

MOON Framework: An Emotionally Adaptive Voice Interaction Model for Older Adults

Hasan Sagga¹, Richard Stone²

Management Information Systems Department, University of Jeddah, Jeddah, Saudi Arabia¹

Human-Computer Interaction, Iowa State University, Ames, USA¹

Industrial and Manufacturing Systems Engineering Department, Iowa State University, Ames, USA²

Abstract—As the world experiences a rapidly aging population, it has become a pressing design challenge to ensure that emerging digital technologies remain understandable, supportive, and meaningful for older adults. One of the most promising modalities for promoting inclusive interaction is the Voice User Interface (VUI), supported by artificial intelligence (AI) and speech recognition. However, most existing VUIs emphasize functional accuracy and overlook users' emotional conditions, such as anxiety, confidence, or frustration, which significantly affect long-term engagement and adoption. To address this gap, this study introduces a structured and adaptive model for developing emotionally intelligent voice interfaces, namely MOON Framework (Model-Observation-Optimization-Nurture). This framework integrates demographic and linguistic profiling (Model), real-time emotional perception (Observation), adaptive vocal modulation (Optimization), and feedback-driven learning (Nurture) to create a closed-loop system capable of adjusting dynamically to user experience. Grounded in Affective Computing Theory, Socioemotional Selectivity Theory (SST), and Adaptive User Modeling, MOON Framework conceptualizes empathy as a measurable and computable dimension of VUIs. Unlike existing affective systems that rely primarily on visual or physical expressiveness, MOON demonstrates that vocal empathy, achieved through modulation of tone, cadence, and speech rate can foster comparable emotional attunement. Its cyclical design transforms emotion from a passive observation into an active driver of system adaptation. Focusing on familiarity and confidence specifically in older adults, MOON provides both a theoretical foundation and a practical framework for creating emotionally inclusive AI technologies that promote trust, engagement, and long-term well-being.

Keywords—Voice User Interfaces; emotional adaptivity; affective computing; older adults; human-computer interaction; empathy; accessibility; adaptive systems; MOON Framework

I. INTRODUCTION

Inclusive technology design is now a matter of greater concern as the world population is aging. VUIs are a potentially effective option to enhance the accessibility of older adults (intuitive, hands-free interface) [1], [2], [3], [4]. However, even with their potential, the majority of the existing systems are not sensitive to affective states, including confusion, confidence, or frustration, and react to verbal commands without sensibilities [3]. The aspect of trust and continued use of technology is highly affected by older adults based on emotional factors, and these concepts should be addressed to bring about the concept of real inclusivity [2], [5], [6].

Some of these reasons, besides the technological constraints, include the emotional and contextual constraints

which influence the manner in which the older adults use the computing systems [3], [6], [7]. The modern interface design practice as explored in the previous work [8] tends to presume homogeneous sensorial and cognitive abilities and undervalues the emotional aspect of adoption of technology. The literature review has shown that the confidence and familiarity of older users decrease when the digital systems do not recognize the user's emotional clues and do not change their communication style in response [2], [3]. Based on these findings, the current study presents those findings in a formalized form of a structured model of interaction that does not view emotion as a by-product, but as a variable of central concern in design. This continuation of the earlier research based on empirical findings into the development of formal frameworks makes this current work a conceptual follow-up of the previous research [3], [7].

In this respect, it is more crucial to reconsider how ubiquitous, emotionally responsive technologies have the potential to make the aging demographic more engaged and well. With computing being integrated into everyday practices, such as smart houses, social interaction, and health care, its usefulness with the elderly population lies in how successfully it can incorporate emotional sensitivity into communication [2], [3], [7]. Computing is everywhere, and this is extraordinary in possibilities except that, unless designed in an emotionally aware way, it will only perpetuate digital exclusion. MOON Framework (Model, Observation, Optimization, and Nurture) is a response to this dilemma as it offers a methodical, reproducible approach to the integration of empathy in voice-based systems to make such technologies easy, meaningful, and trustworthy to various users.

The research problem stated in this study relates to the lack of emotional intelligence in the existing Voice User Interfaces and the subsequent gap of accessibility and sustained engagement for older adults. In response to this, in this study, the following research questions are asked: 1) How can, in voice interaction systems, emotional adaptivity be formally modeled?, 2) What theoretical elements and interaction mechanisms will be needed to integrate empathy into VUIs? and 3) How can such a framework help promote familiarity, confidence and long-term use to older users? Accordingly, the purpose of this research is to construct a structured theoretical framework (MOON), which conceptualizes emotional adaptivity in voice interaction and to describe its phases of design, its underlying principles and its implications for practice. The importance of this work is to make a step towards inclusive design by shifting emotion from being a secondary consideration in design to being a

primary driver of accessibility to create emotionally responsive AI systems for trust, engagement, and well-being among older adults.

The structure of this study includes an introduction, related work, conceptual foundation, framework development methodology (MOON Framework), discussion, and conclusion and future work.

II. RELATED WORK

The study of emotionally intelligent systems has developed rapidly over the past three decades, particularly in affective computing, psychology, and HCI. Prior research [9], [10], demonstrated that the integration of emotion recognition within computing enables both functional accuracy and affect-sensitive communication. Recent advances in emotional artificial intelligence have expanded these foundations, yet most models remain focused on embodied robots and visually interactive agents rather than auditory modalities [11], [12], [13]. As a result, voice-based systems remain comparatively underdeveloped in their responsiveness to emotion, a gap that motivates the design of MOON Framework. This framework is particularly relevant for older adults, whose interactions with technology involve distinctive emotional and contextual challenges [8], [14]. Accordingly, the following subsections review three areas of prior research, affective computing, VUIs for older adults, and adaptive HCI frameworks, that collectively inform and contextualize the development of MOON.

A. Emotional AI and Affective Computing

Affective computing has long served as the scientific foundation for engineering systems capable of perceiving, interpreting, and responding to human emotions. Recent systematic reviews, such as those by Wang et al. (2022)[15], highlight how the field has progressed from basic emotion detection toward more sophisticated affective modeling. Early research typically focused on improving multimodal emotion recognition by combining facial expressions, gestures, physiological cues, and vocal signals to increase accuracy and robustness [14], [16]. As computational methods evolved, interactive systems began to incorporate adaptive mechanisms that enabled them to mirror, regulate, or guide users' emotional states-transitioning toward socially oriented artificial intelligence capable of more human-like engagement [17], [15].

