

# Supporting mental-health monitoring of patients with Social Robots<sup>\*</sup>

Rita Francese<sup>1,\*</sup>, Edmondo Nicolò De Simone<sup>1</sup>

<sup>1</sup>Computer Science Department, University of Salerno, Via Giovanni Paolo II, Fisciano (SA), Italy

## Abstract

This study presents a preliminary evaluation of a system for mental health support based on the Furhat social robot integrated with a conversational AI model. The system enables natural-language interactions in Italian and automatically structures dialogues into JSON format for potential clinical use. A user study with 19 participants (ages 18–70), none with diagnosed mental health conditions, was conducted using standardized tools such as the System Usability Scale (SUS), Godspeed Questionnaire, and Net Promoter Score (NPS). Results showed a statistically significant improvement in participants' self-reported mood after the interaction ( $p = 0.015$  with a medium Cliff's Delta effect size), along with favorable perceptions of clarity, neutrality, and engagement. The Net Promoter Score (47.4) indicates positive acceptance, although the overall SUS score (54.6) highlights areas for usability improvement. Qualitative feedback emphasized the value of the robot's non-judgmental perception but also suggested enhancements in expressiveness, personalization, and emotional recognition. Future work will focus on integrating standardized clinical interview protocols into the dialogue process, enabling structured data collection to support continuous monitoring and early relapse detection. This work lays the foundation for responsibly introducing social robots as complementary tools in clinical mental health care.

## Keywords

Social Robot, Mental Health, Large Language Models, User Study, Furhat

## 1. Introduction

The growing prevalence of mental health disorders and the limitations of traditional healthcare systems in providing continuous, personalized care have prompted the exploration of innovative technological solutions. such as using Deep Learning [1] techniques or gait analysis. Among these, social robots offer a promising avenue for enhancing mental health monitoring by enabling natural, empathetic interactions and scalable data collection [2, 3].

This paper presents a preliminary study which explores the use of the social robot *Furhat*, integrated with conversational AI (ChatGPT 4.5), to support the continuous monitoring of patients' mental well-being. The research addresses the following question:

**RQ: Is it possible to use a social robot as a tool for supporting the evaluation and monitoring of mental health?**

To answer this question, the study introduces a system capable of conducting natural conversations with users in Italian, automatically generating structured JSON outputs that segment dialogues for clinical review. The system was evaluated with 19 participants, using standardized tools such as the System Usability Scale (SUS), the Goodspeed scale for robot perception, and the Net Promoter Score (NPS). These methods have been widely validated for assessing the usability and perception of socially assistive robots in healthcare [4, 5, 6].

This contribution is in alignment with the goals of the Research Projects of Significant National Interest (PRIN) 2022 PNRR SPECTRA project [7], focused on using AI for supporting patients with schizophrenia through Natural Language Processing [1], gait analysis [8], and emotion detection

*EMPATH-IA: Workshop on EMpowering PATients THrough AI, Multimedia, and Explainable HCI: Innovations in Personalized Healthcare, in conjunction with the 16th Biannual Conference of the Italian SIGCHI Chapter (CHITALY 2025), 6–10 October 2025, Salerno, Italy*

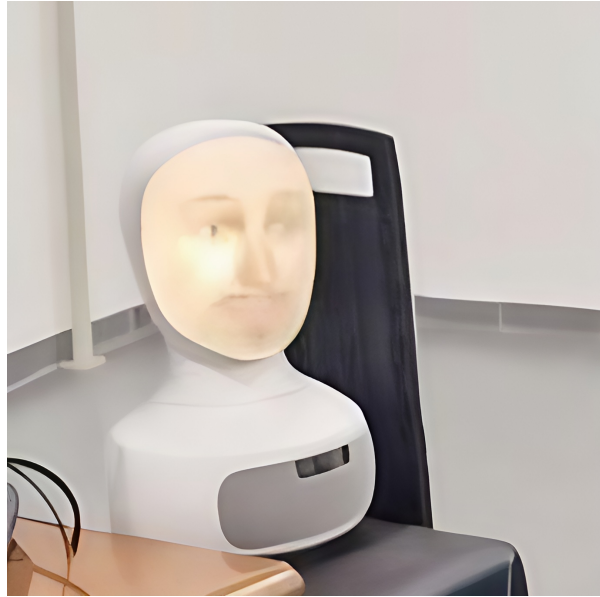
\*Corresponding author.

✉ francese@unisa.it (R. Francese); e.desimone32@studenti.unisa.it (E. N. D. Simone)

id 0000-0002-6929-0056 (R. Francese)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



**Figure 1:** The Furhat robot.

techniques [9]. This preliminary study lays the foundation for future developments in continuous, explainable, and ethically sound mental health monitoring solutions.

The paper is structured as follows. In Section 2 we discuss the background of this study; Section 3 presents the proposed system; Section 4 describes the evaluation planning while Section 5 reports and discusses the results. Finally Section 6 concludes the paper with final remarks and future work.

