

Original Paper

# A Conversational Platform (Okaya) for Multimodal Digital Biomarkers of Fatigue, Cognition, and Mental Health: Feasibility Observational Study

Matthew So<sup>1</sup>, BS, MS; Michael Sobolev<sup>2</sup>, PhD; Gregory Menvielle<sup>1</sup>, AA, BA, MA

<sup>1</sup>Okaya, Boulder, CO, United States

<sup>2</sup>Cornell Tech, New York, NY, United States

**Corresponding Author:**

Matthew So, BS, MS

Okaya

540 Mohawk Drive

Boulder, CO 80303

United States

Phone: 1 9259841028

Email: [matthew@okaya.me](mailto:matthew@okaya.me)

## Abstract

**Background:** Collection of multimodal data (video, audio, and text) can yield digital biomarkers relevant to mental health, fatigue, and cognition. However, the feasibility and signal characteristics in operational populations remain underexplored.

**Objective:** The objectives of this study were to (1) extract an evidence-based library of vision, speech, and language features; (2) assess the feasibility of a fully remote conversational platform (Okaya) for collecting analyzable multimodal data; and (3) conduct preliminary signal checks for depression, fatigue, and cognition.

**Methods:** Participants were recruited from the US Air Force and US Space Force. All participants completed the Okaya check-in, which included a voice conversation with a large language model. A total of 66 visual, acoustic, and text features were extracted from each interaction between the participant and the large language model. For validation purposes, the study also collected measures of depression (Patient Health Questionnaire–9), fatigue (Cancer Fatigue Scale), and cognition (trail making test). We evaluated the feasibility of the platform and correlation between the extracted features and the validated assessments.

**Results:** A total of 8 unique participants contributed with 62 sessions over a period from March 6, 2025, to August 6, 2025. The platform was deemed feasible as 6 of the 8 participants opted to complete more than one session, and the 3 participants who provided feedback reported high overall experience and usability. From the data perspective, preliminary correlations produced significant results for multiple potential digital biomarkers, including (1) pitch ( $P=.047$ ), volume SD ( $P=.04$ ), volume slope ( $P=.04$ ), automated readability index complexity ( $P=.047$ ), Flesch-Kincaid complexity ( $P=.04$ ), and Gunning Fog complexity ( $P=.04$ ) for depression; (2) pitch ( $P=.009$ ), volume SD ( $P=.007$ ), volume slope ( $P=.02$ ), average F2 formant frequency ( $P=.03$ ), Gunning Fog complexity ( $P=.049$ ), and eyelid droop ( $P=.047$ ) for fatigue; and (3) shimmer ( $P=.03$ ) for cognition. We also observed how features varied over time among participants with multiple sessions.

**Conclusions:** The conversational and artificial intelligence-enabled platform was feasible among an operational sample as a method to collect multimodal data correlated with depression, fatigue, and cognition. These results align with those for previously discovered digital biomarkers of mental health, fatigue, and cognition and inform the development of personalized models for each user while detecting anomalies in a remote monitoring setting.

*JMIR Form Res* 2026;10:e87054; doi: [10.2196/87054](https://doi.org/10.2196/87054)

**Keywords:** digital biomarkers; multimodal sensing; computer vision; speech acoustics; natural language processing; artificial intelligence; AI

## Introduction

Mental health, fatigue, and cognitive functioning are critical to overall well-being and operational performance, yet their assessment often relies on intermittent self-report or clinician-administered instruments. Such methods, while clinically validated, can be burdensome, subjective, and limited in their ability to capture dynamic changes over time [1-9]. Mobile health technology provides an alternative solution due to the ability of apps and sensors to collect high-fidelity and high-frequency data pertaining to activity, behavior, symptoms, and cognition [10-12]. With increasing demands for continuous and remote monitoring—especially in occupational and high-performance settings—digital biomarkers have emerged as a promising avenue for objective and scalable assessment of psychological and cognitive states [13-16].

Recent advances in computer vision, speech acoustics, and natural language processing have enabled the extraction of behavioral and physiological signals from everyday digital interactions. Multimodal sensing that integrates video, audio, and language data provides an opportunity to quantify indicators of affect, fatigue, and cognitive load with greater ecological validity than traditional laboratory tests [11,17]. Prior research has demonstrated associations between acoustic features (eg, pitch variation [18-23], jitter [18,19,24-26], and spectral harmonicity [25-28]), facial expression dynamics (eg, eye gaze [29-32] and microexpressions [33]), and linguistic markers (eg, sentiment [28,34,35], complexity [28,36], and emotional valence [28,35,37-41]) with symptoms of depression, stress, and cognitive decline. However, most existing work has been conducted in tightly controlled environments, and less is known about the feasibility of collecting such data in naturalistic or operational contexts [42].

Conversational platforms, especially those powered by large language models, introduce new opportunities for naturalistic data capture [43-46]. By engaging users through dialogue, these systems can simultaneously elicit and record multimodal signals—speech, facial behavior, and language—while maintaining a familiar and low-burden interaction format [47-52]. Such interfaces align with the increasing integration of conversational artificial intelligence (AI) into health, wellness, and performance domains, offering a pathway toward longitudinal, user-centered monitoring of mental and cognitive health. Nevertheless, empirical evaluation of these systems' feasibility and usability remains limited, particularly in real-world populations such as operational and occupational groups [53,54]. These groups can also be more susceptible to experiencing mental and cognitive health challenges due to their job functions [55,56].

The objective of this study was to evaluate the feasibility of a fully remote conversational platform, Okaya, for collecting analyzable multimodal data (video, audio, and text) in an operational population from the US Air Force (USAF) and Space Force. Specifically, we sought to determine whether participants would engage with the platform across multiple sessions and whether the system could reliably extract a comprehensive set of visual, acoustic, and linguistic features suitable for future digital biomarker modeling of depression, fatigue, and cognition. Finally, we present a case study for the design of a risk score based on a model of extracted digital biomarkers to illustrate the future potential of the platform.

## Methods

### Study Population

This study was part of an existing Small Business Innovation Research Direct to Phase II contract with the USAF, and a pool of potential participants from the 71st Special Operations Squadron, provided by the USAF, was introduced to the Okaya platform as part of a pilot program. Participation was entirely optional. A total of 8 unique users participated in the study, contributing 62 sessions (median 2.5, range 1-38, IQR 1.5-7.5). Participants provided data during the period from March 6, 2025, to August 6, 2025.

