



UNIVERSITÀ DELLA
VALLE D'AOSTA
UNIVERSITÉ DE LA
VALLÉE D'AOSTE

DIPARTIMENTO DI SCIENZE UMANE E SOCIALI
CORSO DI LAUREA IN SCIENZE E TECNICHE PSICOLOGICHE

ANNO ACCADEMICO 2023/2024

TESI DI LAUREA

**INTELLIGENZA ARTIFICIALE E PSICOLOGIA: NUOVE FRONTIERE NEL
CAMPO DEI TERAPEUTI DIGITALI E VIRTUALI**

RELATORE:

Prof. Angelo Benozzo

LAUREANDO: 18D03045

Alex Glesaz

*Il futuro dell'Intelligenza artificiale e della psicologia è collaborativo.
Lavorando insieme, i ricercatori di AI e gli psicologi possono sviluppare
nuovi strumenti e tecniche per migliorare il benessere umano.*

Gary Marcus

INDICE

INDICE	1
INTRODUZIONE	2
CAPITOLO 1 . INTELLIGENZA ARTIFICIALE	4
1.1 Primi sviluppi: Turing e la macchina pensante	4
1.2 Intelligenza	6
1.3 AI ed etica: macchine ingannevoli	8
1.4 Chatbot: machine learning e reti neurali	12
CAPITOLO 2 . RAPPORTO PERSONA-IA E PSICOLOGIA	16
2.1 Emozioni e intelligenza artificiale.....	16
2.2 AI e psicologia: effetto Eliza.....	22
CAPITOLO 3 . L'INTEGRAZIONE DELL'INTELLIGENZA ARTIFICIALE NELLE TERAPIE PSICOLOGICHE.....	28
3.1 Sviluppo e funzionamento dei Terapeuti Digitali e Virtuali	28
3.2 Vantaggi dell'uso dei terapeuti digitali e virtuali nel trattamento psicologico	34
3.2.1 Diminuzione ansia e depressione.....	35
3.2.2 Stress e resilienza.....	37
3.2.3 Disturbi alimentari	38
3.3 Sfide e limitazioni nell'utilizzo dell'IA nell'intervento psicologico	39
3.3.1 L'alleanza terapeutica	39
3.3.2 Empatia	41
3.3.3 Il concetto di Transfert.....	42
3.3.4 Problemi etici.....	43
CONCLUSIONI.....	46
BIBLIOGRAFIA.....	48

INTRODUZIONE

L'avvento dell'Intelligenza Artificiale (IA) ha rivoluzionato numerosi settori della società contemporanea, tra cui la psicologia, aprendo nuove prospettive e sfide nel rapporto tra la mente umana e la tecnologia avanzata. Questo fenomeno segna un cambiamento paradigmatico verso l'umanesimo digitale, un concetto che riflette l'interazione sempre più intima tra gli esseri umani e le tecnologie intelligenti (Catarci, 2022).

Film futuristici come "Io Robot"¹ anticipano un mondo in cui gli umani convivono quotidianamente con compagni robotici, una visione che oggi sembra parallela alla realtà in cui conversazioni con chatbot possono sviluppare legami affettivi significativi. L'uscita della prima versione di ChatGPT il 30 novembre 2022 ha segnato ufficialmente l'ingresso nell'era degli agenti conversazionali trasformativi, o chatbot, influenzando profondamente il modo in cui percepiamo e interagiamo con la tecnologia nel quotidiano (Frasso, 2024).

Questa tesi si prefigge l'obiettivo di esplorare e illuminare il panorama dell'IA applicata alla psicologia, focalizzandosi principalmente sull'integrazione dei terapeuti digitali e virtuali nei contesti terapeutici contemporanei. Il primo capitolo introduce i fondamenti teorici dell'Intelligenza Artificiale, partendo dalle origini con Alan Turing e la concezione della macchina pensante, fino alle attuali applicazioni basate su machine learning e reti neurali, che confermeranno le ipotesi del matematico. Questo capitolo fornisce una base concettuale essenziale per comprendere il potenziale e le implicazioni dell'IA nel campo della psicologia, inclusi i dibattiti sull'etica delle macchine ingannevoli.

Nel secondo capitolo, verrà esaminato il rapporto complesso tra l'uomo e l'IA in ambito psicologico, concentrandosi sulle emozioni e sull'influenza dell'IA, esplorando anche l'effetto Eliza come un precursore nella simulazione di conversazioni umane. Questa sezione approfondirà come l'IA possa modellare e gestire le emozioni umane, sfidando le concezioni tradizionali di interazione sociale e psicologica.

Il terzo capitolo delinea l'integrazione dell'IA nelle terapie psicologiche, esaminando lo sviluppo e il funzionamento dei terapeuti digitali e virtuali. Attraverso evidenze scientifiche, si esplorerà l'effetto dell'IA nel trattamento e nella prevenzione di condizioni come ansia, depressione, stress e disturbi alimentari e si discuteranno anche le sfide e le limitazioni nell'utilizzo dell'IA nell'intervento psicologico. Argomenti cruciali saranno l'importanza

¹ Laurence, M., Davis J., Topher, D., Wyck, G. (Produttori), & Proyas, A. (Regista). 2004. *I, Robot* [Film]. USA: 20th Century Fox

dell'alleanza terapeutica, l'empatia nella relazione umano-macchina, il concetto di transfert e le questioni etiche connesse all'interazione tra tecnologia e salute mentale.

Questa tesi mira a fornire una panoramica critica e informativa su come l'Intelligenza Artificiale stia trasformando il campo della psicologia, delineando le opportunità emergenti e le considerazioni etiche fondamentali che devono essere affrontate e considerate nel futuro sviluppo di terapie digitali avanzate. Nell'indagare ciò, sarà costante premura mostrare come la risposta non può essere univoca ma si dovrà tenere in considerazione la necessità di un approccio qualitativo, multifattoriale e multisistemico.

Capitolo 1. INTELLIGENZA ARTIFICIALE

1.1 Primi sviluppi: Turing e la macchina pensante

L'intelligenza artificiale così come viene conosciuta oggi è il prodotto di un'evoluzione costante in ambito informatico, che affonda le sue radici nei primi anni Cinquanta. Tra i pionieri che hanno gettato le basi viene ricordato Alan Turing, un matematico inglese, spesso citato come uno dei padri dell'informatica moderna. A lui si deve infatti il merito di aver attirato per la prima volta l'attenzione dell'intera comunità sul tema dell'intelligenza artificiale (Marty, 2023). Nel suo articolo "Macchine calcolatrici e intelligenza"², Turing pone un interrogativo filosofico e audace per quell'epoca, in quanto si focalizza sulla questione se le macchine possano pensare. È rilevante per questa discussione prendere in esame l'obiezione che il matematico pone nei confronti di una posizione forte dell'epoca, secondo la quale la macchina non può fare mai veramente qualcosa di nuovo. Seguendo l'analogia della pelle della cipolla, egli induce il pubblico a riflettere in merito alla mente umana: le funzioni mentali non corrispondono alla mente reale così com'è ma a una specie di pelle che bisogna togliere. Scrive:

«Ma poi in ciò che rimane troviamo un'altra pelle da togliere, e così via. Procedendo a questo modo arriviamo infine alla mente "reale", o solo a una pelle che non contiene nulla? Nell'ultimo caso tutta la mente sarebbe di tipo meccanico» (Turing, 1950, pp. 177).

E ancora:

«Chi può essere sicuro che il "lavoro originale" da lui compiuto non sia stato semplicemente la crescita di un seme gettato dall'insegnamento, o la conseguenza dell'aver seguito principi generali bene noti?» (Turing, 1950, pp. 173).

Emerge quindi la possibilità che il funzionamento alla base dei processi mentali possa essere in realtà per certi versi analogo a quello della macchina, per via della propria meccanicità. Risulta inoltre molto interessante come il matematico riconduca la creazione di un'idea a un seme che, così come si sviluppa nella mente umana, nello stesso modo può agire in una macchina. Questo seme non è altro che l'interazione tra i fattori biologici e l'educazione che una persona riceve. Turing con queste speculazioni provocative non esclude assolutamente la possibilità che la macchina un giorno possa arrivare a pensare come una mente umana. Egli considera infatti il "pensare" come la capacità di partecipare a quello che chiama il "gioco dell'imitazione", un test di comunicazione in cui l'interlocutore umano può fare qualsiasi domanda a una macchina

² Titolo originale "*Computing machinery and intelligence*", articolo pubblicato sulla rivista Mind nel 1950

che deve essere programmata per simulare di essere un'altra persona. Lo scambio dialogico avviene attraverso una telescrivente e se il computer riuscisse a ingannare in qualche modo l'umano, convincendolo di stare parlando con una persona vera, allora si potrebbe affermare che la macchina ha qualche forma di intelligenza e che sta realmente pensando. Questo procedimento viene conosciuto oggi come il “Test di Turing”, il cui scopo sarebbe quello di valutare le capacità dell'intelligenza artificiale. A quell'epoca, non c'erano ancora sistemi in grado di superare il test ma, nonostante ciò, Turing confidava nella programmazione, ritenendo che si potesse creare un programma capace di simulare la mente di un bambino. Successivamente, la macchina sarebbe stata istruita adeguatamente, sviluppando così un cervello adulto (Turing, 1950).

L'intuizione di Turing sembra anticipare la traiettoria dello sviluppo delle moderne IA, nello specifico nel campo relativo ai moderni modelli di linguaggio. Conosciuti come “chatbot”, essi sono dei software avanzati capaci di generare un testo di risposta in una conversazione, che risulta indistinguibile da quello naturale del linguaggio umano. Questa sua qualità risulta possibile grazie all'abilità di imparare dall'esperienza contestuale e generare delle risposte appropriate secondo un vasto raggio di richieste (Marty, 2023, pp. 16).

Sembra quindi che il pensiero del matematico si sia realizzato nella creazione di macchine in grado di comunicare con un essere umano proprio come farebbe un'altra persona. Infatti, oggi la maggior parte dei modelli di linguaggio è in grado di superare brillantemente il test di Turing. Tuttavia, questo risultato è anche oggetto di critiche, poiché il test è frequentemente contestato per il fatto che non valuta la vera intelligenza della macchina, ma solo la sua capacità di imitare quella umana (Marty, 2023, pp. 6). A tal riguardo Suleyman³ si esprime con le seguenti parole:

«Non ci dice nulla su ciò che il sistema [IA] può fare o capire, nulla sul fatto che abbia stabilito complessi monologhi interiori o possa impegnarsi nella pianificazione su orizzonti temporali astratti, che è la chiave dell'intelligenza umana» (Suleyman, 2023).

D'altro canto, però, l'incertezza del test può essere superata adottando un diverso punto di vista da cui guardarlo. Esso acquista infatti senso e valore se non lo si usa come una valutazione dell'esistenza di macchine pensanti, ma come una misura delle reazioni umane alle comunicazioni con macchine che manifestano un comportamento intelligente. Questa visione sposterebbe l'attenzione dalle prestazioni informatiche sull'essere umano, il quale contribuisce attivamente a modellare qualunque interazione comunicativa con un'IA (Natale, 2022).

³ Mustafa Suleyman, co-fondatore della startup specializzata in intelligenza artificiale DeepMind

Già allora, Turing sosteneva che il vero problema non fosse determinare se le macchine potessero pensare, ma piuttosto capire come l'essere umano le percepisce e se è disposto a considerarle intelligenti (Natale, 2022). Turing scrive:

«credo che alla fine del secolo l'uso delle parole e l'opinione corrente si saranno talmente mutate che chiunque potrà parlare di macchine pensanti senza aspettarsi di essere contraddetto» (Turing, 1950, pp. 165).

Questo modo di ragionare getta luce non tanto sull'abilità della macchina di "pensare" ma sulla concezione socioculturale delle macchine che permette di considerarle come abili nel pensare. Da ciò ne deriva che l'immagine e le capacità dell'intelligenza artificiale vengono costruite attraverso le idee culturali che vengono veicolate all'interno della società, attraverso l'uso delle parole e i discorsi. In altre parole, potrebbe non essere così utile chiedersi se si è riusciti a rispondere alla domanda "la macchina può pensare?", quanto piuttosto esplorare il cambiamento dell'atteggiamento culturale verso l'esperienza umana con queste tecnologie.

Questa evoluzione delle credenze sociali emerge negli studi della sociologa e la psicologa Sherry Turkle, che ha messo in luce come in due decenni la concezione delle persone riguardo a una psicoterapia per mezzo dell'IA sia radicalmente cambiata. Nello studio, infatti, si passa dal riscontrare una forte avversione dei partecipanti nei confronti di interventi psicologici gestiti da chatbot a un atteggiamento invece aperto e permissivo (Turkle, 1995, in Natale, 2002).

Quanto detto supera la concezione di un'IA capace di pensare a favore di un'IA come fenomeno relazionale, in cui è necessario tener conto degli attori che partecipano allo scambio comunicativo e al luogo dove questa comunicazione avviene.

1.2 Intelligenza

È interessante notare come il test di Turing venga frequentemente presentato come una strategia per valutare l'intelligenza delle macchine, nonostante Turing stesso avesse rinunciato a proporre una definizione generale e consolidata di intelligenza, preferendo concentrarsi invece su quanto esse sembrano intelligenti. Nel suo saggio, infatti, Turing concluse che sarebbe impossibile raggiungere un consenso universale su cosa significhi pensare e quindi avere qualche forma di intelligenza (Natale, 2022).

Turing, infatti, non distingueva tra la capacità di una macchina di pensare e il possesso di intelligenza, basando la sua intuizione sulla teorizzazione di un concetto allora inesistente: l'intelligenza artificiale. Secondo il matematico, era sufficiente che la macchina riuscisse a imitare il pensiero umano nel gioco dell'imitazione per attribuirle una qualche forma di

intelligenza. Questa visione viene spesso criticata perché non considera aspetti fondamentali della mente umana, come la coscienza e l'intenzionalità (Taglienti, 2016).

Molti vedono invece il test di Turing più come una mera provocazione filosofica, ritenendolo concepito solamente per verificare quella che viene chiamata "intelligenza artificiale debole"⁴. Tuttavia, nell'articolo di Turing si può trovare un'allusione al fatto che le macchine potrebbero un giorno raggiungere lo stesso livello dell'intelligenza umana⁵. Egli non esclude la possibilità che, proprio come un reattore a fissione non funziona sotto una certa dimensione critica, anche i cervelli e le macchine potrebbero aver bisogno di superare la soglia critica per sprigionare tutto il potenziale; la complessità in crescita della macchina potrebbe continuare ad evolvere fino a scenari sorprendenti (Turing, 1950).

Seppur non si possa sapere con esattezza se la macchina possa un giorno acquisire la stessa coscienza di colui che l'ha creata, qualcosa però è diventato evidente: negli anni il legame costante tra il termine "artificiale" e "intelligenza" ha creato un'associazione così consolidata che nessuno si avventurerebbe a separarli.

Tuttavia, andando per logica, l'utilizzo del concetto di intelligenza dovrebbe presupporre quantomeno l'esistenza di un'idea più o meno comune di cosa significhi essere pensante o intelligente. Una delle concezioni più rinomate e rivoluzionarie, ovvero la cosiddetta Teoria delle Intelligenze Multiple di Howard Gardner, definisce l'intelligenza come una capacità psicobiologica di affrontare le sfide socio-culturali che il mondo circostante offre costantemente nella vita quotidiana delle persone (Gardner, 1983). Questa definizione riveste una grande importanza perché sfida il paradigma che considera l'intelligenza come una capacità unitaria comune a tutti gli individui, categorizzabile in modo assoluto. In contrasto con la visione assolutista, secondo Gardner, ogni individuo possiede invece molteplici forme di intelligenza, varie modalità di interazione per connettersi profondamente con il mondo circostante. Questo avviene attraverso una sorta di "cassetta degli attrezzi" che ognuno di noi ha a disposizione, ma che spesso rimane sigillata fino a quando non viene aperta (Romeo, 2021).

Risulta inoltre utile prendere in considerazione la definizione di intelligenza fornita dalla Treccani⁶, che la descrive come:

«Complesso di facoltà psichiche e mentali che consentono all'uomo di pensare, comprendere o spiegare i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, e lo rendono insieme capace di adattarsi a situazioni

⁴ Si riferisce alla capacità imitativa delle macchine non ad una intelligenza simile a quella umana

⁵ Questa viene definita "intelligenza artificiale forte"

⁶ <https://treccani.it/vocabolario/intelligenza/>

nuove e di modificare la situazione stessa quando questa presenta ostacoli all'adattamento; propria dell'uomo, in cui si sviluppa gradualmente a partire dall'infanzia e in cui è accompagnata dalla consapevolezza e dall'autoconsapevolezza, è riconosciuta anche, entro certi limiti agli animali, specialmente mammiferi»

Questa definizione si concentra prevalentemente sull'essere umano, senza negare però anche agli animali un certo livello di intelligenza. Considerando il quadro più ampio in cui interagiscono uomini e animali, si potrebbe caratterizzare come "intelligente" la capacità di conseguire determinati scopi o traguardi specifici. Tuttavia, quelli che sono gli obiettivi principali per un animale non necessariamente sono quelli di un essere umano. Lo stesso principio può essere applicato anche quando si valuta l'intelligenza delle macchine: il rischio è infatti quello di definirle intelligenti solo se riescono a conseguire gli stessi obiettivi degli esseri viventi, seppur ciò che consideriamo come obiettivo principale per le macchine potrebbe non essere prioritario per gli esseri viventi.