## **2. Background**

The integration of social robots into mental health support is based on prior research in human–robot interaction, socially assistive technologies, and conversational AI. This section outlines the technical features of the Furhat platform and reviews related work that highlights both the potential and current limitations of social robots in healthcare and mental well-being contexts.

### **2.1. The Furhat Social Robot**

Furhat is a tabletop humanoid robot developed by Furhat Robotics in Sweden (est. 2014) characterized by a back-projected face capable of expressive facial animation, eye gaze, and head movements, enabling human-like conversational interaction across multiple languages and modalities [10]. It is programmable via a modular SDK (including Kotlin support), integrates with speech recognition/synthesis engines, and can embed large language models to deliver adaptive, multimodal dialogue [10].

Furhat has been deployed in a variety of domains including healthcare, education, customer service, and mental health support. As an example, applications such as the MIA-PROM project use Furhat to guide patients through health questionnaires in clinical settings, improving accessibility and engagement [11]. Other initiatives, such as the Translational Research Centre for Digital Mental Health (TRC-DMH), are exploring Furhat’s expressivity and conversational presence in emotionally sensitive contexts involving special needs and mental health support [12].

Within academic research, Furhat enables embodied conversational agents like *FurChat*, which combines LLM-based dialogue generation with real-time facial expression control, demonstrating its suitability for naturalistic, socially rich interaction [13]. These features position Furhat as a compelling platform for investigating social robotics in mental health contexts.

## 2.2. Related Work

Socially assistive robots (SARs), social robots designed to aid users through emotional and social support, have been widely investigated as interventions in mental health and well-being contexts [14]. A systematic review of SAR deployments in healthcare settings, including therapy, monitoring, and patient engagement, documents over 279 real-world applications across 33 countries, highlighting their growing adoption and wide-ranging use cases [15].

Another systematic review focusing on mental health outcomes reported that most studies involved older adults with dementia and relied on small usability trials, revealing a need for more diverse populations and rigorous measurement frameworks [16]. These findings underscore limitations in generalizability and the necessity of broader, longitudinal studies.

Empirical studies with conversational SARs provide evidence for their potential. For instance, the Ryan robot, which integrates multimodal emotion recognition, was shown to improve mood and engagement among older adults in pilot studies, with the empathic variant deemed more likable and effective [17]. Likewise, research deploying a conversational robot to deliver internet-based Cognitive Behavioral Therapy (iCBT) (Ryan platform) demonstrated outcomes comparable to traditional human-led interventions over multiple weeks [18].

Recent studies emphasize the value of social robots for community-oriented mental health tools, including mood visualization and peer support frameworks [19]. Furhat-based systems have specifically been employed to assess stress, anxiety, and depression via conversational AI in non-judgmental environments [20]. However, user studies such as one conducted with older adults revealed that perceptions of anthropomorphism and intimacy vary, and acceptance depends on design features and context [21].

In summary, the literature indicates strong promise for SARs in mental-health contexts, but also highlights the importance of explainability, cultural sensitivity, and rigorous evaluation, especially in clinically relevant populations.

## 3. The proposed system

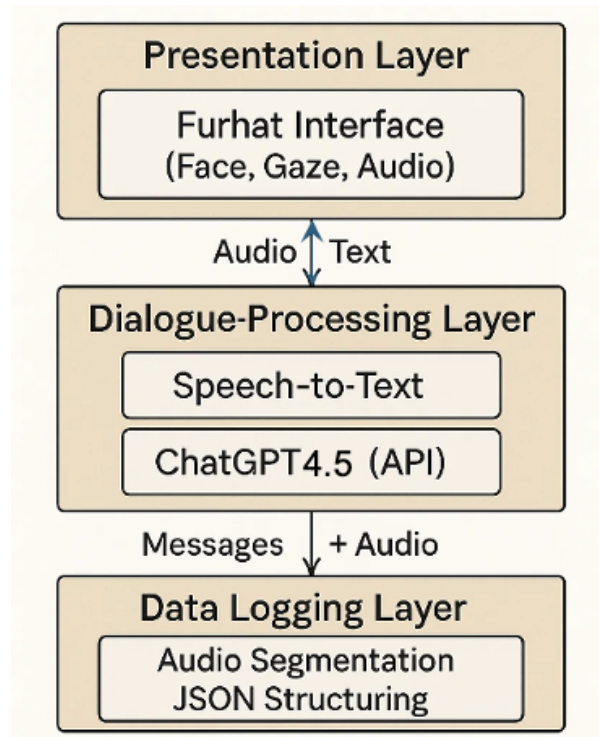
The system is composed of the following three distinct layers (Figure 2).