### Ethical Considerations

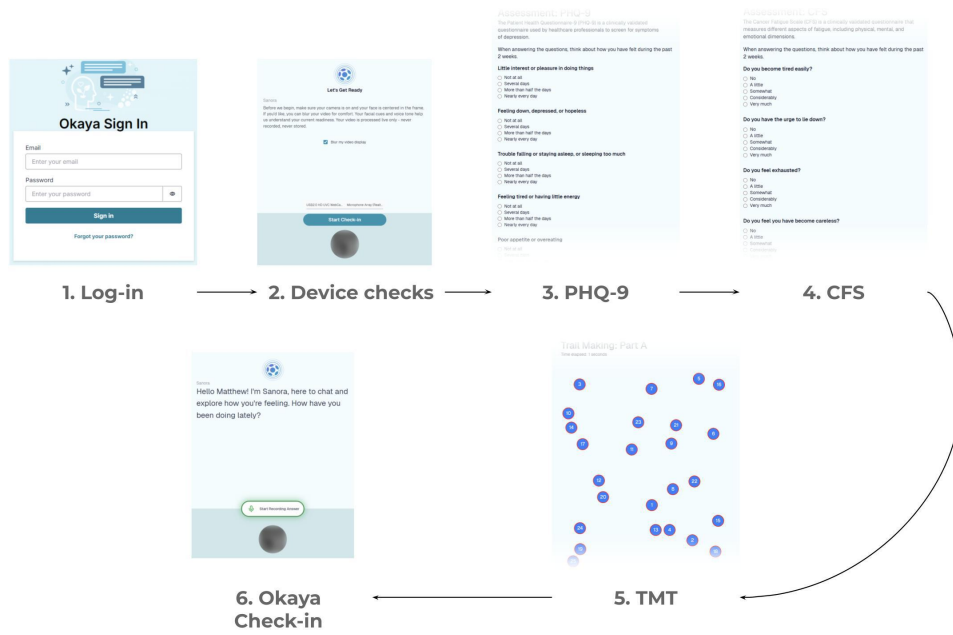
Ethics approval for this study was obtained from Sterling Institutional Review Board (approval 12300-GMenvielle), and all participants signed an informed consent statement that specified anonymized access to research records, potential risks, and voluntary participation. No data were collected except for email addresses, which were required to access the Okaya platform. No demographic or sample composition information was collected due to requirements from the USAF. Participants received no compensation for taking part.

### Procedure and Data Collection

#### Overview

Completion of Okaya check-ins [57] was entirely voluntary. Data collection took place on the Okaya website, and users were allowed to select the device they preferred to access the website with, such as smartphones, tablets, or laptops. The browser used to access the website was also free to select by the user. Once users had successfully acknowledged the consent statement and registered for an Okaya account, they could complete a workflow as described in [Figure 1](#). Participants were asked to complete 3 assessments: the Patient Health Questionnaire-9 (PHQ-9) [58], Cancer Fatigue Scale (CFS) [59], and trail making test (TMT) [60].

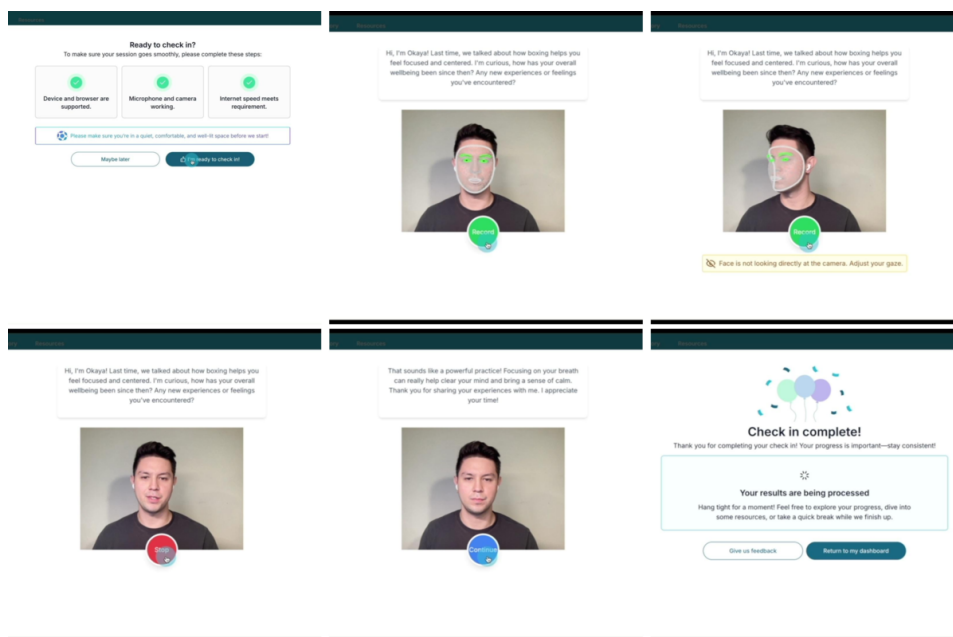
**Figure 1.** Illustration of the Okaya platform and data collection. The 6 steps for a successful sample collection include log-in, device checks, Patient Health Questionnaire–9 (PHQ-9), Cancer Fatigue Scale (CFS), trail making test (TMT), and Okaya check-in.



Upon successful completion of all 3 assessments, users were presented with a text and voice prompt from a large language model (using OpenAI’s GPT-4o-mini model with a custom system prompt, referred to as “Sanora”), shown in Figure 2. Users were first required to meet the device and browser requirements, grant permission to access the microphone and camera, and meet an upload speed requirement of 2 Mb/s. Video was recorded continuously, and users could choose between webcam video, blurred video, a landmark representation of their face, an outline of their face, or a set of axes denoting the location of their face, designed to accommodate users with differing levels of comfort with being recorded.

Although video was shown on the interface, raw video was not sent to our servers; instead, a set of facial landmarks were automatically extracted using a proprietary machine learning model, and 478 landmarks were output, each with an x, y, and z coordinate, as well as 52 “blendshapes,” each representing various facial expressions. The check-in interface, as shown in Figure 2, included the text of Sanora’s prompt at the top; the representation of the user’s webcam video at the center; and a control button, either “Record,” “Stop,” or “Continue,” at the bottom of the screen.

**Figure 2.** Screenshots of the Okaya check-in. Permission was obtained from the user depicted.



After passing device checks and clicking “I’m ready to check-in,” Sanora would generate a prompt for the user to

respond to asking about their general mental well-being, emotions, or experiences they had had recently. Upon

clicking “Record” to record a response to Sanora, the user’s microphone started recording, and an automatic transcription service from Amazon Transcribe started transcribing their speech until the user clicked “Stop” and stopped recording. Both the audio recording and transcribed text were sent back to the Okaya servers. Sanora then used the transcribed text to generate a response. This process continued for 4 turns (5 messages from Sanora and 4 responses from the participant). Upon conversation completion, the user clicked “Continue” to submit the session for analysis, after which a completion screen would indicate that the results were being processed. They would then be redirected to the research dashboard.

All back-end processing took place on Amazon Web Services infrastructure, with communication between the website (front end) and back end occurring via Representational State Transfer application programming interface calls. Data were also stored on Amazon Web Services infrastructure.

## Feature Extraction

From the landmarks, audio, and transcribed text, we extracted 66 features based on the literature. A full list and descriptions of each implemented feature can be found in [Multimedia Appendix 1](#), whereas summaries are shown in [Tables 1-3](#).