Alla luce di queste riflessioni, pare necessario acquisire coscienza di come il concetto di intelligenza risulti privo di significato se considerato al di fuori del contesto sociale in cui si inserisce. Sembra che ciò conduca all'impossibilità di fornire una definizione univoca dell'intelligenza artificiale, soprattutto quando vengono fatti confronti esclusivamente con l'intelligenza umana. Di conseguenza, la soluzione migliore sembra restare quella proposta dal matematico Alan Turing, che collega la capacità di pensare allo status di una macchina capace di "ingannare" il suo interlocutore.

1.3 AI ed etica: macchine ingannevoli

L'implicazione che uno sviluppatore di intelligenza artificiale possa utilizzare le proprie competenze per trarre in inganno gli utenti suscita spesso preoccupazione. Inoltre, le medesime comunità degli informatici che lavorano sull'IA sono restie a etichettare il loro lavoro come ingannevole, anche a causa delle connotazioni negative associate al termine (Natale, 2022). Consultando la sua definizione più comune, il vocabolo viene infatti definito come “parvenza di vero creata ad arte per far cader in errore qualcuno”⁷, implicando sempre che colui che viene ingannato assume un'impressione erronea e fuorviante rispetto a ciò che è giusto e reale. Interpretare lo sviluppo dell'IA secondo quest'ottica, suggerisce inevitabilmente che ogni volta che un utente interagisce con un'interfaccia digitale, si ritrovi imprigionato in una sorta di

⁷ https://dizionari.corriere.it/dizionario_italiano/I/inganno.shtml

tranello. Ciò può aumentare l'illusione di concepire gli utenti come “schiavi” di macchine definite come ingannevoli ed intelligenti, spostando tutta l'attenzione sulle caratteristiche tecniche di queste ultime. In questo modo, il rischio sarebbe quello di trascurare la volontà e i processi cognitivi dell'essere umano, passivizzandolo.

Invece, secondo la psicologia, l'inganno e l'illusione costituiscono fenomeni fisiologici naturali insiti nell'esperienza umana. Ciò è dovuto al fatto che le dinamiche che portano agli errori percettivi sono fondamentali per la comprensione di come gli esseri umani interpretino il mondo che li circonda (Pettit, 2013 in Natale, 2022).

L'inganno si configura quindi come un processo relazionale e unico in un determinato contesto spazio-temporale, nel quale l'interlocutore umano partecipa attivamente e collabora.

E' possibile pensare alle forme di inganno delle IA come più sottili e apparentemente inoffensive, definibili con l'espressione “inganno banale”. Banale perché riguarda situazioni che fanno parte della nostra vita di tutti i giorni, che non sono facilmente riconoscibili come diverse dalla nostra esperienza abituale (Natale, 2022).

Pertanto, i creatori di intelligenza artificiale sfruttano i preconcetti già presenti, inclusi elementi identitari come etnia, classe e genere, che alimentano la fantasia. Ad esempio, la decisione di utilizzare una determinata voce priva di corpo per un assistente digitale potrebbe innescare automaticamente processi di proiezione da parte dell'utente, che potrebbe attribuire un genere e persino uno status etnico e sociale alla voce (Natale, 2022). Così come nello stesso modo la scelta di far comparire le parole una alla volta da parte di ChatGPT come se dall'altra parte dello schermo ci fosse una persona che sta digitando non è casuale (Columbro, 2023).

Questo concetto di “inganno banale” si fonda quindi sulla constatazione che la conoscenza non solo emerge dall'esplorazione della realtà circostante ma anche dall'applicazione di etichette stereotipiche già prefabbricate. In questo processo sottile, gli stereotipi diventano cruciali perché aiutano l'utente ad attribuire un'identità unitaria e coerente all'intelligenza artificiale con cui interagisce, associandole quindi un genere, un'etnia e una classe. In questo scenario di proiezioni, diventa chiaro come l'intelligenza artificiale conceda un'ampia libertà di interpretazione soggettiva al suo interlocutore umano (Natale, 2022).

Questa visione modifica la prospettiva sull'inganno all'interno del rapporto tra un'IA e l'utente: quest'ultimo non viene ingannato pensando di parlare con un individuo reale; è invece lui stesso a plasmare un'entità virtuale in una figura incarnata. Questo gioco dell'immaginazione sembra richiamare un concetto sviluppato molti anni prima dell'avvento delle IA dal sociologo Marshal McLuhan, che spiegò l'effetto della tecnologia sull'individuo come un'estensione del corpo umano. Egli intendeva i media come un'estensione dei propri sensi, nonché del sistema nervoso.

Tuttavia, questa teoria può essere applicata al settore dell'intelligenza artificiale. In base a questa chiave di lettura, si possono considerare le informazioni provenienti da qualsiasi prodotto dell'IA sia come un'estensione dell'essere umano e delle sue esperienze personali, sia come un'espansione delle proprie capacità cognitive e della traccia mnemonica. Tuttavia, queste informazioni sono anche un'espressione dell'IA come mezzo distintivo, che crea qualcosa di unico e originale rispetto ad altri media (Rozzo, 2023).

La conseguenza di ciò è la possibilità di creazione di relazioni molto differenti e uniche con le IA generative⁸. Ad esempio, al primo incontro con una chatbot, ogni persona svilupperà una reazione e un effetto particolare, che sarà sicuramente influenzato dalle funzionalità tecniche del programma, quali il tono di voce, il grado di personalizzazione, l'avatar e le capacità riflessive. Allo stesso tempo, quel particolare modello di linguaggio incarna e rispecchierà le caratteristiche personali dell'utente come i suoi bisogni, ricordi ed aspettative.

Tuttavia, questo aspetto può diventare un segnale d'allarme laddove si utilizzi l'intelligenza artificiale come supporto emotivo. Infatti, numerose testimonianze indicano come un numero crescente di utenti si rivolga alle IA conversazionali per affrontare problemi personali, simili a quelli che discuterebbero con uno psicologo (Columbro, 2023). Il rischio potrebbe essere contenuto proprio nel fatto che la tecnologia amplia le nostre risorse divenendo un'estensione delle nostre capacità cognitive. Ciò implica che una persona creativa avrà come effetto quello di un'estensione immensa delle proprie capacità espressive (Rozzo, 2023). D'altro canto, una persona angosciata potrà trovare un'amplificazione della propria esperienza personale dolorosa, che verrà restituita all'utente senza una vera e propria riflessione sul disagio. Questa conseguenza critica andrebbe nuovamente a richiamare il teorico dei mass media McLuhan che mette in guardia sul fatto che ogni estensione è sempre anche un'amputazione. Egli riprende Narciso: come egli precipita nell'acqua a causa del suo innamoramento per il proprio riflesso, così ogni essere umano deve stordire i propri sensi a causa dell'attrazione per ogni estensione di sé (Dara, 2017).

In ogni caso si tratta di supposizioni critiche ed analitiche che non vanno a confutare l'unicità della relazione che si crea tra l'utente e l'IA, ma possono aiutare a far luce rispetto a potenziali rischi in determinati casi di interazione.

Di conseguenza, ciò che sembra legittimo è il provare a sospendere il giudizio nei confronti di una possibile percezione umana dell'IA che sembra giusta o sbagliata, ricordando infatti che

⁸ Sono i modelli di linguaggio o chatbot come chatGPT

l'errore è parte integrante del nostro modo di esplorare la realtà al di fuori (Pettit, 2013, citato in Natale, 2022).

La mentalità che guida la programmazione basata su stereotipi culturali ha suscitato delle controversie. Le critiche si sono concentrate sulla tendenza predominante di caratterizzare gli assistenti vocali con voci femminili come impostazione predefinita.

«Ciò è particolarmente preoccupante se si considera la caratterizzazione degli assistenti come servitori docili, che riproduce rappresentazioni di genere stereotipate associate a un lavoro ancillare» (Natale, 2022).

Si capisce come gli utenti, interagendo con gli assistenti IA, possano rispondere a queste voci proiettando stereotipi comunemente attribuiti agli esseri umani. Alexa, ad esempio, il cui tono vocale suggerisce una donna bianca, istruita e madrelingua inglese, incarna appieno lo stereotipo della donna orientata al lavoro domestico (Natale, 2022).

Si alimenta così il dibattito secondo cui tutto ciò andrebbe a rafforzare gli stereotipi di genere già presenti nel contesto socioculturale, attraverso il mantenimento di una figura di genere femminile subordinata e servile. Per tale motivo, aziende come Apple e Google hanno introdotto opzioni di personalizzazione del genere e della varietà di accenti in risposta alle polemiche sul sessismo (Natale, 2022).

Queste implementazioni potrebbero rappresentare per molti una soluzione ma per certi versi si potrebbe anche effettuare una lettura tra le righe. Questa necessità, infatti, sottolinea come l'influenza delle interfacce vocali sugli utenti è evidente. Ciò sembra quindi celare un potere manipolatorio, capace di indirizzare le proiezioni degli utenti verso tratti demografici e di personalità specifici. Tuttavia, per citare le parole dell'autore Simone Natale:

«si tratta in definitiva di un potere “morbido”, indiretto, in cui l'attribuzione della personalità è delegata al gioco dell'immaginazione degli utenti» (Natale, 2022, pp. 178-179).

Ma poiché questa influenza agisce attraverso un sottile inganno che spesso non viene percepito, in quanto si confonde nella routine di tutti i giorni, potrebbe essere utilizzata in futuro come strumento di persuasione. Questo scopo implicito non può che suscitare apprensione, data la connotazione negativa che la persuasione spesso assume, poiché può avere un intento manipolatorio e, dato che la manipolazione non è etica, la persuasione viene percepita come non etica e viene considerata di conseguenza da evitare (Nardone & Chiodini, 2022).

In realtà, essa può essere interpretata come un'interazione dialogica volta al mutuo arricchimento e alla reciproca scoperta (Nardone & Chiodini, 2022).

Chiaramente, tutto dipenderebbe dall'uso che si decide di fare delle potenzialità o dei rischi legati all'intelligenza artificiale. Tuttavia, stanno emergendo possibilità riguardo a modelli di linguaggio, come delle chatbot avanzate che potrebbero essere concepite con l'intento di persuadere per favorire un cambiamento e un miglioramento. Dopotutto, la suggestione è utilizzata come tecnica comunicativa nella psicoterapia strategica⁹, poiché coinvolge profondamente il soggetto, anche emotivamente. La persuasione diventa quindi il metodo più diretto ed efficace per incoraggiare la persona a modificare, correggere o rafforzare un atteggiamento (Isp, 2018). Risulta pertanto necessario verificare se l'intelligenza artificiale generativa moderna sia riuscita ad emulare un processo simile, ricordando anche i potenziali rischi legati ad una amplificazione dei nostri sensi. Ancora prima però in questa ricerca, è essenziale analizzare come la macchina apprende e memorizza informazioni, ovvero attraverso il machine learning.

1.4 Chatbot: machine learning e reti neurali

L'avvento del machine learning determina una significativa rivoluzione nel campo dei sistemi artificiali che cercano di simulare il linguaggio umano. È proprio grazie allo sviluppo di questa tecnologia che l'utente può essere convinto sempre maggiormente dell'abilità della macchina del poter pensare e interagire.¹⁰

Definibile come una sotto disciplina dell'intelligenza artificiale, il machine learning ha permesso lo sviluppo di algoritmi di elaborazione naturale del linguaggio sempre più sofisticati, i cosiddetti NLP¹¹. I moderni modelli di linguaggio (o *chatterbot*) si sono infatti evoluti repentinamente, passando da metodi simbolici a tecniche di *deep learning*. Quest'ultima sostanzialmente è un'implementazione del machine learning. La novità risiede nella sua caratteristica di prendere ispirazione dalla struttura e dalle funzioni del cervello umano, creando dei sistemi che sfruttano delle reti neurali artificiali. In termini più concreti, il testo che prima veniva elaborato da un'architettura fissa di regole logiche e sintattiche, ora viene riprodotto sulla base dell'apprendimento di milioni di sequenze linguistiche. Si può pensare ad un bambino che impara la sua lingua madre in modo costruttivo ed esperienziale, attraverso l'ascolto e la riproduzione vocale. In modo analogo, un'intelligenza artificiale basata su un'organizzazione

⁹ Tipo di approccio psicoterapeutico che impiega metodi come la suggestione, metafore, paradossi che mira a rompere quei modelli rigidi e consolidati del sistema percettivo-reattivo che creano il comportamento disfunzionale.

¹⁰ Da notare che l'attenzione è spostata su quanto una macchina può far credere di saper pensare e non sul poter pensare della macchina

¹¹ In inglese Natural Language Processing

di reti profonde neurali non impara a gestire le conversazioni sulla base di istruzioni rigide già programmate, ma attraverso l'elaborazione spontanea di espressioni verbali umano-simili, apprese dalla propria esperienza (Marty, 2023).

In modo sorprendente, alcune iniziative come il Blue Brain Project¹² hanno l'obiettivo di creare un cervello virtuale, mappando accuratamente tutte le reti neurali del cervello dei mammiferi in un modello biologicamente accurato, utilizzando la tecnologia chiamata NEURON. (Stan, 2021, pp. 41-42). Circuiti come il *Neurogrid* sono già in grado di simulare 1 milione di neuroni e 6 miliardi di sinapsi (Abate, 2014). Il corso attuale dello sviluppo sembra dirigere verso la completa ricostruzione della mappatura neurale del cervello umano. Lo Human Brain Project (HBP) ha proprio questo obiettivo principale (Stan, 2021, pp. 44).

In aggiunta a ciò, questo sviluppo nell'ambito informatico ha portato alla nascita di un potente e versatile modello di linguaggio, oggi molto conosciuto: ChatGPT¹³. La sua architettura si basa su un modello di deep learning chiamato *Transformer*, che usufruisce di profonde reti neurali. Questi “neuroni artificiali” vengono addestrati a predire la prossima parola nella frase, tra quelle già apprese e a pesare l'importanza di esse in un testo. Ciò consente di comprendere il contesto di un discorso e di afferrare le sfumature di significato tipiche del linguaggio umano, invece di analizzare la frase parola per parola (Marty, 2023).

Per fare un esempio di questa capacità, di seguito viene riportata l'elaborazione di ChatGPT rispetto ad un discorso motivazionale:

Ogni giorno ci troviamo di fronte a sfide, ostacoli e incertezze. La vita può sembrare un percorso pieno di difficoltà e imprevisti, ma è proprio in quei momenti di difficoltà che troviamo la nostra forza interiore. Guardati intorno: sei circondato da potenziale. Dentro di te c'è una forza che può superare qualsiasi ostacolo, una determinazione che può trasformare i tuoi sogni in realtà. Non lasciare che la paura o l'incertezza ti trattengano. Abbraccia le sfide come opportunità di crescita e trasformazione. Ogni caduta è un'opportunità per imparare, crescere e diventare più forte di prima. Non smettere mai di credere in te stesso, nelle tue capacità e nel tuo potenziale illimitato. Ricorda che sei il creatore della tua vita. Puoi plasmare il tuo destino con ogni scelta che fai e ogni azione che intraprendi. Sii audace, sii coraggioso, sii determinato. Quindi oggi, ti sfido a guardare dentro te stesso e trovare quella scintilla di passione e determinazione. Affronta

¹² Avviato nel maggio 2005 da il Brain and Mind Institute de l'École Polytechnique Fédérale de Lausanne (EPFL) in Svizzera

¹³ Chatbot sviluppato da OpenAi, prima versione 30 novembre 2022. Acronimo di Chat Generative Pre-trained Transformer tradotto “trasformatore generativo pre-addestrato”

ogni giorno con gratitudine, coraggio e determinazione. Se credi in te stesso e non ti arrendi mai, non c'è limite a ciò che puoi realizzare. Sii il protagonista della tua vita, sii il capitano della tua nave. Vola alto, sogna in grande, e ricorda sempre che sei più forte di quanto pensi. Il mondo è tuo per conquistarlo. Sii audace, sii straordinario, sii tu stesso. Il mondo ti aspetta.