For older adults, emotionally intelligent technologies hold particular importance, as empathetic interactions have been shown to improve comfort, motivation, and engagement. Studies in socially assistive robotics and expressive interfaces demonstrate that emotionally responsive behaviors, whether delivered through facial expressions, gestures, or vocal modulation, can enhance confidence and interest among older users [17], [18], [19]. However, despite these promising effects, multimodal expressiveness poses limitations in scalability, as it often relies on a visible embodiment or spatially dependent channels, making it less practical for voice-only or remote communication scenarios [20], [21].

Recent advances in affective computing have therefore prompted a shift toward auditory-based emotional systems.

These approaches leverage prosodic cues, such as intonation, pace, pausing structure, and timbre, to simulate empathy and convey affective intent without relying on a visual body [15], [21]. Research further shows that speech-based emotion recognition can achieve high accuracy and support emotionally attuned interaction, providing strong justification for frameworks like MOON that prioritize vocal adaptivity over multimodal complexity [21], [22].

B. Voice User Interfaces and Older Adults

Voice interaction is widely recognized as an intuitive and accessible communication modality, particularly for individuals with visual, motor, or literacy limitations. Older adults benefit substantially from VUIs because spoken communication reduces reliance on complex graphical interfaces, which may be difficult to navigate due to age-related perceptual changes [1], [2], [23], [24]. However, these benefits depend heavily on how well the system manages both linguistic and emotional complexity. Prior research in HCI has emphasized that VUIs designed for older users must account for their unique emotional needs and challenges, particularly regarding trust, frustration tolerance, and communication clarity [3], [8].

Despite this potential, commercial VUIs such as Amazon Alexa and Google Assistant remain predominantly command-based and rarely adapt their vocal style to users' affective states. This limitation is especially problematic for older adults, whose emotional attitudes toward technology are often shaped by anxiety, self-efficacy, and perceived control [2], [5], [25]. For example, recurring recognition errors may lead to frustration or embarrassment, reducing trust and willingness to continue using the system. Without mechanisms to sense or respond to such emotional dynamics, even well-designed systems risk disengaging those who may benefit from them most.

The tension between functional accuracy and emotional attunement is a persistent challenge highlighted across empirical studies of older adults and VUIs. Prior work shows that while older users appreciate the convenience of voice assistants, the interaction is often described as insufficiently empathetic or personalized to their needs [2], [26]. Conversely, emotional warmth and social presence have been shown to be stronger predictors of long-term use than usability alone [27], [28]. These findings underscore a longstanding design gap: VUIs may perform well technically yet fail emotionally. Addressing this mismatch requires shifting from task-only design toward supporting emotional continuity, ensuring the system maintains confidence, reassurance, and attunement throughout interaction [8].

MOON Framework is proposed as a response to this challenge by embedding emotional adaptivity into the operational core of VUI behavior. Rather than treating emotion as an optional layer, MOON emphasizes empathy, familiarity, and confidence as primary design variables, offering a structured approach to closing the gap between functional accessibility and emotional responsiveness in voice interaction for older adults.

C. Adaptive and Personalized HCI Frameworks

The concept of adaptation has long been central in HCI, where user modeling and personalization aim to enhance usability, engagement, and long-term relevance. Adaptive interfaces, as characterized by Fischer (2001), modify their behavior based on the user's actions and learned preferences. However, the vast majority of adaptive HCI systems remain cognitively adaptive, adjusting task difficulty, content delivery, or interface layout, rather than emotionally adaptive. These systems respond to observable user actions but typically fail to account for underlying affective states such as stress, hesitation, or frustration [15], [29], [30]. This distinction is particularly important in the context of older adults, whose emotional experience during interaction cannot be fully captured through performance metrics alone. An older user may complete a task successfully and yet experience anxiety or uncertainty. A cognitively adaptive interface would interpret this as a success, while an emotionally adaptive system would detect subtle affective cues, such as tone, pitch, or pacing, that indicate discomfort and respond accordingly.

Recent models in adaptive HCI have explored multimodal emotion recognition as a means of improving responsiveness. Studies have examined systems capable of detecting emotions through facial expressions, voice, gestures, or physiological signals, demonstrating strong real-time detection accuracy [15], [29], [31]. However, very few of these frameworks extend adaptation across entire interaction sessions. Most systems are episodic: they sense emotion, adapt once, and then reset. They lack mechanisms to preserve emotional history, remember prior states, or build familiarity over time. This limitation is particularly relevant for older adults, whose emotional comfort increases when systems demonstrate consistency and memory of prior interactions. MOON Framework addresses this challenge within its Nurture phase, which uses outcomes from each conversational cycle to shape the system's future behavior. Rather than offering one-step adjustments, MOON supports both intrasessional and intersessional adaptation, building a longitudinal emotional model that evolves with continued use. Its feedback loop is conceptually more advanced than classical adaptive systems, which are typically unidirectional ("input → output → end"). MOON instead integrates continuous emotional learning, ensuring that empathy becomes a dynamic quality that improves with repeated interactions. As the system accumulates experience, it generates increasingly refined adaptive responses aligned with human-centered AI principles, ultimately becoming more emotionally sensitive and personally attuned to the user [32].

III. CONCEPTUAL FOUNDATION

MOON Framework extends the conceptual model developed by Sagga and Stone (2025) [8], in which emotional familiarity and user confidence were identified as central determinants of older adults' participation in digital technologies. While prior research has emphasized the value of emotional attunement for trust, usability, and engagement, these findings have largely been qualitative in nature. The present work advances this line of inquiry by formalizing those insights into a structured computational model. Specifically, MOON Framework operationalizes empathy as both an

input, captured through measurable emotional cues, and an output expressed through adaptive system behavior. This conceptualization builds on foundational perspectives in affective computing together with SST, which explains the growing emotional relevance of interactions for older adults [33], and principles from adaptive user modeling. Integrating these theories provide a unified methodological basis for systems capable of perceiving emotion, adapting vocal responses accordingly, and refining their behavior through iterative feedback, thereby positioning emotional inclusiveness not merely as a design principle but as an ongoing interactive learning process.