**Table 1.** Mapping features for depression.

Feature	Theoretical reasoning	References	Implemented features
<b>Visual features</b>			
Flat affect	A predominant lack of facial expressiveness can be observed in some individuals with depression. This might manifest as a “blank” or “emotionless” appearance even in situations that would typically provoke an emotional response.	[33]	Affect_measure
Facial expression	“Sad” facial expressions, such as less frequent smiling, have been associated with the severity of depression.	[31,40,61-63]	Mouth_curvature and eyebrow_droop
Reduced eye contact	People with depression might avoid making eye contact, which can be a sign of feelings of worthlessness or guilt.	[29-32]	Gaze_down_dist, gaze_x_dist, and gaze_y_dist
Slow movements	Psychomotor retardation can manifest as slowed facial movements or reactions in those with depression.	[61]	Movement_speech_measure
<b>Audio features</b>			
Monotone speech	Individuals with depression might exhibit a lack of variability in pitch, leading to speech that sounds monotonous.	[18-22,27]	Pitch and pitch_std
Reduced speech volume	Speaking more softly or with less projection than usual can be indicative of depression.	[18,22,64-67]	Vol, vol_std, and ealvi
Reduced vocal prosody	A decrease in the rhythmic and melodic aspect of speech can be indicative of depression.	[18-20,24,25,28,64,68]	Jitter, shimmer, timbre, and formant
Long pauses	Depression might result in more increased pauses between words or sentences, reflecting hesitancy, slowed thinking, or difficulty in organizing thoughts.	[18,21,68,69]	Audio_pauses
<b>Transcript features</b>			
Long pauses	Depression might result in more increased pauses between words or sentences, reflecting hesitancy, slowed thinking, or difficulty in organizing thoughts.	[18,21,68,69]	Transcript_pauses
Slow speech rate	People with depression might speak more slowly, potentially	[22,23,68-71]	Words_per_s

Feature	Theoretical reasoning	References	Implemented features
Reduced responsiveness	reflecting slowed cognitive processing. A person with depression might be less verbally responsive in conversations, potentially taking longer to reply or offering shorter answers.	[21,22,69,70]	Response_latency and transcript_len
Decreased complexity and length of speech	Speech in individuals with depression might be less complex in terms of vocabulary and sentence structure, and individuals might be less talkative overall.	[28,34-37,39,40,67]	Complexity, emotion_keyword_prop, and sentiment_score

**Table 2.** Mapping features for fatigue.

Feature	Theoretical reasoning	References	Implemented features
<b>Visual features</b>			
Droopy eyelids	The muscles around the eyes may begin to sag due to tiredness, causing the eyelids to droop.	[72]	Eyelid_droop
Decreased blink rate	Fatigue can lead to a reduced rate of blinking.	[73-77]	Blinks_per_s and blink_len
Yawning	While yawning is a natural behavior, frequent yawning can be an overt sign of fatigue or drowsiness.	[76,78]	Yawns_per_s and yawn_len
Reduced expressiveness	Fatigue might cause an individual to have fewer facial movements or expressions.	[72]	Mouth_curvature
<b>Audio features</b>			
Decreased volume	A tired individual might speak more softly or with less energy than when they are well rested.	[8,79]	Vol and vol_std
Decreased pitch variability	A fatigued voice might sound more monotonous, with less variation in pitch.	[23,25]	Pitch and pitch_std
Flatter voice profile	Decreased variation and width of the voice's spectral profile can be indicative of fatigue.	[25,79]	Timbre and formant
Increased pauses	There might be more frequent and longer pauses between words or sentences, reflecting slowed cognitive processing or the need to gather thoughts.	[23,79,80]	Audio_pauses
<b>Transcript features</b>			
Increased pauses	There might be more frequent and longer pauses between words or sentences, reflecting slowed cognitive processing or the need to gather thoughts.	[23,79,80]	Transcript_pauses
Slow speech rate	The overall rate of speech might decrease when a person is tired.	[23,80]	Words_per_s
Shortened responses	Fatigued individuals might offer shorter answers or engage less in conversation.	[23]	Transcript_len
Impaired memory recall	Fatigued individuals might struggle to remember certain words, names, or details, leading to more frequent use of filler words such as "um" or "uh."	[80]	Hesitations_per_s

**Table 3.** Mapping features for attention and cognition.

Feature	Theoretical reasoning	References	Implemented features
Visual features			
Decreased blink rate	Blinking might increase when in periods of high attention demand during cognitive tasks.	[81]	Blinks_per_s
Eye movements	Eye movements outside the line of focus might be reduced when in a period of visual attention.	[82]	Eye_movements_per_s
Audio features			
Volume	Speech volume can indicate differences in arousal and fatigue.	[26]	Vol, vol_std, vol_range, and volume_slope
Pitch and pitch variability	Pitch might be less varied during periods of low engagement or emotional expressivity.	[26,83]	Pitch and pitch_std
Voice timbre	Voice might be breathier or more strained when lacking attention.	[26]	Jitter, shimmer, and timbre
Transcript features			
Reduced responsiveness	Response latency has been shown to enable the detection of the Stroop effect (delay in reaction time).	[84]	Response_latency

The webcam capture rate varied depending on a number of factors, including the user's device and browser. As this capture rate was not controllable by the Okaya platform, the analysis had to handle landmark frames captured at nonregular times. Generally, the capture rate was 30 to 60 frames per second, but they were usually not captured at a consistent rate. All sessions were interpolated to a capture rate of 30 frames per second, where each time step was a linear interpolation between the 2 neighboring frames. Features were extracted in Python (Python Software Foundation) using the libraries *SciPy* and *NumPy*.

Similarly, audio capture rate varied per user, so all audio recordings were resampled to a sampling frequency of 22,050 Hz. Features were extracted in Python using the libraries *Librosa*, *NumPy*, *SciPy*, *Parselmouth*, and *Statsmodels*.

Text features were extracted in Python using the libraries *NumPy*, *Textacy*, and *Natural Language Toolkit* (Team NLTK). The exact implementation of each feature extractor is omitted as the details of the Okaya platform are proprietary.

Several temporal aggregations were used to calculate feature values: averaged per second, per frame, per response, or per detected event. The exact aggregation and unit used for each feature are listed in [Multimedia Appendix 1](#).

## Results

### Engagement and Usability Results

Three participants provided qualitative and survey feedback following completion of their Okaya sessions. Full survey responses can be found in [Multimedia Appendix 2](#). Overall, preliminary user satisfaction with the conversational platform was given a rating of between 4 and 5 out of 5 on overall experience and usability among all participants who provided feedback (3/3). These participants described the platform as intuitive and easy to navigate and reported feeling adequately

supported during setup and use. One participant highlighted the responsiveness of the support team and appreciated having direct access to human assistance when needed.

Motivations for engagement included curiosity about the technology and the opportunity for structured self-reflection during check-ins. One participant noted that the conversations prompted personal insight and self-awareness, describing the process as a valuable moment to "intentionally reflect" on life aspects discussed during the session. Common barriers to more frequent engagement included time constraints, insufficient or poorly timed reminders, and occasional concerns about privacy when speaking in shared environments. Notably, all participants who completed the survey expressed overall confidence in data privacy and platform security, each giving this aspect a rating of 5 out of 5.

