EMPATHIES: HUMAN AND DIGITAL BODIES

An interdisciplinary approach to enhancing human–conversational agent interaction

ECAs, nonverbal communication, human-computer interaction, design, AI

EMPATIE:
CORPI UMANI E DIGITALI

Un progetto interdisciplinare per migliorare l'interazione tra persone e agenti conversazionali

ECAs, comunicazione non verbale, human-computer interaction, design, AI

Alessia Nicoletta Marino [1], David Landi [2], Enrico Randellini [2]

[1] Università degli Studi di Siena / Università della Campania Luigi Vanvitelli
[2] QuestIT s.r.l

alessianicolettamarino@unicampania.it, d.landi@quest-it.com, randellini@quest-it.com

Abstract

L'evoluzione degli assistenti virtuali incarnati (ECAs) apre nuove prospettive nella comunicazione uomo-macchina, combinando linguaggio verbale e non verbale per favorire un'interazione più empatica. Il seguente studio adotta un approccio interdisciplinare per sviluppare un algoritmo capace di riconoscere gesti di saluto, accordo e disaccordo. La ricerca si fonda sulla raccolta di dati e sull'impiego di stimoli emozionali, come espressioni facciali degli avatar e musica, per suscitare risposte non verbali utili all'addestramento del sistema. Pur non percependo emozioni umane, gli ECAs possono offrire un'empatia razionale, ottimizzando l'esperienza utente e favorendo interazioni più fluide e naturali. Gli avanzamenti tecnologici permettono ai designer di avere nuovi strumenti per progettare interazioni efficaci.

The evolution of embodied conversational agents (ECAs) opens new perspectives in human-computer communication, combining verbal and non-verbal language to foster more empathetic interaction. The following study adopts an interdisciplinary approach to develop an algorithm capable of recognizing gestures of greeting, agreement, and disagreement. The research is based on data collection and the use of emotional stimuli, such as avatars' facial expressions and music, to elicit non-verbal responses useful for training the system. Although ECAs do not experience human emotions, they can provide rational empathy, optimize the user experience and enabling smoother, more natural interactions. Technological advancements give designers new tools for creating effective interactions.

Introduzione

L'empatia è la capacità di "partecipare alla posizione dell'altro", cioè comprendere e condividere lo stato emotivo delle altre persone. La comunicazione non verbale attraverso posture, gesti, espressioni facciali e tono della voce può trasmettere vicinanza emotiva e comprensione. Ad esempio, adottare la stessa postura dell'interlocutore può rafforzare il senso di connessione, così come un sorriso può influenzare positivamente l'esperienza emotiva di chi lo riceve (Heylighen et al., 2019). Questa stretta relazione tra empatia e linguaggio corporeo non riguarda solo le interazioni umane, ma ha un impatto significativo anche nel design delle tecnologie. Le qualità empatiche sono infatti fondamentali nel processo di design, perché permettono di individuare bisogni espliciti e impliciti per sviluppare soluzioni che rispondano alle esigenze e aspettative delle persone (Andresa et al., 2019). Con l'evoluzione tecnologica, i designer possono implementare questi aspetti anche nelle interazioni uomo-macchina, rendendole più naturali. Per ottenere questo risultato, è fondamentale adottare un modello comunicativo bidirezionale e multimodale che replichi la comunicazione umana, andando oltre il solo scambio verbale e includendo anche segnali emotivi veicolati da voce, espressioni facciali e postura (Costantini et al., 2019). Tra le tecnologie che incarnano queste caratteristiche troviamo gli Embodied Conversational Agents (ECAs), assistenti virtuali dall'aspetto umanoide. Il loro corpo virtuale può assumere forme diverse, ad esempio può essere realistico o cartoonesco, può mostrare solo il volto oppure l'intera figura. Il corpo umano, dal punto di vista cognitivo viene considerato come una forma di "intelligenza situata" capace di trasmettere segnali verbali e non, portatori di significato (Cassell et al., 2000). Grazie al loro aspetto e alle loro funzionalità, gli ECAs non si limitano alla comunicazione text-based o dialogue-based, ma integrano canali non verbali come espressioni facciali, movimenti oculari, gesti e postura (Loveys et al., 2020). Oltre all'aspetto, dispongono di strumenti avanzati per "percepire" e interpretare gli stati emotivi dell'interlocutore. Sensori e algoritmi che gli permettono di riconoscere il parlato, le espressioni facciali (FER), le emozioni, la postura (Saxena et al., 2020) e la risposta cutanea (Dzedzickis et al., 2020).