A. Affective Computing: From Recognition to Empathy

Affective computing provides the technological foundation on which MOON Framework is built. Introduced by Rosalind Picard (1997) [34], the field defines computational systems capable of recognizing, interpreting, and simulating human emotions. Early studies in affective computing primarily focused on emotion detection, classifying user states as positive, negative, or neutral based on physiological and behavioral cues [10]. Over time, however, research expanded toward affective adaptation, in which emotion functions not only as an input signal but also as a determinant of system behavior [14].

Within this evolution, MOON Framework aligns with the second wave of affective computing research that emphasizes emotion-aware adaptation rather than simple recognition [14], [15]. The framework extends traditional detection mechanisms toward what may be described as emotion reciprocity: the system's ability to respond empathetically to a user's affective state. In voice interaction, this reciprocity is achieved by identifying patterns in speech, such as pitch shifts, rhythm changes, speech rate, and pause length, and using these cues to modulate the system's own vocal output [14], [35].

A central contribution of MOON to the field is its shift toward unimodal empathy, relying solely on the auditory channel rather than multimodal combinations involving facial, physiological, or textual data. Multimodal systems, although powerful, often introduce complexity and accessibility issues, particularly for older adults who may be uncomfortable with cameras or wearable sensors [15], [36]. By leveraging prosodic cues alone, MOON demonstrates that meaningful empathy can be achieved using minimal but highly informative vocal signals, an approach supported by evidence that listeners are remarkably skilled at interpreting emotion from voice even without visual indicators [35].

In this sense, MOON reframes affective computing as an interactive emotional dialogue. The user communicates through affective speech patterns, the system perceives and interprets these signals, and the system's vocal output is shaped accordingly through adaptive modulation [14], [15]. This interaction mirrors natural human conversation, where subtle variations in tone, rhythm, and cadence convey interest, concern, or attentiveness. By formalizing this exchange within a computational cycle, MOON transforms emotional intelligence from an abstract design ideal into a functional component of everyday voice-based interaction.

B. Socioemotional Selectivity Theory (SST): The Psychology of Emotional Relevance

Socioemotional Selectivity Theory provides the psychological grounding for which emotion-centered design is essential for older adults. Developed by Carstensen (2006) [33], SST argues that increased awareness of limited time horizons with age leads to a motivational shift: older adults prioritize emotionally meaningful goals, stable relationships, and affective fulfillment over novelty exploration. This shift gives emotional well-being a central role in shaping attention, memory, and decision-making, as supported by later research in lifespan development [37], [38].

Applied to human-technology interaction, SST suggests that older adults may approach digital systems not merely as tools but as potential social partners that can offer comfort, reassurance, and emotional stability. As a result, systems that are sensitive to affective cues are more likely to foster trust, credibility, and long-term engagement. Conversely, emotionally neutral or insensitive systems can be perceived as cold, burdensome, or cognitively demanding [39], [40].

MOON Framework integrates SST by emphasizing emotional congruence, the alignment between the user's affective state and the system's communicative style. Through prosodic modulation, a VUI can mirror or complement the user's emotional tone, thereby enhancing psychological comfort and interpersonal synchrony. For example, frustration may be alleviated by a calm, tolerant vocal response, whereas a challenging task may benefit from an encouraging and steady tone [41], [42]. SST further indicates that older adults actively regulate their emotions and prefer interactions that maintain positive affect rather than those that introduce excitement or novelty. This insight reinforces one of MOON's core design principles: adaptation should aim at stabilizing the user's emotional experience, not arousing it. Thus, the system's goal is not to demonstrate human-like exuberance but to deliver reliability, reassurance, and emotional predictability, qualities that align naturally with the motivational priorities of aging adults [41], [42].

By incorporating socioemotional selectivity directly into its structure, MOON Framework reframes emotion as a functional design variable rather than an incidental feature of interaction. Emotional adaptivity becomes both a design strategy and an ethical responsibility, one that supports the psychological well-being of users while maintaining technological clarity and accessibility.

C. Adaptive User Modeling: Learning from Interaction

Affective computing explains the mechanisms through which systems detect and interpret emotion, while Socioemotional Selectivity Theory provides the psychological justification for why systems should attend to these affective states. Adaptive user modeling complements these perspectives by offering the computational structure through which emotional information can be integrated over time. Rooted in human-computer interaction and cognitive science, adaptive modeling involves constructing dynamic representations of users that evolve as systems accumulate interaction data. Whereas traditional user models rely on static profiles, such as age, gender, or stated preferences, adaptive models

continually update their parameters based on observed behavior and contextual cues, allowing the system to refine its understanding of the user across sessions.

MOON Framework extends adaptive user modeling into the affective dimension. Its Model and Nurture phases reflect this structure: the Model phase establishes an initial baseline built from demographic and linguistic characteristics, while the Nurture phase incrementally updates this profile using emotional and behavioral outcomes from prior interactions. Each cycle contributes to a gradual recalibration of the system's emotional responses, enabling MOON to simulate a human-like trajectory of experiential learning and emotional attunement.

A key distinction of MOON is its use of affective feedback as a learning signal, rather than relying solely on conventional performance indicators. Traditional adaptive algorithms measure success through error rates, task duration, or accuracy, but MOON incorporates emotional variables, such as hesitation, confidence, frustration, or satisfaction, as informative cues. This approach transforms affective experience into quantitative data, enabling the system to refine not only its task strategies but also its communicative behaviors through adaptive prosody.

Through this mechanism, MOON achieves what can be described as affective adaptivity, an ability to optimize its emotional intelligence by continuously observing user affect. This methodology integrates computational adaptivity with psychological empathy, creating a form of emotional learning that parallels reinforcement learning, but grounded in affective signals rather than rewards. Over time, the system becomes increasingly sensitive to individual users, reinforcing familiarity, personalization, and trust in long-term interactions [14].

D. Integrative Theoretical Perspective

Affective computing, SST, and adaptive user modeling collectively form a cohesive theoretical foundation for emotionally adaptive voice interaction within VUI design. Affective computing provides the algorithmic infrastructure required for emotion recognition and expressive modulation, enabling systems to detect and interpret affective cues. SST clarifies why such adaptations are particularly meaningful for older adults, explaining that motivational priorities shift toward emotional fulfillment, stability, and relational significance with age. Adaptive user modeling enables these insights to be operationalized through continuous learning mechanisms that refine the system's understanding of the user over time.