Suggestions for improvement focused on enhancing convenience and personalization, including text-based reminders with direct log-in links, greater mobile accessibility, and shorter baseline tasks prior to the conversational check-in. Survey participants also expressed interest in receiving personalized summaries or progress visualizations of their responses. All the participants who provided feedback indicated interest in participating in future sessions using the Okaya platform, reinforcing the platform's acceptability and potential for longitudinal engagement within operational populations.

### Correlation Analysis

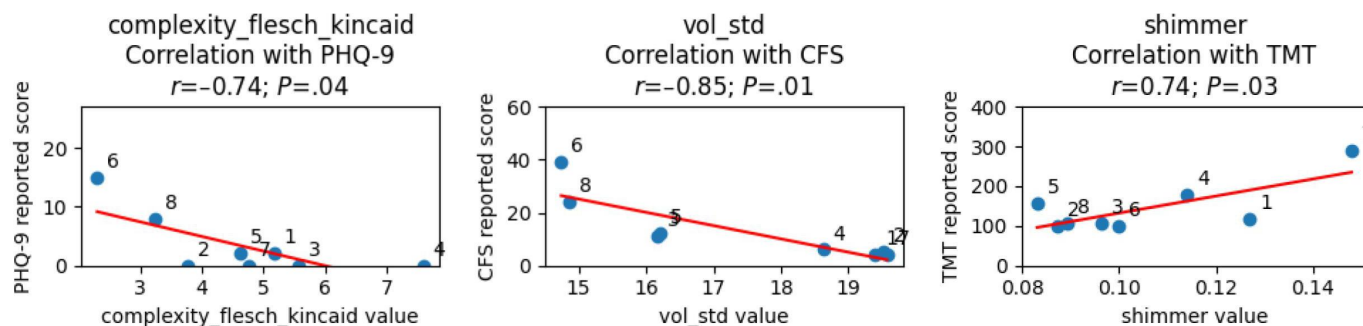
#### Overview

Pearson correlations and their associated significance are listed below. Although every participant was included in the dataset and most completed multiple sessions, we only took into account the earliest sample for each participant for the correlation analysis. The earliest sample was used to illustrate the use of an initial sample as a baseline, upon which personalized models can be built, illustrated in the case

study below. We also observed low intraparticipant variability for each clinical measure, which hindered the application of more sophisticated statistical models at this stage. For each feature, outliers (defined as any measurement more than 2

SDs from the mean) were removed. The most significant features are plotted against the associated measures in Figure 3.

**Figure 3.** Most significant features for each measure and their associated correlation. CFS: Cancer Fatigue Scale; PHQ-9: Patient Health Questionnaire-9; TMT: trail making test.



### Correlations for Depression

Exploratory correlation analyses revealed several significant associations between multimodal features and depressive symptom severity as measured using the PHQ-9. Acoustic and linguistic indicators showed the strongest relationships. In this small sample specifically, pitch ( $r = 0.71$ ;  $P = .047$ ) and volume slope ( $r = 0.78$ ;  $P = .04$ ) were positively correlated with PHQ-9 scores. This suggests preliminary patterns and warrants further exploration into predictive modeling to validate that greater variability and upward trajectories in vocal intensity may correspond to higher self-reported depression levels. In contrast, volume SD was negatively correlated ( $r = -0.73$ ;  $P = .04$ ), encouraging a possible exploration into how participants with greater depressive symptoms may exhibit reduced dynamic range in speech.

Textual complexity metrics—including the automated readability index ( $r = -0.71$ ;  $P = .047$ ), Flesch-Kincaid complexity ( $r = -0.74$ ;  $P = .04$ ), and Gunning Fog index ( $r = -0.72$ ;  $P = .04$ )—also demonstrated negative correlations with PHQ-9 scores; thus, the individuals in our preliminary dataset reporting higher depressive symptoms tended to produce shorter or linguistically simpler utterances during conversational interactions. Due to the low sample size, it is impossible to say whether there are any actual meaningful correlations; however, it does warrant further exploration with larger sample sizes and more resilient analyses. Due to our limited sample size, individual correlations are also not expected to be reliable or generalizable at this stage. Together, these relationships support the feasibility of deriving speech- and language-based digital biomarkers of depression from naturalistic conversational data.

### Correlations for Fatigue

Analysis of correlations between multimodal features and fatigue as measured using the CFS identified several significant acoustic, visual, and linguistic markers. Among acoustic features, pitch ( $r = 0.84$ ;  $P = .009$ ), volume slope ( $r = 0.84$ ;  $P = .02$ ), and average F2 formant frequency ( $r = 0.76$ ;  $P = .03$ ) were positively correlated with fatigue severity, whereas volume SD ( $r = -0.85$ ;  $P = .007$ ) showed a strong

negative association. This warrants further investigation into how individuals reporting higher fatigue might exhibit flatter, less dynamically modulated vocal patterns, consistent with decreased speech energy and prosodic variation observed in fatigue-related speech studies.

Additionally, the linguistic Gunning Fog index was inversely correlated with CFS scores ( $r = -0.71$ ;  $P = .049$ ), indicating that higher fatigue levels were associated with simpler or less complex speech production. One visual feature, eyelid droop ( $r = 0.71$ ;  $P = .047$ ), also demonstrated a significant positive association with fatigue, consistent with prior literature linking ocular and facial muscle changes to tiredness. Together, these findings highlight the sensitivity of both speech acoustics and facial metrics to self-reported fatigue, supporting their potential use as digital biomarkers for real-time fatigue monitoring. However, due to the limited sample size and analysis, individual associations should not be interpreted as predictive or applicable to general populations at this stage.

### Correlations for Attention and Cognition

Exploratory analyses examining associations between multimodal features and cognitive performance as measured using the TMT identified a single significant relationship. The acoustic feature shimmer, which captures cycle-to-cycle variability in vocal amplitude, was positively correlated with TMT scores ( $r = 0.74$ ;  $P = .03$ ). This association points to greater irregularity in voice amplitude when participants had longer task completion times. Further research into how lower cognitive efficiency might potentially relate to fatigue-related motor or attentional variability in speech production may be worthwhile. Although preliminary, this finding aligns with those of prior literature linking changes in vocal stability to fluctuations in attention and executive functioning. These results indicate that fine-grained acoustic features derived from conversational speech may offer a feasible, low-burden proxy for cognitive performance monitoring in remote or operational settings. Once again, although the exact associations found should not be treated as reliable or generalizable at this stage, this does suggest a potential future study with larger samples and analyses.

