Krumscheid et al, Gerontechnology (2026) 25:3
https://doi.org/10.4017/gt.2026.25.3.1222.5

A multimodal chatbot-based system for monitoring and enhancing well-being in older adults: A proof-of-concept design

Mario Krumscheid[a], Aktas Mahmut[a], Jonas Besler[a], Laurin Goell[a], Madline Lutz[a], Ricarda Neumann[a], Valdrit Preteni[a], Pierre Eric Tah Hine[a], Katja Bochtler PhD[a]
[a]Faculty of Computer Science, Kempten University of Applied Sciences, Kempten, Germany

**Corresponding author: katja.bochtler@hs-kempten.de*

Background: Global demographic trends toward aging populations highlight growing mental health challenges among older adults. Traditional methods for assessing psychological well-being are often infrequent, inaccessible, and subject to self-report biases, underscoring the need for innovative monitoring solutions.
Research aim: The paper presents a proof-of-concept design for a multimodal chatbot-based system intended to support continuous monitoring of psychological well-being in older adults. It explores how natural language processing (NLP) and speech analysis technologies can be integrated to identify emotional states and enable proactive support.
Methods: The proposed system integrates daily conversational interactions via an empathetic chatbot, employing automated textual sentiment analysis through a fine-tuned Large Language Model (LLM) and prosodic speech analysis to assess emotional states. These analyses are complemented by regular subjective self-assessments to create comprehensive and personalized well-being profiles.
Results: The developed prototype shows the conceptual feasibility of unobtrusive monitoring and early detection of mental distress. The proposed design further includes mechanisms for personalized recommendations and triggering interventions when significant declines in well-being are detected.
Conclusions: The current proof-of-concept requires future empirical validation of its usability, reliability, and accuracy. Further research should also address technical refinements, ethical considerations, and broader cultural adaptability.

Keywords: elderly well-being, conversational agent, natural language processing, mental health monitoring

Introduction

The global demographic shift toward an aging population presents significant challenges, particularly concerning the psychological well-being and mental health of older adults. As people age, they often encounter increased risks of social isolation, loneliness, and depression, all of which can significantly undermine their quality of life and lead to poorer physical health [1]. Recent research has underscored loneliness and diminished subjective well-being as critical public health issues, strongly linked to psychological distress and even accelerated physical health decline among older adults [2].

Assessment of well-being in the elderly

Psychological well-being is a multifaceted construct encompassing emotional, mental, and social health. It is often defined in broad terms, for example, the World Health Organization (WHO) defines quality of life as "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns," a concept influenced by one's physical health, psychological state, level of independence, social relationships, and environment [3]. Similarly, well-being has been described as a state "in which the individual is able to develop in their potential, work productively and creatively, build strong and positive relationships with others, and contribute to their community" [4]. These definitions highlight that well-being is multidimensional, touching on life satisfaction, interpersonal relations, and broader quality-of-life considerations. In discussing well-being, it is useful to distinguish subjective well-being, an individual's self-appraised satisfaction and feelings, from objective well-being, which refers to external life circumstances (e.g. income, health status) that may influence one's quality of life. Modern approaches emphasize that subjective self-report is essential, as objective indicators alone are not able to capture the subjective interpretation of an individual's situation. Accordingly, reliable measurement of psychological well-being has become of high importance in health economics and social science research.

To quantify well-being, researchers have developed numerous standardized questionnaires, each focusing on different dimensions of quality of life and mental health (see [5]). Widely established examples include the WHOQOL-BREF (assesses perceived quality of life in physical, psychological, social, and environmental domains), BBC Subjective Well-Being Scale (captures a broad spectrum of well-being, integrating physical health, psychological, and relationship aspects), Ryff's Psychological Well-Being Scales (evaluates positive psychological functioning across dimensions such as autonomy, personal growth, and purpose in life), Warwick-Edinburgh Mental Well-Being Scale (measures positive mental health attributes like optimism, usefulness, and relaxation) and the Life Satisfaction and Happiness Scales (short measures assessing cognitive judgments of overall life satisfaction and subjective happiness).