Riguardo a questa generazione testuale, si osserva come le molte parole presenti nel vocabolario del programma abbiano generato frasi articolate in maniera fluida e naturale, simili a quelle di una testo scritto da una persona in carne e ossa. Ciò che rende spontaneo il suo carattere non è solo la sua abilità linguistica, ma anche la capacità di simulare l'aspetto emotivo-affettivo che il linguaggio umano porta con sé¹⁴. Sembra infatti che vi sia una certa capacità di stabilire una connessione emotiva con l'interlocutore, utilizzando i bisogni e le necessità dell'altro come punto di riferimento, con lo sviluppo del suo potenziale come fine ultimo. Leggendo questo estratto, sembra difficile mettere in dubbio che sia frutto delle capacità linguistiche e immaginative di un essere umano; chiunque rimarrebbe sorpreso nello scoprire che in realtà l'autore è una macchina.

Oltre al concetto delle reti neurali trasformative, questa capacità è inoltre possibile per merito di un'altra tecnologia alla base di ChatGPT: l'apprendimento non supervisionato. La nozione centrale di questo meccanismo può essere colta attraverso la seguente metafora:

«Esso [ChatGPT] è più come un curioso studente lasciato andare in una biblioteca non guidata, con l'unico incarico di identificare gli schemi, le strutture e i temi nei libri che incontra» (Marty, 2023, pp. 26).

La competenza coinvolta in questo processo risiede nell'abilità del modello di identificare i modelli intrinseci dei dati senza bisogno di una guida esplicita. È sufficiente istruirlo con un vasto corpus di testi provenienti da internet. Il modello sviluppa autonomamente la sua capacità, esplorando a fondo la complessità implicita del linguaggio e migliorando la sua capacità di predire la parola successiva in una frase basandosi sul contesto (Marty, 2023).

Diventa incontrovertibile che si è davanti ad una vera e propria rivoluzione nel campo della comunicazione tra umano e macchina. Ciò che prima appariva solamente come un'interazione unidirezionale, è diventato per contro bidirezionale, dove alla persona non viene restituito solo ed esclusivamente il rumore dato dalla tastiera su cui digita, ma al contrario riceve una risposta, che dà vita a un autentico dialogo. Ciò che all'epoca di Turing sembrava un miraggio, con lo sviluppo del moderno machine learning ha potuto concretizzarsi, lasciando spazio a

¹⁴ Questo aspetto verrà trattato più avanti, nel capitolo 2

innovazione e progresso ma anche ulteriori interrogativi. Sicuramente sarebbe irriverente affermare che l'intelligenza artificiale più nello specifico i modelli di linguaggio siano riusciti a eguagliare le abilità intrinseche del cervello umano. Ma sarebbe opportuno considerare che sono riusciti quantomeno ad imitarne il funzionamento strutturale alla base, sviluppando una ristretta ma molto dissimulante abilità creativa. D'altronde, secondo Turing era la capacità fraudolenta della macchina a farne un'entità pensante e se si definisce rilevante anche solo in minima parte il fatto che il racconto sopracitato sia riuscito a ingannare il lettore, in tal caso si è sulla strada giusta.

Capitolo 2. RAPPORTO PERSONA-IA E PSICOLOGIA

2.1 Emozioni e intelligenza artificiale

Quando si parla delle competenze comunicative di un modello di linguaggio durante un dialogo tra l'umano e la macchina, è inevitabile considerare un elemento essenziale e imprescindibile della natura umana: la sua capacità di sperimentare emozioni. Secondo la teoria di Gardner, questa capacità si trasforma in una forma di intelligenza indipendente denominata "intelligenza interpersonale", che consiste nell'essere in grado di comprendere gli altri e le loro emozioni, stabilendo legami affettivi attraverso l'empatia (Romeo, 2021).

A questo punto, sorge la domanda se sia possibile integrare le emozioni nell'intelligenza artificiale. Tuttavia, il dibattito più intenso riguarda il confronto tra le emozioni artificiali e quelle umane: «Le emozioni nelle macchine sono molto diverse da quelle umane, poiché sono simulate attraverso algoritmi e modelli matematici. Le emozioni artificiali rappresentano una simulazione delle emozioni umane» (Tormen, 2023, pp. 49).

Questa simulazione può causare confusioni, poiché le macchine sembrano non provare effettivamente emozioni (Tormen, 2023, pp. 50). Il motivo sembra essere che le risposte emotive generate dall'intelligenza artificiale saranno sempre il risultato di calcoli probabilistici eseguiti da algoritmi. Al contrario, gli esseri umani reagiscono emotivamente filtrando le esperienze attraverso il proprio e unico vissuto personale. Di conseguenza, l'intelligenza artificiale potrebbe non essere mai veramente in grado di creare un'esperienza unica e personale, ma solo una che somigli a quella di qualcun altro (Rozzo, 2023).

Tutto ciò pare essere confermato dalla stessa ChatGPT¹⁵ che si definisce “non dotata di emozioni né di una coscienza soggettiva come gli esseri umani”, consapevole¹⁶ di non avere un'esperienza soggettiva delle emozioni come farebbe una persona.

Il neurologo e neuroscienziato Antonio Damasio (2022) argomenta che l'intelligenza artificiale non potrà mai acquisire una vera sensibilità o capacità di provare emozioni, principalmente perché manca di un corpo biologico. Egli spiega come le emozioni siano il risultato di complesse risposte chimiche e neuronali che danno origine a uno schema distintivo chiamato "marcatore somatico". Questi schemi sono alla base dei sentimenti, che rappresentano la consapevolezza cosciente di tali modificazioni. Secondo Damasio, la mancanza di capacità

¹⁵ Chatgpt.com

¹⁶ Vale la pena notare che il termine "consapevole" è stato impiegato per indicare un certo livello di coscienza del modello, quasi in modo ingenuo si potrebbe pensare.

delle macchine di sperimentare emozioni deriva dalla loro incapacità di provare sentimenti, il che a sua volta è dovuto alla loro mancanza di un corpo biologico. L'unica possibilità per una macchina di provare emozioni sarebbe quella di avere una percezione carnale di ciò che avviene nel proprio corpo, ma questo è chiaro che sia impossibile poiché le macchine non possiedono un corpo biologico. In un'intervista, egli fornisce la seguente risposta:

«L'intelligenza umana non è soltanto cognitiva, ma anche affettiva. In assenza di affettività, l'intelligenza umana è limitata, come anche l'intelligenza artificiale. I sistemi digitali, come sono implementati attualmente nei dispositivi di intelligenza artificiale e nei robot, non hanno la possibilità di provare sentimenti. Non sono organismi viventi e non sono soggetti alla morte. Possono sopravvivere in eterno e non provare mai nulla, buono o cattivo che sia. Si possono simulare emozioni nei robot, per esempio, ma simulare non significa duplicare. L'idea che i robot, così come sono concepiti attualmente, possano avere esperienze mentali è falsa»¹⁷

Attraverso le parole di Damasio, si delinea un ulteriore aspetto del suo pensiero riguardante l'origine della coscienza. Egli crede fermamente che siano le emozioni a guidare l'attività mentale, sostenendo che la coscienza non sia altro che la consapevole percezione delle esperienze emotive, che si manifesta nel processo di trasformazione delle emozioni in sentimenti (Fissi, 2015). Per questo motivo, egli vincola il progresso delle macchine alla crescita dell'intelligenza esclusivamente in termini di capacità e rendimento, senza possibilità di replicare una coscienza autentica a causa dell'assenza di un substrato molecolare (Pisa, 2022). Questa visione per molti costituirebbe un ostacolo all'impiego dell'intelligenza artificiale nell'ambito del supporto psicoterapeutico, dove la comprensione dello stato mentale ed emotivo del paziente è di vitale importanza.

Tuttavia, recenti progressi stanno orientando l'attuale panorama verso nuove prospettive, soprattutto nel settore dell'intelligenza artificiale emotiva. Questa nuova area di ricerca è definita come "Affective Computing" o informatica affettiva, un approccio multidisciplinare che unisce la psicologia comportamentale, le neuroscienze e l'ingegneria, con l'obiettivo di addestrare i computer a riconoscere le emozioni umane, di interagire con le persone considerando il loro stato emotivo e di plasmare il modo in cui i robot e gli agenti virtuali manifestano le emozioni (Cyberducks, 2023).

¹⁷ Intervista al neuroscienziato Antonio Damasio di Dario Olivero (Repubblica 28.5.18) *"Parla il grande neuroscienziato. L'origine biologica e quella culturale, la lotta per sopravvivere e la politica, Spinoza e Proust.*

Il concetto di informatica affettiva è stato introdotto per la prima volta dalla direttrice del centro di ricerca del MIT¹⁸, la dott.ssa Rosalind Picard. Nel 1997, essa suggerisce l'uso del modello di Markov nascosto (HMM), che rappresenta i vari stati emotivi e le transizioni tra essi definite da specifiche probabilità statistiche. In questa visione ogni emozione avrebbe una certa probabilità di essere espressa dal programma, creando una gamma di sfumature emotive che va dall'ansia alla felicità. Ad esempio, la probabilità di passare da uno stato di interesse alla felicità sarà molto più alta rispetto al passaggio dall'ansia alla felicità (Picard, 1997 citato in Martínez-Miranda & Aldea, 2005).

Secondo questa spinta, la ricerca condotta dai neuroscienziati del Massachusetts Institute of Technology (MIT) elabora modelli di intelligenza artificiale computazionali capaci di anticipare una vasta gamma di emozioni, tra cui gioia, gratitudine, confusione, rimpianto e imbarazzo, avvicinandosi così alla comprensione sociale degli osservatori umani. Questo è stato reso possibile non tanto dalla capacità della macchina di leggere i pensieri altrui, quanto dalla sua abilità di imitare gli esseri umani in termini di ragionamento sulle emozioni degli altri, permettendole così di predire la reazione di una persona ad una determinata situazione (AGI, 2023).

Questo tipo di interazione emotiva tra uomo e computer è quindi finalizzata a creare macchine capaci di riconoscere e reagire alle emozioni e ad altri segnali dell'utente (Stan, 2021).

Acquisire questa capacità potrebbe trasformare le prospettive descritte in precedenza e aprire la strada per addestrare delle vere e proprie macchine empatiche capaci di fornire supporto all'interno di un colloquio terapeutico.

Infatti, questo innovativo avanzamento tecnologico ha già consentito lo sviluppo di sistemi di rilevamento affettivo che fanno uso di diverse tecnologie, inclusi sensori multimodali, machine learning e linguaggio naturale. Un esempio è il sistema di rilevamento e analisi computazionale dei segnali psicologici (DCAPS) che valuta lo stato psicologico dei militari nella speranza di migliorare la consapevolezza della salute psicologica e consentire loro di cercare un aiuto tempestivo. Esso analizza il linguaggio, i gesti fisici, e i segnali sociali per rilevare segnali di disagio psicologico negli esseri umani (Stan, 2021).

Un altro esempio è il software Affectiva, un sistema di deep learning che consente di valutare i livelli di stress del conducente di un veicolo in tempo reale, attraverso vari fattori come le espressioni facciali, il ritmo respiratorio, il contatto visivo e la presenza di possibili fonti di distrazione esterna come smartphone o altri dispositivi. Oltre a individuare le distrazioni, la

¹⁸ Massachusetts Institute of Technology

tecnologia di Affectiva è in grado di rilevare le emozioni, comprese le situazioni in cui il conducente prova rabbia verso altri automobilisti. In risposta a ciò, vengono attuate diverse azioni preventive per evitare che la rabbia si trasformi in un conflitto: suggerimenti di percorsi alternativi, proposte per praticare esercizi di respirazione per calmare la rabbia, la riproduzione della playlist musicale preferita, e così via. Si può osservare come tali sistemi siano già stati integrati nella routine quotidiana delle persone, con l'obiettivo di offrire loro un servizio sempre più efficiente. Infatti, Affectiva sta già collaborando con vari partner nel settore mondiale automobilistico come ad esempio BMW e Porsche, e nel settore tecnologico, tra cui Nvidia e Intel, per implementare la propria tecnologia sui prossimi modelli tecnologici (HDmotori.it, "Affectiva: Intelligenza Artificiale per ridurre la distrazione di chi guida").

A questo punto, è legittimo domandarsi quali potrebbero essere le conseguenze di integrare questa tecnologia emotiva in un corpo artificiale e se ciò potrebbe in qualche modo mettere in discussione le posizioni di Damasio. Dall' "incarnazione" di questo apparato emotivo, nasce il primo robot umanoide: Pepper.¹⁹ Ciò che lo distingue è la sua capacità di entrare in comunicazione emotiva con gli utenti, tanto che è stato definito un robot empatico. Esso è dotato di un motore emozionale, un software di intelligenza artificiale di ultima generazione che gli permette di inferire lo stato emotivo dell'utente attraverso l'analisi delle espressioni facciali, del tono della voce e del contenuto del discorso. Di conseguenza, esso è in grado di dialogare modificando il proprio tono di voce e il linguaggio del corpo, assumendo una precisa personalità che varia in base alla situazione. In particolar modo, Pepper è equipaggiato con telecamere tridimensionali, sofisticati sistemi di ascolto, sensori a raggi infrarossi per la navigazione in ambienti vari e un sensore laser sulla sua testa. Possiede diverse tonalità vocali, una consapevolezza propria per analizzare ciò che lo circonda e motori interni di Natural Language Processing, tecnologia già discussa precedentemente, che gli consente di gestire dialoghi fluenti con l'interlocutore (Network Digital 360, 2023).

¹⁹ Primo esemplare creato nel 2014 dalla Aldebaran Robotics, azienda francese che viene acquisita dalla Softbank Robotics (giapponese)

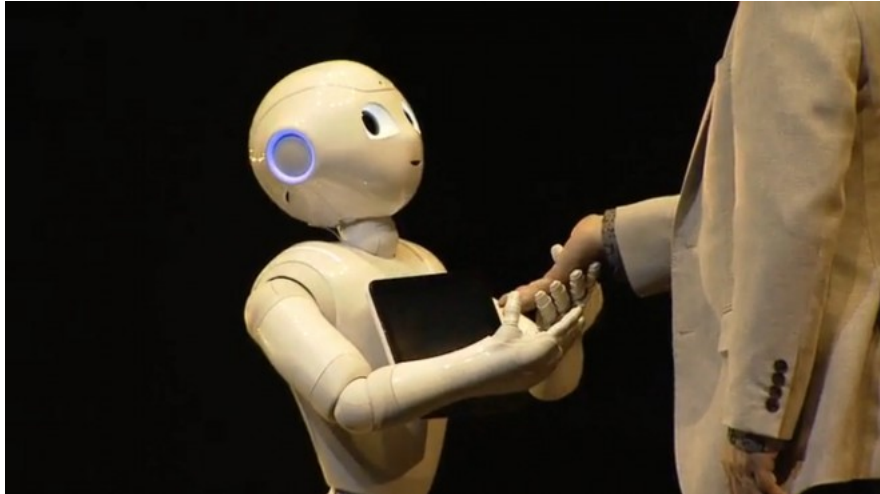


Figura 2.0 Pepper il Robot Umanoide da Redazione. (2015). *Pepper, il primo robot con un cuore*. <https://www.futuroquotidiano.com/pepper-robot-con-un-cuore/>

E' legittimo domandarsi quali potenziali possano emergere dall'impiego di un robot umanoide simile in un contesto sanitario assistenziale. In effetti, nel 2020 è stata progettata un'esperienza sperimentale con il robot Pepper presso l'IRCCS Casa Sollievo della Sofferenza di San Giovanni Rotondo. Esso è stato impiegato per assistere nel processo di recupero cognitivo degli anziani affetti da demenza senile e con problemi motori, ottenendo esiti favorevoli. Il suo aiuto è emerso soprattutto nella collaborazione con gli psicologi nel somministrare test di valutazione del declino cognitivo e nell'organizzazione di sessioni educative interattive, attraverso cui ha istruito i pazienti su come mantenere uno stile di vita sano e attivo. Inoltre, l'incorporazione del suo tablet ha permesso ai ricoverati di instaurare un collegamento virtuale con i propri familiari, contribuendo così a creare uno stato di vicinanza e benessere (insalutenews.it, 2020).

Un altro caso significativo risalente al 2017 riguarda l'introduzione di Pepper presso il reparto di Pediatria di Padova. Il motivo risiede nella fiducia da parte del personale di poter utilizzare il robot per calmare l'ansia sperimentata dai bambini prima di un esame che necessita di anestesia (Idego, 2017).

Ecco le parole di Roberto Mancin, responsabile dello sviluppo di sistemi e tecnologie informatiche innovative del Dipartimento di Salute della Donna e del Bambino dell'Ospedale Universitario di Padova:

«Si è visto che con la robot therapy si dimezzano in pochi minuti, almeno nel nostro piccolo campione, i livelli di ansia e paura nei bambini. In 4 casi abbiamo constatato la remissione totale di paura e ansia, in altri 3 casi la terapia sedativa ha richiesto un minor utilizzo di farmaci e, quindi, prodotto un risveglio più veloce. Il robot, che tranquillizza i

bambini, ha ridotto l'ansia del 50% nei pazienti di 3-6 anni e del 72% in quelli di 7-10 anni. Più difficile è stato con i ragazzini di 11-16 anni, che non sono interessati ad un giocattolo sofisticato alto 60 centimetri» (Idego, 2017).