These theoretical principles converge within MOON Framework, which structures emotional interaction into a logical cycle of recognition, alignment, adaptation, and reinforcement. Emotion functions simultaneously as an input to system behavior, informing how the VUI adjusts its speech rate, tone, or cadence, and as an output that shapes the user's subsequent emotional experience. This reciprocity mirrors interpersonal relationships, where emotional exchange deepens familiarity, trust, and attunement. Within the domain of aging and technology, this synthesis represents a significant step toward designing interfaces that are not only accessible and

intuitive but also emotionally supportive and sensitive to the psychological needs of older adults.

IV. FRAMEWORK DEVELOPMENT METHODOLOGY: MOON FRAMEWORK

The proposed MOON Framework represents a systematic and iterative approach to designing emotionally adaptive VUIs. It integrates the technological mechanisms of affective computing with adaptive user modeling and the psychological principles of SST, providing a unified methodology for emotion-centered interaction design [2], [3], [43]. Within this framework, emotion is treated as a fundamental organizing principle: VUIs are expected not only to recognize and interpret users' affective states but also to adjust their communicative behavior in ways that enhance familiarity, confidence, and long-term engagement among older adults [2], [3].

Fig. 1 presents a comparative summary table of the most relevant frameworks addressing older adults' emotional and interactional needs, including those by [2], [3], [17]. The table contrasts these approaches with MOON Framework across several thematic dimensions, emotional needs, personalization capabilities, long-term engagement mechanisms, user involvement, AI-based emotion recognition, comparison with younger users, and structural organization. This comparison highlights both the conceptual lineage and the unique contribution of MOON as a comprehensive, emotion-centered model.

From an architectural perspective, MOON functions as a closed-loop emotional adaptation cycle composed of four phases: Model, Observation, Optimization, and Nurture. The cycle begins with a static baseline representation (Model), which incorporates demographic, linguistic, and familiarity-related parameters. During interaction, the system gathers and analyzes vocal and emotional indicators (Observation), enabling it to identify cues such as stress, hesitation, or confidence. These cues serve as inputs to the Optimization phase, where the system modulates its vocal output, adjusting speech rate, tone, cadence, or accent, to align with the user's emotional state. Finally, the Nurture phase evaluates the interaction outcomes, updating the internal user model to support longitudinal emotional learning. The resulting feedback informs the next iteration of the Model phase, allowing the system to evolve continuously as it accumulates affective experience.

This cyclical form reflects natural human communication patterns, where individuals form mental models of others, interpret emotional feedback, adjust their behavior accordingly, and refine their understanding over time. Viewed mathematically, MOON operationalizes this process as a data-driven feedback loop, treating emotional cues as a continuous input stream that iteratively updates the system's internal state. Structurally, the framework's modular design allows seamless integration with existing speech-recognition pipelines, conversational systems, and dialogue-management infrastructure. Moreover, its adaptability enables deployment across various domains, including e-services, healthcare, education, and smart-home environments.

Aspect/Theme	MOON Framework	Liu et al. (2023)	Abdollahi et al. (2022)	Hu et al. (2024)
Emotional Needs	Central, multi-stage focus (Model, Nurture)	Central, especially companionship and enjoyment	Central, via empathic robot and mood measurement	Central, via co-design and well-being
Personalization	Iterative, AI-driven adaptation (Observation, Optimization)	Discussed, but not iterative	Via multimodal emotion recognition and adaptive dialogue	Emerges from co-design, not always AI-driven
Long-term Engagement	Explicit "Nurture" phase for ongoing adaptation and companionship	Not explicit	Not explicit	Not explicit
User Involvement	Observation and feedback integrated throughout	Survey-based, empirical	User mood and engagement measured	Co-design sessions with older adults
AI/Emotion Recognition	Multimodal, continuous, and central to framework	Not a primary focus	Central: facial and speech sentiment analysis	Not a primary focus
Comparison to Younger Adults	Explicitly claims older adults benefit more from MOON's approach	Focused only on older adults	Focused only on older adults	Focused only on older adults
Framework Structure	Four-stage, cyclical, holistic (Model–Observation–Optimization–Nurture)	Factor-based model (TAM extension)	Experimental comparison (empathic vs. non-empathic robot)	Co-design methodology

Fig. 1. MOON Framework vs. Related frameworks [2], [3], [17].

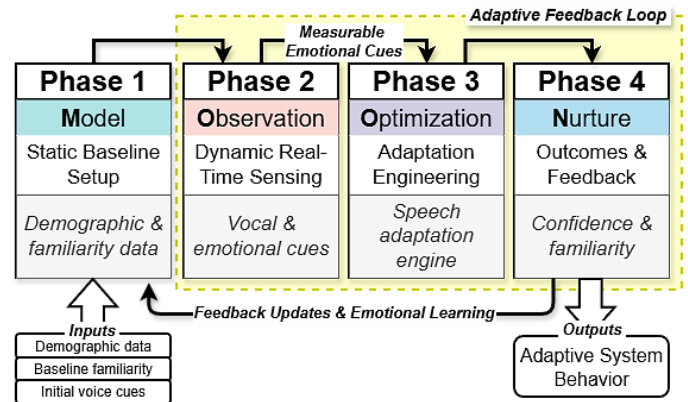


Fig. 2. MOON Framework phases and adaptive feedback cycle.

Fig. 2 illustrates the operation of the model in simplified and high-level steps. The sequential process, as a linear diagram, is running left to right and starts from phase one (Model); phase two (Observation); phase three (Optimization); to ends with phase four (Nurture), with a feedback arrow indicating a curved arrow (Adaptive Feedback Loop), which runs back to the left to the Model phase. Moreover, the description of core activities, base configuration, real-time sensing, adaptive modulation, and outcome assessment are briefly described in each compartment of the illustration. The arrows between the compartments point to the direction of data flow and affective interdependence and the feedback arrow highlights continuous recalibration. Generally, the diagram summarizes the main idea of the framework and the self-healing empathic cycle that integrates detection, response, and evolution. The model will be explained in detail for each stage

as follows:

1) Phase 1 - Model (Establishing the emotional baseline): The Model phase forms the foundational layer of MOON Framework. In this stage, the system constructs an initial user profile incorporating demographic, linguistic, and psychological attributes relevant to voice interaction. These attributes may include age, primary language, familiarity with digital technology, baseline confidence, cognitive preferences, and previously observed emotional tendencies in spoken communication. Based on this information, the system generates baseline parameters, such as preferred speech rate, pitch range, vocabulary complexity, and turn-taking rhythm, that guide the VUI's default communicative style.