### Descriptive Summary of Significant Associations

To complement the quantitative analyses, we conducted a descriptive review of the distribution of significant multimodal features across clinical domains. As shown in Table 4, depression-related features were primarily linguistic and acoustic in nature, with no significant visual markers identified. Fatigue exhibited the broadest feature coverage, encompassing 1 visual variable, 4 acoustic variables, and 1 linguistic variable. When examining fatigue subscales, distinct modality patterns emerged: the physical and affective

fatigue components were each associated with multiple acoustic features, whereas the cognitive subscale was linked to both visual and acoustic indicators. In contrast, cognitive performance as measured using the TMT was primarily associated with a single acoustic feature, reflecting a narrower signal profile. It is important to note that feature set size also had an effect; for example, fatigue measures displayed a clear feature representation across audio features, but audio features also made up a larger proportion of the total features compared to visual and text features. There is no clear evidence of any specific feature modality showing broader significance or sensitivity.

**Table 4.** Count of significant features for each subscale of the clinical measures and feature type out of the total number of features tested.

Clinical measure and subscale	Feature categories, n (%)		
	Visual features (n=13)	Audio features (n=28)	Text features (n=25)
PHQ-9 <sup>a</sup>	0 (0.0)	3 (10.7)	3 (12.0)
CFS <sup>b</sup>	1 (7.7)	4 (14.3)	1 (4.0)
Physical	1 (7.7)	4 (14.3)	3 (12.0)
Affective	0 (0.0)	5 (17.9)	0 (0.0)
Cognitive	1 (7.7)	5 (17.9)	0 (0.0)
TMT <sup>c</sup>	0 (0.0)	1 (3.6)	0 (0.0)
Part A	0 (0.0)	0 (0.0)	1 (4.0)
Part B	0 (0.0)	1 (3.6)	0 (0.0)

<sup>a</sup>PHQ-9: Patient Health Questionnaire–9.

<sup>b</sup>CFS: Cancer Fatigue Scale.

<sup>c</sup>TMT: trail making test.

### Risk Score Case Study

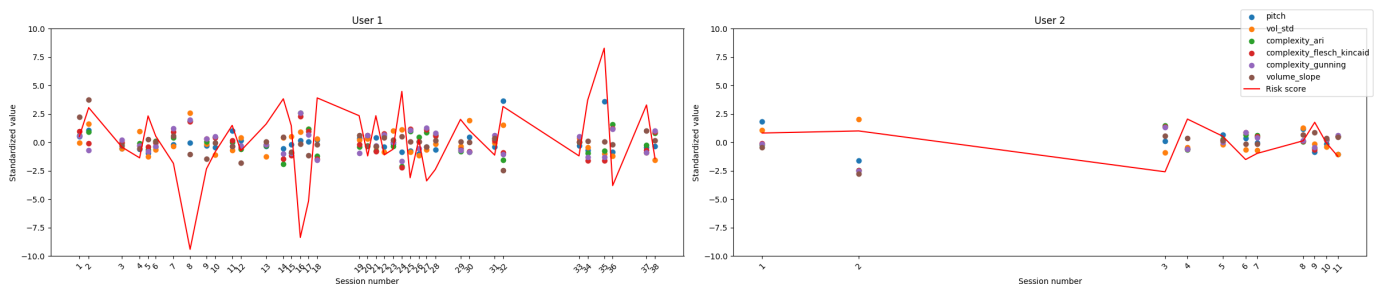
To explore the feasibility of individualized monitoring, we conducted case studies on participants who completed multiple Okaya sessions. For each individual, a composite “risk score” was calculated by standardizing the significant multimodal features ( $\tau$ -scoring within participants) and summing the deviations from the mean, with directionality determined by the feature’s correlation with the clinical measure. This approach provided a demonstration of an interpretable, participant-specific index of deviation from baseline functioning across time; however, due to the low sample size and unstable correlations, it is purely used as a demonstration in this case, not a proof of actual meaningful capability.

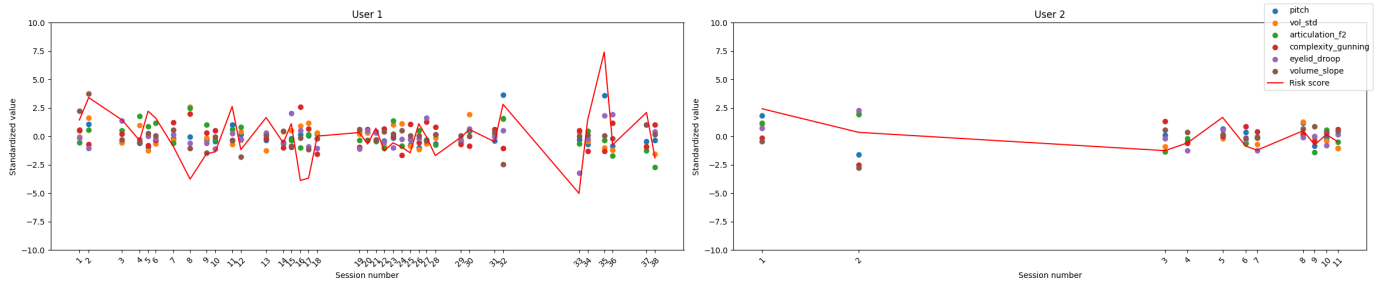
An example use case can use a longitudinal visualization of these risk scores to reveal dynamic fluctuations in both

depression- and fatigue-related indexes across sessions. For example, in participants with higher variability, increases in risk score can correspond to periods of increased self-reported symptom severity, whereas lower or stable values can align with more consistent well-being reports. Although exploratory, these case studies illustrate how multimodal conversational data can support within-person anomaly detection and temporal tracking of psychological and physiological states. This example provides a demonstration of how individualized, feature-based composite scores may offer a viable framework for early identification of changes in mood, fatigue, or cognitive function in remote monitoring contexts.

Data on depression and fatigue for the 2 participants with the most sessions are shown in Figures 4 and 5, respectively (TMT only exhibited 1 significant feature and, as such, a composite score was not meaningful).

**Figure 4.** Risk score case study for depression.



**Figure 5.** Risk score case study for fatigue.

## Discussion

### Principal Findings

We found several multimodal features that correlated with depression, fatigue, and cognition, some of which extended beyond patterns reported in prior literature. As expected, depression was characterized by reduced variability in vocal intensity and lower linguistic complexity, consistent with psychomotor slowing and affective flattening. Fatigue demonstrated a broader multimodal signature involving both acoustic and visual domains, including reduced volume variability, higher pitch, and greater eyelid droop, reflecting decreased energy and arousal. Notably, volume slope, a feature we did not find in the literature, was significant for both depression and fatigue, suggesting that individuals with depression or fatigue might have noticeable patterns in how speech volume progresses over time, especially during utterances provided on the Okaya platform. Additionally, complexity, a feature previously linked to depressive speech, was found to correlate with fatigue, indicating that depression and fatigue may share some similar effects on word choice and diversity [28,36]. Together, these findings highlight the potential for future study of established digital biomarkers in the Okaya platform and the emergence of cross-domain signals that may bridge affective, cognitive, and fatigue-related processes.