Il progetto

Il progetto mira a fornire agli ECAs strumenti innovativi per rendere le interazioni sempre più naturali, avvicinandole a quelle che tipicamente

Introduction

Empathy is the capacity to understand and vicariously experience another person's perspective or situation. Nonverbal communication through posture, gestures, facial expressions, and tone of voice can convey emotional closeness and understanding. For example, adopting the same posture as one's interlocutor can strengthen the sense of connection, just as a smile can positively influence the emotional experience of the person receiving it (Heylighen et al., 2019). This close relationship between empathy and body language is not limited to human interactions but also has a significant impact on technology design. Empathic qualities are in fact fundamental in the design process, as they allow the identification of both explicit and implicit needs to develop solutions that meet people's requirements and expectations (Andresa et al., 2019). With technological evolution, designers can implement these aspects in human–machine interactions as well, making them more natural. To achieve this, it is essential to adopt a bidirectional and multimodal communication model that replicates human communication, going beyond mere verbal exchange and including emotional signals expressed through voice, facial expressions, and posture (Costantini et al., 2019). Among the technologies that embody these characteristics are Embodied Conversational Agents (ECAs), virtual assistants with a human-like appearance. Their virtual body can take on different forms. For instance, it can be realistic or cartoon-like, and it may display only the face or the entire figure. From a cognitive perspective, the human body is considered a form of "situated intelligence" capable of transmitting both verbal and nonverbal signals that carry meaning (Cassell et al., 2000). Thanks to their appearance and functionalities, ECAs go beyond text-based or dialogue-based communication, integrating nonverbal channels such as facial expressions, eye movements, gestures, and posture (Loveys et al., 2020). Beyond appearance, they are also equipped with advanced tools to "perceive" and interpret the interlocutor's emotional states. These include sensors and algorithms that enable them to recognize speech, facial expressions (FER), emotions, posture (Saxena et al., 2020), and skin conductance response (Dzedzickis et al., 2020).

Project Overview

The project aims to provide ECAs with innovative tools to make interactions increasingly natural, bringing them closer to those that typically

si instaurano tra persone. Per raggiungere questo obiettivo, vorremmo dotare questi sistemi della capacità di comprendere lo stato d'animo dell'interlocutore rispetto alla conversazione, così da elaborare risposte più adeguate. Dopo una serie di brainstorming tra i membri dell'Università di Siena e l'azienda partner del progetto QuestIT, sono stati individuati tre atteggiamenti sociali chiave, la cui implementazione potrebbe migliorare il design di questi sistemi: i saluti, i gesti di accordo e quelli di disaccordo. L'uso del saluto consentirebbe di avviare la conversazione in modo più naturale, simulando un incontro tra persone. I gesti di accordo segnalano una condivisione di opinioni, che di solito porta a un'alleanza, un impegno alla cooperazione e un atteggiamento positivo reciproco. Al contrario, il disaccordo implica generalmente conflitto, mancanza di cooperazione e un atteggiamento negativo (Bousmalis et al., 2013). Disporre di algoritmi in grado di riconoscere questi segnali offrirebbe ai designer un'indicazione preziosa sul coinvolgimento dell'ascoltatore, permettendo di avviare e concludere una conversazione in modo più naturale. Infine, queste inferenze rispettano il quadro normativo previsto dall'AI Act, che invece vieta di dotare gli ECAs di sistemi per il riconoscimento delle emozioni, salvo nei casi previsti dal regolamento (https://artificialintelligenceact.eu/ai-act-explorer/).

Stato dell'arte

L'elaborazione dei segnali sociali è un campo di ricerca interdisciplinare che combina scienze sociali, design e ingegneria per studiare e comprendere i comportamenti umani attraverso strumenti computazionali (Vinciarelli et al., 2009). Per rendere possibile questa analisi, è necessario addestrare algoritmi di intelligenza artificiale in modo che siano in grado di riconoscere ed interpretare tali segnali. Tuttavia, affinché un algoritmo riesca ad apprendere, ha bisogno di essere esposto a grandi quantità di dati, raccolti all'interno di dataset strutturati. Esistono diversi corpus e database, che hanno al loro interno anche dati etichettati come gesti di accordo e disaccordo. Ad esempio, un dataset di video ed annotazioni creato per lo studio delle interazioni sociali è Canal9, ottenuto dalle registrazioni dei dibattiti politici trasmessi dalla televisione svizzera (Bousmalis et al, 2013). Oppure AMI e AMIDA, che raccolgono dati multimediali e una classificazione dei gesti provenienti da riunioni di lavoro in Inghilterra. Questi dataset sono stati pensati per implementare sistemi che facilitino la cooperazione durante i meeting aziendali (https://

occur between people. To achieve this goal, we intend to equip these systems with the ability to understand the interlocutor's mood in relation to the conversation, thereby enabling them to generate more appropriate responses. Following a series of brainstorming sessions between members of the University of Siena and the project's partner company QuestIT, three key social behaviors were identified whose implementation could enhance the design of these systems: greetings, gestures of agreement, and gestures of disagreement. The use of greetings would allow conversations to begin in a more natural way, simulating a human encounter. Gestures of agreement indicate a sharing of opinions, which usually fosters alliance, commitment to cooperation, and mutual positive attitudes. Conversely, disagreement generally implies conflict, lack of cooperation, and a negative attitude (Bousmalis et al., 2013). Implementing algorithms capable of recognizing these signals would provide designers with valuable insights into the listener's engagement, thereby enabling more natural ways of initiating and concluding a conversation. Finally, these inferences comply with the regulatory framework established by the AI Act, which prohibits equipping ECAs with emotion recognition systems, except in cases explicitly allowed by the regulation (https://artificialintelligenceact.eu/ai-act-explorer/).

