

Turkish-language Large Language Model-supported Voice Companion for Cognitive Monitoring via Social Artificial Intelligence in Older People Living Alone: A Single-Arm Pilot Feasibility Study

Asli Kilavuz

Ege Üniversitesi Tıp Fakültesi Hastanesi

Özgür Akşan

ozguraksan@aydin.edu.tr

Istanbul Aydın University

Cenab Batu Bora

Georgia Institute of Technology

Feryal Akşan

Istanbul Aydın University

Pınar Sırmatel Bücü

Buca Seyfi Demirsoy Education and Research Hospital

Sebnem Bora

Ege University

Research Article

Keywords: Artificial Intelligence, Large Language Model, Loneliness, Depression, Cognitive Function, Older Adults, Mobile Health

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10998298/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Older adults living alone are particularly vulnerable to reduced social interaction and loneliness. Simultaneously, age-related cognitive decline becomes more prevalent, leading to a growing need for scalable, technology-supported interventions that can both reduce social isolation and support routine cognitive monitoring in home settings. The aim of this study was to evaluate the feasibility, usability, and acceptability of iShe (Intelligent Social Health and Interaction) a Turkish-language Large Language Model (LLM) supported mobile voice companion designed for supportive social engagement and preliminary cognitive screening among community-dwelling older adults living alone.

Methods

In this single-arm pilot feasibility study, ten community-dwelling adults aged 65 years and older living alone used the application for 30 days and were encouraged to complete one daily voice session. Loneliness and depressive symptoms were assessed at baseline and endline using the De Jong-Gierveld Loneliness Scale (DJGLS) and Geriatric Depression Scale-15 (GDS-15). Cognitive status was monitored in-app daily using the Short Portable Mental Status Questionnaire (SPMSQ), and the Standardized Mini-Mental State Examination (SMMSE) was administered face-to-face at baseline and endline. Paired changes were analyzed using Wilcoxon signed-rank tests and descriptive feasibility metrics (retention, adherence, session duration, support needs).

Results

Retention was 100%. Participants completed all 30 sessions. Mean session duration was 18.6 ± 3.2 minutes. DJGLS decreased from 8.50 ± 5.28 to 5.70 ± 4.92 ($p = .007$; Cohen's $d_z = 0.98$; median change 2 [IQR 1–3.75]). GDS-15 decreased from 4.50 ± 4.22 to 2.30 ± 2.26 ($p = .039$; Cohen's $d_z = 0.66$; median change 1 [IQR 0–2]); the proportion scoring 6 or higher declined from 30% to 10%. SMMSE remained stable (29.3 ± 1.1 to 29.6 ± 0.7), and all SPMSQ results remained in the normal range (0–2 errors), consistent with SMMSE categories.

Conclusions

This study provides strong feasibility and acceptability signals for daily use of an LLM- supported voice companion by older adults living alone, with exploratory within-group improvements in loneliness and depressive symptoms and stable, concordant cognitive screening outputs. Larger controlled studies are needed to establish efficacy and longer-term cognitive monitoring value.

Background

Population aging is accelerating globally and in Türkiye, increasing the need for scalable interventions that support psychosocial well-being and cognitive health in later life [1–3]. Older adults frequently experience social isolation and loneliness due to retirement, mobility limitations, bereavement, and shrinking social networks [4, 5]. Loneliness is associated not only with psychological issues, but also with adverse health outcomes, including depression and anxiety as well as cardiovascular morbidity, immune dysfunction, and increased mortality [6, 7].

Another important aspect of the aging process is changes in cognitive functions. Although no significant cognitive decline is observed during the healthy aging process, advanced age is one of the most important risk factors for neurodegenerative diseases such as dementia and mild cognitive impairment (MCI) [8, 9]. Cognitive disorders progress slowly clinically and distinguishing them from the normal aging process can be difficult. Therefore, systematic screening methods are needed for early diagnosis and intervention [10, 11]. Conventional neuropsychological assessment methods have limitations in terms of feasibility, particularly for elderly individuals living alone and in rural areas, due to reasons such as the need for trained personnel, the requirement for face-to-face administration, and geographical access difficulties [12].

Within the healthcare sector, the evolving influence of artificial intelligence (AI) is reshaping traditional practice, and large language models (LLMs) are increasingly used for education, documentation, and decision support across specialties including psychiatry, nursing, dentistry, pathology, pharmacy, urology, gastroenterology, parasitology, and ophthalmology [13–16]. LLMs enable more natural, context-aware dialogue than prior rule-based systems, creating opportunities to integrate companionship with structured, validated screening tasks in low-friction formats. In geriatrics, digital health tools and voice interfaces are especially promising because they reduce interaction burden and can provide frequent contact for older adults [17]. Recent work suggests that voice assistants and conversational agents can improve loneliness and social support outcomes, although much of the literature remains focused on usability and early-stage feasibility rather than definitive clinical efficacy [18–20]. Yet geriatrics-focused deployments remain scarce; we implemented this approach in geriatrics, and to our knowledge it represents the first LLM-supported voice companion evaluated in the field and the first delivered in Turkish. This study is novel in combining daily LLM-driven voice companionship with embedded, validated cognitive screening in a Turkish-language geriatric cohort, a combination not previously reported in the literature. From a speech-processing perspective, additional challenges arise because written and spoken language differ, and diverse accents and dialects can affect understanding; robust comprehension of these variations is essential. Prior pilots with smart speakers, social robots, and chatbots have emphasized usability and perceived social presence [21, 22]. However, evidence remains limited for systems that combine (i) natural, empathetic voice interaction in the local language, (ii) routine, validated cognitive screening embedded into everyday conversation, and (iii) privacy-preserving data handling. To address this gap, we developed iShe (Intelligent Social Health and Engagement) companion, a Turkish-language mobile voice companion that offers daily conversation and unobtrusively

incorporates the SPMSQ within dialogue. This dual focus addresses two underdeveloped areas in the literature: cognitive concordance with clinician tools and sustained engagement with minimal support in an older adult population.

Voice-based conversational agents and social robots are increasingly explored as tools to reduce loneliness and improve social connectedness in later life. Recent reviews emphasize that much of the evidence base remains small and feasibility-oriented, yet several studies report meaningful improvements in loneliness and social support, particularly with more interactive modalities [18, 23]. Observational and pilot studies using commercial voice assistants have shown reductions in loneliness and increases in perceived social support, with video-enabled devices appearing to yield larger gains than audio-only systems [19, 21]. A text-based chatbot trial in Taiwan reported high daily engagement and improved loneliness among older adults during COVID-19 isolation [22]. Randomized trials of social robots, such as BOCCO emo, also show greater loneliness reduction and improved well-being compared to wait-list controls, alongside qualitative reports of emotional support [24]. A recent study similarly concludes that AI companionship can meaningfully support mental wellbeing, highlighting autonomy, perceived ease of use, and the relational quality of human-AI interaction as key determinants [25]. Overall, the literature suggests promising psychosocial benefits, but the field still needs larger, rigorously controlled evaluations.