These instruments differ in their specific emphasis, ranging from physical health to emotional or social well-being, but all provide validated quantitative assessments of subjective well-being.

NLP and language features for psychological state detection

Language is a rich source of information about a person's psychological state. Natural language processing (NLP) techniques enable the systematic analysis of textual data (from spoken or written language) to detect subtle linguistic patterns associated with mental health, mood, or well-being [6]. By examining what people say (word content, semantics) and how they express themselves (grammar, complexity, sentiment), NLP can uncover markers of emotional state, depression, anxiety, and related factors that might not be obvious in casual observation.

A substantial body of research has identified linguistic markers that correlate with depression and other mood disturbances. These include vocabulary choices, grammatical patterns, and even the focus of conversation. For example, studies have found that individuals experiencing depression tend to use disproportionately high rates of first-person singular pronouns (e.g., "I", "me") and negative emotion words (e.g., "sad", "hate"), alongside a higher frequency of absolutist terms like "always", "nothing" or "completely". Such absolutist words appear especially indicative, as they track the severity of the affective disorder better than general negative emotion words [7]. These linguistic signals align with known cognitive patterns in depression, such as self-focused attention and black-and-white thinking. By contrast, positive emotion words and references to others might be less prevalent in depressed individuals' language. Table-based content analyses and lexicon counts have repeatedly confirmed these trends across diverse samples and contexts. Modern NLP approaches leverage such markers in combination with machine learning to classify or predict mental states. Traditional techniques often rely on tools like Linguistic Inquiry and Word Count (LIWC), a program that quantifies text in psychologically relevant categories (pronouns, affect words, social words, etc.), to derive feature sets for prediction. Simple logistic regression models based on LIWC features have achieved above-chance detection of conditions like depression and anxiety from writing or transcripts [8], though performance can be modest. More recently, deep learning methods have been applied to textual data for mental health screening. Transformer-based language models (which consider word context and sequence) can detect nuanced language patterns that simpler bag-of-words approaches might miss. In one example, a fine-tuned Transformer model improved generalized anxiety disorder detection from speech transcripts (AUROC ≈0.64) compared to a LIWC-based model (AUROC ≈0.58).

Beyond individual studies, there is growing evidence that NLP on large-scale natural language data (e.g. social media posts, online journals, therapy transcripts) can effectively flag mental health issues. Researchers have demonstrated that text-derived features from social media and transcribed speech can be used to detect depression and even stress levels. For example, aggregating a person's Twitter or forum posts and analyzing linguistic tone, topics, and sentiment has enabled early detection of depression in otherwise undiagnosed individuals, often aligning with clinical PHQ-9 depression scores reported by those users [9]. The advantages of NLP-based analysis are that it can be done passively, at scale, and objectively. It does not rely on self-disclosure in a clinical setting; instead, it can draw insights from natural communication that a person is already producing. However, it is worth noting that language-based detection works best when algorithms are carefully tuned to context (to avoid misinterpreting slang, cultural expressions, or topic effects) and when used in conjunction with human judgment.

Speech prosody and acoustic markers of psychological state

In addition to linguistic content, the acoustic qualities of speech, intonation, rhythm, volume, and other prosodic features, carry significant information about a speaker's emotional and mental state. Prosody (tone of voice) often reveals feelings that words alone might con-

ceal, for example a trembling, high-pitched voice may signal fear or anxiety, whereas a flat, monotone voice can suggest sadness or depression. Researchers in speech analysis and affective computing have long capitalized on this, developing algorithms to recognize emotions from audio recordings by extracting vocal features like fundamental frequency (pitch), intensity (loudness), speaking rate, and voice quality measures [10]. These features collectively form a kind of "paralinguistic signature" of different emotions. High arousal emotions such as anger or panic typically produce more extreme prosodic modulation (e.g. raised pitch, increased volume, faster tempo), while low arousal states like sadness yield slower speech with lower overall pitch and energy.

