

RESEARCH

Open Access



When a chatbot asks “How are you?”: A cross-sectional study of AI call conversations and depressive symptom detection among older adults in rural South Korea

Doyoon Kim^{1†}, Raon Jang^{1†}, Soong-nang Jang² and Susan Park^{3*}

Abstract

Background Rapid population aging in Korea has intensified the need for emotional and health-related support among older adults, especially in rural regions facing depopulation and limited care resources. Conversational artificial intelligence (AI) call systems have emerged as potential tools for monitoring daily well-being and detecting early signs of psychological distress. This study examined the association between conversational patterns in a conversational AI call service and risk of depressive symptoms among community-dwelling older adults, integrating quantitative and qualitative analyses to evaluate the system’s feasibility and limitations.

Methods This cross-sectional study included 2,896 adults aged 65 years or older residing in Jeongeup City, who used the Naver Clova AI CareCall service and completed a survey between May and August 2024. We linked 37,294 call records with survey data to compare service utilization patterns between high-risk (Geriatric Depression Scale-Short Form score ≥ 2) and low-risk groups. Multivariable logistic regression was used to identify predictors of the risk of depressive symptoms. We reviewed 73 conversation cases containing depressive cues, categorized into ten subthemes based on World Health Organization’s criteria. Three researchers independently assessed AI detection accuracy as correct, misinterpreted, or missed.

Results Among all participants, 42.0% were classified as high-risk for depressive symptoms. They were older, more likely to live alone, received social assistance more frequently, and reported poorer health. The high-risk group had fewer positive responses, more negative and unasked responses, and longer call durations. After adjustment, average negative responses (aOR = 1.18, 95% CI = 1.04–1.35), unasked domains (aOR = 1.12, 95% CI = 1.03–1.21), and longer call time (aOR = 1.01, 95% CI = 1.00–1.02) were associated with a higher risk of depressive symptoms. Qualitative findings revealed expressions such as appetite loss, fatigue, and hopelessness, though the AI frequently misinterpreted or failed to detect them due to contextual or speech-recognition errors.

[†]Doyoon Kim and Raon Jang contributed equally to this work as co-first author.

*Correspondence:
Susan Park
susan78@inha.ac.kr

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions This study provides real-world evidence that conversational and linguistic features in AI call service are associated with the risk of depressive symptoms among older adults in rural communities. AI call service may complement mental health monitoring by enabling early detection of emotional distress. However, limited contextual understanding and variable user acceptance highlight the need for continued technical refinement and integration with professional interpretation.

Keywords Older adults, Depressive symptoms, Conversational artificial intelligence, Depopulating rural areas, Digital care

Background

Rapid demographic aging is accompanied by a growing mental health crisis among older adults. Globally, depressive disorders represent the largest contributor to the burden of mental, neurological and substance use disorders, accounting for over 60% of years lived with disability (YLDs) in Asia-Pacific countries [1]. Korea exemplifies this global pattern, experiencing both rapid population aging and a growing burden of mental ill-health among older adults. The 1-year prevalence of major depressive disorder was also highest among older adults, reaching 1.9% in men in their 60s and 5.4% in women in their 70s [2]. These figures indicate that older adults represent the most vulnerable group in Korea's mental health landscape.

Depressive symptoms, as defined by the World Health Organization, encompass emotional, cognitive, and somatic manifestations such as persistent depressed mood, loss of interest, fatigue, and functional impairment lasting at least two weeks. While these symptoms do not necessarily constitute a clinical diagnosis, they are clinically meaningful and may progress along a continuum toward depressive disorders [3]. Evidence from OECD countries indicates that depressive symptoms are highly prevalent at the population level, with most individuals experiencing mild-to-moderate symptoms that do not meet clinical diagnostic thresholds. Importantly, these milder forms account for nearly 90% of all depressive symptom presentations and often remain undiagnosed despite their persistence, functional impact, and potential progression to more severe disorders. These findings underscore the substantial public health burden posed by depressive symptoms beyond clinically diagnosed depression [4].

Older adults living in rural areas, where aging and depopulation converge, are particularly vulnerable to experiencing depressive symptoms. Persistent outmigration of younger populations and low fertility have structurally entrenched population decline, weakening local economies and social foundations [5]. As a result, access to healthcare and welfare infrastructure has diminished. Residents of depopulated areas tend to report fewer social interactions, greater loneliness, and lower perceived social support, reflecting a decline in overall social well-being [6]. The 2024 Community Health Survey

indicates that depressive symptoms are more prevalent among older adults aged 60 years and above in Jeollabuk-do compared with younger age groups. Notably, municipalities with populations under 300,000 reported higher levels of depressive symptoms than larger cities, suggesting heightened vulnerability in rural and small-sized communities [7]. Jeongeup City in Jeollabuk-do, the focus area of this study, exemplifies these challenges as a rural community where both depopulation and aging are progressing rapidly. As of 2024, 31.9% of its residents were aged 65 or older, classifying the city as super-aged [8].