Unlike traditional personalization methods, where user profiles remain static, the Model phase in MOON is intentionally dynamic. The baseline acts as a flexible emotional template that evolves as additional interaction data is accumulated. This approach allows personalization to move beyond surface-level demographic adjustment toward a deeper understanding of each user's communicative disposition and emotional inclinations [44]. For example, an initial model may indicate that an older adult with cognitive challenges responds better to slower speech. Over time, as the system detects increased confidence and reduced hesitation, it may adjust and gradually increase conversational pacing, creating a natural sense of progress and continuity [44].

By establishing both the technical parameters and the emotional expectations for future interactions, the Model phase prepares the framework for meaningful adaptivity across subsequent stages.

2) Phase 2 – Observation (Detecting emotion through voice): The Observation phase functions as the system's perceptual layer. During this stage, the VUI continuously analyzes the user's speech, extracting real-time emotional and behavioral cues. Prosodic and linguistic indicators, such as speech rate, pitch variability, vocal intensity, pause duration, and lexical choice, serve as measurable signals of the user's affective state [45]. For example, frequent pauses or a narrowing pitch range may indicate hesitation or uncertainty, whereas increased pitch, smoother pacing, or reduced pausing may reflect confidence or engagement.

To interpret these cues, the system employs emotion-recognition algorithms trained on diverse speech corpora, incorporating age-specific datasets when available. These models classify incoming signals into broad emotional categories such as calm, frustrated, confused, or satisfied [45]. Importantly, MOON Framework emphasizes subtlety over strict categorization. Rather than assigning users to rigid emotional labels, the system aims to interpret contextual and gradational affective cues. This approach aligns with evidence that emotional expression in older adults varies due to physiological changes, cultural influences, and communication styles; thus, MOON relies on probabilistic estimation rather than deterministic classification.

The Observation phase operates bidirectionally, attending to both the content of the user's utterances and the manner in which they are delivered. The system evaluates timing, phrasing, hesitation markers, and repetition patterns to infer the user's emotional state and confidence level. For instance,

repeated requests for clarification, delayed responses, or conditional phrasing may suggest confusion or reduced self-assurance. By developing this interpretive capacity, the Observation phase enables subsequent stages of the framework to respond with empathy rather than mechanical precision, ensuring that adaptation is sensitive, context-aware, and aligned with the user's affective needs.

3) Phase 3 – Optimization (Adapting through empathic response): The Optimization phase is the functional core of MOON Framework. In this stage, the system transforms the affective cues detected in the Observation phase into adaptive vocal and conversational responses. This process relies on two interconnected mechanisms: emotion interpretation, which infers the user's affective state, and speech adaptation, which adjusts vocal parameters and dialogic strategies to align the system's behavior with the user's emotional needs.

A significant component of Optimization involves adjusting speech speed and prosodic characteristics. Extensive research demonstrates that speech rate plays a critical role in older adults' comprehension, comfort, and confidence during interaction. Classic psycholinguistic studies show that older adults experience a sharper decline in processing accuracy when listening to fast-paced or dense speech, particularly at higher rates of presentation [46]. Complementing this, research on synthetic speech indicates that both younger and older adults prefer listening rates in the 150–200 wpm range, with older listeners showing reduced comfort as speed increases [47]. More recent large-scale analyses confirm that speech and articulation rates naturally slow with age, reinforcing that older adults process speech more effectively at moderated rates [48]. Additionally, work in VUI and TTS design shows that slower or adjusted synthetic speech improves intelligibility and reduces listening effort for older adults [49]. Together, these findings highlight cognitive-auditory constraints and expectations for slower, rhythmically balanced speech, supporting MOON's principle that dynamic speech-rate adjustment enhances emotional comfort and comprehension.

For example, if the Observation module detects increased frustration, the Optimization phase may slow the system's speech, soften tonal elements, and produce clearer, segmented instructions, for instance, "Let's go through this step together. I will explain it slowly". Conversely, if indicators of engagement or confidence are detected, the system can adopt a slightly faster rhythm or use more complex phrasing to maintain challenge and interest. This adaptive modulation shifts the interaction from rigid command execution to a context-responsive conversational exchange.

Importantly, the Optimization phase extends beyond prosodic adjustments. It also refines dialogue content, timing, and confirmation strategies. The system may shorten prompts to reduce cognitive load or elaborate instructions when clarity is needed. The overall goal is to maintain emotional consistency, ensuring that system behavior remains attuned to the user's affective condition [17], [50].

Through continual fine-tuning, the Optimization phase demonstrates the framework's principle of affective symmetry, each change in the user's emotional state is met with an appropriate, context-sensitive adjustment in the system's

communicative stance.

4) *Phase 4 – Nurture (Learning through feedback)*: The Nurture phase completes the adaptation cycle through three primary steps: evaluating interaction outcomes, incorporating user feedback, and refining the emotional model established in Phase 1. The purpose of this phase is to assess the effectiveness and emotional impact of the system's responses so that subsequent iterations of the model are grounded in experiential evidence. Objective performance indicators include task success rates, interaction duration, and the frequency of clarification requests. Subjective indicators include emotional responses such as confidence, comfort, and familiarity. These subjective responses may be inferred indirectly through vocal cues or determined directly through brief reflective prompts such as "Was that explanation clear for you?" This dual-measure approach ensures that both quantitative and qualitative aspects of user experience contribute to model refinement [17], [51], [52].

The information collected in the Nurture phase is analyzed to identify patterns of success or strain. For example, if consistent relief or satisfaction is detected following slower speech delivery, the system will incorporate slower cadence into the baseline parameters during the next Model phase. Through repeated interactions, this feedback mechanism enables the system to build an emotional memory, mirroring human experiential learning, and turning each new encounter into an extension of an ongoing emotional relationship between the system and the user [51], [52].

The continuous connection between the Nurture and Model phases distinguishes MOON Framework from other adaptive designs by establishing a sustained, longitudinal empathy loop. As emotional learning accumulates across interactions, the system progresses from reactive empathy, responding to immediate cues, to anticipatory empathy, where it can predict when emotional support may be needed based on prior patterns of interaction [51], [52].