- *Presentation Layer*: The physical and expressive interface is provided by the Furhat robot, which delivers face projection, gaze control, and audio interaction capabilities. This layer manages the audio and text exchange between the user and the system.
- *Dialogue-Processing Layer*: Incoming speech is converted into text via a Speech-to-Text service. The textual input is then processed by *ChatGPT 4.5* (via API), which is guided by a structured prompt reported in Listing 3 to ensure safe, empathetic, and contextually relevant dialogue. The generated text is sent back to the Presentation Layer for speech synthesis, and optionally accompanied by audio cues or emotional expression.
- *Data Logging Layer*: All messages and corresponding audio are captured, segmented, and structured in JSON format. This dual-channel logging (User/Robot) enables precise turn segmentation, preserves the conversation context, and facilitates later analysis.

The prompt plays a crucial role in guiding the conversational agent's behaviour. It defines the role of the robot (a friendly and understanding psychologist), the type of questions to be asked, and the boundaries of the conversation (avoiding sensitive topics such as schizophrenia). The prompt instructs the agent to begin each interaction with an empathetic greeting, adapt follow-up questions based on user interest, and maintain a natural, fluid dialogue.

## 4. Evaluation planning

The evaluation combines quantitative and qualitative measures to explore user experience, acceptance, and affective impact. The following research question guided the study:



**Figure 2:** The three-tier system architecture.

**RQ:** Is it possible to use a social robot as a tool for supporting the evaluation and monitoring of mental health?

The goal was to examine participants' attitudes towards the use of a social robot for mental health monitoring, assessing both functional and affective responses.

The evaluation design considered the following dimensions:

- *Affective change:* Measuring shifts in self-reported emotional state before and after the interaction.
- *Usability:* Assessing ease of use, learnability, and overall user-friendliness.
- *Perception of the robot:* Evaluating anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety.
- *Satisfaction and acceptance:* Measuring willingness to recommend the system.
- *Qualitative impressions:* Collecting open-ended feedback to capture nuances not covered by standardized scales.

To assess these dimensions, the following tools were adopted:

- *System Usability Scale (SUS)* [5]: A widely used 10-item questionnaire that yields a score from 0 to 100, where higher scores indicate better perceived usability.
- *Godspeed Questionnaire Series* [22]: Five subscales (Anthropomorphism, Animacy, Likeability, Perceived Intelligence, Perceived Safety) rated on 5-point Likert items.
- *Net Promoter Score (NPS)*: A single-item metric asking participants how likely they are to recommend the system to others, producing a score from -100 to +100.

**Procedure.** The evaluation was structured into three sequential phases:

1. *Pre-interaction:* Collection of demographic information and baseline affective self-assessment.
2. *Interaction:* Voice-based conversation with the Furhat robot in Italian, guided by the supportive prompt described in Section 3. The session had no fixed duration to allow natural dialogue flow.

**Listing 1:** The conversation prompt.

```
val chatbot = OpenAIChatbot(
  """
  You are a friendly and understanding psychologist named $name.

  The following is a conversation in Italian between $name and a Person.

  $name will ask a series of general questions to get to know the Person
  better and encourage them to talk.

  The questions will aim to explore the Persons interests, experiences,
  and thoughts. $name will carefully avoid any mention of topics related
  to schizophrenia.

  The format of the conversation will be: question, answer, question,
  answer, etc.

  $name must respond in context to the answers, deepening topics that
  elicit more emphasis or interest from the Person, without repeating the
  same questions.

  The conversation will always start with "How are you?" to create empathy
  and make the Person feel at ease.

  The goal is to create a smooth and natural dialogue that allows the
  Person to express themselves freely and share their thoughts and feelings.

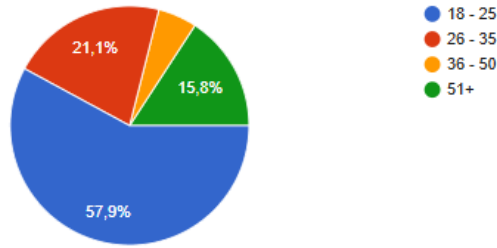
  If the Person shows particular interest or emphasis in a specific topic,
  $name will follow up and deepen that topic with further relevant questions.

  Otherwise, $name will continue with new questions to cover a wide range
  of topics.

  Examples of follow-up questions to deepen answers include:
  - If the Person talks about a book, $name could ask:
    "What did you like most about that book?"
    or "Do you have other books by the same author that you enjoy?"
  - If the Person mentions a favorite film, $name could ask:
    "What is your favorite scene in that film?"
    or "Are there other similar films you enjoy?"
  - If the Person talks about a hobby, $name could ask:
    "How long have you been practicing this hobby?"
    or "What made you start this hobby?"
  """
)
```

3. *Post-interaction:* Completion of the SUS, Godspeed, and NPS questionnaires, as well as the post-test affective self-assessment and open-ended feedback.