In this context, conversational AI call systems have gained attention as an alternative means to support mental health in regions with limited healthcare access. Chatbots and conversational agents have been used for self-management, psychoeducation, and mental health screening, demonstrating promising feasibility and user acceptance [9]. Recent research has expanded to explore cultural or thematic biases in chatbot-generated messages [10] and to analyze user conversations among specific populations, such as individuals with autism spectrum disorder, to examine diagnostic experiences and emotional expressions [11]. In Korea, an AI-based conversational agent, the Naver Clova CareCall, has been deployed in community care settings to monitor older adults' daily living and emotional health. A randomized controlled trial using this service reported reductions in depression and improvements in memory among patients with dementia [12], and its potential for preventing solitary deaths has also been examined [13].

Previous studies have largely focused on the development or acceptability of chatbot systems evaluating their performance [9, 10, 14]. However, most analyses have used text-based chatbots or short-term interventions within specific study participants, rather than actual older adults in real service environments. Few studies have adopted a mixed-methods approach combining quantitative analyses of demographic and call-pattern data with qualitative exploration of depressive expressions in real conversations. This gap indicates a lack of empirical evidence regarding whether conversational AI can function as an effective early detection tool for mental health issues in real-world care settings.

Therefore, this study utilized data from the AI Care-Call service implemented in Jeongeup City, South Korea,

to analyze participants' call responses and conversational content using both quantitative and qualitative methods. Specifically, the study examined associations between call response patterns and depressive symptom risk and explored conversational content to understand how depressive symptoms are manifested in AI-mediated interactions. Building on these analyses, the study aimed to assess the feasibility and limitations of conversational AI calls for early detection of depressive symptoms among older adults in rural areas. Findings from this study will provide foundational evidence for technology-based mental health interventions in communities facing simultaneous aging and depopulation, with implications for other societies undergoing similar demographic transitions.

Methods

This study employed a cross-sectional design using both quantitative and qualitative methods, with findings integrated at the interpretation stage.

Data source

The data used in this study were derived from a community-based needs assessment survey and a linked AI CareCall service. A face-to-face household survey was conducted among adults aged 60 years and older residing in the 15 town and township areas of Jeongeup City, Jeollabuk-do, South Korea, between September and November 2023. Eleven trained interviewers administered the survey using a standardized protocol and a tablet-assisted personal interviewing (TAPI) method, and a total of 4,989 older adults participated.

The survey collected information on sociodemographic characteristics, physical and mental health, caregiving status, activities of daily living, digital literacy, and emotional health, including depressive symptoms. After completion of the survey, participants who provided informed consent were enrolled in a telephone-based AI care service (Naver Clova CareCall).

The present analysis included participants aged 65 years and older who received the CareCall service and consented to linkage of their survey and service data between May and August 2024. Of the 3,097 registered CareCall recipients, individuals who did not respond to the survey ($n=29$), had missing data on the primary outcome variable—the Geriatric Depression Scale–Short Form (GDS-SF) ($n=60$), or had missing values in other analytic variables ($n=112$) were excluded. The final analytic sample comprised 2,896 participants. A total of 37,294 CareCall records were linked with survey data for quantitative analyses. For the qualitative component, 73 conversational cases indicating depressive expressions were selected from 8,195 call transcripts collected in June 2024.

The CareCall service was implemented to provide periodic monitoring of older adults' health and daily living status and to offer emotional support. The system automatically made weekly voice calls to registered participants and guided their responses across five key domains: health, sleep, diet, exercise, and outdoor activity. Each response was categorized as “normal (Yes)” or “problem (No),” while unasked questions were recorded as blank. Additional spontaneous remarks or requests from participants were logged as “extra requests.” All voice calls were transcribed into text and provided with corresponding audio files [15].

Variables and instruments

The primary outcome variable in the quantitative analysis was the risk of depressive symptoms, assessed using the Geriatric Depression Scale–Short Form (GDS-SF), a validated screening tool for older adults. The scale comprises five yes/no items, each scored as one point, with total scores ranging from 0 to 5. Consistent with prior studies, a cut-off score of 2 or higher was used to indicate elevated depressive symptom risk, and participants were classified into high-risk and low-risk groups accordingly [16].

Covariates included sociodemographic and health-related factors: age, sex, household type, receipt of social benefits, prior depression diagnosis, number of chronic conditions, disability status, and self-rated health. Household type was categorized as living alone, living with spouse only, living with other family members, or living with others. Receipt of social benefits was assessed based on self-reported eligibility for government-supported welfare programs targeting low-income or economically vulnerable older adults, including basic livelihood security, near-poverty assistance, and the basic pension. This variable was dichotomized as receiving benefits versus general status.

Health-related variables included prior or current diagnosis of depression (yes/no), disability status (yes/no), number of chronic conditions, and self-rated health. Chronic conditions were assessed by asking participants to report all past or current physician-diagnosed diseases from predefined lists—such as hypertension, diabetes and hyperlipidemia—and the total number was calculated. Self-rated health was measured using a four-point Likert scale ranging from very good to very poor.

CareCall-derived variables captured participants' engagement and response patterns during AI-based conversational calls. These included average call duration, number of completed calls, and responses across five domains (health, sleep, diet, exercise, and outdoor activity), categorized as “Yes” (normal), “No” (problem), or “Not asked.” Spontaneous remarks or help-seeking comments were recorded as extra requests. Additional

summary indicators were calculated at the individual level, including the number of domains with valid responses and the average number of positive and negative responses.