V. DISCUSSION

This model also builds on our prior research [8], which demonstrated that emotional familiarity and perceived empathy strongly influence older adults' confidence and readiness to use technology. That earlier work identified a clear gap in VUI design for older users: while existing systems could reliably interpret commands, they rarely adjusted to the user's affective state. MOON Framework addresses this gap by transforming emotional and behavioral cues into measurable system behaviors through dynamic feedback and machine learning. In doing so, the framework converts previous descriptive findings into a functional architecture that can be systematically evaluated and refined over time.

The introduction of MOON Framework represents a meaningful advancement in emotionally intelligent interaction design. Its cyclical architecture integrates previously fragmented perspectives from affective computing, adaptive user modeling, and accessibility research into a cohesive system. More importantly, MOON reframes how emotional intelligence is conceptualized in artificial systems. Rather than treating empathy as a superficial feature or optional enhancement, the framework positions empathy as

an operational variable, one that can be measured, interpreted, and used to shape system behavior. As a result, learning becomes bidirectional: system behavior is shaped by user experience, and user experience evolves in response to adaptive system behavior.

By treating emotion simultaneously as a form of data and a design principle, the framework introduces significant theoretical, practical, and methodological implications for HCI and related fields. It demonstrates how affect can function not only as a user-state descriptor but also as a driver of interactional adaptation, long-term personalization, and system evolution. This dual conceptualization offers a pathway toward more responsive, inclusive, and psychologically attuned technologies for aging populations and beyond.

A. Theoretical Implications

Theoretically, MOON Framework expands the scope of affective computing by introducing a new paradigm of affective adaptivity. Traditional affective systems detect emotion but remain largely reactive, responding to affective cues without allowing emotion itself to shape system development. MOON reconceptualizes emotion as a feedback signal capable of conditioning the system over time. This principle is operationalized in the Nurture-to-Model feedback loop, which converts emotional responses into iterative learning signals. The framework thereby reverses the conventional view of emotionally intelligent AI: rather than imitating emotion, the system progressively learns emotion through accumulated interaction.

This transformation aligns with broader shifts in cognitive science toward models of dynamic cognition, which emphasize intelligence as an adaptive process rather than a repository of preprogrammed rules or static knowledge. In this view, emotional intelligence in AI is not a fixed capability but an evolving function that develops through ongoing engagement. MOON therefore bridges computational theories of learning with psychological theories of emotion regulation, offering a conceptual framework in which system-level empathy deepens over time in a manner analogous to human social cognition, where emotional memory and recurring interactions enhance interpersonal understanding.

The framework also contributes to gerontechnology and aging research by explicitly integrating SST into technical design. Unlike accessibility approaches that frame aging primarily in terms of sensory or motor decline, MOON treats aging as an emotional opportunity. Older adults place greater value on emotionally meaningful, familiar, and confidence-enhancing experiences. Technologies that adapt emotionally are therefore likely to resonate more deeply than those that merely reduce task difficulty. From this perspective, MOON embodies a form of emotionally congruent design, in which the system aligns with the motivational patterns and affective priorities of older adults.

Finally, the framework advances theoretical understanding of digital empathy. Machine empathy has traditionally been viewed as either artificial (scripted emotional mimicry) or unattainable. MOON introduces a third category: computational empathy. Unlike surface-level mimicry seen in many chatbots, computational empathy involves

identifying emotional signals, interpreting them into adaptive behavior, and refining that behavior through feedback. This renders empathy a scientifically measurable, assessable, and improvable construct, deepening theoretical insight into how emotional involvement can be modeled, operationalized, and studied within AI systems.

B. Design and Practical Implications

In addition to its theoretical contributions, MOON Framework offers significant implications for design practice, particularly in domains where emotional trust, familiarity, and long-term engagement are essential. In assistive and accessibility-oriented technologies, MOON provides a structured method for embedding emotional inclusiveness into interactions with older adults. Rather than treating empathy as an optional feature added to the technical pipeline, the framework offers designers a conceptual blueprint for integrating empathy directly into the system's core architecture.

At the design level, MOON can be implemented as four modular components within a VUI system. First, the Model phase can be realized through user-profiling algorithms that incorporate demographic, linguistic, and affective characteristics. Second, the Observation phase can employ real-time speech analytics to detect affective cues based on prosodic features such as pitch, rhythm, and vocal energy. Third, the Optimization phase enables adaptive speech synthesis through dynamic adjustments to tone, cadence, and phrasing. Finally, the Nurture phase functions as a data-aggregation and learning module, transforming emotional feedback into parameter updates that refine the system over successive sessions. Together, these modules create a conversational ecosystem that evolves with the user, fostering familiarity, confidence, and emotional comfort.

Importantly, the emotional history accumulated through the Nurture phase becomes part of the system's adaptive continuity, allowing the VUI to interact with the user in a manner reminiscent of a familiar partner, one that "remembers" past interactions and responds accordingly. In this sense, MOON offers not only a technical roadmap but also a design philosophy and ethical foundation for emotionally intelligent technology. It reframes empathy not as an illusion of humanness, but as a quantifiable and trainable system operation rooted in psychological and computational principles. By grounding emotional adaptivity in scientific mechanisms, MOON helps establish empathy as a legitimate component of system design rather than a metaphor or aesthetic enhancement. Ultimately, the framework signals a new generation of adaptive systems, machines capable of learning not only about users, but also from users, while responding in ways that respect their feelings and needs.

VI. CONCLUSION AND FUTURE WORK

This study has introduced MOON Framework as a holistic approach for designing emotionally adaptive VUIs that support older adults. By integrating principles from affective computing, Socioemotional Selectivity Theory, and adaptive user modeling, the framework redefines empathy not as an aspirational quality but as an organized and

quantifiable process. Through MOON, the field of human-computer interaction advances toward demonstrating that emotional intelligence can be computationally implemented in voice-based systems without requiring visual embodiment or physical expressiveness.

The framework offers a pathway for scalable, psychologically grounded, and emotionally inclusive design by emphasizing voice, the most natural and accessible communication modality. Its cyclical structure allows continuous refinement of empathic responses through user feedback, enabling affective cues to directly inform behavioral adjustment. In this way, MOON positions emotional intelligence as a viable and practical design variable rather than an abstract ideal.