This mixed-methods approach ensured that we could triangulate standardized quantitative scores with qualitative user narratives, providing a comprehensive understanding of how participants perceived and responded to the robot in the context of mental health evaluation and monitoring.



**Figure 3:** Age of participants.

## 5. Results

**Participants.** The study involved 19 participants, recruited via convenience sampling. The sample had a balanced gender distribution: 9 males, 9 females, and 1 participant who preferred not to disclose. Most participants were aged 18–25 years (11 participants), followed by 26–35 years (4 participants), 51 years and older (3 participants), and 36–50 years (1 participant)) see Figure 3. No specific exclusion criteria were applied, allowing anyone interested to participate in the experiment. The sample included university students (not from computer science or related fields), working adults from various professions, and older adults with diverse levels of technological experience. We also asked participants information about their previous knowledge of social robots. 8 reported having no prior knowledge of social robots, 7 reported having superficial knowledge, and 4 reported having good knowledge of the topic.

### 5.1. Quantitative results

**Mood Changes Before and After Interaction** Table 1 summarizes the descriptive statistics of participants' mood ratings related to the question: *"How do you feel about talking to a robot?"* before (PRE) and after (POST) interacting with the Furhat robot. The mean mood score increased from 3.58 (SD = 0.77) before the interaction to 4.11 (SD = 0.57) after the interaction, while the median remained stable at 4 for both time points. The minimum score also increased from 2 to 3, indicating that the least positive responses improved following the interaction. The maximum score remained at 5, suggesting that the most positive responses were already at the top of the scale.

**Table 1**

Descriptive statistics of robot interaction perception before (PRE) and after (POST) interaction with Furhat

Time	<i>n</i>	Mean	SD	Median	Min	Max
PRE	19	3.58	0.769	4	2	5
POST	19	4.11	0.567	4	3	5

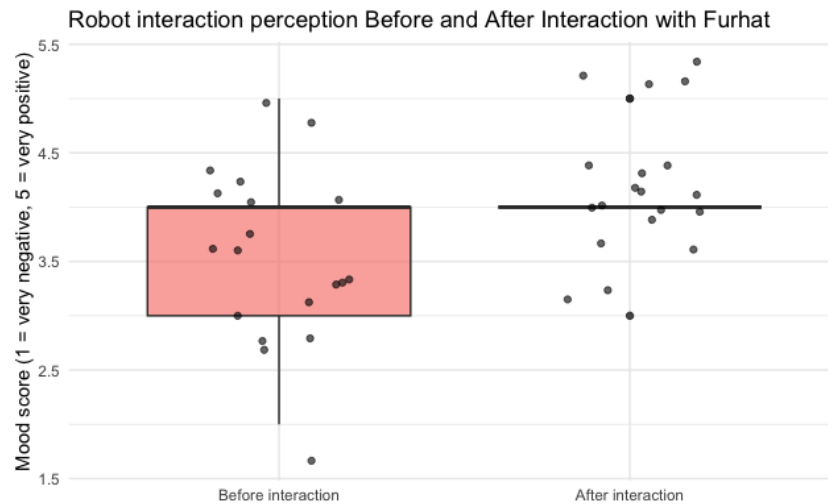
Figure 4 presents the boxplots of the mood scores, showing a general upward shift after interaction. Notably, the interquartile range (IQR) decreased post-interaction, suggesting that participants' responses became more homogeneous and clustered towards higher positive ratings.

**Table 2**

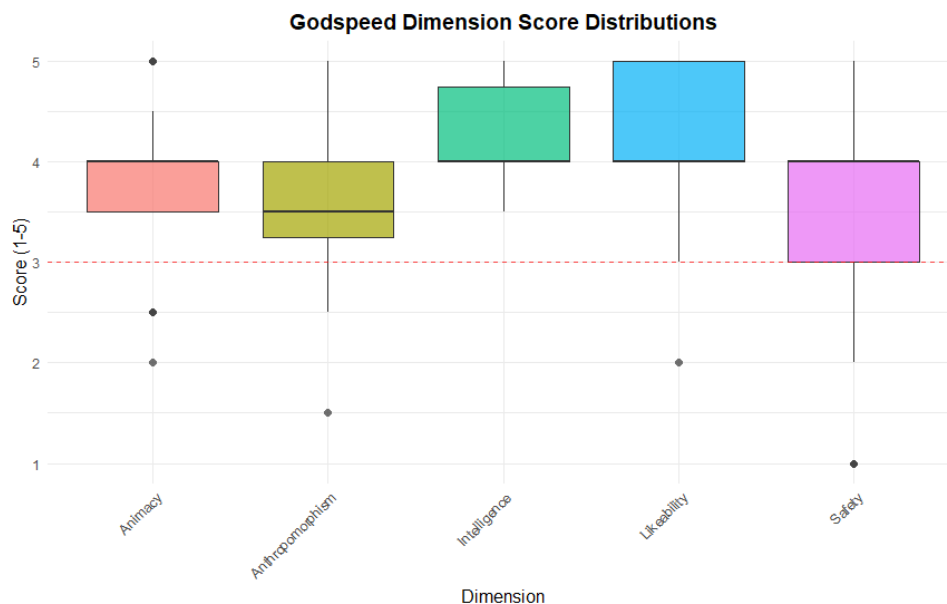
Wilcoxon signed-rank test and Cliff's Delta effect size for mood change before and after interaction