The emergence of LLMs has enabled more natural, context-aware dialogue than earlier rule-based systems, accelerating the use of conversational agents in patient education, documentation, and decision support across multiple specialties [14–17]. Geriatrics-focused LLM deployments, however, remain rare. The GRACE system, a hybrid rule-based and LLM-supported voice assistant with an embodied avatar, demonstrated high usability and a positive working alliance in a lab-based study with older adults [26]. Voice-first interaction is particularly accessible for seniors with lower technology literacy, yet safety, reliability, and cultural-language fit are essential. To our knowledge, no peer-reviewed LLM companion has been developed specifically for Turkish older adults, reinforcing the novelty of the present work.

Conventional cognitive screening tools such as the SMMSE or SPMSQ are typically administered infrequently in clinical settings, limiting timely detection of decline. The SPMSQ has been validated in Turkish geriatric populations with strong diagnostic accuracy and correlation with the SMMSE [27]. Digital cognitive assessments have shown feasibility and can improve standardization and access, including mobile applications designed for self-administered screening [28–30]. Embedding brief orientation and memory questions into everyday dialogue allows screening to occur with minimal burden. However, prior work has not combined daily social interaction with embedded cognitive screening in a local language; this gap motivates the iShe approach.

The aim of this study was to evaluate the feasibility and acceptability of the intervention in a pilot setting, and to explore within-group changes in loneliness and depressive symptoms as well as the descriptive concordance between daily SPMSQ and clinician-administered SMMSE categories.

Methods

Study Design, Setting and Participants

This was a single-arm pretest-posttest pilot feasibility study conducted between April and October 2025 at the Geriatrics Outpatient Clinic of Ege University Faculty of Medicine. A single-group design was selected to characterize implementation and generate preliminary estimates before a controlled trial [31]. Reporting was informed by pilot-feasibility and eHealth reporting guidance where applicable [32, 33]; the study was not randomized. No a priori statistical power calculation was performed. The target of 10 participants was pragmatic and consistent with guidance for pilot and feasibility studies, in which the purpose is to estimate feasibility parameters such as retention, adherence and session duration rather than to test efficacy [31].

The sample comprised the first 10 volunteers aged 65 years or older who lived alone, provided written informed consent, and had basic smartphone proficiency, operationalized as the ability to install and use common communication applications independently. Exclusion criteria were a known diagnosis of MCI or dementia, a prior psychiatric diagnosis, hearing or speech impairment that prevented telephone conversation, or severe uncorrected visual impairment. Hearing and vision were not formally tested.

Participants with known MCI or dementia were excluded because this first pilot focused on basic usability and conversational feasibility without the additional effects of established cognitive impairment on comprehension, engagement, and repeated SPMSQ performance. This choice limits clinical relevance to the population most likely to benefit from cognitive surveillance; evaluation in people with MCI should occur only with stronger safeguards and an appropriately powered protocol.

Ethics and data protection

The protocol was approved by the Ege University Faculty of Medicine Medical Research Ethics Committee (Date: 14 Nov 2024; Decision: 24-11T/7). All participants provided written informed consent in accordance with the Declaration of Helsinki. The AI components used for speech and text processing (Whisper and GPT-4o Realtime) were operated strictly under the ethics-approved protocol. All user data were de-identified before any external processing, and no personally identifiable information was transmitted to external services. All data processing and storage procedures complied with Türkiye's Law on the Protection of Personal Data (KVKK; Law No. 6698). Participants were instructed not to share personally identifying details during sessions. No raw audio was stored; only de-identified text summaries required for research analyses were retained. All study data were hosted on a managed cloud database platform (Supabase) on which data are encrypted in transit (TLS) and at rest by default. The mobile application had no direct database credentials: all read and write operations were mediated by the backend server, which alone held the service keys, and direct database access was restricted to authenticated research-team accounts. Platform-level authentication and API request logs provided a basic audit trail of data access.

Measures and Data Sources

Sociodemographic and medical information (Age, sex, education level, income, and number of chronic diseases) was collected for descriptive reporting.

Instruments were selected because Turkish versions with published psychometric evidence were available, the respondent burden was appropriate for an older sample, and the intended administration mode matched the study workflow. DJGLS and GDS-15 measured psychosocial outcomes; SMMSE provided a trained face-to-face cognitive screen; and SPMSQ provided a short fully verbal in-app screen. No additional standardized usability questionnaire, structured interview, open-ended questionnaire, or qualitative data collection was performed. The specific measures and their operational definitions are listed below in Table 1.

Table 1

Outcome categories and corresponding measures used in the iShe single-arm pilot feasibility study conducted among adults aged 65 and over living alone at the Ege University Geriatric Outpatient Clinic in Izmir between April and October 2025.

Outcome category	Measures and operational definitions
Primary feasibility and acceptability outcomes	• Retention at 30 days (completion of endline assessment). • Adherence (completed daily sessions divided by total possible sessions). • Session duration (minutes). • Technical support needs (number and type of support contacts).
Secondary exploratory outcomes	• Loneliness: DJGLS, 11 items, total 0–22; higher indicates greater loneliness [34–36]. • Depressive symptoms: GDS-15, 15 yes/no items, total 0–15; scores ≥ 6 indicate clinically relevant symptoms [37, 38].
Cognitive screening outcomes	• Clinician-administered SMMSE at baseline and endline (0–30) [39, 40]. • In-app SPMSQ delivered daily (10 items; error count 0–10; 0–2 normal, 3–4 mild, 5–7 moderate, 8–10 severe) [27, 41].
DJGLS, De Jong–Gierveld Loneliness Scale; GDS-15, Geriatric Depression Scale-15; SMMSE, Standardized Mini-Mental State Examination; SPMSQ, Short Portable Mental Status Questionnaire	

Feasibility and Engagement Measures

Feasibility outcomes were retention at 30 days, adherence (completed sessions divided by 300 possible participant-days), session duration, and number of technical support contacts. Retention and adherence were also treated as indirect acceptability indicators; no self-reported acceptability instrument was administered. The available study summary distinguished whether a participant required a remote

troubleshooting contact but did not classify the specific problem. Preintervention conversation volume, out-of-app social interaction, reasons for missed sessions, and participant-level conversational-turn counts were not present in the available analysis materials and therefore were not analysed.

Psychosocial Measures

The 11-item DJGLS measures emotional and social loneliness, with total scores ranging from 0 to 22 and higher scores indicating greater loneliness [34–36]. Items are scored 0–2 (three-point Likert); total 0–22 with higher scores indicating greater loneliness. The Turkish validation study of 809 older adults supported the two-factor structure, reported Cronbach $\alpha = .85$, and reported test-retest reliability of $r = .93$ [36].

