X., Wang, J., Feng, L., Niyato, D., Zhang, R., Kang, J., & Mao, S. (2025). Wireless hallucination in generative ai-enabled communications: Concepts, issues, and solutions. *arXiv preprint arXiv*:2503.06149.

Wardrip-Fruin, N. (2012). *Expressive Processing: Digital fictions, computer games, and software studies*. MIT press.

Waterman, D. A., & Waterman, D. A. (1986). *A guide to expert systems* (Vol. 419). Reading, MA: Addison-Wesley.

Watzlawick, P., Bavelas, J. B., & Jackson, D. D. (2011). *Pragmatics of human communication: A study of interactional patterns, pathologies and paradoxes*. WW Norton & Company.

Weizenbaum, J. (1976). *Computer Power and Human Reason WH Freeman and Co*. San Francisco, USA.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

Weld, H., Huang, X., Long, S., Poon, J., & Han, S. C. (2022). A survey of joint intent detection and slot filling models in natural language understanding. *ACM Computing Surveys*, 55(8), 1-38.

Werbos, P. J. (2002). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550-1560.

Ye, Q., Axmed, M., Pryzant, R., & Khani, F. (2023). Prompt engineering a prompt engineer. arXiv preprint arXiv:2311.05661.

Yeasmin, F., Das, S., & Bäckström, T. (2020). Privacy analysis of voice user interfaces. In Conference of Open Innovations Association, *FRUCT* (Vol. 6).

Yule, G. (1996). *Pragmatics*. Oxford University Press.

Zhang, S., Meng, Z., Chen, B., Yang, X., & Zhao, X. (2021). Motivation, social emotion, and the acceptance of artificial intelligence virtual assistants—Trust-based mediating effects. *Frontiers in Psychology*, 12, 728495.

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). Bias in generative AI. *arXiv preprint arXiv*:2403.02726.

Zhou, K. Q., & Nabus, H. (2023). The ethical implications of DALL-E: Opportunities and challenges. *Mesopotamian Journal of Computer Science*, 2023, 16-21.

Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired*. Routledge. 203-213.

RINGRAZIAMENTI

Grazie, Mami. Grazie, Papi. Grazie, Sis.

Grazie, Michi. *Grazie Dada. Grazie Lily.*

Grazie Noe, Aisha, Ali, Giulia, Agne, Vale, Yas.

Grazie Prof. Cioni, Prof.ssa Baldi, Prof. Simonetta.

Grazie, Puglia.

Grazie, Francia.

E grazie, grazie Firenze.

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020, risorse FSE REACT-EU
Azione IV.4 “Dottorati e contratti di ricerca su tematiche dell’innovazione”
e Azione IV.5 “Dottorati su tematiche Green”