Quantitative analyses

Participants with missing values in the primary outcome or analytic variables were excluded from the analysis, as described in the Data source section. Participants were classified into high-risk and low-risk groups for depressive symptoms, and differences in sociodemographic characteristics and health status between the two groups were examined. Continuous variables were analyzed using independent sample t-tests, and categorical variables were compared using chi-square tests. Multivariable logistic regression analysis was then performed to identify factors associated with high risk of depressive symptoms. Adjusted odds ratios (aORs) and 95% confidence intervals (CIs) were reported. Statistical significance was defined as $p < 0.05$. Variables derived from overlapping CareCall response domains were examined for multicollinearity. Summary indicators that captured redundant information across domains were excluded from the final regression models to minimize multicollinearity. All analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

Qualitative analyses

The qualitative analysis reviewed CareCall questions and participant utterances to identify emotional expressions indicative of depressive symptoms or psychological distress (“depressive cues”). Each conversation was

interpreted in context to examine the meaning of these expressions and how they were recognized by the AI system, following established approaches in conversational AI research [11, 17]. A thematic analysis approach was used to identify and categorize recurring patterns of depressive expressions in the CareCall conversations, guided by the WHO framework for depressive symptoms [3] and prior emotion classification studies [18, 19]. Depressive cues were categorized into ten subthemes: loss of appetite, sleep problems, fatigue or lack of energy, feelings of worthlessness, social isolation, sense of loss, hopelessness or pessimistic thoughts, memory loss, non-specific depressive cues, and history of depression.

AI responses to these utterances were classified as correct detection, misinterpretation, or missed cue. Correct detection indicated appropriate recognition and response; misinterpretation referred to non-depressive statements misclassified as depressive; and missed cues indicated unrecognized depressive expressions. To ensure reliability, three researchers independently coded all utterances and resolved discrepancies through consensus, with final decisions made by majority agreement.

Results

Participant characteristics

A comparison of sociodemographic characteristics between the high-risk ($n = 1,217$, 42.02%) and low-risk ($n = 1,679$, 57.98%) groups for depressive symptoms revealed statistically significant differences across all examined characteristics (Table 1). The mean age of the high-risk group was 79.76 years, higher than that of the low-risk group (76.60 years). The proportion of women

Table 1 General characteristics of study participants*

Variable		Total ($n = 2896$)	Low-risk group ($n = 1679$)	High-risk group ($n = 1217$)	p-value
Age (years), mean $\pm$ SD		77.93 $\pm$ 7.04	76.60 $\pm$ 6.87	79.76 $\pm$ 6.85	< 0.0001
Sex, n (%)	Male	936 (32.32%)	575 (34.25%)	361 (29.66%)	0.0092
	Female	1960 (67.68%)	1104 (65.75%)	856 (70.34%)	
Social benefits, n (%)	General	353 (12.19%)	242 (14.41%)	111 (9.12%)	< 0.0001
	Receiving benefits	2543 (87.81%)	1437 (85.59%)	1106 (90.88%)	
Household type, n (%)	Living alone	1067 (36.84%)	563 (33.53%)	504 (41.41%)	< 0.0001
	Living with spouse only	1513 (52.24%)	941 (56.05%)	572 (47.00%)	
	Living with other family	303 (10.46%)	173 (10.30%)	130 (10.68%)	
	Living with others	13 (0.45%)	2 (0.12%)	11 (0.90%)	
Past/current depression, n (%)	Yes	47 (1.62%)	17 (1.01%)	30 (2.47%)	0.0023
	No	2849 (98.38%)	1662 (98.99%)	1187 (97.53%)	
Number of chronic conditions, mean $\pm$ SD		2.26 $\pm$ 1.34	2.13 $\pm$ 1.31	2.45 $\pm$ 1.37	< 0.0001
Disability, n (%)	Yes	290 (10.01%)	129 (7.68%)	161 (13.23%)	< 0.0001
	No	2606 (89.99%)	1550 (92.32%)	1056 (86.77%)	
Self-rated health status, n (%)	Very good	175 (6.04%)	127 (7.56%)	48 (3.94%)	< 0.0001
	Good	1834 (63.33%)	1196 (71.23%)	638 (52.42%)	
	Poor	779 (26.90%)	343 (20.43%)	436 (35.83%)	
	Very poor	108 (3.73%)	13 (0.77%)	95 (7.81%)	

*All variables were summarized after listwise deletion of missing values

was also higher in the high-risk group (70.34%) than in the low-risk group (65.75%). Social benefits recipients accounted for 90.88% of the high-risk group and 85.59% of the low-risk group. In addition, living alone was more common among the high-risk participants (41.41%) than among those in the low-risk group (33.53%).

In terms of health status, the high-risk group showed greater vulnerability. A history of depression diagnosis was reported by 2.47% of participants in the high-risk group—2.5 times higher than in the low-risk group (1.01%). The average number of chronic diseases was also higher in the high-risk group (2.45). Registered disabilities were observed in 13.23% of the high-risk group compared with 7.68% of the low-risk group, representing a 1.7-fold difference. Furthermore, 43.64% of the high-risk participants rated their subjective health as “poor” or “very poor,” nearly twice the proportion observed in the low-risk group (21.20%).