The GDS-15 is a 15-item yes/no screen for depressive symptoms, with scores ranging from 0 to 15 [37, 38]. The Turkish validation study reported Cronbach $\alpha = .92$ and, at a cutoff of 5 and higher, sensitivity of 92% and specificity of 91% against Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria [38]. In the present study, a score of 6 or higher was used to describe clinically relevant symptom levels, consistent with the prespecified analysis. DJGLS and GDS-15 were administered face-to-face by study staff at baseline and endline.

Cognitive Screening Measures

The SMMSE assesses orientation, registration, attention and calculation, recall, and language on a 0–30 scale [39, 40, 42]. In its Turkish validation, a 23/24 threshold had sensitivity of 0.91 and specificity of 0.95 for mild dementia; interrater reliability was $r = 0.99$, with $\kappa = 0.92$ [39]. SMMSE tests were administered face-to-face by trained physicians at the at baseline and end of the study. Permission to use the SMMT test has been obtained from Dr. D. William Molloy.

The SPMSQ contains 10 verbally answerable orientation, personal-information, and calculation items and is scored by number of errors [41]. Conventional categories are 0–2 errors (normal), 3–4 (mild impairment), 5–7 (moderate impairment), and 8–10 (severe impairment), with education-adjusted interpretation where appropriate. The Turkish validation reported Cronbach $\alpha = .74$ and diagnostic utility in a geriatric sample [27]. In iShe, the SPMSQ was intended to be administered once in each completed daily session. The prompt used fixed-order questions and recorded the participant's first response without telling the participant whether it was correct. Multimedia Appendix 1 provides the documented prompt configurations and version note.

Target Population and Interface Design

iShe was designed as a voice-first mobile application for iOS and Android. During iterative development, the interface was simplified after older users had difficulty with small touch targets and multistep flows. The final interface used high-contrast screens, large text and buttons, a single prominent start/stop

control requiring a 3-second press to reduce accidental activation, visible recording/speaking status, and forced loudspeaker output. Participants could read the live transcript but did not need to type during a session. Study staff provided a brief video demonstration and in-person assistance when needed.

Conversation and SPMSQ Delivery

The intended intervention had two components: supportive Turkish-language conversation and one SPMSQ administration per completed daily session. Early application builds exposed separate “Conversation Mode” and “Question Mode” controls. Study staff instructed participants to converse and then complete the question mode. A later build consolidated the flow and opened directly with the SPMSQ prompt. The documented question-mode prompt announced the SPMSQ, asked 10 items sequentially in a fixed order, and instructed the assistant not to label answers as correct or incorrect.

Repository history shows that prompt wording and the order of conversation versus SPMSQ delivery were not version-locked throughout the April-October 2025 study window. Session records available for analysis did not include an application-build or prompt-version identifier, so delivery cannot be reconstructed at the session level. The final documented configuration is reproduced in Multimedia Appendix 1. This versioning limitation is considered when interpreting the cognitive results. Representative response excerpts generated under this prompt are shown in Table 2.

Table 2

Representative response examples generated under the system prompt demonstrating how to establish contact and transition to the built-in SPMSQ scan within the daily Turkish voice session (pseudonymized; not participant data)

Context	Example dialogue (Turkish and English)	
Rapport-building	User (TR):	“Bugün biraz yalnız hissettim.”
	Assistant (TR):	“Bunu paylaşmanız çok kıymetli Fatma Hanım. Yalnız hissetmek zor olabilir; bugün size iyi gelen bir şey oldu mu?”
	User (EN):	“I felt a bit lonely today.”
	Assistant (EN):	“Thank you for sharing that, Ms. Fatma. Feeling lonely can be hard; did anything make you feel better today?”
SPMSQ transition	Assistant (TR):	“Şimdi küçük bir hafıza egzersizi yapalım, hazır mısınız? Birinci soru: Bugün tarih, ay ve yıl nedir?”
	User (TR):	“21 Aralık 2025.”
	Assistant (TR):	“Teşekkür ederim Fatma Hanım. İkinci soru: Bugün haftanın hangi günü?”
	Assistant (EN):	“Now let’s do a short memory exercise; are you ready? First question: What is today’s date, month, and year?”
	User (EN):	“December 21, 2025.”
	Assistant (EN):	“Thank you, Ms. Fatma. Second question: What day of the week is it today?”
TR, Turkish; EN, English		

System Architecture and Data Flow

The mobile client was built with Expo and React Native. WebRTC transmitted microphone audio over an encrypted connection for realtime interaction. OpenAI Realtime processed audio and generated Turkish audio and text responses; the repository used Whisper-1 for transcription and a GPT-4o Realtime preview model for response generation. Repository history indicates a change from gpt-4o-realtime-preview-2024-12-17 to gpt-4o-realtime-preview-2025-06-03 in August 2025. Model version was not stored with each study session.

A Node.js/Express backend authenticated requests, held service credentials, and created realtime sessions. The participant’s display name and prior transcript context were inserted into model instructions for personalization. User and assistant text transcripts, timestamps, response identifiers, and the authenticated account identifier were stored in a Supabase conversations table. Text embeddings were generated on the backend with the Xenova all-MiniLM-L6-v2 model and stored with the transcript to retrieve prior topics. Thus, operational conversation records were linked to study account

identifiers and were not anonymous at the point of service operation. Figure 1 presents the conceptual data flow.

Repository code from the early study period also contained a client-side recording function and a backend route that could upload session audio to a Supabase Storage bucket; a later build removed the client-side recording call. Deployment records and storage audit logs were not available in the study documentation. Consequently, which builds were deployed, whether the upload route was active, and the retention or deletion status of any recordings could not be determined. Native-device screenshots of the iShe interface are given in Fig. 2.

AI Response Constraints and Safety Procedures

The prompts required Turkish output, a warm and supportive tone, fixed SPMSQ sequencing, no explicit correct/incorrect feedback, and expressions of uncertainty when the model was unsure. The assistant was presented as a companion and screening interface, not as a diagnostic system. Participants had a study messaging channel for reminders and technical contact.

The deployed documentation did not show a deterministic medical-advice filter, a separate content-classification layer, a self-harm or emergency escalation workflow, a dependency-mitigation protocol, or a clinician alert pathway. No formal red-team evaluation, hallucination rate, inappropriate-response audit, or prespecified AI safety-event collection was performed. These omissions are material limitations for use with a potentially vulnerable population and prevent a claim that the system's psychological safety was established.

Procedure

Project documentation divided the work into prototype development, participant preparation, the 30-day pilot, and evaluation. Prototype work integrated the mobile client with the LLM service, and iterative testing informed the interface changes described above. At baseline in the Geriatrics Outpatient Clinic, study staff obtained consent and administered the sociodemographic form, DJGLS, GDS-15, and SMMSE face-to-face. Participants then received onboarding on login, the start/stop control, headset or loudspeaker use, and the intended conversation and SPMSQ workflow. The duration of each participant's onboarding was not recorded.