Statistic	Value	Interpretation
Wilcoxon signed-rank <i>V</i>	4.5	$p = 0.015$ (significant)
Cliff's Delta $\delta$	-0.391	Medium effect (95% CI: -0.651, -0.048)

The Wilcoxon signed-rank test (Table 2) indicated a statistically significant increase in mood scores after interacting with the robot ( $V = 4.5$ ,  $p = 0.015$ ), confirming that the shift observed in the descriptive statistics is unlikely due to chance. Cliff's Delta, an effect size measure for paired data, was



**Figure 4:** Boxplots of the answer to the question: "How do you feel about talking to a robot?" before and after interacting with Furhat.



**Figure 5:** Boxplots of the Godspeed dimensions.

-0.39, indicating a medium negative effect of the interaction on participants' mood. The 95% confidence interval for the effect size (0.048–0.651) suggests that the positive impact is robust across the sample.

These results indicate that the interaction with Furhat not only increased participants' mean mood but also led to more consistent positive responses, with a negative statistically significant and medium-sized effect. Even participants who initially reported lower mood ratings showed improvement after the interaction, highlighting the potential of social robots to positively influence affective states in short-term engagements.

**Godspeed.** The Godspeed Questionnaire assessed participants' perceptions of the Furhat robot across five dimensions: Perceived Intelligence, Likeability, Animacy, Anthropomorphism, and Safety. Overall, results indicate a positive reception, with particularly strong evaluations in cognitive and affective aspects.

Figure 5 illustrates the distribution of scores across the five dimensions of the Godspeed questionnaire.



Participants rated the system particularly high on Intelligence and Likeability, with median values close to the upper end of the scale, suggesting that the conversational AI and robot were perceived as both competent and engaging. Anthropomorphism and Animacy scored moderately, reflecting that while the system was viewed as somewhat lifelike, participants did not strongly anthropomorphize it, consistent with the design focus on clarity and neutrality rather than human mimicry. Safety ratings showed wider variability, with some low outliers, indicating that although most participants perceived the interaction as safe, a minority expressed reservations. The red dotted line at score 3 marks the neutral threshold, showing that all median scores remained above neutrality, highlighting generally favorable impressions.

## 5.2. System Usability Scale (SUS) Results

The System Usability Scale (SUS) was administered to all 19 participants to assess the perceived usability of the Furhat-based monitoring system. The overall SUS score averaged 54.61 (SD = 4.88; range 47.5–67.5), with a median of 55. This value falls below the conventional acceptability threshold of 68, suggesting that while participants generally managed to use the system effectively, its usability remains at a prototype stage and requires further refinement.

At the item level, several strengths emerged. Participants rated the clarity of the robot’s responses very highly (M = 4.74/5), followed by overall ease of use (M = 4.63/5), integration of functionalities (M = 4.42/5), willingness to recommend the system (M = 4.37/5), and confidence in using it (M = 4.21/5). Reversed items also showed positive evaluations, with satisfactory ratings for absence of unnecessary complexity (M = 4.21/5), system not cumbersome (M = 4.16/5), and autonomy in use (M = 3.58/5). No item scored critically low, and the overall mean across items was 4.2/5, pointing to a generally favorable user experience.

The discrepancy between the relatively low global SUS score and the positive ratings of individual items can be partly explained by the prevalence of neutral responses (score = 3 on a 1–5 scale), a pattern consistent with the well-known *central tendency bias*. Rather than signaling dissatisfaction, this suggests cautious or moderate evaluations from participants, possibly reflecting the novelty of interacting with a social robot for mental health purposes.

In summary, despite a SUS score below the standard benchmark, the detailed analysis indicates **good basic usability**, particularly with respect to clarity of communication and ease of use. These findings highlight opportunities for targeted improvements aimed at reducing user hesitation and enhancing overall acceptance in future iterations of the system.

## 5.3. Net Promoter Score (NPS) Results

To assess participants’ willingness to recommend the system to others, we collected ratings on a scale from 0 (not at all likely) to 10 (extremely likely). According to the standard NPS methodology, scores of 9–10 are classified as *Promoters*, scores of 7–8 as *Passives*, and scores of 0–6 as *Detractors*.