This study extends our earlier findings [8] by moving from descriptive evidence to a formalized computational model. Building on the observation that emotional familiarity strongly contributes to older adults' sense of usability and confidence, MOON provides a mechanism for operationalizing and sustaining this familiarity across interactions. Over time, this fosters a longitudinal user-system relationship in which emotion becomes a basis for ongoing learning, approachability, and trust.

Finally, the future of accessible AI, especially for older adults, will depend not only on accuracy or efficiency, but also on the capacity to respond empathetically and respectfully to users' emotional needs. Although the present work establishes the conceptual and methodological foundations of MOON, the framework requires empirical validation to assess its effectiveness in real-world interaction scenarios. This will be the focus of future work.

REFERENCES

- [1] Y. Song, Y. Yang, and P. Cheng, "The investigation of adoption of voice-user interface (vui) in smart home systems among chinese older adults," *Sensors*, vol. 22, no. 4, p. 1614, 2022.
- [2] M. Liu, C. Wang, and J. Hu, "Older adults' intention to use voice assistants: Usability and emotional needs," *Heliyon*, vol. 9, no. 11, 2023.
- [3] X. Hu, S. Desai, M. Lundy, and J. Chin, "Beyond functionality: Co-designing voice user interfaces for older adults' well-being," *arXiv e-prints*, pp. arXiv-2409, 2024.
- [4] B. Stigall, J. Waycott, S. Baker, and K. Caine, "Older adults' perception and use of voice user interfaces: a preliminary review of the computing literature," in *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 2019, pp. 423–427.
- [5] C. Xie, Y. Xie, Y. Wang, P. Zhou, L. Lu, Y. Feng, and C. Liang, "Understanding older adults' continued-use intention of ai voice assistants," *Universal Access in the Information Society*, vol. 24, no. 2, pp. 1687–1699, 2025.
- [6] T. Schroeder, L. Dodds, A. Georgiou, H. Gewald, J. Siette *et al.*, "Older adults and new technology: Mapping review of the factors associated with older adults' intention to adopt digital technologies," *JMIR aging*, vol. 6, no. 1, p. e44564, 2023.
- [7] C. Lee and J. F. Coughlin, "Perspective: Older adults' adoption of technology: An integrated approach to identifying determinants and barriers," *Journal of Product Innovation Management*, vol. 32, no. 5, pp. 747–759, 2015.
- [8] H. A. Sagga and R. Stone, "Older adults and technology design from the hci perspective," *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 7, 2025.
- [9] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for human-robot interaction: Recent advances and future perspectives," *Frontiers in Robotics and AI*, vol. 7, p. 532279, 2020.