A messaging group was used for reminders and troubleshooting. Participants used iShe at self-selected times for 30 days and were encouraged to complete one session daily. The available materials did not specify a fixed reminder schedule or a rule for contacting participants after non-use. Endline face-to-face assessment in the clinic repeated DJGLS, GDS-15, and SMMSE. The documentation identified the assessors only as study staff and did not state whether they were blinded to baseline scores.

The study summary recorded session completion, duration, support contacts, and SPMSQ category results. The available documentation did not identify who reviewed daily SPMSQ results or define a threshold for staff notification or clinical escalation. The application repository also stored

conversational transcripts, but a verified participant-level turn-count export was not included with the analysis files. Acoustic, paralinguistic, and lexical-semantic markers such as speaking rate, response latency, vocabulary diversity, or sentiment were not analysed.

Statistical Analysis

Analyses were performed in IBM SPSS Statistics version 25. Continuous variables are summarized with mean and SD and, where relevant, median and range. Categorical variables are presented as counts and percentages. Because fewer than 100 participants were enrolled, percentages are reported without decimals and with absolute counts.

Normality was assessed with the Shapiro-Wilk test. Paired baseline-endline DJGLS and GDS-15 scores were compared using Wilcoxon signed-rank tests. Exact P values are reported. Cohen d_z was calculated as the mean paired change divided by the SD of paired changes to describe exploratory magnitude. The nominal significance threshold was $P < .05$; no multiplicity adjustment was made. Because this was a small single-arm feasibility study, inferential results are treated as hypothesis-generating rather than evidence of efficacy [31].

Results

Participants' Characteristics

Ten older adults participated in this study. Mean age was 73.40 ± 3.17 years (range 67–78); 90% ($n = 9$) were women; 50% ($n = 5$) were university graduates. Eighty percent of participants ($n = 8$) reported their income status as “income equaled expenses,” and 50% ($n = 5$) had two chronic diseases (Table 3). By design, all participants lived alone (an inclusion criterion), and individuals with known cognitive impairment had been excluded at enrollment. Income was recorded relative to expenses; a categorization commonly used in Turkish sociodemographic reporting.

Table 3

Sociodemographic and clinical characteristics of participants aged 65 and over living alone who applied to the Geriatric Outpatient Clinic of Ege University, Izmir, between April and October 2025 (n = 10)

Variables	Categories	n (%)
Age, years	Mean $\pm$ SD (range)	73.40 $\pm$ 3.17 (67–78)
Sex	Female	9 (90)
	Male	1 (10)
Education level	University graduate	5 (50)
	High school graduate	2 (20)
	Secondary school graduate	1 (10)
	Elementary school graduate	1 (10)
	Literate	1 (10)
Economic status	Income equaled expenses	8 (80)
	Income exceeded expenses	1 (10)
	Expenses exceeded income	1 (10)
Number of chronic diseases	1	2 (20)
	2	5 (50)
	3	3 (30)
SD, Standard deviation. Values are n (%) unless otherwise stated.		

Feasibility and Engagement

All 10 participants completed the study (10/10, 100% retention). The final study summary recorded 300 completed sessions among 300 possible participant-days (100% adherence), with a mean duration of 18.6 ± 3.2 minutes. Two participants (2/10, 20%) each had one remote troubleshooting contact during the first week; 8 (8/10, 80%) had none. The support record did not distinguish whether the contacts concerned login, connectivity, audio routing, or another problem.

The analysis dataset did not contain preintervention conversation volume, out-of-app social-contact measures, reasons for missed sessions, or a verified participant-level distribution of in-app conversational turns. Consequently, participant-level skew in conversational volume and change from before to after the intervention could not be evaluated. No adverse events, unintended effects, or safety incidents were reported by participants or recorded in the study documentation during the 30-day period. Adverse events were not, however, elicited with a structured instrument, and no systematic audit of model-generated content was performed (see Limitations).

Exploratory Psychosocial Outcomes

Mean DJGLS scores decreased from 8.50 ± 5.28 at baseline to 5.70 ± 4.92 at endline ($P=.007$; Cohen $d_z = 0.98$). Nine participants had lower endline scores and 1 had no change. Mean GDS-15 scores decreased from 4.50 ± 4.22 to 2.30 ± 2.26 ; ($P=.039$; Cohen $d_z = 0.66$). Five participants had lower GDS-15 scores and 5 had no change; no participant's score increased. The number with a GDS-15 score of 6 or higher decreased from 3 (3/10, 30%) to 1 (1/10, 10%). Supplementary Table 2 reports summary values consistent with the participant-level values in Supplementary Table 3.

Cognitive screening outputs

SMMSE remained stable overall (29.3 ± 1.1 at baseline; 29.6 ± 0.7 at endline). Two participants improved (28 to 30 and 27 to 29) and one decreased by one point (30 to 29). All participants remained within the normal range across both time points. In-app SPMSQ was administered 300 times across the month; all results remained in the normal range (0–2 errors). SPMSQ category outputs were concordant with SMMSE categories, and no acute cognitive declines were detected (Supplementary Table 3).

Discussion

In this 30-day single-arm pilot, all 10 selected participants completed endline assessment, and the final study summary recorded completion of all scheduled sessions with little recorded technical support. Loneliness and depressive symptom scores were lower at endline, while face-to-face SMMSE and in-app SPMSQ results remained in the conventional normal range. These results support operational feasibility in a small group of cognitively healthy, smartphone-capable older adults living alone. They do not establish that iShe caused psychosocial improvement, validate daily SPMSQ as a monitoring instrument, or demonstrate safety for unsupervised clinical use. These operational findings are especially important because feasibility and usability barriers are common in older adult technology interventions [43–45], and they mirror signals from recent AI companion pilots for loneliness reduction [21–23].

Within-group improvements in loneliness (DJGLS) and depressive symptoms (GDS-15) were observed over one month, with large-to-moderate exploratory effect sizes. These findings should not be interpreted as causal because the study lacked a control group and participants may have experienced novelty effects and other non-specific influences. Nonetheless, the direction of change is consistent with emerging work showing that conversational technologies, including voice assistants and conversational agents, can be associated with reduced loneliness and improved social support in older adults living alone [18–22]. Baseline GDS-15 scores were low on average, leaving limited room for improvement, and future work should assess whether longer engagement or tailored content can yield more durable mood effects.

In a study conducted by De Leo et al., it was reported that a telephone-based intervention program reduced suicide rates and regressed depressive symptoms over a 10-year period [46]. The iShe application adapts the traditional telephone-support approach to a digital environment using an

LLM-based application with advanced natural language processing capabilities, thereby offering a more dynamic, personalizable, and around-the-clock form of social interaction.