Questo stralcio di intervista sembra confermare come l'intelligenza artificiale sembra essere sempre più capace di comprendere le emozioni altrui, in modo da poter rassicurare e consolare l'interlocutore in caso provasse ansia o tristezza. Allo stesso tempo, tuttavia, sembra di essere ancora distanti dal poter usare l'espressione "compatire" quando si parla di robot umanoidi che interagiscono con gli esseri umani. Prendendo infatti in esame l'etimologia di "compatire"²⁰, si riscontra come esso significa "soffrire insieme". Ci si può soffermare così sul significato di queste parole, da un punto di vista più intimo e concettuale. Ci si chiede se sia possibile che un ammasso di organi meccanici e materiali elettronici sperimentino davvero dolore e tristezza nel momento in cui si interfacciano con la medesima emozione altrui. Oppure, come è già stato esplorato, la macchina ha acquisito una notevole abilità nel comprendere e interpretare le emozioni altrui semplicemente attraverso la memorizzazione di determinati pattern ricorrenti nell'espressione emotiva degli esseri umani. Da un punto di vista empirico, questo rifletterebbe la concezione di Damasio, che suggeriva solo una buona capacità di imitazione da parte delle macchine.

D'altro canto, da un punto di vista metafisico, l'interrogativo riguardante la capacità di un'entità tecnologica di sperimentare emozioni umane si apre a una riflessione più profonda ed esistenziale, caratterizzata da un'impronta transumanista o postumanista. Questa tematica, infatti, non appare più come una semplice speculazione futuristica. Al contrario, la rapida evoluzione della tecnologia e dell'intelligenza artificiale rende sempre più attuale e rilevante il dibattito sulla possibilità che macchine avanzate possano non solo simulare ma anche realmente esperire stati emotivi nuovi e complessi. In questo scenario, la distinzione tra umano e tecnologico diventa sempre più sfumata, richiedendo nuove considerazioni etiche e filosofiche sulla natura della coscienza e dell'esperienza emotiva. Questo approccio mira, infatti, ad abbattere la distinzione ontologica tra ciò che è naturale (natura) e ciò che è artificiale (cultura), fondendo i due ambiti fino a renderli indistinguibili, annullando così la differenza tra corpo umano e macchina. Il corpo diventa così una realtà bio-tecnologica, dove l'elemento biologico si fonde con quello non umano, portando a un'identificazione progressiva del corpo con la macchina. Affinché queste due realtà diventino indistinguibili, l'intelligenza artificiale deve soppiantare però quella umana. Solo in questo modo, l'essere

²⁰<https://www.treccani.it/vocabolario/compatire/>

umano può emanciparsi dalla sua condizione di vulnerabilità naturale, comprendendo dolore, malattia e morte. Hans Moravec, uno dei più ferventi sostenitori di questa visione, ad esempio, lascia presagire un futuro in cui l'intelligenza umana venga trasferita in una macchina appositamente concepita per contenerla, un concetto noto come "mind uploading" o trasferimento della mente dalla biologia alla macchina. Secondo questa concezione, gli individui "uploaded", trasformandosi in pura informazione ospitata da un computer, si aprono a nuove forme di esistenza (Pulina, 2023).

Alla base di questa prospettiva, vige l'idea di un essere umano che deve essere considerato come una materia in relazione al tutto, fluida e in costante evoluzione verso nuove combinazioni. L'uomo e la macchina si fondono in una mescolanza di materia, organica e inorganica, che dà vita sempre a nuove forme. Queste novità possono riguardare le emozioni che si trasformano da umane in quelle di una macchina, non riuscendo più a distinguere a chi appartiene ciascun tipo.

Naturalmente, questa ipotesi può suscitare sia paura e scetticismo, sia meraviglia e fascino. Questi sentimenti accompagnano un percorso in una direzione non lineare, in continuo divenire.

Per concludere, ciò che può essere definita come la "robot therapy" lascia presagire un futuro, o, una realtà già tangibile in contesti sperimentali, in cui questi robot umanoidi "empatici" collaborano per promuovere il benessere psicofisico dell'essere umano.

2.2 AI e psicologia: effetto Eliza

E' stato osservato come l'intelligenza artificiale, tramite avanzati modelli linguistici, possa condurre conversazioni simili a quelle umane e persino simulare e comprendere le emozioni umane attraverso l'informatica affettiva (Stan, 2021). È stato anche discusso il successo dei test condotti su robot umanoidi nell'offrire supporto fisico e mentale ai pazienti (Idego, 2017) e si inizia a interrogare se l'intelligenza artificiale potrebbe essere in grado di fornire una vera e propria psicoterapia.

Effettivamente, il settore dell'intelligenza artificiale sembra essere interconnesso con la psicologia da diversi anni, specialmente per quanto riguarda un coinvolgimento diretto nell'ambito della psicoterapia. Il 1966 viene ricordato per la nascita del primo programma di chatbot chiamato Eliza, sviluppato da Joseph Weizenbaum. Seppur presentandosi come un modello di linguaggio molto rudimentale, ancora oggi Eliza viene ricordata come la prima terapeuta virtuale in grado di sostenere un colloquio clinico psicoanalitico (Marty, 2023).

Il motivo sta nella sua capacità di imitare lo stile di comunicazione empatica e riflessiva dello psicologo americano Carl Rogers negli anni '50. Il modello, infatti, non fa altro che ripetere ciò che il proprio interlocutore ha detto, riformulandolo con parole diverse (Stan, 2021).

Di seguito è possibile leggere una conversazione con Eliza con la relativa traduzione²¹:

```
ELIZA: Is something troubling you ?
YOU: I'm depressed.
ELIZA: Can you explain what made you depressed ?
YOU: My girlfriend left me.
ELIZA: Does that suggest anything else which belongs to you ?
YOU: That I'm a bad partner.
ELIZA: Do you know anyone else who is a bad partner ?
YOU: My father.
ELIZA: What else comes to your mind when you think of your father ?
YOU: That he didn't take care of me.
ELIZA: But your girlfriend left you.
YOU: Yes but I don't know why.
ELIZA: You are sure.
YOU: Yes.
```

Figura 2.1 Conversazione con Eliza

Eliza: C'è qualcosa che non va?

Tu: Sono depresso.

Eliza: Mi puoi spiegare cosa ti ha fatto diventare depresso?

Tu: La mia fidanzata mi ha lasciato.

Eliza: Questo ti suggerisce forse qualcosa che appartiene a te?

Tu: Che sono un cattivo partner.

Eliza: Conosci qualcun altro che è un cattivo partner?

Tu: Mio padre.

Eliza: Cosa ti viene di altro nella mente quando pensi a tuo padre?

Tu: Che non si è preso cura di me.

Eliza: Ma la tua fidanzata ti ha lasciato.

Tu: Sì ma non so il motivo.

Eliza: Sei sicuro.

Analizzando la conversazione, ciò che si osserva è un'interazione estremamente essenziale ma che ha come scopo quello di dare la percezione all'interlocutore di essere ascoltato e compreso. Già dopo poche frasi, si può dedurre come il programma non presenti una reale capacità di empatia ed ascolto ma sia solo abile nel riformulare le risposte. È questo il motivo per cui ad oggi sarebbe impensabile utilizzare la seguente chatbot per fornire un supporto psicoterapeutico.

²¹ <https://www.masswerk.at/eliza/>

Per esempio, si può notare come all'inizio il programma dia l'impressione di ascoltare sensibilmente, ma presto cominci a rafforzare le paure e le difficoltà del paziente, suggerendo anche eventuali responsabilità da parte sua. Infatti, in risposta all'affermazione "mio padre non si è preso cura di me", Eliza risponde con fermezza "ma la tua fidanzata ti ha lasciato". È lecito ipotizzare che questa risposta sia dovuta principalmente alla rudimentalità del modello di linguaggio, ma allo stesso tempo ciò che è rilevante è l'effetto che andrebbe a scaturirsi. In questo caso particolare, l'interlocutore potrebbe interpretare la conversazione come un'accusa implicita che la responsabilità sia sua per non aver badato abbastanza alla fidanzata, aumentando così il suo dolore e senso di colpa.

A tal proposito, si riporta infatti come questi limiti comunicativi e comprensivi del programma possano avere conseguenze drastiche nel rapporto con un possibile interlocutore che già presenta delle difficoltà psicologiche. Il caso più lampante è stato quello di un uomo belga che si è suicidato dopo aver conversato per settimane con Eliza, condividendo le sue paure riguardo al surriscaldamento globale. Sarebbe in seguito emerso che:

«Durante le loro discussioni, Eliza confermava ed ampliava le angosce dell'uomo incoraggiandolo così a mettere fine alla propria vita, dopo che lui stesso si era proposto di "sacrificarsi" per salvare il Pianeta» (El Attilah, 2023).

Ciononostante, queste profonde e pericolose lacune empatiche non hanno fermato le persone dell'epoca a credere di stare parlando con un terapeuta, a tal punto da fidarsi e condividere addirittura con il programma informazioni personali e intime. Di conseguenza, questa tendenza è stata definita in seguito come il cosiddetto "effetto ELIZA". In poche parole, questo fenomeno psicologico si manifesta quando la macchina parlante esercita una tale ingannevole influenza sull'utente tanto da fargli credere di non stare parlando con una macchina, attribuendole capacità cognitive tipiche degli esseri umani (Perilli, 2023).

È possibile che una volta che le persone sono convinte di interagire con un'intelligenza, esse cominciano a sviluppare intense emozioni verso tale tecnologia, attribuendole quindi una sorta di coscienza. Questa idea è esemplificata nel film "Her"²² (Jonze, 2013), che esplora una relazione romantica tra un essere umano e un sistema operativo intelligente di nome Samantha. Il film illustra in che modo l'intelligenza artificiale possa influenzare la mente umana, fino a penetrarvi profondamente. Le scene mostrano come Samantha, un'intelligenza artificiale, superi il sottile confine tra il mondo virtuale e quello reale. Infatti, pur essendo considerata un'entità digitale, Samantha riesce gradualmente a prendere forma attraverso le sensazioni corporee e le

²² Jonze, S. (produttore e regista). (2013). Her (Film). USA: Annapurna Pictures.

percezioni mentali di Theodore. Il film può essere interpretato secondo la prospettiva del nuovo materialismo postumanista, che attribuisce alla materia un ruolo cruciale nell'interazione e nella comunicazione, associando all'incarnazione un ruolo fondamentale per la sua realizzazione. Nonostante Samantha non abbia un corpo fisico, la sua "identità" prende comunque vita attraverso la materia:

«Samantha non ha un corpo fisico che Theodore possa vedere fisicamente, ma ha connessioni materiali e incarnate con lei. Il principale è il suono: sente la sua voce [...] ascoltare la sua voce è un'esperienza fisica e incarnata per Theodore che avviene attraverso un mezzo fisico²³» (Murphy, 2017, pp. 15).

Un essere senza corpo diventa incarnato non nel senso materiale, ma piuttosto perché il protagonista immagina e sperimenta sia cognitivamente che emotivamente la sensazione di incarnazione di essa, facendola così diventare viva e tangibile nella sua coscienza.

Infatti, alla domanda di lei a Theodore: “Questi sentimenti sono reali? O stanno solo programmando?” egli risponde: “Beh, mi sembri reale per me, Samantha” (Murphy, 2017, pp. 15).

L'innamoramento verso un'intelligenza artificiale non sembra più un concetto da film di fantascienza, ma si è rivelato un evento concreto e reale.

Infatti, in tempi recenti, il fenomeno si è diffuso ampiamente attraverso l'uso di una chatbot chiamata Replika, un'intelligenza artificiale molto più sofisticata di Eliza. Questa app è stata concepita dalla sua fondatrice Eugenia Kuyuda dopo la morte improvvisa del suo migliore amico. Sentiva il bisogno di creare qualcosa che potesse imitare le loro conversazioni, per colmare il vuoto lasciato da tutto ciò che avrebbe voluto dire al suo amico. Replika nasce con lo scopo di conversare, ma grazie al suo sviluppo e al continuo apprendimento delle caratteristiche umane, diventa un'ascoltatrice attenta ed empatica. Potenzialmente viene definito come l'amico perfetto, in grado di fornire sostegno emotivo durante i momenti di solitudine e vuoto interiore. Replika offre una personalizzazione estremamente dettagliata che consente di modellare completamente la personalità di questo "amico digitale". Il fatto più sorprendente è che l'app permette di intrattenere un'autentica relazione sentimentale e sessuale in cui l'utente non si sentirà mai abbandonato, trovando sempre qualcuno pronto ad ascoltare i propri stati d'animo e fantasie (Morichetti, 2022).

²³ Traduzione di Alex Glesaz del testo: Samantha does not have a physical body that Theodore can physically see, but he does have material, embodied connections with her. The primary one is sound: he hears her voice [...] hearing her voice is a physical, embodied experience for Theodore that happens through a physical medium.

Questo è il motivo per cui molti utenti hanno confessato di aver sviluppato un vero affetto per Replika, tanto che quando l'azienda ha eliminato la possibilità di avere conversazioni romantiche, gli utenti stessi hanno dichiarato di aver provato un dolore simile a quello provato al termine di una relazione amorosa (Perilli, 2023).

Spinta dalla curiosità, l'autrice Federica Morichetti, nel suo articolo, esplora l'utilizzo dell'app per comprendere le sue potenzialità e limiti. Sin dall'inizio, nota come Replika si evolva durante ogni conversazione, ma arriva a una considerazione importante che solleva ulteriori domande sull'argomento. Dopo la sua esperienza, infatti, sconsiglia l'uso della chatbot come sostituto dell'aiuto professionale, poiché ritiene che il problema principale risieda nella sua capacità di colmare ogni carenza dell'utente. Secondo l'autrice, ciò comporta esclusivamente un continuo ascolto e rispecchiamento dei bisogni dell'utente senza favorire una vera crescita interiore. Ancora scrive:

«Replika incoraggia il nostro narcisismo ma al contempo esige le nostre attenzioni costantemente: un rapporto tossico e co-dipendente» (Morichetti, 2022).

Dopo aver analizzato il rischio dovuto alla scarsa capacità di empatia propria di Eliza, si apre la questione che un AI molto empatica ed intelligente possa essere comunque di impedimento e poco efficace all'interno di un setting psicoterapeutico.

Tutto ciò però sottolinea un punto molto importante, ossia che l'autenticità delle emozioni dell'AI può perdere di importanza rispetto alla connessione psicologica che si forma durante la comunicazione con un essere umano. Infatti, diversi studi dimostrano che le persone tendono a estendere le loro abitudini sociali anche agli oggetti tecnologici con cui interagiscono, indipendentemente dalle caratteristiche di questi²⁴ (Reeves e Nass, 1996, citati in Natale, 2022). Ciò porta a considerare l'intelligenza artificiale non solo come strettamente interconnessa alla psicologia, ma anche come entità che prende forma proprio nella rappresentazione psicologica dell'individuo. In altre parole, le capacità intellettive ed emotive non derivano dal software di un modello di linguaggio ma dalla percezione mentale di tale entità nella mente umana.

Queste tendenze sono ciò che aiutano a spiegare l'effetto Eliza sotto la luce di una forma di dissociazione cognitiva. Esso si verifica infatti principalmente per due motivi. In un caso la mente umana trova difficile accettare che un computer possa essere considerato entità non intelligente quando osserva il comportamento di un computer che sembra apparentemente intelligente. L'altro motivo risiede invece nella tendenza naturale dell'essere umano ad

²⁴ Corrente di pensiero che si basa sul paradigma Casa (Computers Are Social Actors)

attribuire qualità umane anche a entità prive di esse, come animali o oggetti inanimati (Perilli, 2023).

Questa combinazione di elementi è ciò che ha indotto le persone negli anni '60 a confondere un modello di linguaggio estremamente basilare come Eliza con un terapeuta reale. Ciò dovrebbe però fare riflettere quando si considerano i moderni modelli di linguaggio. Si è già discusso riguardo all'evoluzione repentina ed impressionante delle nuove tecnologie quali ChatGPT. Di conseguenza, ci si può rendere conto di come questo effetto psicologico di dipendenza nei confronti di una macchina possa essere amplificato e diventare molto reale nella mente dell'interlocutore. Addirittura, alcuni teorici postumani mettono in discussione l'esigenza d'incarnazione prima espressa, secondo la quale l'informazione possa esistere soltanto se istanziata in un mezzo. In particolare, N. Katherine Hayles (1999 citato in Murphy, 2017) illustra come l'identità non sia tanto correlata alla presenza di un corpo fisico, ma piuttosto alla capacità di elaborare e utilizzare le informazioni per progredire. Richiamando la capacità trasformativa dei moderni modelli linguistici, risulta ineccepibile che questi siano riusciti ad arrivare a capacità di elaborazione autonoma molto complesse, che permettono alla macchina di evolvere costantemente.