- [10] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from eeg-based brain-computer interface using machine learning: a comprehensive review," *Neural Computing and Applications*, vol. 34, no. 15, pp. 12 527–12 557, 2022.
- [11] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, "Voice in human-agent interaction: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–43, 2021.
- [12] S. Chamishka, I. Madhavi, R. Nawaratne, D. Alahakoon, D. De Silva, N. Chilamkurti, and V. Nanayakkara, "A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling," *Multimedia Tools and Applications*, vol. 81, no. 24, pp. 35 173–35 194, 2022.
- [13] K. Seaborn, T. Sekiguchi, S. Tokunaga, N. P. Miyake, and M. Otake-Matsuura, "Voice over body? older adults' reactions to robot and voice assistant facilitators of group conversation," *International Journal of Social Robotics*, vol. 15, no. 2, pp. 143–163, 2023.
- [14] D. R. Faria, A. L. Godkin, and P. P. da Silva Ayrosa, "Advancing emotionally aware child-robot interaction with biophysical data and insight-driven affective computing," *Sensors*, vol. 25, no. 4, p. 1161, 2025.
- [15] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, vol. 83, pp. 19–52, 2022.
- [16] G. Udaheureka, K. Djouani, and A. M. Kurien, "Multimodal emotion recognition using visual, vocal and physiological signals: a review," *Applied Sciences*, vol. 14, no. 17, p. 8071, 2024.
- [17] H. Abdollahi, M. H. Mahoor, R. Zandie, J. Siewierski, and S. H. Qualls, "Artificial emotional intelligence in socially assistive robots for older adults: a pilot study," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 2020–2032, 2022.
- [18] A. Rusch, "Artificial emotional intelligence's potential in improving social wellness of older adults," in *Proceedings of the AAAI Symposium Series*, vol. 5, no. 1, 2025, pp. 293–295.
- [19] O. E. Lee, K. O. Nah, E. M. Kim, N. G. Choi, and D.-H. Park, "Exploring the use of socially assistive robots among socially isolated korean american older adults," *Journal of Applied Gerontology*, vol. 43, no. 9, pp. 1295–1304, 2024.
- [20] M. R. Lima, M. Wairagkar, M. Gupta, F. R. y Baena, P. Barnaghi, D. J. Sharp, and R. Vaidyanathan, "Conversational affective social robots for ageing and dementia support," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 4, pp. 1378–1397, 2021.
- [21] V. Paul, "Technical review: Emotional intelligence in voice ai - enhancing human-like interactions," *European Modern Studies Journal*, 2025.
- [22] T. Anvarjon, Mustaqeem, and S. Kwon, "Deep-net: A lightweight cnn-based speech emotion recognition system using deep frequency features," *Sensors*, vol. 20, no. 18, p. 5212, 2020.
- [23] S. Kim *et al.*, "Exploring how older adults use a smart speaker-based voice assistant in their first interactions: Qualitative study," *JMIR mHealth and uHealth*, vol. 9, no. 1, p. e20427, 2021.
- [24] A. Pradhan, A. Lazar, and L. Findlater, "Use of intelligent voice assistants by older adults with low technology use," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 27, no. 4, pp. 1–27, 2020.
- [25] X. Cao, H. Zhang, B. Zhou, D. Wang, C. Cui, and X. Bai, "Factors influencing older adults' acceptance of voice assistants," *Frontiers in Psychology*, vol. 15, p. 1376207, 2024.
- [26] R. Zhong and M. Ma, "Effects of communication style, anthropomorphic setting and individual differences on older adults using voice assistants in a health context," *BMC geriatrics*, vol. 22, no. 1, p. 751, 2022.
- [27] V. K. Jones, M. Hanus, C. Yan, M. Y. Shade, J. Blaskewicz Boron, and R. Maschieri Bicudo, "Reducing loneliness among aging adults: the roles of personal voice assistants and anthropomorphic interactions," *Frontiers in public health*, vol. 9, p. 750736, 2021.
- [28] R. A. Marziali, C. Franceschetti, A. Dinculescu, A. Nistorescu, D. M. Kristály, A. A. Moşoi, R. Broekx, M. Marin, C. Vizitiu, S.-A. Moraru *et al.*, "Reducing loneliness and social isolation of older adults through voice assistants: literature review and bibliometric analysis," *Journal of medical Internet research*, vol. 26, p. e50534, 2024.
- [29] R. V. Aranha, C. G. Corrêa, and F. L. Nunes, "Adapting software with affective computing: a systematic review," *IEEE Transactions on Affective Computing*, vol. 12, no. 4, pp. 883–899, 2019.
- [30] E. Cambria, S. Poria, A. Hussain, and B. Liu, "Computational intelligence for affective computing and sentiment analysis [guest editorial]," *IEEE Computational Intelligence Magazine*, vol. 14, no. 2, pp. 16–17, 2019.
- [31] J. Han, Z. Zhang, M. Pantic, and B. Schuller, "Internet of emotional people: Towards continual affective computing cross cultures via audiovisual signals," *Future Generation Computer Systems*, vol. 114, pp. 294–306, 2021.
- [32] R. Zall and M. R. Kangavari, "Comparative analytical survey on cognitive agents with emotional intelligence," *Cognitive Computation*, vol. 14, no. 4, pp. 1223–1246, 2022.
- [33] L. L. Carstensen, "The influence of a sense of time on human development," *Science*, vol. 312, no. 5782, pp. 1913–1915, 2006.
- [34] R. W. Picard, "Affective computing," 1997.
- [35] M. Khan, A. El Saddik, F. S. Alotaibi, and N. T. Pham, "Aad-net: Advanced end-to-end signal processing system for human emotion detection & recognition using attention-based deep echo state network," *Knowledge-Based Systems*, vol. 270, p. 110525, 2023.
- [36] P. S. Tomar, K. Mathur, and U. Suman, "Unimodal approaches for emotion recognition: A systematic review," *Cognitive Systems Research*, vol. 77, pp. 94–109, 2023.
- [37] L. L. Carstensen and H. E. Hershfield, "Beyond stereotypes: Using socioemotional selectivity theory to improve messaging to older adults," *Current Directions in Psychological Science*, vol. 30, no. 4, pp. 327–334, 2021.
- [38] S. T. Charles and E. J. Urban, "Socioemotional selectivity theory/future time perspective," *The encyclopedia of adulthood and aging*, pp. 1–5, 2015.
- [39] P. Joshi, A. Kononova, and S. Cotten, "Understanding older adults' preferences for and motivations to use traditional and new ict in light of socioemotional selectivity and selection, optimization, and compensation theories," *International Journal of Communication*, vol. 14, p. 20, 2020.
- [40] S. H. Bardach, E. K. Rhodus, K. Parsons, and A. K. Gibson, "Older adults' adaptations to the call for social distancing and use of technology: insights from socioemotional selectivity theory and lived experiences," *Journal of Applied Gerontology*, vol. 40, no. 8, pp. 814–817, 2021.
- [41] K. M. Livingstone and D. M. Isaacowitz, "Situation selection and modification for emotion regulation in younger and older adults," *Social psychological and personality science*, vol. 6, no. 8, pp. 904–910, 2015.
- [42] M. Wirth, A. Voss, and K. Rothermund, "Age differences in everyday emotional experience: Testing core predictions of socioemotional selectivity theory with the miva model," *The Journals of Gerontology: Series B*, vol. 78, no. 7, pp. 1152–1162, 2023.
- [43] Y. Ma, Y. Zhang, D. Fu, S. Z. Portales, D. Kragic, and M. Fjeld, "Advancing user-voice interaction: Exploring emotion-aware voice assistants through a role-swapping approach," 2025. [Online]. Available: <https://arxiv.org/abs/2502.15367>
- [44] Z. Yan, V. Dube, J. Heselton, K. Johnson, C. Yan, V. Jones, J. Blaskewicz Boron, and M. Shade, "Understanding older people's voice interactions with smart voice assistants: a new modified rule-based natural language processing model with human input," *Frontiers in Digital Health*, vol. 6, p. 1329910, 2024.
- [45] S. ZEWAR, S. Zhiyong, and A. Adnan, "Enhancements in immediate speech emotion detection: Harnessing prosodic and spectral characteristics," *International Journal of Innovative Science and Research Technology (IJISRT) IJISRT24APR872*, pp. 1526–1534, 2024.
- [46] P. A. Tun, A. Wingfield, E. A. Stine, and C. Meccas, "Rapid speech processing and divided attention: processing rate versus processing resources as an explanation of age effects," *Psychology and aging*, vol. 7, no. 4, p. 546, 1992.
- [47] B. Sutton, J. King, K. Hux, and D. Beukelman, "Younger and older adults' rate performance when listening to synthetic speech," *Augmentative and Alternative Communication*, vol. 11, no. 3, pp. 147–153, 1995.

- [48] C. Fougeron, F. Guitard-Ivent, and V. Delvaux, "Multi-dimensional variation in adult speech as a function of age," *Languages*, vol. 6, no. 4, p. 176, 2021.
- [49] Y. Kim, S.-m. Lee, M.-k. Choi, S.-m. Jung, J. E. Sung, Y. Lee *et al.*, "The effects of speakers' age on temporal features of speech among healthy young, middle-aged, and older adults," *Phonetics and Speech Sciences*, vol. 14, no. 1, pp. 37–47, 2022.
- [50] B. Irfan, A. Narayanan, and J. Kennedy, "Dynamic emotional language adaptation in multiparty interactions with agents," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 2020, pp. 1–8.
- [51] R. Sanjeewa, R. Iyer, P. Apputhurai, N. Wickramasinghe, and D. Meyer, "Empathic conversational agent platform designs and their evaluation in the context of mental health: systematic review," *JMIR Mental Health*, vol. 11, p. e58974, 2024.
- [52] K. M. Kokorelias, A. Grigorovich, M. T. Harris, U. Rehman, L. Ritchie, A. M. Levy, K. Denecke, and J. McMurray, "Longitudinal coadaptation of older adults with wearables and voice-activated virtual assistants: Scoping review," *Journal of Medical Internet Research*, vol. 26, p. e57258, 2024.