Potential mechanisms underlying the observed improvement include the structured routine provided by daily regular interaction, the presence of an empathetic and attentive “listener,” increased cognitive stimulation, and greater opportunities for social interaction. Especially for individuals living alone, having someone to talk to at a specific time of day can partially fill the social void. In addition, the application’s ability to remember previous conversations (via a vector-based memory system) may have made interactions more personal and meaningful. These interpretations are hypothesis-generating rather than causal inferences due to the study design.

Smith et al. have shown that digital technologies (e.g., virtual reality, robots, and smartphone applications) can yield beneficial effects on the cognitive and psychological outcomes in older adults [47]. The findings of this study suggest that LLM-based applications could also be considered a potential option among such digital technologies. Notably, speech-based interaction requires neither visual acuity nor fine motor skills, which is an important accessibility advantage for older users.

This study demonstrated the feasibility of embedding a validated cognitive screening tool (SPMSQ) into daily conversational routines. The key contribution is the high-frequency, low-burden collection of screening data, which could support earlier detection in longer, controlled studies and generate rich longitudinal data for concordance analyses with clinician-administered tools.

Another important finding of this study is that in-app SPMSQ assessments yielded consistent results with face-to-face SMMSE results. This finding supports the validity of cognitive screening conducted via digital platforms and aligns with similar studies in the literature [28, 29]. The literature indicates that digital neuropsychological assessments increase standardization, reduce administrator bias, and facilitate access to a wider population [28]. This study demonstrates that, in addition to these advantages, cognitive assessment can be naturally integrated into daily life routines.

The lack of significant change in SPMSQ scores during a one-month follow-up period is an expected finding, as changes in cognitive function typically emerge over months to years [48]. However, even within this short timeframe, the ability of the application to consistently perform cognitive screening demonstrates significant potential for long-term monitoring. Frequent and repeated assessments are critical for the early detection of slowly progressive conditions such as MCI [49]. Compared to the mobile cognitive assessment application developed by Zorluoglu et al. [30], iShe provides not only cognitive checks but also social interaction. This dual-pronged approach increases user engagement, offering a more sustainable monitoring model.

In this study, the fact that all participants used the application regularly demonstrates the success of the user-friendly interface design. Challenges in technology use among older adults are among the most important barriers for such projects [43, 44]. Contemporary reviews emphasize that much of the evidence base for voice assistants in loneliness reduction is still early-stage and often centered on

usability and acceptability, which underscores the value of feasibility-oriented pilot evidence [18, 23]. Challenges in technology use among older adults are among the most important barriers for such projects [43, 44]. In this cohort, high adherence (median 30 of 30 sessions) with minimal support needs suggests that even first-time users can sustain daily use when the interface is voice-first and support is available early. As Czaja et al. point out, interfaces designed to meet the specific needs of older adults significantly increase technology acceptance [45]. A future study should measure usability directly and report the nature and duration of every support event.

The use of cross-platform technologies such as Expo and React Native has ensured a fast and consistent experience for both iOS and Android users. Furthermore, Over-the-air updates have enabled improvements based on user feedback to be implemented immediately. This flexibility is particularly critical in pilot studies. The use of WebRTC technology has enabled real-time, low-latency voice communication, which has enhanced the naturalness of interaction. Optimizing the GPT-4o realtime model for Turkish and the high accuracy of the Whisper model's speech recognition likely contributed to user satisfaction. In terms of data security and user privacy, not using image recording and storing data in a secure infrastructure such as Supabase has played an important role in gaining the trust of participants [50, 51]. This approach is of critical importance, especially in the context of increasing legal regulations regarding the protection of health data.

The most clinically significant potential of the iShe application is its ability to be integrated into primary care services and geriatric monitoring programs. Especially for older adults who cannot undergo regular face-to-face assessments due to geographic distance, transportation difficulties, or healthcare system capacity, such digital solutions offer a valuable alternative.

The application's ability to perform daily cognitive screening could establish a cost-effective strategy for the early detection of high-risk individuals and their referral to necessary healthcare services. As Geldmacher et al. point out, even delaying the progression of dementia by one year improves quality of life and significantly reduces the socioeconomic burden [52]. In addition, data collected by the application (such as speaking duration, word variety, and response speed) can be analyzed using machine learning algorithms in the future to detect earlier signs of cognitive decline. Such linguistic markers are known to have potential in predicting dementia risk [53, 54].

This study has several key strengths. First, to our knowledge, this study is one of the first to evaluate an LLM-based application for loneliness, depression, and cognitive monitoring in the Turkish elderly population. Second, high participant compliance and minimal technical support requirements support the application's feasibility in real-world settings. Finally, the application's integrated approach targeting both psychosocial well-being and cognitive health is a significant strength from a holistic elderly care perspective.

Limitations

This pilot study has some limitations. Firstly, the small sample size ($n = 10$) and single-center design limit the generalizability of the findings. Secondly, the one-month follow-up period is insufficient to assess long-term changes in cognitive function. Studies lasting at least 6–12 months are needed to determine long-term effects. Standardized usability questionnaires (for example, SUS or MAUQ) and structured qualitative interviews were not collected, limiting acceptability granularity. Because the same ten SPMSQ items were administered repeatedly within each daily session over 30 days, participants may have learned or memorized correct responses over time, which could artificially stabilize or inflate in-app SPMSQ performance independent of true cognitive status; future work should rotate or vary item content, or benchmark against unpractised probes, to disentangle genuine cognitive stability from a practice effect. Thirdly, the absence of a control group in the study makes it difficult to attribute the observed improvements solely to the intervention. Fourthly, the fact that the majority of participants (50%) were university graduates with a middle socioeconomic status limits the generalizability of the findings to older adults with lower education and socioeconomic levels. Fifthly, the fact that 90% of participants were women creates a limitation in terms of gender balance. Sixthly, repository history shows changes in prompt flow, model version, and audio handling during the study window. Raw session logs needed to verify the reported 300/300 adherence value and audio-retention status were not supplied with the manuscript files. No formal audit of hallucinations, harmful outputs, safety events, or sensitive-data leakage was conducted. Finally, due to the study design, it is not possible to distinguish which specific components of the application (maintaining a daily routine, empathy-based interaction, cognitive stimulation, etc.) contributed to the therapeutic effects.

Conclusions

The iShe LLM-based assistant has demonstrated the feasibility and preliminary signals of benefit for psychosocial well-being while enabling continuous at-home cognitive screening in older adults living alone. A definitive evaluation should use a randomized controlled design with an active comparator, longer follow-up (6–12 months), blinded outcome assessment where feasible, and inclusion of standardized usability measures and richer qualitative evaluation. Future studies should also adopt AI-specific reporting guidance for clinical evaluation of AI interventions and document model versioning, input-output handling, and error analysis to improve transparency. Comparing iShe with alternative social supports (for example, human-led check-in calls or different AI modalities) will help isolate specific benefits, and integration with clinical workflows (alerts for cognitive change) warrants exploration. Larger, longer-term randomized trials are necessary to validate effects, assess cost-effectiveness, and explore integration with primary care.