The findings of this feasibility study indicate that the Okaya conversational platform is both usable and acceptable for collecting multimodal data in an operational setting. Participants demonstrated relatively sustained engagement, with most completing multiple sessions, suggesting that the platform's conversational format and remote accessibility supported ongoing participation. Usability feedback further underscored this positive experience—participants rated the system highly in terms of ease of use, clarity of instructions, and overall satisfaction. Qualitative feedback revealed that users found the interaction intuitive and meaningful, often describing the check-in as an opportunity for reflection rather than a burdensome task. Importantly, participants expressed high confidence in data privacy and security, which is critical for user trust in AI-enabled health technologies [85-87]. At the same time, feedback highlighted opportunities for improvement, including more flexible scheduling options, reminders, and greater mobile accessibility. Together, these findings suggest that Okaya's design successfully balanced data collection rigor with user comfort and autonomy, supporting its feasibility for repeated deployment in diverse

and high-performance populations. These results provide an encouraging foundation for the platform's next phase of validation and refinement, emphasizing scalability, personalization, and long-term adherence.

The findings from the case study align closely with the long-term vision for the Okaya platform. The observed fluctuations in individualized risk scores and their correspondence with self-reported symptom changes demonstrate the platform's potential for longitudinal, within-person monitoring—a central principle behind the Okaya Index and the simplified risk score we presented. Rather than relying on evidence from population-based studies, the Okaya platform emphasizes personalized baselines and interprets deviations relative to each individual's unique behavioral and physiological profile. This individualized approach reflects the platform's goal of enabling early identification of shifts in emotional, cognitive, or fatigue states. As the platform evolves, integration of these multimodal risk scores into the Okaya Index will enable continuous refinement of trajectories across time and populations, ultimately supporting adaptive, AI-driven insights that can inform both clinical decision-making and support for operational populations ranging from pilots to firefighters. This trajectory aligns with the goal of advancing proactive mental health and performance optimization in real-world settings.

From a clinical and research perspective, these findings illustrate the promise of conversational, multimodal sensing as a foundation for precision mental health and digital phenotyping. By capturing subtle changes in voice, facial behavior, and language over time, systems such as Okaya can complement traditional self-report and clinician-administered tools, providing a richer, continuous view of mental and cognitive functioning. Such individualized, data-driven assessments have potential applications across health care, occupational performance, and behavioral health monitoring, particularly in cases in which real-time insight and early intervention are critical. Importantly, the Okaya platform's emphasis on *nondiagnostic* and *privacy-preserving* analytics aligns with emerging standards for responsible AI in health technology [13,88,89]. As digital biomarkers move closer to clinical integration, continued validation across diverse populations and conditions will be essential to ensure generalizability, interpretability, and equity [85,86,90]. Ultimately, this study of the Okaya platform represents a step toward an ethically grounded, adaptive ecosystem for mental health monitoring—one that transforms multimodal

data into actionable insights while safeguarding user trust and autonomy.

This study has several limitations that should be considered when interpreting the findings. First, the sample size was small and limited due to our focus on operational populations, which are harder to recruit, constraining generalizability and the ability to model interindividual variability. Thus, the correlations we report are not meant to be statistically resilient or generalizable at this stage. Therefore, these findings preclude reliable statistical inference of robust estimates of the correlations in these data and, rather, suggest future avenues for discovery. Second, although validated instruments such as the PHQ-9, CFS, and TMT were used for comparison, these measures may not fully capture the temporal dynamics or subtle within-person fluctuations that multimodal digital biomarkers are designed to detect. Traditional self-report and performance-based assessments tend to be episodic and relatively stable, which may underestimate moment-to-moment variability in affect, fatigue, and cognition. Future studies should incorporate higher-resolution clinical measures—such as ecological momentary assessment or brief daily self-reports—that provide richer temporal data

and enable stronger coupling between behavioral features and self-reported states [1-3,87,91]. Integrating these temporally aligned measures would allow for more precise modeling of intraindividual change and enhance the ecological validity of the derived digital biomarkers. Finally, while the Okaya platform demonstrated feasibility and acceptability in this study, further validation with larger, more diverse samples and extended longitudinal follow-up is necessary for predictive modeling and design of personalized remote monitoring systems.

## Conclusions

The conversational and AI-enabled platform (Okaya) was feasible among an operational sample from the US Air and Space Force as a method for collecting multimodal data correlated with depression, fatigue, and cognition. Future work will examine larger samples with repeated measures to assess the test-retest reliability and predictive validity of multimodal digital biomarkers. Ultimately, the goal of the platform is the development of personalized models for each user while detecting anomalies in a remote monitoring setting.

## Acknowledgments

The authors would like to thank Sara Ghandehari, MD, MCSO, FACHE, as well as Nick Uram, PsyD, for their invaluable contributions toward the development of the Okaya platform. The authors are also grateful to Will Jackson for his assistance in literature review. The authors declare the use of generative artificial intelligence (GenAI) in the research and writing process. According to the Generative Artificial Intelligence Delegation Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: literature search and systematization, code generation, code optimization, process automation, creation of algorithms for data analysis, data analysis, and visualization. The GenAI tool used was GPT-4o via the GitHub Copilot platform. Although the code design and functions were entirely defined by the authors, GitHub Copilot was used to generate autocomplete suggestions as the code was being written. The generated code was used for correlation calculations, risk score case study calculation, and generation of all shown figures. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

This study was funded by SmartTec Inc to fulfill deliverables under a Direct to Phase II award from AFWERX (Department of the Air Force; research and innovation).

## Data Availability

The datasets generated and analyzed during this study are not publicly available due to their use in the development of a proprietary platform as well as confidentiality requirements from the US Department of Defense. Public release of the data could enable reverse engineering of the platform or unauthorized commercial use. However, unidentified data may be available from the corresponding author on reasonable request.

## Authors' Contributions

Conceptualization: M So (lead), GM (equal), M Sobolev (equal)

Data curation: M So (lead), GM (equal)

Formal analysis: M So (lead), M Sobolev (equal)

Funding acquisition: GM

Investigation: M So (lead), M Sobolev (equal)

Methodology: M Sobolev (lead), M So (equal), GM (supporting)

Resources: M So (lead), GM (supporting)

Software: M So

Validation: M So (lead), M Sobolev (supporting)

Visualization: M So

Writing—original draft: M Sobolev (lead), M So (supporting)

Writing—review and editing: M So (lead), M Sobolev (equal), GM (supporting)

---

## Conflicts of Interest

M So and GM are employed by and hold financial interest in SmartTec Inc, who develops the Okaya platform, and could benefit from research outcomes. M Sobolev receives consulting fees from SmartTec Inc.

---

## Multimedia Appendix 1

Feature list.

[[DOCX File \(Microsoft Word File\), 13 KB-Multimedia Appendix 1](#)]

---

## Multimedia Appendix 2

Survey responses.