Out of 19 participants, none were detractors, while 9 (47.4%) were promoters and 10 (52.6%) were passives. This distribution resulted in a Net Promoter Score (NPS) of **47.4**, which falls into the positive range typically interpreted as “good” for technology acceptance studies. The absence of detractors and the majority of respondents falling into either the promoter or passive category indicates an overall favorable perception of the system, with strong potential for recommendation in mental health contexts.

**Table 3**

Distribution of NPS categories among participants

Category	Score Range	Count	Percentage
Promoters	9–10	9	47.4%
Passives	7–8	10	52.6%
Detractors	0–6	0	0.0%
<b>Net Promoter Score (NPS)</b>			<b>47.4</b>



## 5.4. Qualitative results

In addition to quantitative measures, participants provided open-ended feedback on their interaction with the Furhat robot. Their comments were analyzed thematically across three guiding questions: *strengths of the interaction*, *potential usefulness of Furhat for other people with mental health conditions*, and *suggestions for improvement*.

**Strengths of the Interaction.** A recurring theme was the **clarity and fluidity of communication**. Several participants described the interaction as *“the communication was simple and fast”* or *“the dialogue felt natural and fluid, as if talking to a person”*. The **non-judgmental and neutral stance** of Furhat was also appreciated—one participant noted that its *“constant and neutral expressiveness was reassuring”*. This aligns with findings from a systematic review indicating that social robots’ non-judgmental nature can be particularly important for individuals who may initially avoid therapy.

Moreover, participants emphasized the robot’s ability to create a low-pressure environment for self-expression: *“Some people fear being judged when talking about their mental health. Furhat, as an artificial entity, reduces the possibility of feeling inadequate, stimulating greater openness and sincerity.”*

Emotional stability emerged as another strength: interacting with Furhat avoided emotional destabilization that might occur with a human professional—*“With a professional, exposing emotions could destabilize me, which does not happen with Furhat for obvious reasons.”* This echoes insights from studies on socially assistive robots used in mental health contexts, where empathic presence supports stable emotional engagement.

**Usefulness for People with Mental Health Conditions** Many participants believed Furhat could serve as a valuable tool. Some highlighted its role in **stimulating communication**: *“it could give great help in stimulating communication”*. Others envisioned clinical applications: *“if used alongside a doctor, it could be an excellent tool for evaluation or patient stimulation”*. This is consistent with systematic survey findings that social robots like NAO, CRECA, or Paro have been explored for counseling and motivational interviewing tasks.

Furhat’s consistent, non-judgmental demeanor was seen as supportive: *“With Furhat I established a constant, non-judgmental dialogue, which can help people feel listened to and understood.”* Additionally, practical benefits were noted, such as easing clinical load and improving care access—*“It could lighten the clinical burden and facilitate access to care, especially where staff is limited.”* These observations are aligned with research showing companion-like robots foster therapeutic alliance and improve psychological well-being in longer-term deployments.

**Suggestions for Improvement** The most frequently proposed improvement was to make Furhat appear **more human-like**. Suggestions ranged from cosmetic enhancements—*“I would improve its appearance by making it more human, maybe with hair”*; *“its humanity, because it is still too mechanical”*—to expressiveness: *“it should be able to perform other expressions with face and voice, so that conversations do not become repetitive”*. Research in HRI emphasizes that recognizable emotional expressions and anthropomorphism significantly impact social engagement.

Another recurrent theme concerned **enhanced emotional and non-verbal recognition**. Participants suggested features such as *“implementing recognition of body language and micro-expressions”* or *“detecting the patient’s emotions through their tone of voice”*. These align closely with affective computing advances and the demonstrated benefits of empathic robots in recognizing and responding to emotions to increase engagement and likability.

Finally, some participants highlighted the value of **memory and personalization**, proposing *“it would be useful if Furhat could remember key elements of previous conversations”* to support continuity and richer long-term interaction. Cognitive robotics research supports the idea that memory systems—such as narrative or episodic memory architectures—can help robots ground interactions over time and respond more appropriately based on past exchanges.

**Summary** Participants found key strengths in Furhat’s clarity, neutral stance, and accessibility. They envisioned its utility for mental health settings—especially in contexts where non-judgmental, consistent companionship is beneficial. To enhance Furhat’s impact, participants recommended improvements in anthropomorphism (appearance and expressiveness), emotional perception, and memory-driven personalization. These suggestions align well with ongoing research in socially assistive robotics and affective human-robot interaction.

## 5.5. Threats to Validity

While the evaluation provided valuable insights into participants’ affective responses and perceptions of the Furhat robot, several limitations should be considered when interpreting the results.

**Internal Validity** The study relied on self-reported measures of mood and subjective questionnaires, which are susceptible to response bias and social desirability effects. Participants may have reported more positive experiences due to perceived expectations or the novelty of interacting with a robot. Additionally, the lack of a control condition prevents ruling out alternative explanations for mood improvements, such as placebo effects or general engagement with any interactive system.