Literature Review

The processing of social signals is an interdisciplinary field of research that combines social sciences, design, and engineering to study and understand human behaviour through computational tools (Vinciarelli et al., 2009). For this analysis to be possible, artificial intelligence algorithms need to be trained to recognize and interpret these signals. However, for an algorithm to learn, it must be exposed to large amounts of data, collected within structured datasets. Several corpora and databases exist, some of which include annotated data such as agreement and disagreement gestures. For example, Canal9 is a dataset of videos and annotations created for the study of social interactions, obtained from recordings of political debates broadcast by Swiss television (Bousmalis et al., 2013). Other examples include AMI and AMIDA, which collect multimedia data and classify gestures from business meetings in the United Kingdom. These datasets were designed to support the development of systems that facilitate cooperation during corporate meetings (https://groups.inf.ed.ac.uk/ami/corpus/). Another audiovisual corpus

groups.inf.ed.ac.uk/ami/corpus/). Un altro corpus audiovisivo è SEMAINE, registrato presso l'Università di Belfast, che non solo classifica gesti di accordo e disaccordo, ma include anche la valenza emotiva delle persone che li esprimono. Il dataset è stato sviluppato con l'obiettivo specifico di fornire dati di addestramento per assistenti virtuali capaci di riconoscere e rispondere alle emozioni umane in modo più realistico (McKeown et al., 2010). Tuttavia, bisogna considerare che uno stesso gesto può assumere interpretazioni diverse a seconda del contesto e della cultura d'appartenenza della persona che lo esprime (Vogeley et al., 2010). Data l'assenza di un dataset open-source italiano, abbiamo optato per una raccolta dati locale, in cui le persone hanno simulato specifici gesti di saluto, accordo e disaccordo.

Il setup

Per la raccolta dati, abbiamo cercato di andare oltre un approccio puramente ingegneristico, ponendoci domande su come l'interazione si svolgesse realmente. Il nostro obiettivo era catturare l'esperienza delle persone durante l'interazione con gli ECAs. Per questo motivo, ci siamo interrogati su diversi aspetti, come comprendere come le persone comunicano con questi sistemi. In seguito, abbiamo adottato delle scelte che simulassero l'ambiente in cui persone ed avatar interagiscono. Ad esempio, è stato scelto di registrare con un'inquadratura frontale, poiché gli utenti solitamente li utilizzano tramite PC o smartphone. Oppure abbiamo considerato che alcuni assistenti, come quelli di supporto alle operazioni bancarie, richiedono una comunicazione più formale, mentre altri, come gli assistenti alle vendite nei negozi, adottano un approccio più informale. Considerando tutti questi fattori, abbiamo infine creato un setup su PsychoPy, un software open-source progettato per condurre esperimenti in laboratorio (https://www.psychopy.org/). Il setup prevedeva due scenari, ciascuno con quattro situazioni e tre diverse valenze emotive:
- scenari: informale "L'Amica" e formale "Il Capo" o "Il Datore di lavoro";
- situazioni: saluti, accordo, disaccordo/dubbio con quanto detto dall'ECAs;
- valenze emotive: neutra, positiva di gioia e negativa di tristezza.

I frame presentati su PsychoPy riguardavano uno dei due contesti, che venivano somministrati separatamente. Per ciascuno, erano previste dodici situazioni, ognuna composta da un testo descrittivo del contesto, un avatar-interlocutore virtuale con differenti espressioni che incarnava "L'Amica" o "Il Capo" e, nei casi di valenza positiva

is SEMAINE, recorded at the University of Belfast, which not only classifies agreement and disagreement gestures but also includes the emotional valence of the individuals expressing them. This dataset was developed with the specific goal of providing training data for virtual assistants capable of recognizing and responding to human emotions in a more realistic way (McKeown et al., 2010). Nevertheless, it should be acknowledged that the same gesture may be subject to different interpretations depending on the context and the cultural background of the individual performing it (Vogeley et al., 2010). Given the absence of an open-source Italian dataset, we opted for a local data collection effort in which participants simulated specific gestures of greeting, agreement, and disagreement.

Experimental Setup

For the data collection, we aimed to go beyond a purely engineering approach by asking questions about how the interaction unfolded. Our goal was to capture people's lived experience while interacting with ECAs. For this reason, we reflected on several aspects, such as understanding how individuals communicate with these systems. Subsequently, we made design choices to simulate the environment in which people and avatars interact. For example, we decided to record from a frontal perspective, since users typically engage with ECAs through a PC or smartphone. We also considered that certain assistants, such as those supporting banking operations, require more formal communication, whereas others, like sales assistants in retail settings, adopt a more informal approach. Taking all these factors into account, we created an experimental setup using PsychoPy, an open-source software designed for laboratory-based studies (https://www.psychopy.org/). The setup included two scenarios, each comprising four situations and three different emotional valences:
- scenarios: informal ("The Friend") and formal ("The Employer");
- situations: greetings, agreement, disagreement/doubt regarding the ECA's statements;
- emotional valences: neutral, positive (joy), and negative (sadness).

The frames presented in PsychoPy referred to one of the two contexts, which were administered separately. For each context, twelve situations were provided, each consisting of a descriptive text outlining the context, a virtual avatar–interlocutor displaying different expressions and embodying either "The Friend"

e negativa, una musica di supporto per favorire l'immedesimazione nella situazione.