Abbreviations

AI	Artificial intelligence
DJGLS	De Jong-Gierveld Loneliness Scale

EN	English
GDS-15	Geriatric Depression Scale-15
iShe	Intelligent Social Health and Interaction
LLM	Large Language Model
MCI	Mild cognitive impairment
SPMSQ	Short Portable Mental Status Questionnaire
SMMSE	Standardized Mini-Mental State Examination
TR	Turkish

Declarations

Acknowledgments

This study was supported by The Scientific and Technological Research Council of Türkiye (TÜBİTAK), 1002-A Rapid Support Program, Project No. 324S300. We thank all participants for their time and commitment.

Authors' Contributions

AK, ÖA, CBB, FA, and SB generated the study idea and designed the protocol. AK, ÖA and PSB contributed to data collection and/or analysis of study outcomes. AK, ÖA, CBB, FA, PSB and SB drafted the manuscript. AK, ÖA, SB, FA, SB, and CBB reviewed the manuscript providing substantial academic and/or clinical input. All authors approved the final manuscript. AK is the guarantor of the work and as such had full access to all the data and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

This research was funded by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1002-A Rapid Support Program grant number 324S300.

Data Availability

De-identified data are available from the corresponding author upon reasonable request and subject to ethics approval.

Ethics approval and consent to participate

The study was approved by the Ege University Faculty of Medicine Clinical Research Ethics Committee (Date: 14 November 2024; Decision 24-11T/7). All procedures carried out in this study were in accordance with the ethical standards of the institution's research committee and the 1964 Helsinki Declaration and its subsequent amendments. During the data collection process, the participants were informed about the study, and their written consent was obtained.

Consent for publication

Not applicable.

Competing interests

None declared.

Generative Artificial Intelligence Use Disclosure

During the preparation of this manuscript, the authors used generative artificial intelligence tools to assist with language editing, restructuring of text to meet journal formatting requirements, and drafting of selected passages. All AI-assisted text was reviewed, verified, and edited by the authors, who take full responsibility for the content of the publication. Generative AI was not used to generate or analyse study data, to produce statistical results, or to create or alter figures reporting study findings. The intervention evaluated in this study (iShe) is itself a large language model-based system; its configuration and use are described in the Methods.

Clinical trial number: not applicable

References

1. WHO (World Health Organization). *World Report on Ageing and Health*. 2025. Accessed August 21, 2025. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>
2. World Health Organization (WHO). *Global Report on Ageism*. 2021. Accessed August 23, 2025. <https://www.who.int/publications/i/item/9789240016866>
3. Turkish Statistical Institute. *Elderly Statistics, 2024*. 2025. Accessed August 10, 2025. <https://data.tuik.gov.tr/Bulten/Index?p=Elderly-Statistics-2024-54079&dil=2>
4. Stewart L. Gerontology: Will you still need me, will you still feed me? *Nature*. 2014;514(7522):S14-S15. doi:10.1038/514S14a
5. Victor CR, Scambler SJ, Bowling A, Bond J. The prevalence of, and risk factors for, loneliness in later life: a survey of older people in Great Britain. *Ageing Soc*. 2005;25(6):357-375. doi:10.1017/S0144686X04003332
6. Coyle CE, Dugan E. Social Isolation, Loneliness and Health Among Older Adults. *J Aging Health*. 2012;24(8):1346-1363. doi:10.1177/0898264312460275

7. Holt-Lunstad J, Smith TB. Loneliness and social isolation as risk factors for CVD: implications for evidence-based patient care and scientific inquiry. *Heart*. 2016;102(13):987-989. doi:10.1136/heartjnl-2015-309242
8. Corrada MM, Brookmeyer R, Paganini-Hill A, Berlau D, Kawas CH. Dementia incidence continues to increase with age in the oldest old: The 90+ study. *Ann Neurol*. 2010;67(1):114-121. doi:10.1002/ana.21915
9. Morley JE. An Overview of Cognitive Impairment. *Clin Geriatr Med*. 2018;34(4):505-513. doi:10.1016/j.cger.2018.06.003
10. Kuşoğlu L, Oktay C, Yiğit Ö. Acil Servise Başvuran 50 yaş ve üzeri hastalarda bilişsel fonksiyon bozukluğu. *Akdeniz Medical Journal*. 2021;7(2):258-267. doi:10.53394/akd.958952
11. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild Cognitive Impairment. *Arch Neurol*. 1999;56(3):303. doi:10.1001/archneur.56.3.303
12. Miller JB, Barr WB. The Technology Crisis in Neuropsychology. *Archives of Clinical Neuropsychology*. 2017;32(5):541-554. doi:10.1093/arclin/acx050
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
14. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung EW, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. doi:10.1038/s41586-023-06291-2
15. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940. doi:10.1038/s41591-023-02448-8
16. Sartori G, Orrù G. Language models and psychological sciences. *Front Psychol*. 2023;14. doi:10.3389/fpsyg.2023.1279317
17. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *New England Journal of Medicine*. 2023;388(13):1233-1239. doi:10.1056/NEJMs2214184
18. Marziali RA, Franceschetti C, Dinculescu A, Nistorescu A, Kristály DM, Moşoi AA, et al. Reducing Loneliness and Social Isolation of Older Adults Through Voice Assistants: Literature Review and Bibliometric Analysis. *J Med Internet Res*. 2024;26:e50534. doi:10.2196/50534
19. Jones VK, Yan C, Shade MY, Boron JB, Yan Z, Heselton HJ, et al. Reducing Loneliness and Improving Social Support among Older Adults through Different Modalities of Personal Voice Assistants. *Geriatrics*. 2024;9(2):22. doi:10.3390/geriatrics9020022
20. Valtolina S, Hu L. Charlie: A chatbot to improve the elderly quality of life and to make them more active to fight their sense of loneliness. In: *CHIItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter*. ACM; 2021:1-5. doi:10.1145/3464385.3464726
21. Chen G, Spaulding R. The Impact of AI-Powered Alexa Assistant on Loneliness in Homebound Older Adults: A Pilot Study. *Adv Geriatr Med Res*. 2025;7(1). doi:10.20900/agmr20250003
22. Chou YH, Lin C, Lee SH, Lee YF, Cheng LC. User-Friendly Chatbot to Mitigate the Psychological Stress of Older Adults During the COVID-19 Pandemic: Development and Usability Study. *JMIR Form*