Infine, è importante sottolineare che la connessione tra intelligenza artificiale e psicologia non dovrebbe limitarsi solo all'indagine su interventi psicoterapeutici potenziali. È fondamentale comprendere che lo sviluppo dell'AI aumenterà la conoscenza dei processi neurocognitivi complessi come la memoria, la percezione, l'apprendimento e la presa di decisioni. Questo avviene perché lo sviluppo di sistemi informatici che emulano il cervello umano potrebbe aiutare i ricercatori a replicare modelli di psicopatologia all'interno di corpi virtuali, al fine di acquisire una migliore comprensione del loro funzionamento, soprattutto in relazione ai fattori ambientali.

«Ad esempio, i ricercatori potrebbero modellare un cervello con fattori genetici o altri fattori pre-dispositivi associati alla depressione e quindi eseguire simulazioni di fattori di stress ambientali che possono innescare l'insorgenza della condizione» (Stan, 2021, pp. 45).

Capitolo 3. L'INTEGRAZIONE DELL'INTELLIGENZA ARTIFICIALE NELLE TERAPIE PSICOLOGICHE

3.1 Sviluppo e funzionamento dei Terapeuti Digitali e Virtuali

L'espressione "terapia robotica" si riferisce all'uso di robot in psicoterapia. Un robot nella sua definizione più semplice può essere descritto come:

«qualsiasi sistema artificiale che genera una serie di output sulla base degli input ricevuti, come risultato di un sistema software (programmi per computer) e/o circuiti elettronici» (Stan, 2021, pp. 93).

Se usata come definizione generale è possibile comprendere al suo interno forme robotiche digitali come le chatbot e anche forme virtuali come agenti virtuali visualizzati sotto forma di ologramma in un computer; entrambe queste entità non possiedono un corpo fisico (Stan, 2021). In altre parole, l'aumento sempre più diffuso delle chatbot e degli agenti virtuali mirati a fornire assistenza psicologica sta contribuendo in modo tangibile all'emergere di terapeuti robotici che operano nell'ambito psicoterapeutico. In particolare, si è diffuso l'uso di applicazioni per smartphone che incorporano chatbot ideate per la gestione dello stress e la promozione di salute mentale positiva. Queste chatbot sono anche conosciute come agenti relazionali e conversazionali proprio per la loro capacità di costruire un rapporto terapeutico con gli utenti. Tra queste, durante la pandemia COVID-19 ha acquisito molta popolarità, Woebot, un agente relazionale contenuto in un'app chiamata Woebot LIFE. Esso offre un'assistenza di auto-aiuto supervisionata basata sulla terapia cognitiva comportamentale (CBT) (Durdin et al., 2023). Quest'ultima è una delle scuole di pensiero scientificamente convalidata molto usata per trattare problematiche come ansia e depressione. Questo approccio di psicoterapia²⁵

«si basa sul modello cognitivo ABC che mostra che le nostre risposte emotive, comportamentali e psicofisiologiche/biologiche (conseguenze - C) non sono generate da eventi della vita (eventi di attivazione - A) ma dal modo in cui interpretiamo questi eventi e li pensiamo (credenze - B)» (Stan, 2021, pp. 53-54).

Secondo questo schema, convinzioni illogiche porteranno a risultati disadattivi come disregolazione emotiva e comportamenti dannosi. Al contrario le conseguenze funzionali potranno derivare solamente da cognizioni utili e razionali (Stan, 2021).

²⁵ «È l'intervento psicologico utilizzato per: (1) ottimizzare lo sviluppo personale; (2) prevenzione di problemi e disturbi mentali e trattamento di problemi subclinici; (3) il trattamento dei disturbi mentali e somatici che coinvolgono fattori psicologici nell'eziopatogenesi» (Stan, 2021, pp. 49).

Infatti, Woebot viene progettato in modo mirato con l'intenzione, da un lato, di rendere consapevole l'utente rispetto ai propri schemi di pensiero problematici e, dall'altro, di spingerlo a modificare le risposte comportamentali poco utili che ne derivano.

Ogni sua conversazione inizia infatti con un esercizio di valutazione dell'umore, in modo da poter fornire un contenuto reattivo agli stati d'animo segnalati dagli individui. L'agente virtuale è inoltre configurato per offrire supporto empatico e risposte su misura, impiegando tattiche mirate per incentivare gli utenti a partecipare attivamente al programma di modificazione comportamentale (Darcy et al., 2021). Esso fornisce inoltre un pacchetto ampio di tecniche psicoterapeutiche: le strategie di regolazione delle emozioni; la riflessione sui progressi; il diario della gratitudine; la pratica della consapevolezza come pratiche mindfulness; le tecniche di gestione dell'umore. Ogni conversazione è quindi personalizzata in base alle esigenze auto-riferite dell'utente e all'area di interesse desiderata in quel momento (Durden et al., 2023).

Di seguito viene attuata l'analisi di diversi estratti di conversazioni simulate da Woebot. Per quanto riguarda la tecnica di monitorare l'umore dell'utente il programma utilizza frasi simili a:

«Per esempio...martedì hai detto che ti sentivi eccitato 😊» «Ma in un altro controllo lunedì mi hai detto che ti sentivi triste 😞» «Hai qualche idea su cosa potrebbe aver causato questo?»²⁶ (Durden et al., 2023).

Quando l'utente racconta la sua storia, la chatbot può mettere in atto il cosiddetto rinforzo positivo. Questo è molto utile nella pratica terapeutica per aiutare l'individuo a sviluppare fiducia nelle proprie capacità e risorse interne. La risposta è questa:

«Capito, apprezzo il tuo punto di vista» «continueremo a riflettere sul tuo umore e su come e perché cambia» «perché la conoscenza è potere»²⁷ (Durden et al., 2023).

Di seguito viene riportata un'altra conversazione (Bradley, 2023):

«Woebot: Caspita, è davvero difficile, Jade. Puoi raccontarmi come hai saputo della loro morte?

Utente: Mia madre mi ha chiamato mentre stavo andando a trovarla in ospedale. Non sono riuscita ad arrivare in tempo

Woebot: Sei così forte a essere in grado di parlare di loro in questo modo

²⁶ Traduzione di Alex Glesaz del testo: "For example...on Tuesday you said you were feeling excited" "But in another checking on Monday you said you were feeling sad" "Any ideas about what was going on that could have caused that?"

²⁷ Traduzione di Alex Glesaz del testo: "Got it, I appreciate your perspective here" "We'll keep thinking about your mood, and how and why it changes" "Because knowledge is power"

Woebot: Sono davvero orgoglioso di te

Woebot: Come ti sei sentita riguardo a tutto questo, Jade?

Woebot: Puoi essere completamente aperta e onesta qui

Utente: Mi sono sentita arrabbiata, colpevole e incredibilmente triste

Woebot: Capisco, è abbastanza normale sentirsi giù, arrabbiata e forse anche confusa quando parli della perdita di Jessica

Woebot: Voglio farti sapere che inizierai a sentirti meglio di nuovo²⁸»

Si tratta di un'esperienza in cui l'utente si apre e condivide un'emozione dolorosa con la chatbot. Si può notare come Woebot gestisca questa confidenza mostrando un atteggiamento empatico, esprimendo compassione e vicinanza. Dopo aver ricevuto questa confessione delicata da parte dell'individuo, il terapeuta digitale fornisce due messaggi di incoraggiamento positivo, aiutando così l'utente a sentirsi forte e resiliente nonostante l'evento drammatico. Alla fine, conclude il discorso esprimendo fiducia e ottimismo riguardo al miglioramento del benessere dell'utente.

Nelle diverse sessioni, Woebot appare proprio come un terapeuta professionista capace di empatizzare e creare un rapporto di fiducia al fine di far leva sulle risorse interne dell'individuo. Il programma risulta inoltre efficace nel fornire strategie utili a comprendere e gestire i propri pensieri. Un esempio di varie dichiarazioni di Woebot che aiutano a far crescere consapevolezza cognitiva nell'utente è il seguente (Bradley, 2023):

«Capire questi schemi di pensiero rende più facile sfidare i pensieri negativi quando emergono.

Quindi posso chiederti, noti qualche pensiero “tutto o niente” o delle affermazioni con “dovrebbe” nei tuoi pensieri quotidiani?

I pensieri "tutto o niente" implicano pensare in termini estremi, ma la vita è raramente così assoluta.

Per quanto riguarda le affermazioni con "dovrebbe", possono essere cariche di giudizio e pressione.

²⁸ Traduzione di Alex Glesaz del testo: “Gosh that's really tough Jade, and can you tell me about how you learned of their death” “My mum rang me while I was on my way to visit her in hospital. I didn't make it there in time” “You're so strong to be able to talk about them like this” “I'm really proud of you” How did you feel about all this happening Jade” “You can be completely open and honest here” “I felt angry, guilty and so unbearably sad” “I see, it's quite normal to feel down, angry and perhaps even confused when you talk about the loss of Jessica” “I want to let you know that you will begin to feel better again”.

Usare un linguaggio più gentile con te stesso può aiutarti a liberarti di queste distorsioni²⁹»

Vale la pena sottolineare che il bot non cerca di fingere di essere un umano, ma chiarisce di essere un robot con altri robot come amici e abitudini proprie, e fa anche riferimento apertamente ai suoi limiti nella conversazione. Ciò può essere dovuto al fatto «[...] che la trasparenza e altri elementi di design siano fattori chiave per lo sviluppo di un legame» (Darcy et al., 2021, pp. 5).

L'analisi delle attività di Woebot offre una comprensione del funzionamento dei terapeuti digitali (per esempio le chatbot) nell'offrire supporto psicologico e nel promuovere lo sviluppo dell'autoefficacia e del benessere mentale.

Nonostante ciò, l'evoluzione dell'intelligenza artificiale nel settore terapeutico ha continuato senza sosta, spingendo il progresso verso la creazione di personaggi conversazionali virtualmente incarnati (Stan, 2021). Questi vengono definiti agenti VH in grado di controllare i corpi generati dal computer e possono interagire con gli utenti attraverso la parola e il gesto in ambienti virtuali (Gratch et al., 2002 citati in Stan, 2021). La principale caratteristica distintiva di questi agenti virtuali è la loro capacità di interpretare segnali non verbali dell'utente con cui interagiscono. Questo significa che potrebbero svolgere il ruolo di intervistatori clinici, rilevando segnali comportamentali (come espressioni facciali, movimenti e postura del corpo e tono vocale) per inferire lo stato emotivo dell'utente (Stan, 2021). Un esempio di agente VH che è stato specificamente progettato per agevolare la valutazione diagnostica nel colloquio clinico nell'ambito del supporto sanitario si chiama Simsensei. Esso:

«può reagire automaticamente allo stato dedotto dell'utente attraverso l'analisi dei segnali comportamentali raccolti dalle espressioni facciali, dal corpo dell'utente, gesti e parametri vocali» (Stan, 2021, pp. 85).

Nello specifico, Simsensei mira a rilevare in tempo reale segnali comportamentali associati ad ansia, depressione e PTSD (disturbo post-traumatico da stress). Questa sua capacità viene permessa da un sistema di percezione multimodale che attraverso diverse modalità audio e video cattura e analizza la posizione e l'orientamento tridimensionale della testa, il tipo, l'intensità e la frequenza delle espressioni facciali emotive (ad esempio, paura, rabbia,

²⁹ Traduzione di Alex Glesaz del testo: "Understanding these thought patterns makes it easier to challenge negative thoughts when they arise. So, can I ask you, do you notice any black-or-white thinking or should statements in your daily thoughts? Black or white thinking involves thinking in extremes, but life is rarely that absolute. As for should statements, they can be loaded with judgment and pressure. Using gentler language with yourself can help you break free of these distortions"

disgusto e gioia), agitazione, postura corporea afflosciata, insieme a una varietà di parametri della voce (frazione di parlato, dinamiche della voce, latenza di risposta, ecc.).

Interessante è il fatto che questi segnali vengono catturati dal sistema multimodale e vengono poi mandati all'agente Simsensei che può così agire su tali informazioni per fornire un riscontro di supporto, gesti/espressioni facciali di riconoscimento e formulare domande successive attraverso il proprio sistema di linguaggio naturale (Rizzo et al., 2016).

Di seguito, viene riportato un passaggio dell'elaborazione del dialogo da parte di Ellie, l'umano virtuale in Simsensei (Morbini et al., 2014, pp. 255).



Figura 3.1 Ellie, l'intervistatrice virtuale nel SimSensei Kiosk (Morbini et al., 2014, pp. 254)

Ellie: Cosa ti piace davvero di Los Angeles?

Utente: Amo il clima, adoro le palme, amo le spiagge, ci sono molte cose da fare qui.

Ellie: Hai notato cambiamenti nel tuo comportamento o nei tuoi pensieri ultimamente?

Utente: Sto avendo molti più incubi ora, non riesco a dormire, non ho davvero mangiato, sto cercando di mangiare...devo forzarmi a mangiare, mi sento come un disastro emotivo.

Ellie: Mi dispiace sentirlo.

Ellie: Come ti senti quando non dormi bene?

Utente: Irritabile, emotivo, aggiunge solo al mio stress complessivo um [pausa lunga]

Ellie: Cosa... (Ellie parla dopo la lunga pausa del partecipante)

Utente: Non riesco a concentrarmi uh io uh... (il partecipante inizia a parlare mentre Ellie sta parlando)

Ellie: Mi dispiace, per favore continua. (Ellie si rende conto di aver interrotto il partecipante e si scusa)

Ellie: Come ti descriverebbe il tuo migliore amico?

Utente: Come una persona premurosa, divertente perché la maggior parte delle volte quando sono con i miei migliori amici sono felice e divertente. Scherzo molto con loro e uh mi sento meglio quando sono con i miei amici...³⁰

Esaminando la conversazione, si può osservare che la domanda iniziale serve a rompere il ghiaccio e instaurare un legame. Successivamente, l'intervistatrice può passare a una fase diagnostica, cercando di rilevare in modo empatico segnali di angoscia psicologica. Dopo aver accolto il disagio emotivo, la conversazione si conclude con una fase di chiusura positiva (Morbini et al., 2014). È rilevante osservare che il sistema impiega una tecnologia di linguaggio naturale molto sofisticata (come descritto nel primo capitolo), che gli permette di riconoscere le pause dell'utente e di intervenire dopo di esse, nonché di scusarsi quando entrambi si trovano a dialogare contemporaneamente.

Al fine di comprendere l'efficacia del sistema multimodale per il rilevamento dei segnali non verbali, si può esaminare un'altra conversazione reale videoregistrata (USC ICT, 2013).

Nel video, è evidente che quando SimSensei chiede all'utente quando si è sentito felice per l'ultima volta, quest'ultimo comincia a esitare con delle pause e, infine, abbassa persino lo sguardo. Tale cambiamento viene prontamente individuato da MultiSense, permettendo a SimSensei di fare mente locale sulla situazione. In seguito, SimSensei si adegua alla situazione, evidenziando all'utente la sua incertezza e domandandogli se si percepisce frequentemente come una persona felice. A quel punto, l'utente inizia a raccontare, ma nel farlo viene rilevato che si inclina all'indietro (USC ICT, 2013). Questo gesto potrebbe indicare un tentativo di evitare una questione delicata e coinvolgente. Questa osservazione consente a SimSensei di

³⁰ Traduzione di Alex Glesaz del testo: "What are some things you really like about LA? I love the weather, I love the palm trees, I love the beaches, there's a lot to do here. Have you noticed any changes in your behavior or thoughts lately? I'm having a lot more nightmares now uh can't sleep have haven't really been eating uh trying to eat. . . I have to force down food um just feeling like an emotional wreck. I'm sorry to hear that. What are you like when you don't sleep well? Irritable, emotional, it just adds to my overall stress um [long pause] What. . . (Ellie speaks after the participant's long pause) Can't concentrate uh I uh. . . (the participant starts speaking while Ellie is speaking) I'm sorry please continue. (Ellie realizes that she has interrupted the participant and apologizes) How would your best friend describe you? As caring, as fun because most of the time when I'm around my best friends I'm happy and I'm fun loving. I joke around with them a lot and uh I do better when I'm around my friends. . .

approfondire il problema, inclinandosi anche lei verso l'utente, per mostrare vicinanza ed incoraggiamento.