Raccolta dati

I dati sono stati raccolti in ambito universitario ed aziendale. Prima di iniziare, ai partecipanti veniva consegnato un consenso informato. Successivamente, svolgevano una sessione di prova per familiarizzare con l'ambiente PsychoPy. Nel corso dell'esperimento, veniva richiesto di immedesimarsi negli scenari proposti: "L'Amica" e "Il Capo", leggere attentamente le frasi-contesto, immaginare una risposta e premere il tasto space per registrarla. A quel punto, si apriva una finestra cattura video della durata di 3 secondi, durante la quale dovevano registrare la loro reazione non verbale. Dopo aver completato entrambi gli scenari, veniva somministrato un questionario sotto forma di scala Likert da 1 a 5 punti, in cui dovevano attribuire un punteggio per ogni stimolo. L'obiettivo delle domande era comprendere cosa avesse influenzato maggiormente le loro risposte non verbali, come l'espressione dell'avatar, le frasi-contesto o la musica. Inoltre, veniva domandato se avessero riscontrato delle difficoltà nel fornire le risposte. La raccolta dati si è svolta su base volontaria e ha coinvolto 52 partecipanti, di cui 26 uomini e 26 donne, che hanno registrato 26 video ciascuno. Le sessioni si sono svolte su più giornate, concordate in base alla disponibilità dei partecipanti, tra giugno e ottobre 2024.

Pulizia e addestramento dati

Dopo la raccolta dei dati, prima della fase di addestramento, è stato necessario un processo di data preparation, condotto sia con strumenti di machine learning che manualmente. I video raccolti sono stati ridimensionati in un formato quadrato e successivamente segmentati in clip della durata di 1,5 secondi. Per garantire che nessun gesto venisse troncato, il taglio dei video è stato effettuato sovrapponendo i frame: ogni clip comprendeva 0,5 secondi precedenti al segmento originale, con sequenze da 0 a 1,5 secondi, da 1 a 2,5 secondi e da 1,5 a 3 secondi. Dopo questa fase, i video sono stati rinominati e sottoposti a una pulizia manuale per eliminare i tempi che non contenevano gesti o espressioni rilevanti. Una volta ottenuto il dataset pulito e strutturato, è iniziata la fase di addestramento del modello, realizzato dall'azienda.

Discussione

Per sviluppare soluzioni tecnologicamente efficaci, è fondamentale addestrare gli

or "The Employer," and in the cases of positive or negative valence, a supporting musical track to facilitate emotional immersion in the situation.

Data collection

Data were collected in both university and corporate settings. Before beginning, participants were provided with an informed consent form. They then completed a trial session to familiarize themselves with the PsychoPy environment. During the experiment, participants were asked to immerse themselves in the proposed scenarios ("The Friend" and "The Employer"), carefully read the context sentences, imagine a response, and press the space bar to record it. At that point, a three-second video capture window opened, during which they were required to record their nonverbal reaction. After completing both scenarios, a questionnaire was administered in the form of a 5-point Likert scale, where participants assigned a score to each stimulus. The purpose of the questions was to understand what most influenced their nonverbal responses, such as the avatar's expression, the context sentences, or the accompanying music. Participants were also asked whether they had encountered any difficulties in providing their responses. Data collection was conducted on a voluntary basis and involved 52 participants, 26 men and 26 women, each of whom recorded 26 videos. The sessions took place across multiple days, scheduled according to participants' availability, between June and October 2024.

Data cleaning and training

After the data collection phase and prior to training, a data preparation process was required, carried out using both machine learning tools and manual methods. The collected videos were resized into a square format and subsequently segmented into 1.5 second clips. To ensure that no gesture was cut off, video segmentation was performed with overlapping frames: each clip included 0.5 seconds preceding the original segment, resulting in sequences from 0 to 1.5 seconds, 1 to 2.5 seconds, and 1.5 to 3 seconds. Following this step, the videos were renamed and manually cleaned to remove segments that did not contain relevant gestures or expressions. Once the dataset was cleaned and structured, the model training phase, developed by the company, was initiated.

Discuss

To develop technologically effective solutions, it is essential to train algorithms with data

algoritmi con dati rappresentativi dei casi reali che il sistema dovrà gestire. Per questo motivo la configurazione della raccolta dati riveste un ruolo cruciale nell'acquisizione di informazioni realmente rilevanti. A tal proposito, è consigliabile predisporre un setup che integri anche gli aspetti progettuali da implementare. In questo senso, una base di partenza possono essere sessioni di brainstorming o focus group volte a individuare le situazioni chiave. Ad esempio, può risultare utile definire un ambiente sperimentale che riproduca scenari realistici e progettare stimoli in grado di simulare specifiche interazioni. Dai risultati del questionario, compilato da 46 partecipanti su 52, è emerso che, nel contesto sperimentale, le espressioni facciali di gioia e tristezza comunicate dall'avatar hanno favorito risposte non verbali più efficaci rispetto ad espressioni neutre.
Anche l'elemento sonoro, utilizzato per facilitare l'immedesimazione in situazioni di gioia o tristezza, ha avuto un impatto significativo sull'interazione complessiva.
Le descrizioni delle frasi-contesto hanno influenzato le risposte dei partecipanti. Anche in questo caso, le descrizioni neutre come 'Saluta l'amica', sono risultate meno incisive rispetto a quelle che includevano un contesto e un'indicazione emotiva.
Alcuni partecipanti hanno segnalato difficoltà nell'immedesimarsi e nell'esprimere determinati stati d'animo, evidenziando come le espressioni non verbali tendono a manifestarsi in modo spontaneo. A supporto di questa osservazione, un commento raccolto afferma: "È stato molto difficile perché solitamente le espressioni non verbali sorgono in modo spontaneo; dover riflettere su quale sarebbe stata la mia risposta in quei contesti ha richiesto un certo sforzo cognitivo."