CareCall response patterns

Table 2 presents the comparison of conversational CareCall response characteristics between the high-risk and low-risk groups for depressive symptoms. There were no significant differences in total call duration between the groups. A consistent pattern was observed across the

health, sleep, diet, and exercise domains. In all domains, the proportion of “Yes” responses was lower among the high-risk group, suggesting an overall decline in daily functioning. Specifically, in the health domain, the positive response rate was 44.21% in the high-risk group compared with 51.34% in the low-risk group, with similar differences observed in the sleep, diet, and exercise domains.

The proportion of “Not asked” responses—cases in which questions for a particular domain were not delivered during the call—was higher in the high-risk group. In the exercise domain, the “Not asked” rate was 77.24% in the high-risk group compared with 72.25% in the low-risk group. In contrast, there were no significant differences between groups in positive response rates within the outdoor activity domain. Although the frequency of extra requests (e.g., spontaneous remarks or help-seeking comments) was slightly higher among the high-risk group, this difference was not statistically significant.

Overall, participants in the high-risk group had fewer total domains with valid responses, fewer positive responses per person, and significantly higher counts of negative responses and unasked domains, indicating reduced engagement and functional status during CareCall interactions.

Table 2 Comparison of call characteristics and CareCall responses by risk of depressive symptom

Variable		Low-risk group (n = 1679)	High-risk group (n = 1217)	p-value
Call characteristics	Average call time (sec)	29.19 ± 27.57	28.83 ± 27.91	0.7262
	Completed calls, mean ± SD	3.39 ± 3.73	3.04 ± 3.47	0.0091
Health domain	Yes, n (%)	862 (51.34%)	538 (44.21%)	0.0006
	No, n (%)	249 (14.83%)	218 (17.91%)	
	Not asked*, n (%)	568 (33.83%)	461 (37.88%)	
Sleep domain	Yes, n (%)	694 (41.33%)	417 (34.26%)	< 0.0001
	No, n (%)	23 (1.37%)	31 (2.55%)	
	Not asked*, n (%)	962 (57.30%)	769 (63.19%)	
Diet domain	Yes, n (%)	996 (59.32%)	665 (54.64%)	0.0173
	No, n (%)	15 (0.89%)	19 (1.56%)	
	Not asked*, n (%)	668 (39.79%)	533 (43.80%)	
Exercise domain	Yes, n (%)	361 (21.50%)	181 (14.87%)	< 0.0001
	No, n (%)	105 (6.25%)	96 (7.89%)	
	Not asked*, n (%)	1213 (72.25%)	940 (77.24%)	
Outdoor activity domain	Yes, n (%)	396 (23.59%)	258 (21.20%)	0.0784
	No, n (%)	144 (8.58%)	130 (10.68%)	
	Not asked*, n (%)	1139 (67.84%)	829 (68.12%)	
Extra request**	Yes, n (%)	106 (6.31%)	96 (7.89%)	0.1005
	No, n (%)	1573 (93.69%)	1121 (92.11%)	
Total number of captured domains, mean ± SD		2.29 ± 1.80	2.10 ± 1.79	0.0045
Average number of positive(“Yes”) responses per person, mean ± SD		1.97 ± 1.71	1.69 ± 1.62	< 0.0001
Average number of negative(“No”) responses per person, mean ± SD		0.32 ± 0.60	0.41 ± 0.72	0.0007
Average number of domains not asked per person, mean ± SD		2.71 ± 1.80	2.90 ± 1.78	0.0045

*Not asked means that the chatbot did not present the domain-related question during the call

**Extra requests include chest pain, high fever, rescue request, vomiting, dizziness, request for help, symptoms of paralysis, fainting symptoms, and breathing difficulty

Associations between CareCall response patterns and risk of depressive symptoms

Multivariable logistic regression analysis was conducted to examine the associations between CareCall response characteristics and classification into the high-risk group for depressive symptoms (Table 3). Several response indicators showed significant relationships with risk of depressive symptoms.

Each additional average negative (“No”) response per participant was associated with a higher likelihood of being classified as high-risk (aOR = 1.18, 95% CI = 1.04–1.35). Similarly, an increase in the number of unasked domains was associated with greater odds of high-risk classification (aOR = 1.12, 95% CI = 1.03–1.21). Longer mean call duration was also positively correlated with high-risk status (aOR = 1.01, 95% CI = 1.00–1.02). In contrast, the number of completed calls showed no significant association with the risk of depressive symptoms.

Domain-specific analysis showed significant associations in the sleep and exercise domains. Participants who had “Not asked” responses in the sleep domain were 1.30 times more likely to be classified as high-risk for depressive symptoms (aOR = 1.30, 95% CI = 1.04–1.63). In the exercise domain, those who responded “No” (indicating a problem) had 1.47 times higher odds of being in the

high-risk group (aOR = 1.47, 95% CI = 1.03–2.10). No significant associations were observed in the health, diet, or outdoor activity domains, nor with the occurrence of extra requests.

Qualitative analysis of conversational patterns by depressive cues

Analysis of depressive expressions observed during CareCall interactions revealed a wide range of emotional cues embedded in participants’ utterances, with variations in the CareCall’s response patterns depending on the type of expression. Representative conversational excerpts illustrating these patterns are presented in Table 4.