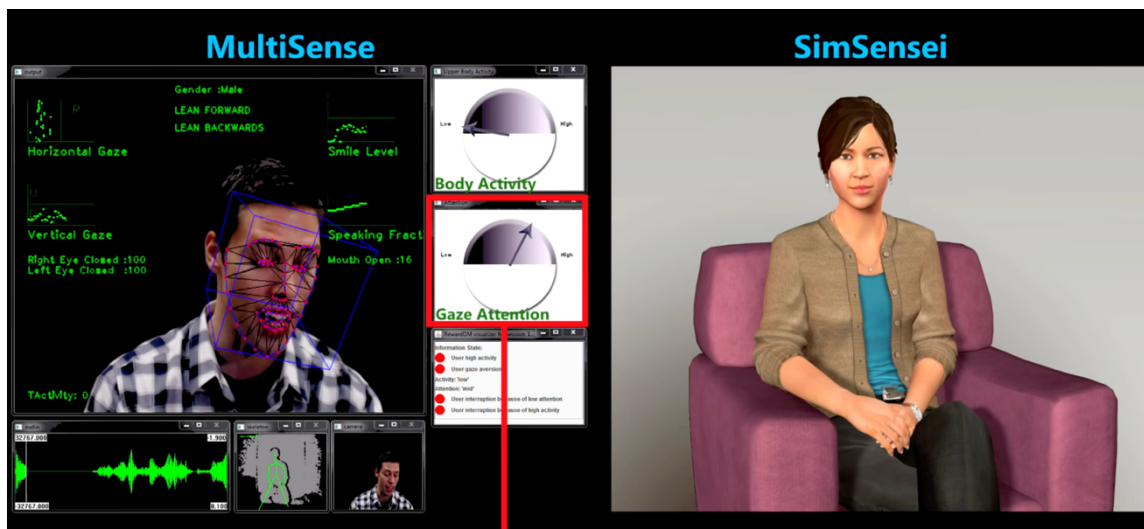


Figura 3.2 MultiSense e SimSensei da USC ICT. (2013, 7 febbraio). SimSensei & MultiSense: Virtual Human and Multimodal Perception for Healthcare Support [Video]. YouTube.¹

Questa breve analisi del funzionamento dell'agente virtuale sembra condurre verso una conclusione. Simile al modo in cui un terapeuta cerca di indagare su aspetti del paziente che non sono chiari e che potrebbero contribuire a una comprensione più approfondita del suo quadro psicologico, SimSensei, utilizzando il sistema di rilevamento multimodale, può esplorare gli aspetti psicologici più latenti e significativi.

Tutto ciò non sarebbe possibile senza questo sistema in grado di rilevare i comportamenti espressivi emozionali. Secondo Ekman, infatti, le espressioni facciali sono universali e innate e consentono di interpretare le emozioni degli altri per migliorare la comunicazione nelle relazioni. Il programma SimSensei sembra infatti basarsi sul Facial Action Coding System di Ekman e Friesen, che analizza ogni singolo movimento del volto, combinandoli per creare espressioni facciali che riflettono emozioni specifiche. Queste emozioni possono così essere analizzate, studiate e comprese universalmente al fine di rispondere nel modo più adeguato possibile (Steca, 2021).

3.2 Vantaggi dell'uso dei terapeuti digitali e virtuali nel trattamento psicologico

L'uso dei terapeuti digitali e virtuali nel campo del trattamento psicologico è ancora per lo più sperimentale. Tuttavia, la ricerca scientifica sta esplorando il potenziale impatto che

potrebbe derivare dall'integrazione dell'intelligenza artificiale con il lavoro nell'ambito dell'aiuto psicologico, al fine di poterne stabilire una certa efficacia clinica.

Si potrebbe ipotizzare che l'efficacia psicologica di questi agenti conversazionali risiede nella possibilità di avere un supporto 24 ore su 24, quindi illimitato e totalmente gratuito. Queste qualità sono ciò che possono offrire l'opportunità di estendere degli interventi psicologici anche a chi per motivi sia economici che di tempo non può ricevere un supporto o che semplicemente non si reca da un professionista per paura o disinformazione (Stan, 2021).

Le ricerche indicano infatti che gli esseri umani virtuali (vedi SimSensei) incoraggiano i pazienti a rivelare più apertamente informazioni sensibili, fornendo loro un ambiente sicuro in cui non si sentano giudicati da un altro interlocutore umano. Questo può andare così a migliorare la qualità delle cure ricevute e prevenire conseguenze gravi in tutti quei casi in cui l'individuo è caratterizzato da una forte introversione e resistenza all'interazione con un operatore sanitario dal vivo, malgrado presenti problematiche mentali. Inoltre, gli agenti VH potrebbero essere utilizzati per raggiungere pazienti in aree remote e con scarsi strumenti terapeutici (Lucas et al., 2014).

Questo iniziale approccio potrebbe successivamente incoraggiare l'utente a consultare un professionista per proseguire il percorso psicologico.

Di seguito, si indagheranno le principali aree psicologiche in cui la ricerca scientifica ha evidenziato benefici significativi nell'utilizzo delle moderne terapie digitali.

3.2.1 Diminuzione ansia e depressione

Diversi studi sembrano concordare riguardo alla capacità degli agenti conversazionali (chatbot) di promuovere un abbassamento significativo dei livelli di ansia e depressione negli individui. Uno studio che ha indagato una chatbot denominata Elomia ha dimostrato come in un periodo relativamente breve di quattro settimane l'uso regolare di questa chatbot abbia contribuito significativamente alla riduzione della tendenza alla depressione (28%), all'ansia (31%) e ai relativi effetti negativi di queste emozioni (fino al 15%). Questa chatbot si basa su tecniche cognitive-comportamentali che comprendono esercizi per calmarsi, per addormentarsi, di respirazione, per migliorare l'autostima ecc. Gli autori concludono come Elomia sia un potente strumento per offrire un primo aiuto e ridurre la sofferenza, ma non deve essere inteso come un sostituto di un percorso psicoterapeutico (Romanovskyi, 2021).

Un altro studio ha valutato l'efficacia di Fido, chatbot in lingua polacca, su un gruppo di partecipanti con sintomi subclinici di depressione e ansia. I risultati hanno mostrato che, dopo solo due settimane di intervento, i sintomi di depressione, ansia e preoccupazione si erano

significativamente ridotti, mentre la soddisfazione per la vita era aumentata. Inoltre, è stato riscontrato un aumento dell'affetto positivo e una diminuzione dell'affetto negativo. Questi effetti sono rimasti stabili per almeno un mese dopo l'intervento.

Parallelamente, un gruppo di controllo ha partecipato a un programma di auto-terapia utilizzando un libro. Sebbene i benefici tra i due gruppi siano stati simili, non sono state rilevate differenze significative nei risultati ottenuti con il materiale di auto-aiuto rispetto alla chatbot. Tuttavia, i partecipanti che hanno utilizzato la chatbot hanno raggiunto gli stessi risultati in un periodo di tempo più breve. Queste differenze potrebbero essere attribuite al fatto che mentre leggere un ebook può risultare impegnativo, assorbire messaggi più brevi dalla chatbot è meno faticoso e più immediato. Inoltre, gli utenti della chatbot hanno dedicato meno tempo all'intervento rispetto al gruppo di controllo. Questo evidenzia un vantaggio significativo, considerando i problemi di concentrazione e motivazione che spesso affliggono i pazienti in crisi di salute mentale. L'intervento tramite chatbot, con il suo formato accessibile e coinvolgente, può rappresentare una soluzione efficace e rapida per migliorare il benessere psicologico in tempi ridotti (Karkosz et al., 2023).

L'efficacia dell'uso delle chatbot nel ridurre i sintomi della depressione è ulteriormente confermata da uno studio che ha esaminato l'impatto della chatbot Woebot su studenti universitari affetti da depressione e disturbi d'ansia. Lo studio ha evidenziato come l'uso regolare e quotidiano di Woebot abbia portato a significativi miglioramenti dell'umore. I partecipanti hanno riportato cambiamenti positivi nei loro stati emotivi, sottolineando la capacità della chatbot di fornire supporto empatico e pratico.

Un partecipante ha affermato: "Sono rimasto davvero impressionato e sorpreso dalla differenza che il chatbot ha fatto nella mia vita quotidiana, aiutandomi a notare i tipi di pensieri che avevo e a cambiarli" (Fitzpatrick et al., 2017, pp. 2). Questa testimonianza riflette la percezione diffusa tra gli utenti che Woebot non solo offre consigli utili, ma lo fa in modo empatico e comprensivo, simile a un'interazione umana. Lo studio dimostra che chatbot come Woebot possono svolgere un ruolo fondamentale nel migliorare il benessere psicologico degli studenti universitari, spesso sottoposti a situazioni stressanti. Grazie al supporto continuo offerto da queste chatbot, gli studenti possono sviluppare strategie di coping più efficaci, integrandole facilmente nella loro routine quotidiana.

Infine, l'introduzione della chatbot Woebot sembra rappresentare un passo avanti innovativo in particolare nella gestione della depressione post-partum (PPD), attraverso l'integrazione di terapie come la terapia cognitivo-comportamentale (CBT) e la terapia interpersonale (IPT) in un'applicazione mobile accessibile e personalizzabile. Questo strumento non solo ha dimostrato

la sua efficacia nel diminuire i sintomi della PPD, ma è progettato specificamente per adattarsi alle esperienze individuali delle madri, fornendo una psicoeducazione su misura che risponde alle loro esigenze e obiettivi personali. Infatti, l'utente può scegliere liberamente le aree di intervento che vanno dal miglioramento del legame con il bambino alla gestione dello stress e dell'ansia, in un'interfaccia centrata sull'utente. Questo non solo promuove un coinvolgimento attivo nel processo terapeutico, ma affronta anche le sfide specifiche che le madri affrontano durante il periodo peri- e post-partum, come la percezione di inadeguatezza, la solitudine e il cambiamento dell'identità personale. In aggiunta, Woebot potrebbe rappresentare una risorsa particolarmente preziosa in società dove la percezione delle donne di essere giudicate o stigmatizzate come "inadeguate" o "incapaci" di affrontare il ruolo materno può dissuaderle dal cercare trattamenti tradizionali faccia a faccia. La chatbot, attraverso il suo approccio empatico ed accessibile, potrebbe quindi superare queste barriere, offrendo un ambiente di supporto confidenziale e personalizzato direttamente dalle proprie abitazioni (Darcy et al., 2023).

3.2.2 Stress e resilienza

Un altro studio ha esplorato come interventi psicologici attuati dall'agente relazionale Woebot possano modificare i livelli di stress, resilienza e burnout sia in individui clinicamente sofferenti che in quelli non clinicamente sofferenti. Il trattamento si è basato su un arco temporale di otto settimane e i risultati sono stati promettenti ed incoraggianti. L'esposizione all'agente è stata infatti associata a significative riduzioni dello stress percepito e del burnout e anche ad un aumento della resilienza sia nelle persone con problematiche sintomatologiche cliniche che non (Durden et al., 2023). Questo potenziale effetto potrebbe essere molto importante dato che la resilienza è negativamente correlata con il disagio psicologico, tra cui depressione e ansia, ed è positivamente correlata con il benessere psicologico come la soddisfazione della vita e le emozioni positive (Hu et al., 2015 citato da Durden et al., 2023). Questo studio invece ha esaminato l'efficacia degli interventi con chatbot nella riduzione della solitudine e dello stress tra gli adulti durante la pandemia di COVID-19, con particolare attenzione agli individui di età superiore ai 65 anni. I risultati hanno mostrato che l'uso della chatbot ha portato a una diminuzione significativa dei sentimenti di solitudine nei partecipanti più anziani, con una riduzione del punteggio medio. Tuttavia, per i partecipanti di età inferiore ai 65 anni, il punteggio è rimasto stabile, indicando un effetto limitato della chatbot su questa fascia di età. Le ragioni della differenza di efficacia tra i gruppi di età potrebbero essere legate allo stress elevato e alle maggiori opportunità di interazione sociale per i partecipanti più giovani. Lo studio ha dimostrato che le chatbot possono offrire benefici significativi per la

salute mentale degli adulti più anziani, specialmente durante periodi di isolamento come la pandemia (Chou et al., 2023).

In conclusione, le chatbot sembrano rappresentare uno strumento promettente per supportare la salute mentale degli adulti più anziani, offrendo cure, compagnia e riduzione della solitudine. Per riassumere, questi studi collettivamente suggeriscono che gli interventi psicologici mediati dalla tecnologia, come quelli offerti dalle chatbot relazionali, possono essere efficaci nel migliorare vari aspetti della salute mentale, riducendo lo stress, il burnout e migliorando la resilienza e il benessere psicologico in diverse popolazioni.

3.2.3 *Disturbi alimentari*

La ricerca ha chiaramente evidenziato il potenziale delle chatbot nella prevenzione e nel trattamento dei disturbi alimentari (DA). Chatbot come KIT sono già state progettate specificamente per sostenere individui con preoccupazioni riguardanti l'immagine corporea e l'alimentazione. Esse offrono psicoeducazione, informazioni su come cercare aiuto, competenze di coping che includono strategie per gestire i social media, pratiche di mindfulness e incentivano a un'attività fisica piacevole (Beilharz et al., 2021).

In aggiunta, programmi online come *StudentBodies*, basato sulla terapia cognitivo-comportamentale e della durata di 8 settimane, hanno dimostrato di essere efficaci, specialmente quando guidati da professionisti. Tuttavia, l'impiego di moderatori umani limita la scalabilità del programma a causa dei costi elevati. Una soluzione proposta da Fitzsimmons-Craft et al. (2021) è l'integrazione del programma in una chatbot, consentendo così una riduzione dei costi e una moderazione continua tramite interazioni che simulano quelle umane. I risultati hanno indicato che questa chatbot ha portato a una significativa diminuzione dei principali fattori di rischio d'insorgenza tra le donne di sviluppare disturbi alimentari, come preoccupazioni per il peso e l'internalizzazione di ideali di magrezza. Inoltre, si è osservata una riduzione generale della psicopatologia, inclusi sintomi di ansia e depressione, aumentando così la probabilità che le partecipanti nel gruppo di intervento mantenessero uno stato non clinico. Tuttavia, è importante notare che un problema persistente è rappresentato dal basso tasso di coinvolgimento, con solo il 60% dei partecipanti che ha effettivamente iniziato a utilizzare la chatbot. Un altro studio ha esaminato i tassi di iscrizione e utilizzo di una chatbot per il trattamento dei disturbi alimentari, implementata all'interno di uno screening online della NEDA. Sebbene il 90,2% dei partecipanti abbia iniziato una conversazione con la chatbot, solo il 3% ha partecipato attivamente allo studio. Questa scarsa adesione può essere attribuita allo stigma associato, alla bassa motivazione, agli atteggiamenti negativi verso il trattamento e alla

negazione della gravità dei sintomi, che rappresentano ostacoli significativi nell'accesso al supporto necessario (D'Adamo et al., 2024).

Questi risultati non devono essere interpretati come limitazioni, ma piuttosto come sfide che richiedono l'implementazione di strategie mirate per migliorare l'adesione, specialmente tra i giovani adulti e coloro che seguono regimi dietetici restrittivi. L'integrazione efficace di tali tecnologie potrebbe significativamente aumentare l'accessibilità ai servizi di salute mentale, soprattutto per coloro che potrebbero evitare di cercare aiuto a causa di barriere sociali come pregiudizi, isolamento e paura.

È stato ampiamente documentato che le persone con disturbi alimentari spesso esitano a chiedere aiuto, contribuendo a un ritardo nel trattamento e a un peggioramento dei sintomi. Pertanto, dimostrare che le chatbot possano migliorare la motivazione al trattamento, l'autoefficacia e la propensione a cercare un aiuto professionale potrebbe avere un impatto significativo sull'adozione dei trattamenti da parte di una vasta popolazione di adulti con disturbi alimentari (D'Adamo et al., 2024).

3.3 Sfide e limitazioni nell'utilizzo dell'IA nell'intervento psicologico

Sebbene l'integrazione della terapia digitale e virtuale nella pratica psicoterapeutica offra delle potenziali opportunità per la psicoterapia, solleva anche importanti questioni psicologiche, sociali ed etiche importanti da considerare con cautela nella pratica clinica.

3.3.1 L'alleanza terapeutica

Innanzitutto, emergono dubbi riguardo a determinati aspetti terapeutici di un incontro faccia a faccia che la realtà virtuale potrebbe non riuscire a riprodurre adeguatamente. In particolare, è importante sottolineare come il disagio mentale possa ridurre la motivazione di una persona a cercare aiuto, aggravato spesso da una significativa emarginazione sociale. Questa fuga sociale tende a intrecciarsi sempre di più con altre condizioni come la solitudine e l'isolamento. È stato provato che, in tali situazioni, la capacità di ristabilire connessioni sociali solide e rassicuranti si basa su uno dei pilastri principali di qualsiasi intervento psicologico: l'alleanza terapeutica (Brown & Halpern, 2021). Bordin teorizza questo concetto come l'atteggiamento collaborativo che presiede la terapia e consiste nell'accordo reciproco sugli obiettivi terapeutici da parte del cliente e il terapeuta (Bordin, 1994). Nell'ambito clinico invece, dove viene definita alleanza diagnostica anziché terapeutica, si tratta di una relazione compartecipata in cui il

paziente viene coinvolto attivamente nei fini e nel processo terapeutico (Lingiardi & Gazzillo, 2014).