Conclusioni

L'evoluzione delle tecnologie sta aprendo nuove possibilità nello sviluppo di agenti conversazionali incarnati in grado di comunicare sia verbalmente che non verbalmente. Il corpo di questi agenti può trasmettere segnali di connessione e vicinanza emotiva, ma affinché la comunicazione sia realmente bidirezionale, è fondamentale che siano dotati di sistemi capaci di riconoscere lo stato d'animo dell'interlocutore. Nel progetto, è stato adottato un approccio interdisciplinare che combina scienze sociali, design e ingegneria, con l'obiettivo di implementare specifiche funzionalità. In particolare, ci si è concentrati su gesti legati al saluto, all'accordo e al disaccordo all'interno della conversazione. La creazione di un corpus

that are representative of the real-world cases the system will be required to handle. For this reason, the configuration of data collection plays a crucial role in gathering truly relevant information. In this regard, it is advisable to set up a configuration that also incorporates the design aspects to be implemented. A useful starting point can be brainstorming sessions or focus groups aimed at identifying key situations. For example, it may be beneficial to define an experimental environment that reproduces realistic scenarios and to design stimuli capable of simulating specific interactions. The questionnaire, completed by 46 out of 52 participants, revealed that within the experimental context, the avatar's facial expressions of joy and sadness elicited more effective nonverbal responses compared to neutral expressions.

The sound element, used to facilitate immersion in situations of joy or sadness, also had a significant impact on overall interaction.
Participants' responses were influenced by the descriptions of the context phrases. Neutral descriptions (e.g., 'Greet your friend') elicited weaker effects compared to those containing both contextual and emotional cues.
Some participants reported difficulties in identifying with and expressing certain emotional states, highlighting how nonverbal expressions tend to manifest spontaneously. Supporting this observation, one collected comment stated: "It was very difficult because usually nonverbal expressions arise spontaneously; having to think about what my response would have been in those contexts required a certain cognitive effort."

Conclusion

The evolution of technology is opening new possibilities in the development of embodied conversational agents capable of communicating both verbally and nonverbally. The bodies of these agents can transmit signals of connection and emotional closeness, but for communication to be truly bidirectional, it is essential that they are equipped with systems able to recognize the interlocutor's emotional state. In the project, an interdisciplinary approach was adopted, integrating social sciences, design, and engineering, with the goal of implementing specific functionalities. In particular, the focus was placed on gestures related to greeting, agreement, and disagreement within conversations. The creation of a corpus of locally recorded data, implicitly labelled by the participants, can support the training of

di dati registrati localmente e implicitamente etichettati dai partecipanti può supportare l'addestramento degli algoritmi. Per facilitare l'immedesimazione nel contesto, sono stati introdotti stimoli emozionali, come espressioni facciali dell'avatar, musica e frasi-contesto. Tra questi, gli stimoli con connotazioni gioiose o tristi si sono rivelati particolarmente efficaci nell'elicitare risposte non verbali. Questi risultati evidenziano la complessità nel ricreare situazioni in grado di simulare fedelmente l'interazione tra esseri umani e assistenti virtuali, in particolare per quanto riguarda la stimolazione di specifici linguaggi non verbali. Tuttavia, attraverso un'attenta progettazione di scenari e stimoli mirati, è possibile favorire condizioni più vicine alla realtà. Il progetto ha raggiunto una prima fase di sviluppo. Dopo l'addestramento dell'algoritmo, verrà condotto un test per validare l'accuratezza. In seguito, verranno progettate risposte empatiche per consentire agli ECAs di gestire le situazioni individuate, integrando soluzioni di design mirate. Il progetto si presta a diversi ambiti applicativi. In particolare, nel settore sanitario, può promuovere un'assistenza centrata sulla persona, contribuendo all'umanizzazione delle cure in cui la dimensione relazionale assume un ruolo fondamentale (Brusch et al., 2019). Infatti, pur non essendo in grado di percepire realmente le emozioni umane, queste tecnologie possono comunque offrire un'empatia razionale. Secondo Bloom, un'empatia basata sulla comprensione piuttosto che sull'immedesimazione può favorire decisioni più equilibrate e giuste (Bloom, 2023), oltre a migliorare la naturalezza della comunicazione tra esseri umani e agenti virtuali. Inoltre, l'implementazione di queste funzionalità fornisce ai designer nuovi strumenti per modulare i flussi conversazionali, rendendo l'esperienza più simile alla comunicazione human-to-human.