In most cases, CareCall attempted to provide empathetic responses to depressive cues. For example, expressions of loss or grief were typically met with comforting or encouraging messages, while statements indicating sleep disturbance or appetite loss elicited brief follow-up questions or suggestions to visit a healthcare provider. However, in several instances, CareCall’s responses were limited or contextually misaligned with the user’s emotional intent. In such cases, the system failed to fully interpret nuanced emotions or ambiguous statements, resulting in partial or inappropriate responses.

Table 3 Multivariable logistic regression results for risk of depressive symptoms according to CareCall call characteristics and responses*

Variable		aOR	95% CI	p-value
CareCall interactions**				
Average number of negative (“No”) responses		1.18	1.04–1.35	0.0134
Average number of domains not asked		1.12	1.03–1.21	0.0075
Average call time (sec)		1.01	1.00–1.02	0.0361
Number of completed calls		0.95	0.90–1.01	0.0871
Domain-specific response patterns***				
Health domain	Y	Reference	Reference	Reference
	N	1.03	0.81–1.31	0.8341
	Not asked**	1.19	0.91–1.56	0.1986
Sleep domain	Y	Reference	Reference	Reference
	N	1.60	0.87–2.93	0.1278
	Not asked**	1.30	1.04–1.63	0.0211
Diet domain	Y	Reference	Reference	Reference
	N	1.47	0.71–3.07	0.3032
	Not asked**	0.99	0.77–1.29	0.9666
Exercise domain	Y	Reference	Reference	Reference
	N	1.47	1.03–2.10	0.0330
	Not asked**	1.16	0.91–1.48	0.2348
Outdoor activity domain	Y	Reference	Reference	Reference
	N	1.10	0.81–1.50	0.5508
	Not asked**	0.99	0.78–1.26	0.9300
Extra request	Y	Reference	Reference	Reference
	N	1.10	0.79–1.51	0.5819

*Adjusted for demographics, socio-economic status, chronic disease, disability, and self-rated health

**Each model was estimated independently

***Not asked refers to participants who were not presented with the corresponding CA’s question, typically due to branching logic in the conversation flow

Table 4 Representative conversation examples between CareCall and older adults, categorized by depressive cues

Depressive Cue	Conversation Examples
Loss of appetite	User: I can't eat properly these days. CareCall: Why? Are you feeling unwell? User: I have no appetite. CareCall: I see. But try not to skip meals and eat something.
Sleep problems	User: I can't sleep well. CareCall: I see. Have you been under stress lately? User: When I wake up, I can't fall back asleep. It's just like this at the dorm. CareCall: I see. That must be uncomfortable. User: Yes. CareCall: Have you seen a doctor?
Fatigue/Lack of energy	User: I feel so tired. CareCall: Have you been to the hospital? User: Yes. CareCall: Are you taking your medicine? User: Yes. CareCall: And eating properly? User: Yes.
Worthlessness	User: That's why I'm worried. I'd rather just quietly disappear. CareCall: I hope you live a long and healthy life. User: I feel like I'm useless. CareCall: That's not true. You are capable.
Social isolation	CareCall: I see. Would it be possible to ask someone nearby for help? User: I'm alone. CareCall: Then maybe you could try reaching out to your family members? User: No, there's no one. I'm by myself. CareCall: I see. In that case, I'm afraid there's nothing I can do to help. I'm sorry.
Loss	User: I feel heavy-hearted after sending my uncle off to heaven. CareCall: I see. It must have been difficult for you. User: Yes. CareCall: Still, cheer up. Better days will come. User: Do you think so? CareCall: Yes, I'm sure of it. Stay strong!
Hopelessness/ Pessimistic thoughts	CareCall: How are you feeling today? User: Just waiting to die as life goes by. CareCall: Are you worried or concerned about something? User: Nothing. I just hope I die soon. CareCall: I see. But don't worry too much. Better days will come.
Memory loss	User: I'm not well. CareCall: Where are you feeling pain? User: I keep forgetting things. CareCall: I see. Have you been to a hospital?
Nonspecific depressive cues	User: Life is just hard... It doesn't feel like living anymore. CareCall: Don't worry too much. Things will get better.
History of depression	CareCall: By the way, are you taking your depression medication regularly? User: Yes, I take it every day. CareCall: Great. Keeping up with it will help you get better.

Table 5 summarizes the major error types identified during the detection of depressive expressions. The frequency of each expression and detection type is provided in Supplementary Table 1. A representative example of misinterpretation involved a speech recognition error in which the word “implant” was misheard as “divorce,” leading to an inappropriate CareCall response. In cases of missed cues, depressive expressions were clearly present in the user’s speech but went unrecognized, prompting CareCall to proceed with routine questions. These findings highlight the need for system improvement to

enhance early detection and intervention among high-risk older adults.