Applications in older adult populations

Applying NLP and speech-based detection techniques to elderly populations is an area of growing interest. Older adults can particularly benefit from unobtrusive monitoring of mood and well-being, given that they may underreport symptoms or face barriers in access to mental health services [6]. However, there are also unique considerations when working with older voices and language. Late-life depression (LLD) often presents with more somatic and cognitive symptoms than early-life depression, which can complicate diagnosis using standard questionnaires. Voice analysis offers a more objective way to detect depressive signs without solely relying on self-report. At the same time, aging itself brings changes in speech. For example, older adults tend to have a naturally lower harmonics-to-noise ratio (a slightly rougher voice) and may speak more slowly or softly due to health conditions or medications. These age-related factors mean that algorithms must be attuned to distinguishing normal aging effects from genuine psychological distress.

Despite these challenges, recent studies demonstrate the feasibility and utility of speech-based well-being assessment in older populations. An illustrative example is the work by Finze et al., who developed a machine learning model to assess subjective well-being in seniors using natural speech analysis [11]. In their approach, participants' voices were recorded during casual speech tasks, the audio was analyzed purely on acoustic and prosodic characteristics (no speech-to-text transcription), and the model was calibrated against each individual's WHOQOL quality of life questionnaire scores. Remarkably, the system could estimate an older adult's well-being score with a low error margin, essentially predicting WHOQOL-based well-being levels directly from voice patterns. The authors reported that a support vector regression model achieved a mean error around 10.9 (on a 0-100 well-being scale), and was sensitive enough to detect changes in a person's well-being over time. This study highlights several important points. First, it is possible to gauge holistic quality of life from vocal features alone, likely because emotions, energy, and engagement (factors related to well-being) subtly colour one's speech. Second, it holds promise for healthy aging initiatives. Caregivers or digital health tools could monitor elders' spoken interactions (e.g. daily phone calls or voice notes) to detect declines in well-being and intervene earlier.

Gap analysis and positioning of the present work

Although significant progress has been made in NLP-based mental health detection, speech emotion recognition, and digital mental health chatbots, several gaps remain in the current research and application landscape.

A number of conversational agents and chatbots have been developed that provide mental health-related support or monitoring. For example, Wysa is a widely used digital mental health platform that provides an emotionally intelligent conversational agent grounded in evidence-based techniques such as cognitive behavioral therapy (CBT), dialectical behavior therapy (DBT), and meditative exercises to support self-management of anxiety, depression, and other conditions across a global user base. The platform also integrates human coaching and clinical support pathways, aiming to increase accessibility to mental health tools worldwide. Woebot is another established mental health chatbot designed to deliver structured therapeutic content using CBT-inspired conversational flow and mood tracking; empirical research demonstrates that similar CBT-based chatbots can reduce symptoms of anxiety and depression in controlled settings [12] and traditional trials of CBT-based chatbots report positive outcomes in student populations [13]. Tess is a psychological chatbot that provides structured messages and coping strategies via text to support emotional expression and stress regulation, and has been studied for its impact on mood and anxiety and reviewed in narrative analyses of CBT chatbots [14].

Although these systems represent valuable steps towards scalable mental health support, they also illustrate boundaries in current approaches. Many of these chatbots offer predominantly text-based interaction without systematically integrating acoustic or prosodic markers from speech. They also often emphasize therapeutic content

delivery (e.g., CBT exercises, mood tracking) rather than continuous, multimodal emotional and well-being monitoring. Furthermore, while systematic reviews find that AI-based conversational agents can reduce depression and distress symptoms and improve aspects of well-being, overall evidence remains mixed and highlights a need for better-quality designs and long-term studies, especially in older populations [15] and specifically for alleviating loneliness or depressive symptoms among community-dwelling older adults [16].

Despite this growing body of work, integrated multimodal systems that combine continuous NLP-based semantic interpretation, speech acoustic analysis, structured self-report measures, and transparent scoring logic tailored to older adult populations remain comparatively sparse. Many existing chatbots rely on text-based approaches with static decision trees or sentiment analysis and do not incorporate speech-based emotional cues or a composite well-being index. Likewise, personalized trend detection and rule-based intervention thresholds are seldom articulated in existing digital support tools, limiting their capacity for ongoing, adaptive support.