Ringraziamenti

Il progetto si inserisce all'interno del Dottorato di Ricerca di Interesse Nazionale "Design per il Made in Italy: Identità, Innovazione e Sostenibilità", finanziato dall'Unione Europea attraverso i fondi PNRR (M4C2 – "Dalla Ricerca all'Impresa"), Università degli Studi di Siena e co-finanziato da QuestIT s.r.l.

algorithms. To facilitate immersion in the context, emotional stimuli were introduced, such as avatar facial expressions, music, and context-setting phrases. Among these, stimuli with joyful or sad connotations proved particularly effective in eliciting nonverbal responses. These findings highlight the complexity of recreating situations that accurately simulate human–virtual assistant interaction, especially in terms of eliciting specific forms of nonverbal communication. However, through careful scenario design and targeted stimuli, it is possible to create conditions that are closer to real-life experiences. The project has reached an initial stage of development. After the algorithm training, a test will be conducted to validate accuracy. Subsequently, empathic responses will be designed to enable ECAs to manage the identified situations, integrating targeted design solutions. The project lends itself to multiple application domains. In the healthcare sector, it can promote person-centred care and contribute to the humanization of treatment, where the relational dimension plays a key role (Brusch et al., 2019). Indeed, although these technologies cannot truly perceive human emotions, they can still offer rational empathy. According to Bloom, empathy based on understanding rather than emotional identification can foster more balanced and fair decisions (Bloom, 2023), while also improving the naturalness of communication between humans and virtual agents. Moreover, implementing these functionalities provides designers with new tools to shape conversational flows, making the experience closer to human-to-human communication.

Acknowledgment

The project is part of the National PhD Program "Design for Made in Italy: Identity, Innovation, and Sustainability", funded by the European Union through PNRR funds (M4C2 – "From Research to Business"), University of Siena, and co-funded by QuestIT s.r.l.

Bibliografia | References

_Andresa, T. F., & Juanitab, G. T. (2019). Empathic design as a framework for creating meaningful experiences. In Conference Proceedings of the Academy for Design Innovation Management (Vol. 2, No. 1, pp. 908-918).

_Bloom, Paul. (2023) Contro l'empatia. Una difesa della razionalità. Liberlibri

_Bousmalis, K., Mehu, M., & Pantic, M. (2013). Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. Image and vision computing, 31(2), 203-221.

_ Busch, I. M., Moretti, F., Travaini, G., Wu, A. W., & Rimondini, M. (2019). Humanization of care: Key elements identified by patients, caregivers, and healthcare providers. A systematic review. The Patient-Patient-Centered Outcomes Research, 12, 461-474.

_Cassell, J. (2001). Embodied Conversational Agents: Representation and Intelligence in User Interfaces. AI Magazine, 22(4), 67–67. https://doi.org/10.1609/AIMAG.V22I4.1593

_Costantini, S., De Gasperis, G., & Migliarini, P. (2019, June). Multi-agent system engineering for emphatic human-robot interaction. In 2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE) (pp. 36-42). IEEE.

_Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. Sensors, 20(3), 592.

_Heylighen, A., & Dong, A. (2019). To empathise or not to empathise? Empathy and its limits in design. Design Studies, 65, 107-124.

_Loveys, K., Sebaratnam, G., Sagar, M., & Broadbent, E. (2020). The effect of design features on relationship quality with embodied conversational agents: a systematic review. International Journal of Social Robotics, 12(6), 1293-1312.

_McKeown, G., Valstar, M. F., Cowie, R., & Pantic, M. (2010, July). The SEMAINE corpus of emotionally coloured character interactions. In 2010 IEEE international conference on multimedia and expo (pp. 1079-1084). IEEE.

_Saxena, A., Khanna, A., & Gupta, D. (2020). Emotion recognition and detection methods: A comprehensive survey. Journal of Artificial Intelligence and Systems, 2(1), 53-79.

_Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and vision computing, 27(12), 1743-1759.

_Vogeley, K., & Bente, G. (2010). "Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication. Neural Networks, 23(8-9), 1077-1090.