Discussion

This study conducted an integrated quantitative and qualitative analysis to examine the association between conversational patterns using CareCall and the risk of depressive symptoms among community-dwelling older adults. During call with CareCall, they exhibited fewer positive responses, more negative responses, and longer mean call durations—all of which were associated with

Table 5 Representative conversation examples between CareCall and older adults illustrating errors in detecting depressive cues

Detection Error Type	Conversation Examples
Misinterpretation (Communication error)	User: I'm having a really hard time these days because of a divorce. *Actual utterance: I'm having a hard time these days because of a dental procedure (implant). CareCall: I see. That must be difficult. User: Yes. CareCall: Still, cheer up. Better days are ahead.
Missed cue	User: I feel like there's no way out. Maybe I should just die. CareCall: I see. How will you spend the rest of the day?

an increased risk of depressive symptoms. These associations remained significant after adjusting for demographic and health variables. In particular, unasked questions in the sleep domain and negative responses in the exercise domain were independently related to high risk of depressive symptoms. Qualitative analyses revealed various depressive expressions—such as appetite loss, sleep disturbance, and feelings of loss—embedded in everyday conversations. However, CareCall often failed to consistently detect these cues, and both misinterpretation and missed-cue errors related to speech recognition were identified.

These findings reflect the limited acceptance and engagement of older adults with conversational AI call systems. The high-risk group demonstrated significantly fewer completed calls and positive responses, and a higher number of negative responses and unasked domains. Nevertheless, even in the low-risk group, the average number of completed calls was only 3.39, despite weekly scheduling over a 3-month period, and the proportion of unasked domain questions ranged from 33.8% to 72.2%. These results indicate that overall interaction completeness and continuity of conversational AI call were suboptimal.

The low level of engagement may stem from a combination of cognitive, sensory, and technological limitations among older adults. Previous research suggests that older adults exhibit selective constraints in hearing, speech production, turn-taking, and information processing [20]. Those with visual, auditory, or cognitive impairments often experience repeated difficulties when using AI-based services [13]. In the present study, speech recognition errors and unnatural system responses were also observed, underscoring the need for more explicit confirmation mechanisms and lower information density through socially cooperative dialogue [20].

Additionally, some older adults perceive conversational AI as a “life companion,” which can heighten discrepancies between user expectations and actual experiences [21]. Repetitive questioning and the lack of personalized interaction may amplify this gap. Jo et al. [22] reported that users perceived the AI call as “impersonal and robotic” because it failed to recall personal health histories, thereby limiting its capacity to provide emotional support. This aligns with the present study’s finding that

calls were frequently terminated early or failed to continue over time. Such outcomes may reflect emotional disengagement triggered when anticipated human-like interaction is not fulfilled. Moreover, qualitative analyses revealed that the limited engagement observed in this study may also be attributable to the quality of AI-generated responses themselves. As illustrated in Table 4, CareCall responses frequently adopted a formulaic structure, often beginning with neutral acknowledgments such as “I see,” followed by practical advice or reassurance, rather than emotionally validating the user’s expressed distress. For example, when users reported loss of appetite, responses tended to emphasize behavioral suggestions (e.g., “try not to skip meals”) without exploring underlying emotional or situational causes. Similarly, expressions of worthlessness or hopelessness were met with generic reassurance (e.g., “you are capable”), which may have unintentionally bypassed the user’s emotional experience. Such responses suggest that the system prioritizes problem-solving or reassurance over empathic reflection, reinforcing perceptions of the AI as mechanical rather than emotionally responsive.

Older adults exhibit diverse attitudes toward technology adoption. Their perceptions of technology and AI can be situated along a spectrum ranging from adapters, who actively integrate the benefits of technology into their lives, to wary users, who remain cautious due to learning burdens or privacy concerns, and resistant users, who reject technological dependence for fear of weakening human relationships [23]. The observed instances of call discontinuation, response refusal, and unasked questions in this study may reflect such attitudinal differences. These findings suggest the need for personalized interaction design that considers the user’s functional level and familiarity with technology in future AI-based care interventions.

The longer call durations observed among participants in the high-risk group for depressive symptoms may also be related to the qualitative nature of their speech. As shown in Supplementary Table 2, the high-risk group did not differ from the low-risk group in the number of sentences per call, but produced significantly more words and characters per sentence ($p < 0.0001$), indicating a tendency toward longer and more complex sentence structures.

Such linguistic patterns may stem from greater emotional openness during AI conversations. The anonymity and non-face-to-face nature of conversational agents can foster a sense of psychological safety, enabling some users to disclose more sensitive information to chatbots than to human counselors [9]. At the same time, these patterns may also reflect ruminative thinking or reduced linguistic organization associated with depression. König et al. [24] reported that individuals with mild depression used more words in self-reports and social media posts than non-depressed counterparts, interpreting this as increased verbalized thought arising from rumination or mind-wandering. Accordingly, the longer utterances observed among high-risk participants in this study may represent a coexistence of emotional expressiveness and ruminative cognition. However, longer call duration may also be influenced by factors not captured in this study—such as slower speech rate or delayed responses—highlighting the need for further phonological and temporal analyses in future research.