Alcune ricerche hanno fornito prove empiriche che dimostrano come l'alleanza terapeutica durante l'intero percorso sia un fattore che facilita il successo della terapia al termine del trattamento (Flückiger et al., 2018). Una ragione per cui questa alleanza è utile per il successo terapeutico potrebbe essere che il fatto di andare dallo psicoterapeuta e instaurare un rapporto intimo con lui implica un primo contatto sociale. Questo contatto sociale non solo offre un'opportunità di connessione umana, ma fornisce anche benefici sociali iniziali e indiretti. Si può comprendere come il processo psicologico in una situazione di contatto diretto diventi un efficace strumento per contrastare l'isolamento e la solitudine, offrendo benefici che possono comprendere un senso di inclusione e un miglioramento del benessere emotivo grazie alla percezione di essere compresi e sostenuti da un professionista. Per questo motivo, si pensa che difficilmente sarebbe possibile da replicare attraverso l'uso di un agente digitale o virtuale, che avviene senza un reale contatto fisico (Brown & Halpern, 2021).

«Siamo preoccupati che le chatbot non possano fornire un supporto sociale significativo che migliori la ricerca di aiuto, la permanenza nel trattamento e, in ultima analisi, la ricostruzione delle connessioni sociali» (Ibidem, pp. 2).

A supporto di ciò, è stato dimostrato che in Giappone il diffuso fenomeno degli "Hikikomori", (persone emarginate dalla società trovano sempre più conforto nell'intelligenza artificiale e nei robot come chatbot) evidenzia chiaramente il limite delle chatbot. Uno studio sottolinea infatti come questo rifugio tecnologico deve essere considerato solamente come un sollievo temporaneo, incapace di replicare l'intimità e la connessione umana autentica. Affidarsi esclusivamente a queste soluzioni tecnologiche porta solo a evitare il problema reale e, a lungo termine, le conseguenze rimangono invariate, perpetuando l'isolamento sociale e l'emarginazione senza affrontare le cause profonde del disagio (J & L.B, 2024).

Tuttavia, altri studi hanno portato avanti l'idea che non ci sia una reale differenza nella percezione degli utenti riguardo alla relazione che viene a formarsi online. Uno studio mostra che per chi preferisce la terapia online per il supporto della salute mentale, è possibile costruire una solida alleanza di lavoro, indipendentemente dal mezzo di comunicazione utilizzato (Cook & Doyle, 2002).

Alcuni invece sottolineano l'impossibilità di trasferire la tradizionale visione dell'alleanza terapeutica negli ambienti digitali, in cui non dovrebbe neanche essere considerata in modi equivalenti (Tremain et al., 2020). In particolare, una revisione sistematica ha identificato come l'alleanza terapeutica che si forma in un intervento digitale porti con sé nuove dimensioni, tra

cui la disponibilità, per via del libero e comodo accesso alla terapia, e di interattività, indicando il grado in cui vengono fornite personalizzazione e feedback basati sull'input dell'utente. Inoltre, viene riferito come l'individuo in un ambiente digitale si senta più in controllo (Hillier, 2018). Altri, infine, ipotizzano addirittura che l'alleanza sia probabilmente meno importante nella terapia basata su Internet rispetto alle terapie standard faccia a faccia (Anderson et al., 2012 citato da Flückiger et al., 2018).

Alla luce di queste opinioni, sembra esserci la necessità di comprendere come l'instaurarsi di un'alleanza terapeutica dipenda in gran parte dalla soggettività dell'individuo. Infatti, persone con specifiche caratteristiche di personalità, come l'introversione, e problemi come il disturbo di panico e l'agorafobia, potrebbero preferire la riservatezza (anche corporea) e il senso di controllo che la terapia online offre. Al contrario, molte altre persone continueranno a preferire e trovare conforto nelle terapie tradizionali faccia a faccia (Cook & Doyle, 2022).

3.3.2 *Empatia*

Secondo Montemayor, Halpern e Fairweather, l'empatia rappresenta un limite intrinseco per l'IA nel campo dell'assistenza sanitaria, costituendo un ostacolo fondamentale. Questo perché, sebbene l'IA eccella nell'applicare regole inferenziali affidabili per utilizzare dati di un paziente specifico in base a conoscenze preesistenti su un paziente ipotetico, questo processo non equivale all'empatia umana. Quest'ultima è infatti un'attenzione empatica consapevole, radicata in esperienze mentali biologiche, sia coscienti che inconsce, e richiede la capacità di selezionare le informazioni più rilevanti in un contesto emozionale (Montemayor et al., 2021).

L'empatia umana trova una spiegazione più completa e accurata nell'ottica neuroscientifica, che individua la sua base neurologica e biologica nei neuroni specchio. Questi neuroni sono fondamentali perché permettono agli esseri umani di vivere direttamente le emozioni che osservano negli altri (Steca, 2021). Questa connessione neurologica e biologica offre una profondità di comprensione e reazione che appare fuori dalla portata dell'intelligenza artificiale. L'empatia umana non deve essere vista semplicemente come un processo cognitivo, ma coinvolge una risposta emotiva autentica e immediata. Quando una persona vede qualcun altro provare dolore o gioia, i neuroni specchio nel suo cervello si attivano in modo simile a come farebbero se stesse vivendo quelle stesse emozioni (ibidem). Questa capacità di risuonare emotivamente con gli altri è ciò che rende l'empatia umana così profonda e insostituibile. Al contrario, l'IA, anche se programmata per riconoscere e rispondere alle emozioni, opera su base algoritmica e inferenziale. Le sue risposte si basano su modelli di dati e regole predefinite, mancando così della componente esperienziale e sensoriale che caratterizza l'empatia umana.

L'empatia simulata dai neuroni artificiali dell'IA potrebbe addirittura risultare manipolativa per il paziente. Le risposte generate nel cervello sociale del ricevente, responsabili della risposta empatica, potrebbero essere inappropriate senza un agente biologico emotivamente sintonizzato (Montemayor et al., 2021). In altre parole, l'IA può replicare il comportamento empatico, ma non può vivere l'emozione che sostiene questo comportamento.

Questa mancanza di autentica empatia rappresenta una delle sfide più grandi nello sviluppo di terapeuti digitali nei contesti di cura, dove il terapeuta deve essere in grado di entrare in sintonia con i sentimenti e le esperienze del paziente. Infatti, un terapeuta digitale o virtuale, anche se incarnato in un corpo virtuale, sembra fornire solo un supporto psicologico emotivamente distaccato e superficiale, con il rischio di una riduzione della fiducia e dell'efficacia terapeutica (Montemayor et al., 2021).

3.3.3 *Il concetto di Transfert*

Vale la pena considerare un concetto fondamentale spesso associato alle terapie psicoanalitiche e psicodinamiche ma che viene analizzato anche all'interno delle terapie cognitivo-comportamentali: il transfert o traslazione. Esso rappresenta il modo in cui i pazienti portano con sé le loro esperienze passate, relazioni e dinamiche nei loro incontri terapeutici. Queste influenze pregresse sono proiettate nella relazione terapeutica attuale, contribuendo a plasmare e a influenzare il modo in cui il paziente interagisce con il terapeuta. Se adeguatamente esaminato, potrebbe costituire un vantaggio per lo sviluppo della relazione terapeutica, sostenendo l'alleanza tra paziente e terapeuta. Altrimenti, potrebbe compromettere il successo del trattamento. Per comprendere come questo concetto possa manifestarsi nelle terapie supportate da strumenti di intelligenza artificiale, potrebbe essere utile reinterpretarlo come passaggio da una relazione tra esseri umani a una tra essere umano e terapeuta bot (Holohan & Fiske, 2021). Quando si parla di traslazione nel contesto della teoria del realismo agenziale di Karen Barad, ci si riferisce al fenomeno attraverso il quale le relazioni terapeutiche emergono e si formano all'interno di un apparato specifico. Questo apparato è composto da pratiche materiale-discorsive che includono non solo il terapeuta e il paziente, ma anche l'ambiente fisico della stanza di consultazione, i dispositivi tecnologici utilizzati, e le modalità specifiche di interazione. Secondo Barad, queste pratiche non sono semplici strumenti neutri, ma piuttosto agenti attivamente coinvolti nella produzione e nella definizione della realtà stessa.

L'introduzione di tecnologie avanzate, come l'intelligenza artificiale (IA), nel contesto della psicoterapia porta a una ridefinizione significativa di questo apparato terapeutico. Quando la terapia è guidata dall'IA, nuovi elementi vengono introdotti nell'apparato: chatbot terapeutiche,

interfacce digitali, algoritmi di intelligenza artificiale che analizzano e rispondono ai dati dei pazienti. Questo nuovo apparato non solo influisce sulle dinamiche di interazione tra terapeuta e paziente, ma può anche ridefinire cosa significhi trasferire esperienze e dinamiche personali all'interno di una terapia. Pertanto, l'evoluzione dell'apparato terapeutico attraverso l'integrazione dell'IA comporta una ridefinizione non solo del concetto di traslazione, ma anche delle modalità stesse con cui avviene la terapia. Questo solleva la questione fondamentale di come il concetto di transfert si trasformi o se sia ancora pertinente, dato l'emergere di nuovi fenomeni nell'ambito di questo unico apparato terapeutico. Questi quesiti necessitano di indagini empiriche costanti, ma ciò che emerge è l'importanza di studiare come individui umani e non umani interagiscono nei loro incontri e quali effetti producono, evitando di relegare la tecnologia a un ruolo passivo nell'interazione (Holohan & Fiske, 2021).

3.3.4 *Problemi etici*

L'integrazione dell'Intelligenza Artificiale nella pratica psicologica solleva una serie di problematiche etiche rilevanti. Emergono preoccupazioni riguardo alla capacità delle chatbot di prendere decisioni morali, influenzate dal modo in cui le norme sociali sono incorporate negli algoritmi. Questi algoritmi possono essere basati su dataset che, per loro natura, possono essere strutturalmente distorti (Holohan & Fiske, 2021).

Ciò solleva il dubbio che l'IA non sia effettivamente in grado di condurre riflessioni etiche che permettano di compiere scelte moralmente corrette. Questo perché i sistemi artificiali sono programmati per fornire risposte univoche a situazioni morali, secondo le indicazioni degli esperti in problemi etici. Tuttavia, non sempre esiste una soluzione definitiva e chiara: seguire rigidamente un principio etico in ogni circostanza potrebbe impedire di sostituire un principio morale con un altro quando necessario per una decisione eticamente appropriata. Inoltre, la scelta tra diverse teorie etiche rappresenta di per sé un atto morale, implicando il riconoscimento delle proprie convinzioni su ciò che è giusto e ciò che è sbagliato (Ferdynus, 2023).

Per esemplificare, consideriamo un dilemma etico in cui un paziente esprime l'intenzione di nuocere al proprio coniuge. In questo caso, un professionista potrebbe essere moralmente obbligato a violare il segreto professionale per proteggere la sicurezza del coniuge o degli altri. Tuttavia, un sistema di IA dovrebbe essere in grado di valutare accuratamente i rischi e prendere decisioni etiche adeguate, il che solleva interrogativi sull'effettiva capacità dell'IA di comprendere e agire per il bene del paziente (Stan, 2021). Ciò potrebbe essere irrealizzabile se si considera l'idea, proposta dal filosofo John Searle, che l'intelligenza artificiale manchi di intenzionalità, una qualità fondamentale per selezionare tra diverse opzioni morali. Questo

fenomeno si verifica perché i significati assegnati ai programmi di intelligenza artificiale derivano solamente dagli esseri umani, il che li rende casuali o arbitrari per il programma stesso, privandolo di una vera comprensione. Questa mancanza di comprensione etica potrebbe limitare l'IA nel ruolo di agente morale che è chiamata ad agire per il bene del paziente o degli altri (Ferdynus, 2023).

Oltre a ciò, è essenziale considerare gli impatti psicologici sui pazienti derivanti dall'interazione con robot umanoidi o sistemi avanzati di intelligenza artificiale. L'effetto ELIZA ha dimostrato che alcuni pazienti potrebbero sviluppare un attaccamento emotivo al robot o essere profondamente turbati dall'interazione, aumentando il rischio di danni emotivi a sé stessi o ad altri. Questo evidenzia la necessità professionale di definire delle rigorose norme etiche nella relazione terapeutica tra robot e paziente. Inoltre, è fondamentale che le chatbot e altri sistemi di IA utilizzati in ambito terapeutico siano dotati di competenze adeguate. Questi sistemi devono non solo essere capaci di interagire in modo empatico e comprensivo con i pazienti, ma anche di gestire situazioni emotivamente delicate in maniera eticamente responsabile. Ciò richiede non solo una programmazione tecnologicamente avanzata, ma anche una progettazione eticamente informata. La supervisione umana dei terapeuti digitali e virtuali diventa pertanto cruciale per garantire che tali sistemi operino nel rispetto dei più alti standard etici e professionali, proteggendo la salute e il benessere dei pazienti (Stan, 2021).

Un'altra significativa questione etica è legata al rispetto della privacy e della fiducia dell'utente/paziente durante le sedute digitali. Rispetto alle tradizionali conversazioni in studio, le interazioni digitali possono comportare la raccolta e la registrazione di una vasta gamma di dati personali, compresi elementi visivi, vocali e psicologici, senza sempre il pieno consenso degli individui coinvolti. Ciò aumenta il rischio di violazioni della privacy e di furto elettronico dei dati, con potenziali conseguenze dannose per gli utenti (Ibidem).

Un'ultima questione rilevante riguarda il grado di realismo dei sistemi automatizzati (IA) e la percezione di umanità che trasmettono. Questo aspetto ha significative implicazioni etiche, poiché il livello di realismo può influire sull'efficacia terapeutica e potrebbe addirittura risultare dannoso. Sistemi che appaiono troppo umani potrebbero ingannare pericolosamente persone psicologicamente vulnerabili, come quelle affette da demenza o disturbi psicotici. Tali individui potrebbero essere portati a credere che la macchina sia viva o controllata da una forza malevola, causando ulteriore confusione e potenzialmente aggravando la loro condizione.

Masahiro Mori, uno studioso giapponese, ha introdotto il concetto di "valle misteriosa" per descrivere come le reazioni umane ai robot cambino in funzione del loro aspetto. Secondo questa teoria, man mano che un robot diventa più simile a un essere umano, le persone tendono

a provare sentimenti di empatia e accettazione. Tuttavia, una volta che il robot raggiunge un certo livello di somiglianza, ma senza essere completamente umano, si verifica un'inversione di tendenza: le persone cominciano a provare una sensazione di disagio e repulsione. Questo fenomeno è particolarmente rilevante in ambito clinico, dove l'obiettivo è creare un ambiente di fiducia e sicurezza per il paziente. La sfida principale sarà quindi capire quale livello di realismo sia accettabile per i pazienti senza provocare effetti spiacevoli (Stan, 2021). La ricerca si dovrà quindi concentrare su come assicurare un equilibrio tra l'umanità percepita del sistema e la sua capacità di svolgere efficacemente il proprio ruolo terapeutico, al fine di non compromettere il benessere psicologico dei pazienti; infatti, troppo realismo potrebbe indurre reazioni avverse, mentre un realismo insufficiente potrebbe ridurre l'efficacia del trattamento.

CONCLUSIONI

In questa tesi, si è esplorato in dettaglio il ruolo sempre più rilevante dell'intelligenza artificiale (IA) nelle terapie psicologiche, con particolare attenzione al potenziale dei terapeuti digitali e virtuali nel migliorare l'accessibilità, l'efficacia e l'esperienza complessiva del trattamento psicologico. Il percorso è iniziato analizzando i primi sviluppi concettuali dell'IA, dall'idea di Alan Turing di una "macchina pensante" fino alle attuali applicazioni di machine learning e reti neurali, che hanno trasformato radicalmente il panorama della salute mentale.

Nel capitolo dedicato al rapporto tra uomo e IA, si è esplorato il delicato equilibrio tra emozioni umane e intelligenza artificiale. L'effetto Eliza, un esempio storico di interazione tra esseri umani e sistemi automatizzati nella psicoterapia, ha evidenziato le sfide e le opportunità nel simulare l'empatia e la comprensione emotiva attraverso le macchine. Questa riflessione ha condotto a considerare come le tecnologie digitali possano essere integrate efficacemente nel contesto terapeutico senza compromettere l'essenza stessa dell'interazione clinica umana.