The proof-of-concept system proposed in this paper addresses these gaps by presenting a modular multimodal architecture that: (1) embeds conversational AI with NLP-based semantic analysis and speech emotion recognition within a unified interaction flow; (2) combines passive indicators (text and speech features) with structured subjective self-assessment instruments; (3) aggregates multimodal signals into an explicit scoring and decision logic; and (4) is specifically oriented towards supporting older adults living alone. The proposed system aims to enable continuous, accessible, and multimodal well-being monitoring rather than episodic or purely text-based intervention.

Methods

Our proof-of-concept pipeline comprises several modular components designed to continuously and unobtrusively monitor and enhance the subjective well-being of elderly users. Central to this architecture is a conversational AI based on the LLaMA 3.2 (8B) model, which engages users in daily natural language interactions.

Textual data from these interactions are analyzed using the MentaLLaMA model, a fine-tuned version of Meta's LLaMA2 trained specifically for interpretable mental health analysis on social media data. MentaLLaMA was trained on the IMHI dataset containing approximately 105,000 samples across eight mental health tasks. Evaluations indicate that MentaLLaMA performs at or near state-of-the-art accuracy for mental health classification and generates explanations approaching human interpretability, especially when fine-tuned on domain-specific data, thus making it ideal for nuanced psychological assessments of user inputs [17].

For speech-based emotional analysis, we utilize the wav2vec2-IEMOCAP emotion recognition model, fine-tuned on the Interactive Emotional Dyadic Motion Capture (IEMOCAP) dataset, consisting of around 12 hours of emotionally diverse speech from multiple speakers. The wav2vec 2.0 architecture leverages self-supervised learning to extract robust acoustic embeddings, achieving high accuracy on emotion recognition tasks. The model is validated using standard emotion recognition benchmarks such as RAVDESS and IEMOCAP, where it demonstrated superior classification performance compared to traditional and other contemporary models, particularly by exploiting multiple acoustic layers for richer prosodic analysis [18].

To provide a quantitative measure of self-assessed subjective well-being, we implemented the BBC Subjective Well-Being (BBC-SWB) questionnaire into the system prompt of the chatbot, which comprises 24 self-report items distributed across three validated dimensions: psychological well-being, physical health, and relationships. Each question uses a 4-point scale ranging from "not at all" to "extremely". The BBC-SWB scale has shown strong internal consistency (Cronbach's alpha = .944) and good psychometric properties across various populations, demonstrating robust validity and reliability in assessing broad domains of subjective well-being [19].

This combination of conversational interaction, speech and textual emotion recognition, and validated subjective well-being measurement allows comprehensive, nuanced, and reliable tracking of users' emotional states and overall psychological well-being. Moreover, audio-based sensing was specifically selected due to its inherent strengths in privacy, comfort, and user acceptance, enabling detailed analysis of subtle emotional indicators (tone, tempo, pitch) without the intrusive nature of visual or spatial technologies, thereby enhancing long-term user engagement and trust [20].

The current proof-of-concept was implemented in a fully local, on-premise architecture. The large language model inference is executed via a locally hosted Ollama server. Ollama provides

an OpenAI API–compatible interface, allowing seamless substitution of the inference endpoint should cloud-based deployment be required in future implementations. Consequently, migration to a cloud infrastructure would primarily involve modifying the API endpoint configuration without structural changes to the system architecture. Speech-to-text and text-to-speech processing are currently performed client-side using standard browser-based plug-ins, ensuring that raw audio data does not need to be transmitted to external servers for pre-processing. For better speech quality, this data could also be routed to external cloud services. The chosen architecture therefore supports multiple deployment scenarios: (1) fully local on-premise and client-side operation for high-privacy contexts (e.g., assisted living facilities or clinical environments), and (2) scalable cloud-based deployment for broader distribution, should appropriate data protection safeguards (discussed below) be implemented.

The current implementation represents a proof-of-concept system focused on architectural integration and design feasibility. No user-based evaluation has been conducted within the scope of this work.