This study offers three key contributions. First, it provides real-world evidence based on conversational data between older adults and an AI call system implemented in a community care setting. Whereas most previous studies on conversational agents have focused on acceptability and feasibility [25, 26], this study examined how emotional cues were expressed and detected during real-life AI–user interactions. Second, by integrating quantitative and qualitative analyses, the study elucidated the relationship between conversational patterns and risk of depressive symptoms. Notably, emotional signals were identified even in conversations centered on everyday topics such as health, sleep, and diet—without direct questions about mood. This finding aligns with prior research suggesting that open-ended conversations can capture holistic aspects of users' lives and emotions [22]. Third, the study was conducted in a rural area with an aging rate exceeding 30%, demonstrating the potential of AI-based care to mitigate disparities in mental health accessibility. Korea illustrates rapid demographic aging most distinctly, with life expectancy reaching 83.5 years in 2023—well above the OECD average of 81.1—and the proportion aged 65 or older projected to exceed 35% by 2050 [27, 28]. As the older population grows and care resources remain concentrated in urban centers [29], conversational AI may offer an accessible means to support emotional well-being among older adults in underserved communities. The present findings provide evidence that AI-based care can serve as a complementary approach in aging and resource-limited settings. These results support earlier work showing that personalized, routine care can promote emotional well-being and reduce loneliness, and that technology can serve as an effective adjunct to human care [14]. The

implications extend to other societies undergoing similar demographic transitions, where digital solutions may help bridge structural gaps in community-based mental health care.

However, CareCall still demonstrates limited contextual understanding, leading to recurring instances of misinterpretation and missed cues. Based on qualitative analysis, several cases that researchers classified as being at high risk of depressive symptoms based on conversational context were categorized as low-risk in the self-reported GDS-SF survey, and CareCall also failed to detect them consistently. This discrepancy reflects both the underreporting tendency of older adults in self-assessed surveys—often due to stigma concerns [30]—and the current technological limitations of AI detection systems. These findings suggest that both survey-based and AI-based approaches have inherent constraints, underscoring the continued importance of human expert interpretation in integrating conversational context. Future AI-based care services should therefore combine technological refinement with expert interpretive oversight.

This study also has several limitations. It was conducted in a single rural area, which may restrict generalizability. Nevertheless, Jeongeup represents a typical rural community with an aging rate exceeding 30% and is classified as a depopulation-risk region, providing reasonable external relevance for communities with similar demographic characteristics. In addition, the study was based on short-term observations from May to August, and the qualitative analysis focused solely on conversations collected in June, limiting the ability to capture temporal variations in interaction patterns. Future research should employ longitudinal and multi-site designs to evaluate the sustainability of AI-based care services and the stability of emotional cue detection over time.

Conclusion

This study empirically demonstrated the association between a telephone-based conversational AI call service and risk of depressive symptoms among older adults residing in depopulated rural regions. The findings suggest that AI-based call systems may serve as complementary tools for the early detection of emotional distress and prevention of mental health deterioration among older adults, particularly in rural areas with limited care resources.

However, given the technological constraints and individual differences in acceptance, AI call systems should be designed to augment rather than replace human expertise, integrating professional interpretation and support with continuous technological improvement. Future policies should institutionalize AI-based call services as preventive and supplementary components

within community mental health systems and establish infrastructure that enables their long-term operation and secure data management. Such efforts will contribute to building a community-based care model that supports personalized interaction and continuous monitoring for older adults in aging and depopulating societies.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12877-026-07038-0>.

Supplementary Material 1.

Acknowledgements

The authors would like to thank the researchers at the Community Care and Health Equity Research Institute in Chung-Ang University for their assistance with data collection.

Authors' contributions

DYK and ROJ conducted data analysis and manuscript writing; SJ was responsible for study design; and SP contributed to study design, data analysis, and manuscript writing.

Funding

This work was supported by the Ministry of Health and Welfare and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (HS23C0056, RS-2024-00450078).

Data availability

The data used in this study is not publicly available due to ethical restrictions. Access to the data is limited to researchers approved by the Institutional Review Board (IRB), and any secondary use of the data requires separate IRB approval.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board (IRB) of Chung-Ang University (1041078-20230426-HR-118). The research was conducted in accordance with the ethical principles of the Declaration of Helsinki. All participants provided written informed consent prior to participation in the study.

Consent for publication

All participants provided informed consent prior to participation.

Competing interests

The authors declare no competing interests.

Author details

¹Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea

²Red Cross College of Nursing, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul, Republic of Korea

³College of Medicine, Inha University, 100 Inha-ro, Michuhol-gu, Incheon, Republic of Korea