_https://artificialintelligenceact.eu/ai-act-explorer/

_https://groups.inf.ed.ac.uk/ami/corpus/

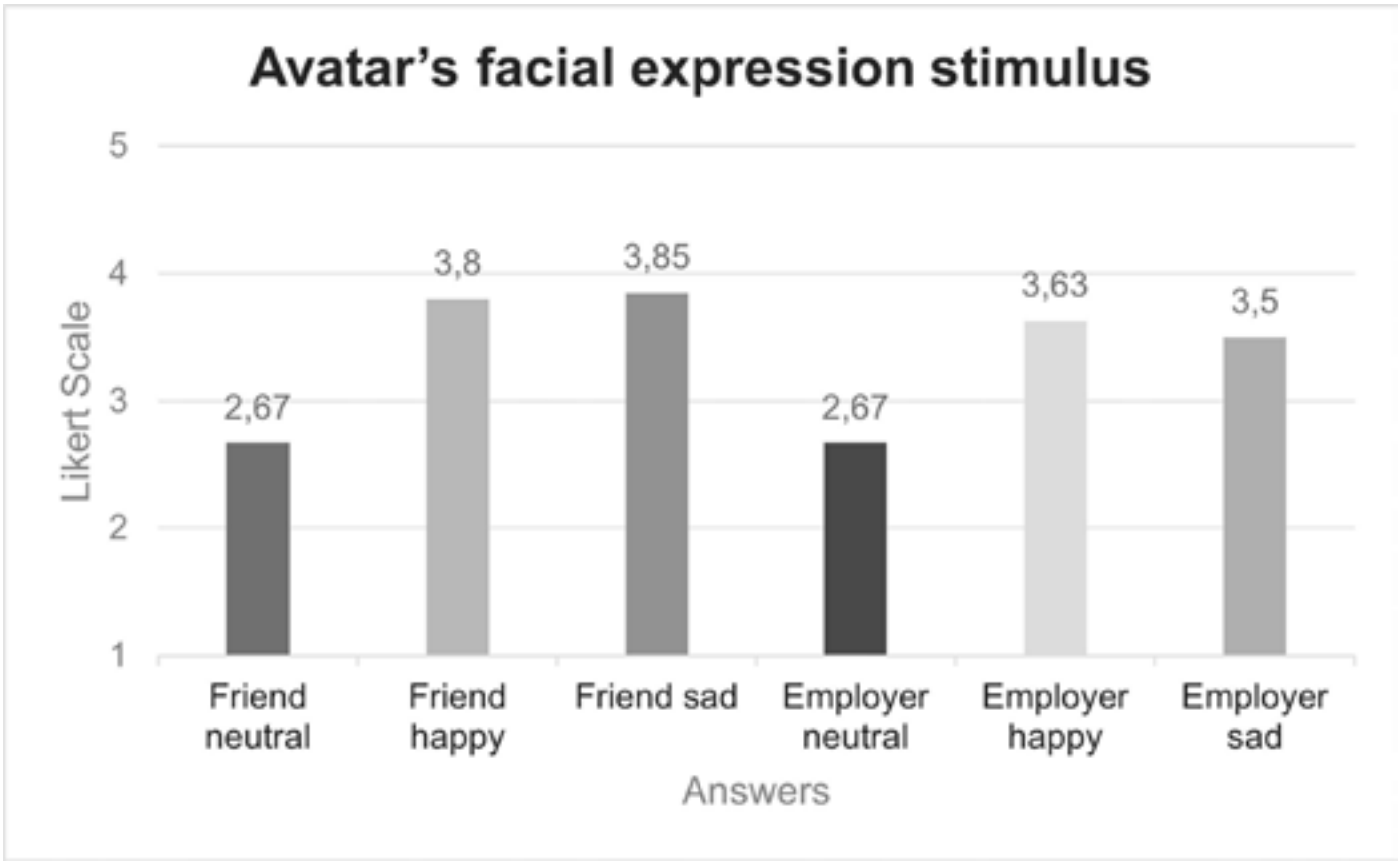
_https://www.psychopy.org/

1

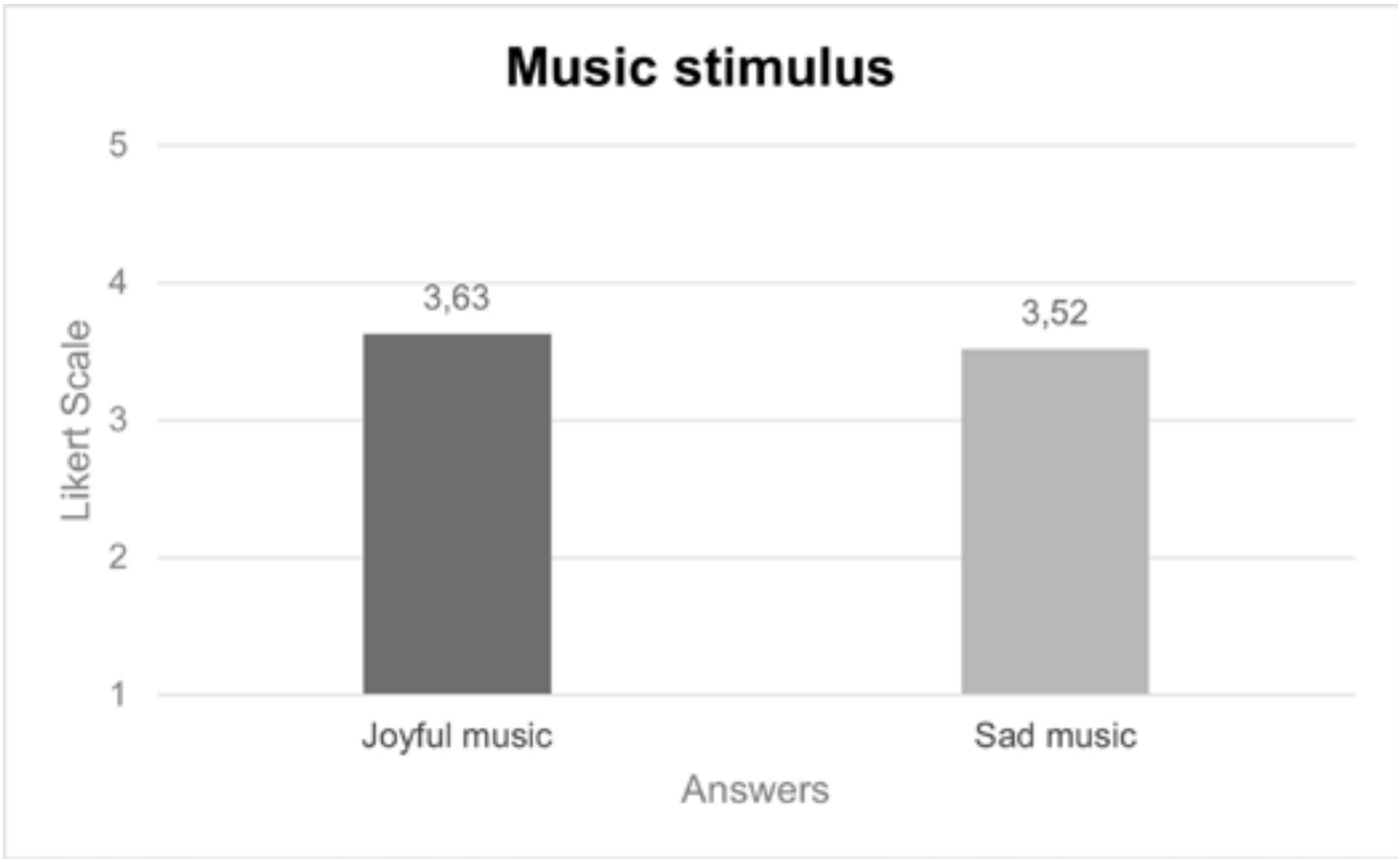
1_Digital Human creato con la piattaforma Algho, sviluppata da QuestIT s.r.l.
2_Fotografia della raccolta dati.
3_ Grafico della media delle risposte allo stimolo emozionale dell'avatar.
4_ Grafico della media delle risposte allo stimolo musicale.
5_ Grafico della media delle risposte allo stimolo della frase contestuale.