Results

This section describes the implemented prototype and its functional components. The proof-of-concept multimodal chatbot system integrates several interconnected modules designed to support the continuous monitoring of older adults' subjective well-being. The core components include: a daily check-in chatbot interface, a prosodic speech emotion recognition module using the speechbrain/emotion-recognition-wav2vec2-IEMOCAP model, an NLP-driven textual analysis module utilizing the klyang/MentaLLaMA-chat-7B model, and a subjective self-assessment component based on a simplified Smiley scale. These modules collectively feed data into a centralized database, from which a weighted average daily well-being score is calculated. The system includes clearly defined thresholds to trigger personalized recommendations or external interventions. *Figure 1* shows a system diagram of our proposed prototype.

Detailed pipeline explanation

The pipeline operates through the following detailed steps:

Daily check-in
Once per day, a chatbot initiates a conversation with the user, prompting them with one randomly selected question from the BBC Subjective Well-Being (SWB) questionnaire. This question is reformulated into a more conversational and empathetic format to foster natural interaction. Users also have the option to proactively initiate conversations by tapping an icon in the chat interface, in which case the chatbot engages in a friendly, empathetic, and supportive dialogue without a predetermined BBC-SWB question.

Speech analysis for emotion detection
To assess the emotional tone of user speech, we employ the aforementioned speechbrain/emotion-recognition-wav2vec2-IEMOCAP model. In our application, each detected emotion is mapped onto a numeric score ranging from 1 (lowest well-being) to 4 (highest well-being). Specifically, positive emotions (excited, happy) are assigned scores of 4, neutral or mildly positive emotions (surprised, neutral) are assigned 3, negative emotions reflecting mild dissatisfaction (frustrated, disappointed) are given a score of 2, and strong negative emotions (angry, sad, fear) are rated as 1.

NLP-Based text analysis
For textual analysis of user interactions, we utilize the klyang/MentaLLaMA-chat-7B model, a large language model based on Meta's LLaMA2-chat-7B foundation. The MentaLLaMA-chat-7B model is designed to analyse mental health conditions and provide reliable explanations for its predictions. It approaches state-of-the-art discriminative methods in correctness and generates high-quality explanations, making it suitable for interpretable mental health analysis in a non-clinical research context.

In our system, the model provides both a qualitative analysis of the user's well-being, including reasoning and context, and a quantitative well-being score from 1 (very poor well-being) to 4 (excellent well-being). The detailed analysis and reasoning from the NLP model are stored for later reference and use in generating personalized recommendations.

Subjective self-assessment
Several times per week (but not necessarily every day to avoid user fatigue), the user is prompted to complete a simple self-assessment using a smiley-based scale. This scale follows the same 1-to-4 point system used in speech and NLP analyses, allowing for straightforward comparison and aggregation into a unified daily well-being score for downstream processing.

Score aggregation
All conversation transcripts, NLP analyses, reasoning, and individual scores from speech and text analysis, as well as subjective self-assessments, are stored in a central database. A Score Engine calculates a weighted daily av-

erage score for overall well-being, combining NLP analysis (40%), subjective self-assessment (40%), and emotional speech analysis (20%).

The weighting scheme was intentionally defined as an initial heuristic configuration, informed by both theoretical and practical considerations within a proof-of-concept context, rather than as an empirically optimized solution. From a theoretical perspective, subjective well-being is fundamentally a self-evaluative construct; therefore, structured self-report was weighted strongly (40%) because it directly captures the individual's own appraisal of psychological state. NLP-based semantic analysis was also assigned substantial weight (40%) because it captures contextualized cognitive and affective patterns expressed in natural language, including self-referential focus, emotional tone, and absolutist thinking.

In contrast, speech emotion recognition was weighted lower (20%) for practical and methodological reasons. First, acoustic emotion classification models are typically trained on acted or laboratory speech datasets and may exhibit reduced robustness in naturalistic or age-diverse conversational settings. Second, the mapping of discrete emotion categories (e.g., "happy," "sad," "angry") onto a unidimensional well-being scale introduces an additional abstraction layer, potentially increasing uncertainty. Third, speech prosody may be influenced by age-related physiological changes, health conditions, or medication effects, which are not necessarily indicative of psychological distress. The reduced weighting therefore reflects a conservative design decision intended to minimize the impact of potential acoustic misclassification while still incorporating complementary affective signals.