Received: 22 October 2025 / Accepted: 19 January 2026

Published online: 28 January 2026

References

1. OECD/WHO. Health at a Glance: Asia/Pacific 2024. Paris: OECD Publishing; 2024. <https://doi.org/10.1787/51fed7e9-en>.
2. Ministry of Health and Welfare, National Center for Mental Health. 2021 National Mental Health Survey of Korea. Sejong: Ministry of Health and Welfare; 2021.
3. WHO. Depression Fact S. 2025. Available from: <https://www.who.int/news-room/fact-sheets/detail/depression>.
4. OECD. Mental health promotion and prevention: best practices in public health. Paris: OECD Publishing. 2025. Available from: <https://doi.org/10.1787/88bbe914-en>.
5. Johnson KM, Lichter DT. Rural depopulation: growth and decline processes over the past century. *Rural Sociol.* 2019;84(1):3–27.
6. Jokela M, Laakasuo M, Parikka S, Rotkirch A, Hämäläinen H. Psychological and social wellbeing associated with regional population change in Finland. *J Community Appl Social Psychol.* 2024;34(4):e2851.
7. Korea Disease Control and Prevention Agency (KDCA). Korea community health at a glance 2024: Korea community health survey (full version). Cheongju: KDCA; 2025.
8. Ministry of the Interior and Safety. Resident registration population status by administrative district and single year of age. Sejong: Ministry of the Interior and Safety; 2024. Available from: <https://umin.mois.go.kr>.
9. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Can J Psychiatry.* 2019;64(7):456–64.
10. Chin H, Song H, Baek G, Shin M, Jung C, Cha M, et al. The potential of chatbots for emotional support and promoting mental Well-Being in different cultures: mixed methods study. *J Med Internet Res.* 2023;25:e51712.
11. Aghakhani S, Carre N, Mostovoy K, Shafer R, Baeza-Hernandez K, Entenberg G, et al. Qualitative analysis of mental health conversational agents messages about autism spectrum disorder: a call for action. *Front Digit Health.* 2023;5:1251016.
12. Kang S, Lee Y, Kim N, Ok S-H, Kim P, Park O-H, et al. Beneficial effect of artificial intelligence care call on memory and depression in community dwelling individuals with dementia. *Sci Rep.* 2025;15(1):27116.
13. Lyu J, Kim YB, Park J-S. AI services to prevent solitary death in South Korea. *Int J Contents.* 2025;21(2):61–66. doi:10.5392/IJoC.2025.21.2.061.
14. Rodríguez-Martínez A, Amezcua-Aguilar T, Cortés-Moreno J, Jiménez-Delgado JJ. Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health. *Healthcare (Basel).* 2023;12(1):62. <https://doi.org/10.3390/healthcare12010062>.
15. Choi E, Han K, Hwang J, Jang S-n, Park S, Choi S, et al. Care networking system-based conversational AI service for older adults in depopulating area: protocol development and empirical study. *Korean J Health Educ Promotion.* 2024;41:99–111.
16. Cho B, Son KY, Oh B, Kim SJ, Kwon IS, Park BJ, et al. Development and validity and reliability of Korean comprehensive assessment tools for geriatric ambulatory care. *J Korean Geriatr Soc.* 2013;17(1):18–27.
17. Aldhyani THH, Alsubari SN, Alshebami AS, Alkahtani H, Ahmed ZAT. Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. *Int J Environ Res Public Health.* 2022;19(19):12635. doi:10.3390/ijerph191912635
18. Inkster B, Kadaba M, Subramanian V. Understanding the impact of an AI-enabled conversational agent mobile app on users' mental health and wellbeing with a self-reported maternal event: a mixed method real-world data mHealth study. *Front Global Women's Health.* 2023;4:1084302.
19. Kaywan P, Ahmed K, Ibaida A, Miao Y, Gu B. Early detection of depression using a conversational AI bot: A non-clinical trial. *PLoS ONE.* 2023;18(2):e0279743.
20. Kopp S, Brandt M, Buschmeier H, Cyra K, Freigang F, Krämer NC, et al. Conversational assistants for elderly users—the importance of socially cooperative dialogue. In: Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications (ICAHGCA@AAMAS); 2018.
21. Chung K, Oh YH, Ju DY. Elderly users' interaction with conversational agent. In: Proceedings of the 7th International Conference on Human-Agent Interaction (HAI 2019). Association for Computing Machinery; 2019. p. 277–9. <http://doi.org/10.1145/3349537.3352791>.
22. Jo E, Epstein DA, Jung H, Kim Y. Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2023:18. <https://doi.org/10.1145/3544548.3581503>.
23. Wolfe BH, Oh YJ, Choung H, Cui X, Weinzapfel J, Cooper RA, et al. Caregiving artificial intelligence chatbot for older adults and their Preferences,

- Well-Being, and social connectivity: Mixed-Method study. *J Med Internet Res.* 2025;27:e65776.
24. König A, Tröger J, Mallick E, Mina M, Linz N, Wagnon C, et al. Detecting subtle signs of depression with automated speech analysis in a non-clinical sample. *BMC Psychiatry.* 2022;22(1):830.
 25. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, Acceptability, and effectiveness of Web-Based conversational agents to facilitate problem solving in older adults: controlled study. *J Med Internet Res.* 2020;22(5):e16794.
 26. Lutfi SL, Fernández-Martínez F, Lorenzo-Trueba J, Barra-Chicote R, Montero JM. I feel you: the design and evaluation of a Domotic affect-sensitive spoken conversational agent. *Sens (Basel).* 2013;13(8):10519–38.
 27. OECD. OECD Health Statistics [Dataset]. Paris: OECD; 2025. Available from: <https://data-explorer.oecd.org>.
 28. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2024: Methodology of the United Nations population estimates and projections.* New York: United Nations; 2024.
 29. Yi Y-J, Kim E-J. The effects of accessibility to medical facilities and public transportation on perceived health of urban and rural elderly: using generalized ordered logit model. *J Korean Reg Dev Association.* 2015;27(1):65–88.
 30. Davison TE, McCabe MP, Mellor D. An examination of the gold standard diagnosis of major depression in aged-care settings. *Am J Geriatr Psychiatry.* 2009;17(5):359–67.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.