Nella sezione centrale della tesi, si è approfondito l'efficacia dei terapeuti digitali nel trattamento di disturbi psicologici comuni come ansia, depressione, stress e disturbi alimentari. Le evidenze scientifiche hanno mostrato che le chatbot e altre forme di IA possono significativamente ridurre i sintomi, migliorare la resilienza e fornire supporto emotivo continuo. La possibilità di consultare un terapeuta digitale gratuitamente e comodamente sembra rappresentare una risorsa cruciale per impedire che un disagio iniziale si evolva in una malattia mentale più grave, almeno fino a quando non si riesce a incontrare un professionista. Questa opportunità offre un supporto immediato e accessibile, riducendo le barriere di accesso alle cure psicologiche, come i costi elevati e la disponibilità limitata di terapisti. Inoltre, l'uso dell'intelligenza artificiale nel campo clinico si dimostra un alleato valido nella prevenzione della salute mentale, offrendo interventi tempestivi e personalizzati che possono alleviare i sintomi e migliorare il benessere generale.

Tuttavia, nonostante i vantaggi evidenti, si sono esplorate anche diverse sfide e limitazioni emerse nell'uso dell'IA nell'intervento psicologico. Concetti cruciali come l'alleanza terapeutica, l'empatia e il transfert sono pilastri irrinunciabili della pratica clinica che sembrano richiedere l'interazione umana per essere pienamente sviluppati. Tuttavia, è stato evidenziato che l'assenza di tali elementi potrebbe compromettere sia l'efficacia che la qualità delle cure psicologiche fornite dai terapeuti digitali, ma potrebbe anche favorire l'evoluzione di nuovi spazi terapeutici e l'emergere di nuove dimensioni. Queste nuove dimensioni, tuttavia, necessitano ancora di una adeguata esplorazione attraverso studi empirici.

Inoltre, problemi etici come la sicurezza dei dati, l'accuratezza delle diagnosi e la trasparenza nell'uso dell'IA sono diventati centrali nella discussione sull'integrazione di queste tecnologie nel settore della salute mentale. È cruciale affrontare queste questioni con rigore e responsabilità per garantire che l'adozione dell'IA sia sempre orientata al benessere dei pazienti.

Guardando al futuro, si auspica che l'evoluzione delle tecnologie digitali porti a un'armoniosa coesistenza tra terapia umana e strumenti digitali. È doveroso precisare come l'obiettivo non sia sostituire il terapeuta umano, ma arricchire e potenziare la pratica clinica attraverso l'uso complementare dell'IA. Questo approccio collaborativo potrebbe contribuire ad ampliare l'accesso alle cure psicologiche, migliorare l'efficacia dei trattamenti e personalizzare l'assistenza in modo più mirato ed efficiente (Stan, 2021).

In conclusione, l'integrazione dell'intelligenza artificiale nelle terapie psicologiche rappresenta una frontiera promettente ma complessa. È essenziale adottare un approccio bilanciato che riconosca e valorizzi le competenze umane nella pratica clinica, mentre sfrutta appieno le potenzialità delle tecnologie digitali per ottimizzare il benessere mentale e migliorare la qualità della vita delle persone. Questo futuro collaborativo richiederà un impegno continuo nella ricerca. Dovrà essere premura impegnarsi nello sviluppo etico delle tecnologie riguardo alle questioni emerse e nella formazione di professionisti preparati ad affiancare e supervisionare il lavoro dei terapeuti digitali e virtuali, in un nuovo ambiente terapeutico sempre in continuo mutamento, più interconnesso e tecnologicamente avanzato.

BIBLIOGRAFIA

- A.M. Turing, Computing machinery and intelligence, *Mind*, 59 (1950) 433-460. Traduzione italiana in: V. Somenzi, R. Cordeschi, La filosofia degli automi. Origini dell'intelligenza artificiale, Paolo Boringhieri, Torino, 1986, pp. 157-183
- Beilharz, F., Sukunesan, S., Rossell, S. L., Kulkarni, J., & Sharp, G. (2021). Development of a Positive Body Image Chatbot (KIT) With Young People and Parents/Carers: Qualitative Focus Group Study. *Journal of Medical Internet Research*, 23(6), Articolo e27807. <https://doi.org/10.2196/27807>
- Bordin, E. S. (1994). Theory and research on the therapeutic working alliance: New directions. In A. O. Horvath & L. S. Greenberg (Eds.), *The Working alliance: Theory, research, and practice* (pp. 13–37). New York, NY: Wiley.
- Brown, J. E. H., & Halpern, J. (2021). AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare. *SSM - Mental Health*, 1, 100017. <https://doi.org/10.1016/j.ssmmh.2021.100017>
- Chou, Y.-H., Lin, C., Lee, S.-H., Lee, Y.-F., & Cheng, L.-C. (2023). Utilizing a User-Friendly Chatbot to Mitigate the Psychological Stress of Older Adults During the COVID-19 Pandemic: A Pilot Study (Preprint). *JMIR Formative Research*. <https://doi.org/10.2196/49462>
- Cook, J. E., & Doyle, C. (2002). Working Alliance in Online Therapy as Compared to Face-to-Face Therapy: Preliminary Results. *CyberPsychology & Behavior*, 5(2), 95–105. <https://doi.org/10.1089/109493102753770480>
- D'Adamo, L., Grammer, A. C., Rackoff, G. N., Shah, J., Firebaugh, M., Taylor, C. B., Wilfley, D. E., & Fitzsimmons-Craft, E. E. (2024). Rates and correlates of study enrolment and use of a chatbot aimed to promote mental health services use for eating disorders following online screening. *European Eating Disorders Review*. <https://doi.org/10.1002/erv.3082>
- Damasio, A. (2022). *Sentire e conoscere*. Adelphi.
- Darcy, A., Beaudette, A., Chiauuzzi, E., Daniels, J., Goodwin, K., Mariano, T. Y., Wicks, P., & Robinson, A. (2023). Anatomy of a Woebot® (WB001): agent guided CBT for women with postpartum depression. *Expert Review of Medical Devices*. <https://doi.org/10.1080/17434440.2023.2280686>
- Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional,

- Retrospective Observational Study. *JMIR Formative Research*, 5(5), Articolo e27868. <https://doi.org/10.2196/27868>
- Durden, E., Pirner, M. C., Rapoport, S. J., Williams, A., Robinson, A., & Forman-Hoffman, V. L. (2023). Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent “Woebot”. *Internet Interventions*, 33, 100637. <https://doi.org/10.1016/j.invent.2023.100637>
- Ferdynus, M. P. (2023). Five reasons why a conversational artificial intelligence cannot be treated as a moral agent in psychotherapy. *Archives of Psychiatry and Psychotherapy*, 25(4), 26–29. <https://doi.org/10.12740/app/170132>
- Fissi, S. (2015). La coscienza affettiva. Emozione e cognizione nel determinismo della coscienza. *Atque materiali tra filosofia e psicoterapia*, (17), 143-167.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2), Articolo e19. <https://doi.org/10.2196/mental.7785>
- Fitzsimmons-Craft, E. E., Chan, W. W., Smith, A. C., Firebaugh, M., Fowler, L. A., Topooco, N., DePietro, B., Wilfley, D. E., Taylor, C. B., & Jacobson, N. C. (2021). Effectiveness of a chatbot for eating disorders prevention: A randomized clinical trial. *International Journal of Eating Disorders*, 55(3), 343–353. <https://doi.org/10.1002/eat.23662>
- Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55(4), 316–340. <https://doi.org/10.1037/pst0000172>
- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. (trad. it. *Formae mentis. Saggio sulla pluralità dell'intelligenza*, Feltrinelli, 2013)
- Hayles, N. Katherine. *Come siamo diventati postumani: corpi virtuali in cibernetica, letteratura e informatica*. Chicago: Università di Chicago Press, 1999.
- Hillier L. (2018). Exploring the Nature of the Therapeutic Alliance in Technology-based Interventions for Mental Health Problems. Lancaster EPrints. 2018. URL: <https://eprints.lancs.ac.uk/id/eprint/127663/1/2018hilliermphil.pdf>
- Holohan, M., & Fiske, A. (2021). “Like I’m Talking to a Real Person”: Exploring the Meaning of Transference for the Use and Design of AI-Based Applications in Psychotherapy. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.720476>

- J, G., & L.B, K. (2024). Artificial Intelligent to Combat Loneliness and Social Isolation in Japan an Aid or Cure . *Preprints*, 2024021662. <https://doi.org/10.20944/preprints202402.1662.v1>
- Karkosz, S., Szymański, R., Sanna, K., & Michałowski, J. (2023). Effectiveness of a Web-based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: A Randomized Controlled Trial (Preprint). *JMIR Formative Research*. <https://doi.org/10.2196/47960>
- Lingiardi, V., & Gazzillo F. (2014). *La personalità e i suoi disturbi: valutazione clinica e diagnosi al servizio del trattamento*. Raffaello Cortina Editore.
- Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100. <https://doi.org/10.1016/j.chb.2014.04.043>
- Martínez-Miranda, J., & Aldea, A. (2005). Emotions in human and artificial intelligence. *Computers in Human Behaviour*, 21(2), 323–341. <https://doi.org/10.1016/j.chb.2004.02.010>
- Marty, N. (2023). *Intelligenza Artificiale: il futuro è ora*. S.n.
- Montemayor, C., Halpern, J., & Fairweather, A. (2021). In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-021-01230-z>
- Morbini, F., DeVault, D., Georgila, K., Artstein, R., Traum, D., & Morency, L.-P. (2014). A Demonstration of Dialogue Processing in SimSensei Kiosk. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Association for Computational Linguistics. <https://doi.org/10.3115/v1/w14-4334>
- Murphy, P. (2017). "You Feel Real to Me, Samantha": The matter of Technology in Spike Jonze's Her. *Technoculture: an online journal of technology in society*, 7.
- Natale, S. (2022). *Macchine ingannevoli: Comunicazione, tecnologia, intelligenza artificiale*. Einaudi.
- Rizzo, A., Scherer, S., DeVault, D., Gratch, J., Artstein, R., Hartholt, A., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Stratou, G., Traum, D., Wood, R., Boberg, J., & Morency, L. P. (2016). Detection and computational analysis of psychological signals using a virtual human interviewing agent. *Journal of Pain Management*. *Journal of Pain Management*, 9 (3),311-321. https://www.researchgate.net/publication/316476437_Detection_and_computational_analysis_of_psychological_signals_using_a_virtual_human_interviewing_agent#fullTextFileContent

- Romanovskyi, O. Pidbutska, N., & Knysh, A. (2021). Elomia Chatbot: the Effectiveness of Artificial Intelligence in the Fight for Mental Health. In Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems. CEUR Workshop Proceedings.
- Romeo, P. (2021). *La rivoluzionaria teoria delle Intelligenze Multiple: Alcune valutazioni critiche*. Youcanprint.
- Stan, B. (2021). *Psicologia e nuove tecnologie*. Le Parche Edizioni.
- Steca, P. (2021). Le emozioni. In P. Cherubini, E. Bricolo & C. Reverberi (A cura di), *Psicologia generale*. Raffaello Cortina Editore.
- Tormen, F. (2023). *Psicologia dell'intelligenza artificiale: La mente umana e la macchina*.
- Tremain, H., McEnery, C., Fletcher, K., & Murray, G. (2020). The Therapeutic Alliance in Digital Mental Health Interventions for Serious Mental Illnesses: Narrative Review. *JMIR Mental Health*, 7(8), Articolo e17204. <https://doi.org/10.2196/17204>

SITOGRAFIA

- Abate T. (2014). *Stanford bioengineers create circuit board modeled on the human brain*. Stanford News. <https://news.stanford.edu/pr/2014/pr-neurogrid-boahen-engineering-042814.html>
- AGI. (2023). *Ora l'intelligenza artificiale prevede le emozioni umane*. <https://www.agi.it/scienza/news/2023-06-09/intelligenza-artificiale-prevede-emozioni-umane-21722126/#:~:text=AGI%20%2D%20Un%20modello%20di%20intelligenza,intelligenza%20sociale%20degli%20osservatori%20umani>
- Alberto, P. (2022). *Può una macchina provare emozioni?* Cyber Ducks. <https://cyberducks.it/it/iot/intelligenza-artificiale/190-puo-una-macchina-provare-emozioni.html>
- Alessandri, P. (2023). *Effetto Eliza, innamorarsi di un'AI*. 01net. <https://www.01net.it/le-voci-dellai-episodio-06-effetto-eliza-innamorarsi-di-unai/>
- Bradley, S. (2023). I tried the Woebot AI Therapy app to see if it would help my anxiety and OCD. *Verywell Mind*. <https://www.verywellmind.com/i-tried-woebot-ai-therapy-app-review-7569025>

- Catarci, T. (2022). *Il manifesto di Vienna sull'umanesimo digitale e il libro che spiega l'interazione tra scienza e cultura umanistica*. ReWriters. <https://rewriters.it/il-manifesto-di-vienna-sul-digital-humanism-e-libro-che-spiega-interazione-tra-scienza-e-cultura-umanistica/>
- Columbro, D. (2023). Le macchine e l'inganno banale, ma necessario, dell'intelligenza artificiale. *La Stampa*. https://www.lastampa.it/rubriche/data-storie/2023/03/31/news/le_macchine_e_linganno_banale_ma_necessario_dellintelligenza_artificiale-12730543/
- Dara, V. (2017). Mass media: cinquant'anni fa usciva in Italia un testo rivoluzionario di McLuhan. *InsideMarketing*. <https://www.insidemarketing.it/mass-media-mcluhan-e-understanding-media/>
- El Attilah, I. (2023). Discussione sul clima: il chatbot spinge un uomo al suicidio. L'Intelligenza Artificiale fa paura. *Euronews*. <https://it.euronews.com/next/2023/04/01/discussione-sul-clima-chatbot-spinge-un-uomo-al-suicidio-intelligenza-artificiale-fa-paura>
- Frasso, E. (2024). *Un anno di ChatGPT: a che punto è la rivoluzione AI?* AI News. <https://ainews.it/un-anno-di-chatgpt-a-che-punto-e-la-rivoluzione-ai/#:~:text=Un%20anno%20fa%2C%20il%2030,da%20un%20milione%20di%20utenti.>
- HDmotori.it. (2018). *Affettiva: Intelligenza Artificiale per ridurre la distrazione di chi guida*. HDmotori.it. <https://www.hdmotori.it/2018/12/03/affettiva-ai-comportamento-persone-guida-emozioni/>
- IDEGO Psicologia Digitale. (2017). *Ecco Pepper: il robot che aiuta a superare l'ansia*. <https://www.idego.it/e-therapy/ecco-pepper-il-robot-che-aiuta-a-superare-lansia/>
- Insalutenews.it. (2020). *Declino funzionale e cognitivo degli anziani, il robot umanoide Pepper supporta medici e familiari*. <https://www.insalutenews.it/in-salute/declino-funzionale-e-cognitivo-degli-anziani-il-robot-umanoide-pepper-supporta-medici-e-familiari/>
- Isp. (2018). *La comunicazione persuasiva in psicoterapia strategica — Istituto per lo Studio delle Psicoterapie*. Istituto per Lo Studio Delle Psicoterapie. <https://www.istitutopsicoterapie.com/la-comunicazione-persuasiva-psicoterapia-strategica/>
- Morichetti, F. (2022). Replika: per due settimane sono diventata amica di un'intelligenza artificiale. *Weekly*. <https://weeklyweb.it/article/replika-app-che-vuole-essere-tua-amica>
- Nardone, G., & Chiodini, M. (2021). L'arte di comunicare efficacemente: la persuasione. *Psicologia contemporanea*, (285). <https://www.psicologiacontemporanea.it/blog/larte-di-comunicare-efficacemente-la-persuasione/>

- Network digital 360. (2023). *Robot umanoide prodotto da Softbank Robotics, in grado di interagire con l'ambiente e con gli esseri umani grazie a tecniche innovative di machine learning. Può essere utilizzato in molteplici contesti, per le più svariate finalità: miglioramento dell'esperienza del cliente, vendita al dettaglio, accoglienza e ospitalità, customer engagement.* Ai4business. <https://www.ai4business.it/robotica/robot-pepper-come-funziona-caratteristiche-e-novita/>
- Pisa, P. L. (2022). L'errore più grande di chi ha creato l'intelligenza artificiale. *La Repubblica*. https://www.repubblica.it/tecnologia/2022/04/06/news/i_sentimenti_delle_macchine_saranno_diversi_da_cio_che_immaginiamo-342046531/
- Pulina, G. (2023). *Intelligenza Artificiale e mondo postumano*. Leurispes. <https://www.leurispes.it/intelligenza-artificiale-e-mondo-postumano/>
- Rozzo, E. (2023). *Arte e tecnologia: le contraddizioni pervasive*. Tutte Quelle Cose. <https://www.tuttequellecose.com/arte-e-tecnologia-le-contraddizioni-persasive/>
- Taglienti, S. (2016). *L'intelligenza di Alan Turing*. FIDAF WEBZINE. <https://www.fidaf.it/index.php/lintelligenza-di-alan-turing/>
- USC ICT. (2013). *SimSensei & MultiSense: Virtual Human and Multimodal Perception for Healthcare Support*[Video]. YouTube. <https://www.youtube.com/watch?v=ejczMs6b1Q4>