Importantly, this weighting configuration is not assumed to be optimal. In future validation studies, weights should be calibrated using empirical data, for example through regression-based fusion models, Bayesian weighting approaches, or machine learning ensemble techniques that learn modality importance based on predictive performance. Longitudinal datasets would additionally allow optimization using outcome-based criteria such as sensitivity to clinically relevant changes in well-being. The current heuristic approach should therefore be interpreted as a proof-of-concept initialization rather than a finalized parameterization.

Furthermore, alternative aggregation strategies are conceivable. Rather than computing a single composite score, modalities could also be monitored independently, with interventions triggered when significant negative trends occur in any one domain. Comparative evaluation of composite versus modality-specific trend detection constitutes an important direction for future research.

Recommendation and intervention

A Recommendation Engine based on the Llama 3.2 instruct model is activated when the user's aggregated seven-day average well-being score falls below 2.5 and the two most recent daily scores are at or below 2. The Recommendation Engine generates personalized advice based on the user's recent conversations, NLP analyses, hobbies, activities, and (where available) calendar entries. This design aims to provide low-threshold, context-aware support (e.g., suggesting previously enjoyed activities or social contact) while remaining non-clinical in nature. These thresholds were defined heuristically based on the four-point structure of the BBC Subjective Well-Being scale and the absence of clinically established cut-off values for continuous conversational monitoring contexts. A score below 2.5 was operationally interpreted as indicating a shift from predominantly positive towards neutral or negative well-being states, while consecutive scores of 2 or lower were treated as a potential downward trend. The additional temporal condition (two consecutive low scores) was included to reduce the likelihood of false alerts triggered by short-term fluctuations. If the average seven-day score falls below 1.5 and the two most recent daily scores remain at or below 2, an Intervention Engine is activated. In this case, the system notifies designated contacts (e.g., caregivers or trusted family members). To support transparency and reduce the risk of misinterpretation, the notification includes a brief rationale derived from recent NLP-based well-being analyses (e.g., recurring negative themes or sustained negative sentiment), rather than only providing a numeric score. However, these cut-offs should not be interpreted as clinically validated diagnostic thresholds; they function as conservative early-warning markers within an exploratory monitoring framework.

Future work should evaluate threshold performance using empirical outcome data. Statistical approaches such as ROC-based optimization, change-point detection, and individualized baseline deviation modeling may enable data-driven calibration of sensitivity and specificity. In particular, personalized adaptive thresholds that account for individual baseline variability are likely to be more appropriate than fixed global cut-offs in older adult populations.

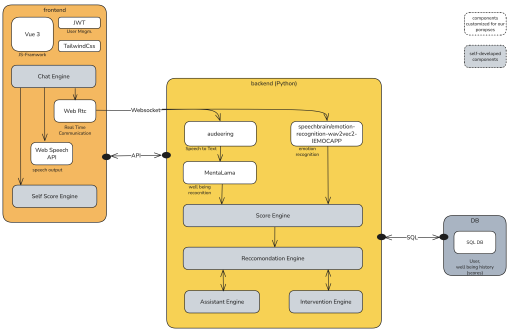

Figure 1. System diagram showing the frontend and backend components of the multimodal chatbot WellBee

Discussion

Compared to similar multimodal systems in recent literature (e.g., Finze et al. [11]; Chou et al. [21]), our approach offers a unique integration of multiple analytical methods, speech-based emotion recognition, NLP-based text analysis, and subjective self-assessment into a single coherent pipeline. Unlike purely NLP or speech-based systems, this comprehensive approach provides richer data for evaluating users' well-being and could support more sensitive detection of well-being-related changes.

The system is designed to provide continuous, proactive monitoring of user well-being through empathetic interactions, minimizing user burden while enhancing the likelihood of early identification of well-being issues. Its modular architecture ensures scalability, adaptability, and ease of further development, enabling rapid integration of newer technologies or alternative analytic methods.

However, the current setup has several limitations inherent to its proof-of-concept stage. Primarily, the arbitrary numeric mapping of emotional labels to well-being scores and the heuristic weighting of individual modalities may introduce bias and reduce the precision of emotional analysis. Although theoretically motivated, these parameters have not yet been empirically optimized and may not reflect the true predictive contribution of each modality.