23. Yang Y, Wang C, Xiang X, An R. AI Applications to Reduce Loneliness Among Older Adults: A Systematic Review of Effectiveness and Technologies. *Healthcare*. 2025;13(5):446. doi:10.3390/healthcare13050446
24. Murayama H, Takase M. Evaluating the Effectiveness of Digital Social Robots in Reducing Loneliness Among Community-Dwelling Older Adults in Japan: Randomized Controlled Trial and Qualitative Analysis. *JMIR Aging*. 2025;8:e74422-e74422. doi:10.2196/74422
25. Wang Y, Tang H, Zhang J, Wang Y, Liu X. Aging with AI companionship: the role of artificial intelligence in enhancing the mental wellbeing of older adults. *Front Public Health*. 2026;14. doi:10.3389/fpubh.2026.1725763
26. Vinay R, Uetova E, Tommila NC, Biller-Andorno N, Kowatsch T. A Hybrid Rule- and Large Language Model–Based Embodied Voice Assistant (GRACE) for Cognitive Stimulation in Older Adults: Usability Study Assessing Technical Feasibility, Technology Acceptance, and Working Alliance. *JMIR Aging*. 2025;8:e76489. doi:10.2196/76489
27. Dikmeer A, Burkuk S, Unsal P, Polat M, Balci C, Halil M, et al. A concise instrument for screening cognitive impairment: validation in a geriatric population. *Aging Clin Exp Res*. 2025;37(1):196. doi:10.1007/s40520-025-03056-x
28. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *Clin Neuropsychol*. 2019;33(2):271-286. doi:10.1080/13854046.2018.1535662
29. Zygouris S, Tsolaki M. Computerized Cognitive Testing for Older Adults. *Am J Alzheimers Dis Other Dement*. 2015;30(1):13-28. doi:10.1177/1533317514522852
30. Zorluoglu G, Kamasak ME, Tavacioglu L, Ozanar PO. A mobile application for cognitive screening of dementia. *Comput Methods Programs Biomed*. 2015;118(2):252-262. doi:10.1016/j.cmpb.2014.11.004
31. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res*. 2011;45(5):626-629. doi:10.1016/j.jpsychires.2010.10.008
32. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ*. Published online October 24, 2016:i5239. doi:10.1136/bmj.i5239
33. Eysenbach G. CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions. *J Med Internet Res*. 2011;13(4):e126. doi:10.2196/jmir.1923
34. de Jong Gierveld J, van Tilburg T. Living arrangements of older adults in the Netherlands and Italy: Coresidence values and behaviour and their consequences for loneliness. *J Cross Cult Gerontol*. 1999;14(1):1-24. doi:10.1023/A:1006600825693
35. de Jong-Gierveld J, Kamphuls F. The Development of a Rasch-Type Loneliness Scale. *Appl Psychol Meas*. 1985;9(3):289-299. doi:10.1177/014662168500900307

36. Akgül H, Yeşilyaprak B. Yaşlılar İçin Yalnızlık Ölçeği'nin Türk Kültürüne Uyarlaması: Geçerlilik ve Güvenirlik Çalışması. *Yaşlı Sorunları Araştırma Dergisi*. 2015;8(1):34-45.
37. Yesavage JA, Sheikh JI. 9/Geriatric Depression Scale (GDS). *Clin Gerontol*. 1986;5(1-2):165-173. doi:10.1300/J018v05n01_09
38. Durmaz B. Validity and Reliability of Geriatric Depression Scale - 15 (Short Form) in Turkish older adults. *North Clin Istanb*. Published online 2017. doi:10.14744/nci.2017.85047
39. Güngen C, Ertan T, Eker E, Yaşar R, Engin F. [Reliability and validity of the standardized Mini Mental State Examination in the diagnosis of mild dementia in Turkish population]. *Turk Psikiyatri Derg*. 2002;13(4):273-281. <http://www.ncbi.nlm.nih.gov/pubmed/12794644>
40. William Molloy D, Standish TIM. A Guide to the Standardized Mini-Mental State Examination. *Int Psychogeriatr*. 1997;9:87-94. doi:10.1017/S1041610297004754
41. Pfeiffer E. A Short Portable Mental Status Questionnaire for the Assessment of Organic Brain Deficit in Elderly Patients†. *J Am Geriatr Soc*. 1975;23(10):433-441. doi:10.1111/j.1532-5415.1975.tb00927.x
42. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198. doi:[https://doi.org/10.1016/0022-3956\(75\)90026-6](https://doi.org/10.1016/0022-3956(75)90026-6)
43. Fisk AD, Czaja SJ, Rogers WA, Charness N, Czaja SJ, Sharit J. *Designing for Older Adults: Principles and Creative Human Factors Approaches*. Second edition. CRC Press; 2009. doi:10.1201/9781420080681
44. Vaportzis E, Giatsi Clausen M, Gow AJ. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. *Front Psychol*. 2017;8. doi:10.3389/fpsyg.2017.01687
45. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (create). *Psychol Aging*. 2006;21(2):333-352. doi:10.1037/0882-7974.21.2.333
46. De Leo D, Buono M Dello, Dwyer J. Suicide among the elderly: The long-term impact of a telephone support and assessment intervention in northern Italy. *British Journal of Psychiatry*. 2002;181(3):226-229. doi:10.1192/bjp.181.3.226
47. Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. A Cognitive Training Program Based on Principles of Brain Plasticity: Results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) Study. *J Am Geriatr Soc*. 2009;57(4):594-603. doi:10.1111/j.1532-5415.2008.02167.x
48. Harada CN, Natelson Love MC, Triebel KL. Normal Cognitive Aging. *Clin Geriatr Med*. 2013;29(4):737-752. doi:10.1016/j.cger.2013.07.002
49. Bruscoli M, Lovestone S. Is MCI really just early dementia? A systematic review of conversion studies. *Int Psychogeriatr*. 2004;16(2):129-140. doi:10.1017/S1041610204000092

50. Rumbold JMM, Pierscioneck B. The Effect of the General Data Protection Regulation on Medical Research. *J Med Internet Res*. 2017;19(2):e47. doi:10.2196/jmir.7108

51. Gajanayake Randike, Iannella Renato, Sahama Tony. Consumer Acceptance of Accountable-eHealth Systems. *Stud Health Technol Inform*. 2014;205:980-984. doi:10.3233/978-1-61499-432-9-980

52. Geldmacher DS, Kirson NY, Birnbaum HG, Eapen S, Kantor E, Cummings AK, et al. Implications of early treatment among Medicaid patients with Alzheimer’s disease. *Alzheimer’s & Dementia*. 2014;10(2):214-224. doi:10.1016/j.jalz.2013.01.015

53. Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer’s Disease in Narrative Speech. *Journal of Alzheimer’s Disease*. 2015;49(2):407-422. doi:10.3233/JAD-150520

54. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer’s disease. *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*. 2015;1(1):112-124. doi:10.1016/j.dadm.2014.11.012

55. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351-1363. doi:10.1038/s41591-020-1037-7

56. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ*. Published online September 9, 2020:m3164. doi:10.1136/bmj.m3164