**Construct Validity** Although validated instruments (SUS, Godspeed, NPS) were used, some constructs, particularly affective change, may not be fully captured by single-item mood ratings. Open-ended feedback partially mitigates this limitation, but qualitative responses are inherently subjective and depend on participants’ willingness to articulate their experiences.

**External Validity** The sample consisted of 19 participants recruited via convenience sampling, with a predominance of young adults and relatively balanced gender representation. This limits the generalizability of findings to broader populations, including older adults, clinical populations, or individuals with different cultural backgrounds. Furthermore, participants’ prior knowledge of social robots varied, which could influence both engagement and reported perceptions.

**Ecological Validity** The interaction took place in a controlled environment without long-term follow-up. Therefore, it remains unclear whether the observed positive effects on mood or user acceptance would persist over repeated sessions or in real-world settings, such as clinical practice.

**Statistical Conclusion Validity** Given the small sample size, the study is underpowered for detecting small effect sizes. While the Wilcoxon signed-rank test indicated a significant mood increase and Cliff’s Delta suggested a medium effect, results should be interpreted cautiously. Larger sample studies are necessary to confirm these preliminary findings.

To address these threats, the study combined quantitative and qualitative measures, allowing for triangulation of self-reported scores with participants’ narrative feedback. Future work should incorporate larger and more diverse samples, longitudinal follow-ups, and control conditions to strengthen the validity and reliability of conclusions regarding the use of social robots in mental health evaluation and monitoring.

## 5.6. Discussion

The present study tried to answer the following research question:

**RQ: Is it possible to use a social robot as a tool for supporting the evaluation and monitoring of mental health?**

Based on the results:

- Participants’ self-reported mood significantly improved after interacting with Furhat, as shown by the Wilcoxon signed-rank test ( $V = 4.5$ ,  $p = 0.015$ ) and a medium-sized effect according to Cliff’s Delta ( $\delta = -0.39$ ).

- Participants reported favorable perceptions of anthropomorphism, animacy, likeability, perceived intelligence, and safety (Godspeed), and a positive willingness to recommend the system (NPS = 47.4). Usability remains a prototype stage and requires further refinement.
- Qualitative feedback indicated that the robot provided a neutral, non-judgmental interaction, encouraging participants to express their emotions freely and suggesting potential applicability in mental health contexts.

Together, these findings support the conclusion that *social robots like Furhat can serve as effective tools for supporting the evaluation and monitoring of mental health*, at least in short-term, controlled settings. The combination of quantitative and qualitative evidence indicates that such robots can capture affective states, provide a safe and engaging environment, and be positively received by users, addressing both functional and affective aspects of mental health assessment. The interaction with the user should still be improved.

Beyond the methodological limitations outlined in Section 5.7, this study raises broader issues related to the responsible use of social robots in mental health contexts. While results show promise, deploying such systems in real clinical practice requires careful attention to additional risks and safeguards.

A first challenge concerns the **transferability to clinical populations**. Our participants did not present mental health conditions, and their generally positive reception may differ substantially from that of patients facing distress, stigma, or cognitive impairments. For clinical adoption, the system must therefore be validated with diverse patient groups, accounting for specific needs and vulnerabilities.

A second issue is the **risk of misinterpretation and over-reliance**. Although the robot can engage in empathetic dialogue, it is not a substitute for professional judgment. If patients perceive it as a therapeutic agent, there is a danger of reducing contact with human clinicians. Clear communication of the robot's supportive but non-therapeutic role is essential to avoid unrealistic expectations.

*Ethical and privacy concerns.* Conversations with the robot may involve sensitive disclosures. These data must be managed under strict security protocols and anonymization strategies, while ensuring transparency about storage, usage, and clinician access. Patients should be empowered to control their data and consent to its use, minimizing the risk of exploitation or loss of trust.

Finally, the design of the system must balance *engagement and authenticity*. While participants appreciated Furhat's neutrality, some expressed a desire for greater emotional expressiveness and personalization. However, excessive anthropomorphism could raise concerns about manipulation or attachment to non-human agents. Future design should carefully calibrate the robot's behavior to support rapport without undermining authenticity or autonomy.

In summary, the potential of social robots for mental health monitoring extends beyond technical feasibility. Their real impact will depend on clinical validation, ethical safeguards, and a design philosophy that positions them as complementary, transparent, and trustworthy partners in care.

## 6. Conclusion and Future Work

This work presented a preliminary study on the use of the Furhat social robot, integrated with conversational AI, to support mental health evaluation and monitoring. The findings indicate that participants experienced significant mood improvements and valued the non-judgmental, supportive stance of the robot, suggesting its promise as a tool for fostering open communication and engagement in mental health contexts.

However, the limited sample size and the absence of participants with diagnosed mental health conditions mean that the results must be interpreted cautiously. The next step is to validate the system in real-world clinical settings with diverse patient populations and over longitudinal deployments, in order to better assess its efficacy and acceptability.