Furthermore, the system relies on pretrained language and speech models that were originally developed on general population datasets rather than specifically on older adult samples. Age-related differences in language use, communication style, and speech acoustics may therefore affect model accuracy. For example, older adults may exhibit slower speaking rates, altered pitch ranges, or different lexical preferences that are not necessarily indicative of psychological distress but could influence model predictions. Without domain-specific fine-tuning or calibration, there is a risk of systematic bias, reduced sensitivity, or false positive classifications in this population.

The absence of user-specific personalization constitutes an additional limitation. Current scoring and threshold mechanisms rely on global parameters rather than individualized baseline modeling. Given that subjective well-being varies substantially across individuals, fixed thresholds may not adequately capture meaningful within-person changes. Personalized baseline adaptation and longitudinal trend modeling would likely provide a more accurate and clinically meaningful monitoring approach.

Ethical considerations also require careful attention. Transparent informed consent procedures, clear communication regarding the non-diagnostic nature of the system, and the ability for users to withdraw participation at any time are essential for responsible deployment. In addition, automated alert mechanisms carry the risk of false alarms or missed detections. False positives may lead to unnecessary concern among caregivers, while false negatives could delay needed support. Therefore, intervention notifications should be interpreted as supportive indicators rather than clinical determinations and ideally be embedded within a human-in-the-loop decision framework. Practically, this could involve routing intervention notifications to a designated caregiver or clinician who reviews recent conversational summaries and trend data before deciding whether outreach or additional assessment is warranted, or what concrete interventions will be forwarded.

Future research should therefore include structured empirical validation involving older adults to assess usability, model accuracy, threshold performance, and user acceptance under real-life conditions. Refinement of emotion detection through age-specific training datasets, domain-adaptive fine-tuning, and potentially federated or on-device personalization approaches may enhance robustness while preserving privacy. Personalized baseline calibration and adaptive thresholds should be explored to better capture intra-individual changes over time.

Conclusion

This paper presented a multimodal chatbot-based system designed as a proof-of-concept for monitoring and enhancing subjective well-being among older adults living alone. The developed pipeline demonstrated the feasibility of integrating empathetic conversational agents,

speech and NLP analyses, and automated recommendation mechanisms. Moving forward, empirical validation studies, enhanced accuracy in emotional analytics, and careful attention to ethical considerations, including data privacy and user consent, are essential steps toward developing a reliable and socially impactful technology to support older adults' mental health and quality of life.

Acknowledgments
This project was supported by Bavarian Center for Digital Health and Social Care, Kempten and AAL Living Lab, Kempten. We appreciate their valuable support throughout this project.

References

[1] National Institute on Aging (2019, April 23). Social isolation, loneliness in older people pose health risks. U.S. Department of Health and Human Services. https://www.nia.nih.gov/news/social-isolation-loneliness-older-people-pose-health-risks

[2] World Health Organization (2021, July 29). Social isolation and loneliness among older people: Advocacy brief. https://iris.who.int/bitstream/handle/10665/343206/9789240030749 eng.pdf?sequence=1

[3] World Health Organization (2012). Measurement of and target-setting for well-being: An initiative by the WHO Regional Office for Europe. https://iris.who.int/bitstream/handle/10665/77932/WHO_HIS_HSI_Rev.2012.03_eng.pdf?sequence=1

[4] Beddington, J., Cooper, C. L., Field, J., Goswami, U., Huppert, F. A., Jenkins, R., Jones, H. S., Kirkwood, T. B. L., Sahakian, B. J., & Thomas, S. M. (2008). The mental wealth of nations. Nature, 455(7216), 1057–1060. https://doi.org/10.1038/4551057a

[5] Zhang, W., Balloo, K., Hosein, A., & Medland, E. (2024). A scoping review of well-being measures: Conceptualisation and scales for overall well-being. BMC Psychology, 12(1), 585. https://doi.org/10.1186/s40359-024-02074-0