[[DOCX File \(Microsoft Word File\), 17 KB-Multimedia Appendix 2](#)]

---

## References

1. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. *Assessment*. Jan 2018;25(1):14-30. [doi: [10.1177/1073191116643164](#)] [Medline: [27084835](#)]
2. Katz MJ, Wang C, Nester CO, et al. T-MoCA: a valid phone screen for cognitive impairment in diverse community samples. *Alzheimers Dement (Amst)*. Feb 2021;13(1):e12144. [doi: [10.1002/dad2.12144](#)] [Medline: [33598528](#)]
3. Singh S, Strong R, Xu I, et al. Ecological momentary assessment of cognition in clinical and community samples: reliability and validity study. *J Med Internet Res*. Jun 2, 2023;25:e45028. [doi: [10.2196/45028](#)] [Medline: [37266996](#)]
4. McHugh RK, Behar E. Readability of self-report measures of depression and anxiety. *J Consult Clin Psychol*. Dec 2009;77(6):1100-1112. [doi: [10.1037/a0017124](#)] [Medline: [19968386](#)]
5. McHugh RK, Behar E. Reply to "Further issues in determining the readability of self-report items: comment on McHugh and Behar (2009)". *J Consult Clin Psychol*. Dec 2012;80(6):1121-1122. [doi: [10.1037/a0030455](#)] [Medline: [23205507](#)]
6. Schinka JA. Further issues in determining the readability of self-report items: comment on McHugh and Behar (2009). *J Consult Clin Psychol*. 2012;80(5):952-955. [doi: [10.1037/a0029928](#)] [Medline: [22924333](#)]
7. Levis B, Benedetti A, Ioannidis JP, et al. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *J Clin Epidemiol*. Jun 2020;122:115-128.e1. [doi: [10.1016/j.jclinepi.2020.02.002](#)] [Medline: [32105798](#)]
8. Ma S, Kang L, Guo X, et al. Discrepancies between self-rated depression and observed depression severity: the effects of personality and dysfunctional attitudes. *Gen Hosp Psychiatry*. 2021;70:25-30. [doi: [10.1016/j.genhosppsy.2020.11.016](#)] [Medline: [33689981](#)]
9. Inoue T, Tanaka T, Nakagawa S, et al. Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. *BMC Psychiatry*. Jul 3, 2012;12:73. [doi: [10.1186/1471-244X-12-73](#)] [Medline: [22759625](#)]
10. Sim I. Mobile devices and health. *N Engl J Med*. Sep 5, 2019;381(10):956-968. [doi: [10.1056/NEJMra1806949](#)] [Medline: [31483966](#)]
11. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA*. Oct 3, 2017;318(13):1215-1216. [doi: [10.1001/jama.2017.11295](#)] [Medline: [28973224](#)]
12. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu Rev Clin Psychol*. May 8, 2017;13:23-47. [doi: [10.1146/annurev-clinpsy-032816-044949](#)] [Medline: [28375728](#)]
13. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digit Med*. 2019;2(1):14. [doi: [10.1038/s41746-019-0090-4](#)] [Medline: [30868107](#)]
14. Dagum P. Digital biomarkers of cognitive function. *NPJ Digit Med*. 2018;1:10. [doi: [10.1038/s41746-018-0018-4](#)] [Medline: [31304295](#)]
15. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. *NPJ Digit Med*. 2019;2:3. [doi: [10.1038/s41746-019-0078-0](#)] [Medline: [31304353](#)]
16. Sobolev M, Gullapalli BT, Rahman T. Advancing the science of digital biomarkers. Presented at: DigiBiom '21: Proceedings of the 2021 Workshop on Future of Digital Biomarkers; Jun 25, 2021. [doi: [10.1145/3469266.3473711](#)]
17. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: a survey. *Pervasive Mob Comput*. Dec 2018;51:1-26. [doi: [10.1016/j.pmcj.2018.09.003](#)]
18. König A, Tröger J, Mallick E, et al. Detecting subtle signs of depression with automated speech analysis in a non-clinical sample. *BMC Psychiatry*. Dec 27, 2022;22(1):830. [doi: [10.1186/s12888-022-04475-0](#)] [Medline: [36575442](#)]
19. Silva WJ, Lopes L, Galdino MK, Almeida AA. Voice acoustic parameters as predictors of depression. *J Voice*. Jan 2024;38(1):77-85. [doi: [10.1016/j.jvoice.2021.06.018](#)] [Medline: [34353686](#)]