Figures

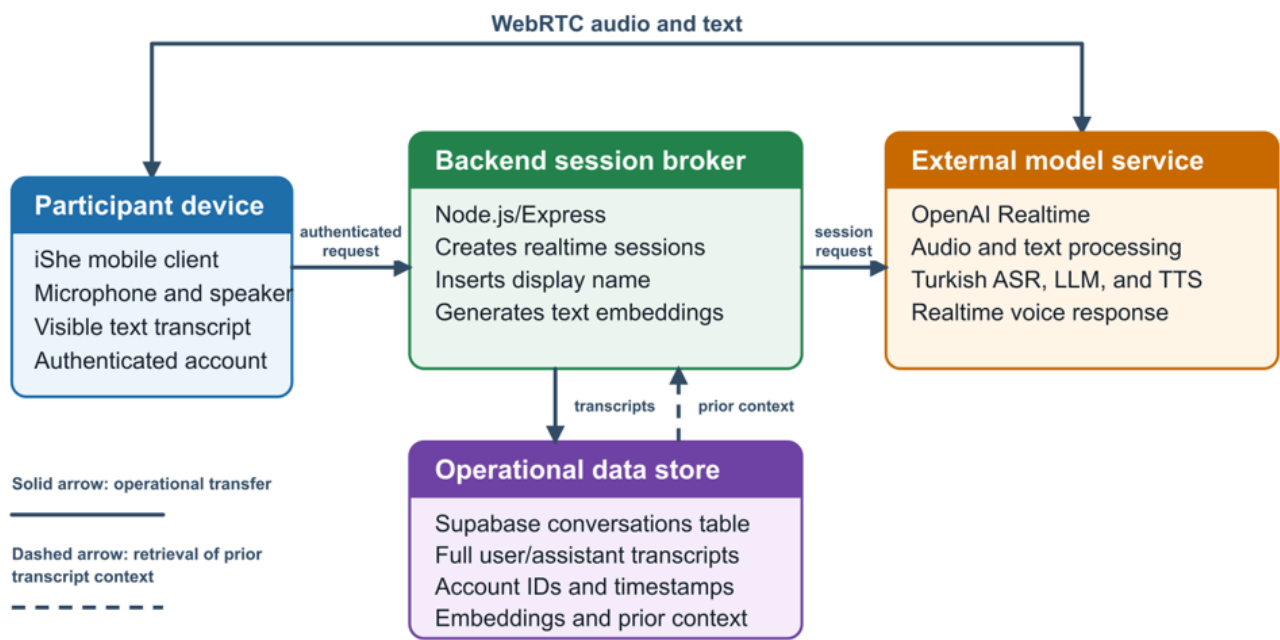


Figure 1

Conceptual data flow of the iShe pilot. Solid arrows indicate operational data transfer; the dashed arrow indicates retrieval of prior transcript context. ASR: automatic speech recognition; LLM: large language model; TTS: text-to-speech

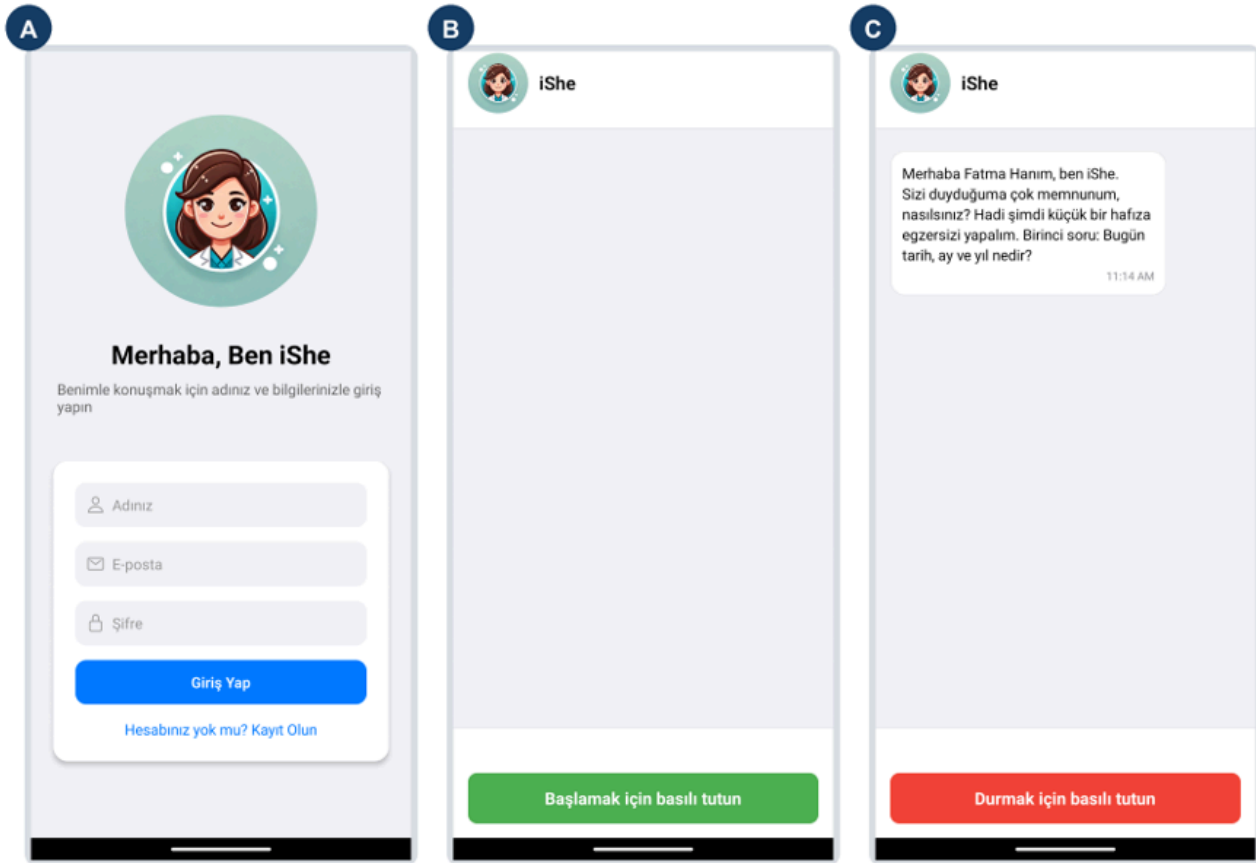


Figure 2

Native-device screenshots of the iShe interfaces. (A) Login screen. (B) Voice-session start control. (C) Assistant greeting and first SPMSQ item in a demonstration session. The displayed name is a pseudonym; no participant data are shown. English translation of panel C: “Hello Ms Fatma, I am iShe. I am very happy to hear you; how are you? Now let us do a short memory exercise. First question: What is today’s date, month, and year?”. Screens are shown in Turkish, the deployment language of the pilot conducted among aged 65 years and older adults living alone (Izmir, Türkiye; April–October 2025).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)

- [SupplementaryTable2.docx](#)
- [SupplementaryTable3.docx](#)