1_Digital Human created with the Algho platform, developed by QuestIT s.r.l.
2_Data collection photograph.
3_ Graph of the average responses to the avatar's emotional stimulus.
4_ Graph of the average responses to the music stimulus.
5_ Graph of the average responses to the context-sentence stimulus.

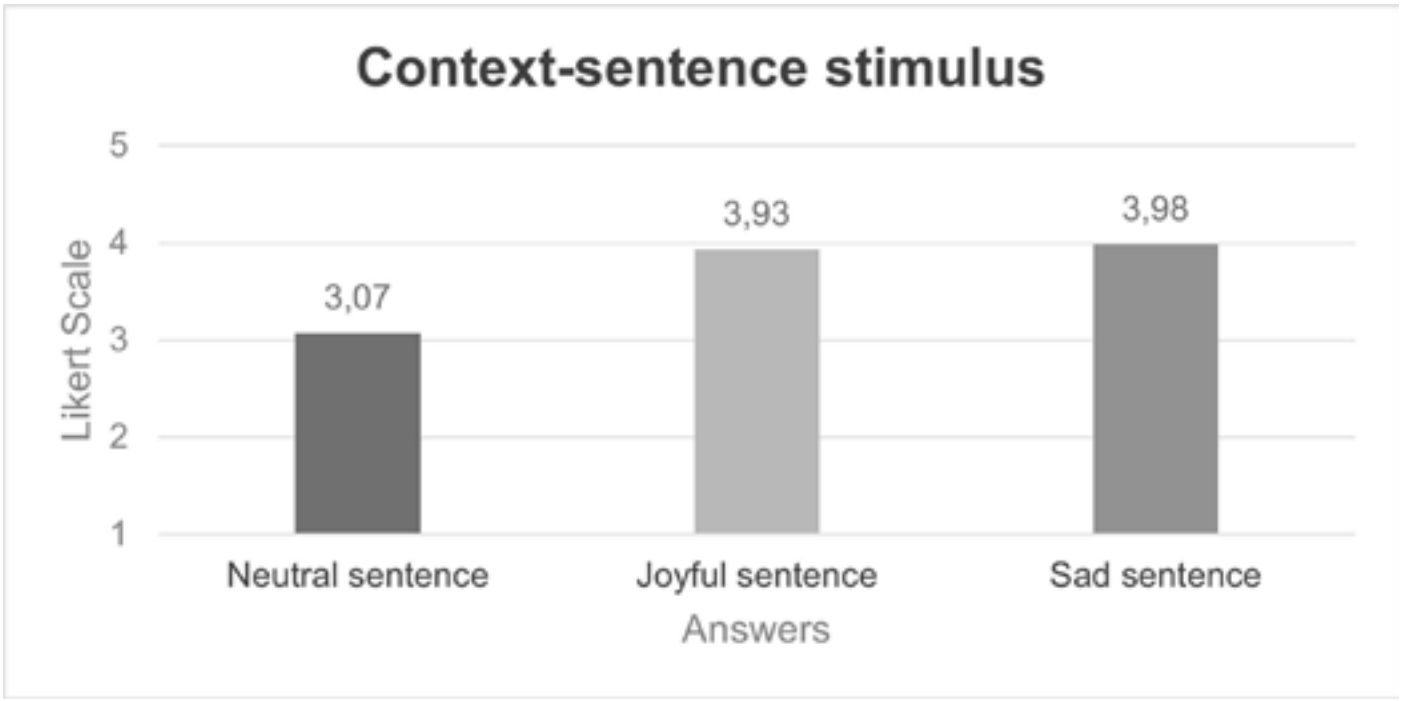
2


Avatar’s facial expression stimulus
Likert Scale
5
4
3
2
1
2,67
3,8
3,85
2,67
3,63
3,5
Friend neutral
Friend happy
Friend sad
Employer neutral
Employer happy
Employer sad
Answers

3


Music stimulus
Likert Scale
5
4
3
2
1
3,63
3,52
Joyful music
Sad music
Answers

4


Context-sentence stimulus
Likert Scale
5
4
3
2
1
3,07
3,93
3,98
Neutral sentence
Joyful sentence
Sad sentence
Answers

5