[6] DeSouza, D. D., Robin, J., Gumus, M., & Yeung, A. (2021). Natural language processing as an emerging tool to detect late life depression. Frontiers in Psychiatry, 12, Article 719125. https://doi.org/10.3389/fpsyt.2021.719125

[7] Yahya, N. H., & Abdul Rahim, H. (2023). Linguistic markers of depression: Insights from English-language tweets before and during the COVID-19 pandemic. Language and Health, 1(2), 36–50. https://doi.org/10.1016/j.laheal.2023.10.001

[8] Teferra, B. G., & Rose, J. (2023). Predicting generalized anxiety disorder from impromptu speech transcripts using context-aware transformer-based neural networks: Model evaluation study. JMIR Mental Health, 10, e44325. https://doi.org/10.2196/44325

[9] Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual based featuring approach for depression detection using machine learning classifiers and social media texts. Computers in Biology and Medicine, 135, Article 104499. https://doi.org/10.1016/j.compbiomed.2021.104499

[10] George, S. M., & Ilyas, P. M. (2024). A review on speech emotion recognition: A survey, recent advances, challenges, and the influence of noise. Neurocomputing, 568, Article 127015. https://doi.org/10.1016/j.neucom.2023.127015

[11] Finze, N., Jechle, D., Faußer, S., & Gewald, H. (2024). How are we doing today? Using natural speech analysis to assess older adults' subjective well being. Business & Information Systems Engineering, 66(3), 321–334. https://doi.org/10.1007/s12599-024-00877-4

[12] Xu, Z., Lee, Y., Stasiak, K., Warren, J., & Lottridge, D. (2025). The digital therapeutic alliance with mental health chatbots: Diary study and thematic analysis. JMIR Mental Health, 12, e76642. https://doi.org/10.2196/76642

[13] Klos, M. C., Escoredo, M., Joerin, A., Lemos, V. N., Rauws, M., & Bunge, E. L. (2021). Artificial intelligence–based chatbot for anxiety and depression in university students: Pilot randomized controlled trial. JMIR Formative Research, 5(8), e20678. https://doi.org/10.2196/20678

[14] Im, C., & Woo, M. (2025). Clinical efficacy, therapeutic mechanisms, and implementation features of cognitive behavioral therapy–based chatbots for depression and anxiety: Narrative review. JMIR Mental Health, 12, e78340. https://doi.org/10.2196/78340

[15] Li, H., Zhang, R., & Lee, Y. C., et al. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. npj Digital Medicine, 6, 236. https://doi.org/10.1038/s41746-023-00979-5

[16] Satake, Y., Costello, H., Naran, N., Ishimaru, D., Ikeda, M., & Howard, R. (2026). Autonomous conversational agents for loneliness, social isolation, depression, and anxiety in older people without cognitive impairment: Systematic review and meta-analysis. Psychological Medicine, 56, e27. https://doi.org/10.1017/s0033291725103073

[17] Yang, K., Zhang, T., Kuang, Z., Xie, Q., Huang, J., & Ananiadou, S. (2024). MentaLLaMA: Interpretable Mental Health Analysis on Social Media with Large Language Models. In Proceedings of the ACM Web Conference 2024, 4489–4500. https://doi.org/10.1145/3589334.3648137

[18] Pepino, L., Riera, P., & Ferrer, L. (2021). Emotion Recognition from Speech Using Wav2vec 2.0 Embeddings. arXiv, abs/2104.03502. https://arxiv.org/abs/2104.03502

[19] Kinderman, P., Schwannauer, M., Pontin, E., & Tai, S. (2011). The development and validation of a general measure of well being: The BBC well being scale. Quality of Life Research, 20(7), 1035–1042. https://doi.org/10.1007/s11136-010-9841-z

[20] Liu, M., Wang, C., & Hu, J. (2023). Older adults' intention to use voice assistants: Usability and emotional needs. Heliyon, 9(11), Article e21932. https://doi.org/10.1016/j.heliyon.2023.e21932

[21] Chou, S.-H., Chandhok, S., Little, J. J., & Sigal, L. (2024). MM-R³: On (in-) consistency of multimodal large language models (MLLMs) (arXiv preprint arXiv:2410.04778). arXiv. https://arxiv.org/abs/2410.04778