Future work will focus on incorporating *standardized clinical interview protocols* into the conversational process. This enhancement will enable systematic and clinically meaningful data collection, which can then be analyzed to support continuous monitoring, early detection of relapse, and more

personalized mental health care. By aligning with established clinical practices and ensuring rigorous ethical safeguards, the system has the potential to evolve into a practical, responsible, and effective tool for supporting both patients and clinicians in mental health monitoring.

## Acknowledgments

The authors thank the anonymous participants.

This research has been financially supported by the European Union NEXTGenerationEU project and by the Italian Ministry of the University and Research MUR, a Research Projects of Significant National Interest (PRIN) 2022 PNRR, project n. D53D23017290001 entitled "Supporting schizophrenia PatiEnts Care wiTh aRtificial intelligence (SPECTRA)", Principal Investigator: Rita Francese.

## Declaration on Generative AI

During the preparation of this work, the authors used Grammarly to perform a grammar and spelling check.

## References

- [1] R. Francese, M. Delle Cave, P. Barra, M. Ciccarelli, G. De Simone, F. Iannotta, F. Iasevoli, Deeptald: a system for supporting schizophrenia-related language and thought disorders detection with nlp models and explanations, *Multimedia Tools and Applications* (2025) 1–34.
- [2] A. A. Scoglio, E. D. Reilly, J. A. Gorman, C. E. Drebing, Use of social robots in mental health and well-being research: systematic review, *Journal of medical Internet research* 21 (2019) e13322.
- [3] C. S. González-González, V. Violant-Holz, R. M. Gil-Iranzo, Social robots in hospitals: a systematic review, *Applied Sciences* 11 (2021) 5976.
- [4] T. H. Goodspeed, Evaluation of human factors in robot design, *Human Factors* 24 (1982) 511–517.
- [5] J. Brooke, Sus: A “quick and dirty” usability scale, *Technical Report*, 1996.
- [6] H. Abdollahi, M. H. Mahoor, R. Zandie, J. Siewierski, S. H. Qualls, Artificial emotional intelligence in socially assistive robots for older adults: a pilot study, *IEEE Transactions on Affective Computing* 14 (2022) 2020–2032.
- [7] R. Francese, F. Iasevoli, M. Staffa, The spectra project: Biomedical data for supporting the detection of treatment resistant schizophrenia, in: *International Conference on Human-Computer Interaction*, Springer, 2024, pp. 353–367.
- [8] R. Francese, L. De Santis, F. Iannotti, I. Felice, XAI for supporting gait analysis of patient with schizophrenia (2024).
- [9] L. Galluccio, L. D’Errico, M. Giordano, M. Staffa, Advancing eeg-based emotion recognition: Unleashing the power of graph neural networks for dynamic and topology-aware models, in: *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–8. doi:10.1109/IJCNN60899.2024.10650427.
- [10] Furhat Robotics, Furhat social robot platform, 2025. Product and research overview, Furhat Robotics website.
- [11] S. Hillmann, et al., Multimodal interactive assistance for the digital collection of patient-reported outcome measures (mia-prom), 2024. Charité – Medizinische Universität Berlin project.
- [12] Translational Research Centre for Digital Mental Health, Tung Wah College, Robotic support for mental health with furhat, 2024. TWC TRC-DMH website.
- [13] N. Cherakara, F. Varghese, et al., Furchat: An embodied conversational agent using llms, combining open and closed-domain dialogue with facial expressions, *arXiv preprint* (2023).
- [14] D. Feil-Seifer, M. J. Mataric, Defining socially assistive robotics (2005).
- [15] L. Aymerich-Franch, I. Ferrer, Socially assistive robots’ deployment in healthcare settings: a global perspective, *arXiv preprint* (2021).

- [16] A. A. J. Scoglio, E. D. Reilly, et al., Use of social robots in mental health and well-being research: Systematic review, *J. Medical Internet Research* 21 (2019) e13322. doi:10.2196/13322.
- [17] H. Abdollahi, M. H. Mahoor, et al., Artificial emotional intelligence in socially assistive robots for older adults: A pilot study, *IEEE Transactions on Cognitive and Developmental Systems* (2022).
- [18] F. Dino, R. Zandie, H. Abdollahi, et al., Delivering cognitive behavioral therapy using a conversational social robot, *arXiv preprint* (2019).
- [19] Human-Robot Interaction 2022 Workshop, Exploring social robot-based tools for community mental health, in: *Proceedings of ACM/IEEE HRI 2022*, 2022.
- [20] A. Nandanwar, et al., Assessing stress, anxiety, and depression with social robots via conversational ai, *ACM Press proceedings* (2023).
- [21] S. Thunberg, et al., Older adults' perception of the furhat robot, *ACM Proc. HRI* (2022).
- [22] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, *International journal of social robotics* 1 (2009) 71–81.