20. Shin D, Cho WI, Park CHK, et al. Detection of minor and major depression through voice as a biomarker using machine learning. *J Clin Med*. Jul 8, 2021;10(14):3046. [doi: [10.3390/jcm10143046](https://doi.org/10.3390/jcm10143046)] [Medline: [34300212](https://pubmed.ncbi.nlm.nih.gov/34300212/)]
21. Alpert M, Pouget ER, Silva RR. Reflections of depression in acoustic measures of the patient's speech. *J Affect Disord*. Sep 2001;66(1):59-69. [doi: [10.1016/s0165-0327\(00\)00335-9](https://doi.org/10.1016/s0165-0327(00)00335-9)] [Medline: [11532533](https://pubmed.ncbi.nlm.nih.gov/11532533/)]
22. Cohen AS, Rodriguez Z, Opler M, et al. Evaluating speech latencies during structured psychiatric interviews as an automated objective measure of psychomotor slowing. *Psychiatry Res*. Oct 2024;340:116104. [doi: [10.1016/j.psychres.2024.116104](https://doi.org/10.1016/j.psychres.2024.116104)] [Medline: [39137558](https://pubmed.ncbi.nlm.nih.gov/39137558/)]
23. Vogel AP, Fletcher J, Maruff P. Acoustic analysis of the effects of sustained wakefulness on speech. *J Acoust Soc Am*. Dec 2010;128(6):3747-3756. [doi: [10.1121/1.3506349](https://doi.org/10.1121/1.3506349)] [Medline: [21218906](https://pubmed.ncbi.nlm.nih.gov/21218906/)]
24. Ozdas A, Shiavi RG, Silverman SE, Silverman MK, Wilkes DM. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Trans Biomed Eng*. Sep 2004;51(9):1530-1540. [doi: [10.1109/TBME.2004.827544](https://doi.org/10.1109/TBME.2004.827544)] [Medline: [15376501](https://pubmed.ncbi.nlm.nih.gov/15376501/)]
25. Gao X, Ma K, Yang H, et al. A rapid, non-invasive method for fatigue detection based on voice information. *Front Cell Dev Biol*. 2022;10:994001. [doi: [10.3389/fcell.2022.994001](https://doi.org/10.3389/fcell.2022.994001)] [Medline: [36176279](https://pubmed.ncbi.nlm.nih.gov/36176279/)]
26. Hamdan AL, Deeb R, Sibai A, Rameh C, Rifai H, Fayyad J. Vocal characteristics in children with attention deficit hyperactivity disorder. *J Voice*. Mar 2009;23(2):190-194. [doi: [10.1016/j.jvoice.2007.09.004](https://doi.org/10.1016/j.jvoice.2007.09.004)] [Medline: [18082369](https://pubmed.ncbi.nlm.nih.gov/18082369/)]
27. Zhao Q, Fan HZ, Li YL, et al. Vocal acoustic features as potential biomarkers for identifying/diagnosing depression: a cross-sectional study. *Front Psychiatry*. 2022;13:815678. [doi: [10.3389/fpsy.2022.815678](https://doi.org/10.3389/fpsy.2022.815678)] [Medline: [35573349](https://pubmed.ncbi.nlm.nih.gov/35573349/)]
28. Arevian AC, Bone D, Malandrakis N, et al. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS One*. 2020;15(1):e0225695. [doi: [10.1371/journal.pone.0225695](https://doi.org/10.1371/journal.pone.0225695)] [Medline: [31940347](https://pubmed.ncbi.nlm.nih.gov/31940347/)]
29. Hinchliffe MK, Lancashire MH, Roberts FJ. Eye-contact and depression: a preliminary report. *Br J Psychiatry*. Nov 1970;117(540):571-572. [doi: [10.1192/bjp.117.540.571](https://doi.org/10.1192/bjp.117.540.571)] [Medline: [5480708](https://pubmed.ncbi.nlm.nih.gov/5480708/)]
30. Suslow T, Hoepfel D, Kersting A, Bodenschatz CM. Depressive symptoms and visual attention to others' eyes in healthy individuals. *BMC Psychiatry*. Mar 6, 2024;24(1):184. [doi: [10.1186/s12888-024-05633-2](https://doi.org/10.1186/s12888-024-05633-2)] [Medline: [38448877](https://pubmed.ncbi.nlm.nih.gov/38448877/)]
31. Fiquer JT, Moreno RA, Brunoni AR, Barros VB, Fernandes F, Gorenstein C. What is the nonverbal communication of depression? Assessing expressive differences between depressive patients and healthy volunteers during clinical interviews. *J Affect Disord*. Oct 1, 2018;238:636-644. [doi: [10.1016/j.jad.2018.05.071](https://doi.org/10.1016/j.jad.2018.05.071)] [Medline: [29957481](https://pubmed.ncbi.nlm.nih.gov/29957481/)]
32. Waxer P. Nonverbal cues for depression. *J Abnorm Psychol*. Jun 1974;83(3):319-322. [doi: [10.1037/h0036706](https://doi.org/10.1037/h0036706)] [Medline: [4844922](https://pubmed.ncbi.nlm.nih.gov/4844922/)]
33. Buyukdura JS, McClintock SM, Croarkin PE. Psychomotor retardation in depression: biological underpinnings, measurement, and treatment. *Prog Neuropsychopharmacol Biol Psychiatry*. Mar 30, 2011;35(2):395-409. [doi: [10.1016/j.pnpbp.2010.10.019](https://doi.org/10.1016/j.pnpbp.2010.10.019)] [Medline: [21044654](https://pubmed.ncbi.nlm.nih.gov/21044654/)]
34. Hadžić B, Ohse J, Alkostanteini ME, et al. Emotional dynamics in semi-clinical settings: speech emotion recognition in depression-related interviews. Presented at: 11th International Conference on Information and Communication Technologies for Ageing Well and e-Health; Apr 6-8, 2025; Porto, Portugal. [doi: [10.5220/0013415700003938](https://doi.org/10.5220/0013415700003938)]
35. Rottenberg J, Gross JJ, Gotlib IH. Emotion context insensitivity in major depressive disorder. *J Abnorm Psychol*. Nov 2005;114(4):627-639. [doi: [10.1037/0021-843X.114.4.627](https://doi.org/10.1037/0021-843X.114.4.627)] [Medline: [16351385](https://pubmed.ncbi.nlm.nih.gov/16351385/)]
36. Trifu RN, Nemeş B, Herta DC, Bodea-Hategan C, Talaş DA, Coman H. Linguistic markers for major depressive disorder: a cross-sectional study using an automated procedure. *Front Psychol*. 2024;15:1355734. [doi: [10.3389/fpsyg.2024.1355734](https://doi.org/10.3389/fpsyg.2024.1355734)] [Medline: [38510303](https://pubmed.ncbi.nlm.nih.gov/38510303/)]
37. Hartnagel LM, Ebner-Priemer UW, Foo JC, et al. Linguistic style as a digital marker for depression severity: an ambulatory assessment pilot study in patients with depressive disorder undergoing sleep deprivation therapy. *Acta Psychiatr Scand*. Mar 2025;151(3):348-357. [doi: [10.1111/acps.13726](https://doi.org/10.1111/acps.13726)] [Medline: [38987940](https://pubmed.ncbi.nlm.nih.gov/38987940/)]
38. Han J, Feng Y, Li N, et al. Correlation between word frequency and 17 items of Hamilton scale in major depressive disorder. *Front Psychiatry*. 2022;13:902873. [doi: [10.3389/fpsy.2022.902873](https://doi.org/10.3389/fpsy.2022.902873)] [Medline: [35592381](https://pubmed.ncbi.nlm.nih.gov/35592381/)]
39. Shinohara S, Nakamura M, Omiya Y, et al. Depressive mood assessment method based on emotion level derived from voice: comparison of voice features of individuals with major depressive disorders and healthy controls. *Int J Environ Res Public Health*. May 19, 2021;18(10):5435. [doi: [10.3390/ijerph18105435](https://doi.org/10.3390/ijerph18105435)] [Medline: [34069609](https://pubmed.ncbi.nlm.nih.gov/34069609/)]
40. Gehricke JG, Shapiro D. Reduced facial expression and social context in major depression: discrepancies between facial muscle activity and self-reported emotion. *Psychiatry Res*. Aug 21, 2000;95(2):157-167. [doi: [10.1016/s0165-1781\(00\)00168-2](https://doi.org/10.1016/s0165-1781(00)00168-2)] [Medline: [10963801](https://pubmed.ncbi.nlm.nih.gov/10963801/)]
41. Cariola LA, Sheeber LB, Allen N, et al. Language use in depressed and non-depressed mothers and their adolescent offspring. *J Affect Disord*. Dec 1, 2024;366:290-299. [doi: [10.1016/j.jad.2024.08.131](https://doi.org/10.1016/j.jad.2024.08.131)] [Medline: [39187178](https://pubmed.ncbi.nlm.nih.gov/39187178/)]
42. Kouba P, Šmotek M, Tichý T, Kopřivová J. Detection of air traffic controllers' fatigue using voice analysis - an EEG validation study. *Int J Ind Ergon*. May 2023;95:103442. [doi: [10.1016/j.ergon.2023.103442](https://doi.org/10.1016/j.ergon.2023.103442)]

43. Torous J, Topol EJ. Assessing generative artificial intelligence for mental health. *Lancet*. 2025;406(10504). [doi: [10.1016/S0140-6736\(25\)01237-1](https://doi.org/10.1016/S0140-6736(25)01237-1)]
44. Cruciani F, Fritsch SG, Cleland I, Rey VF, Nugent C, Lukowicz P. Few-shot human activity recognition using lightweight language models. Presented at: 2025 International Conference on Activity and Behavior Computing; Apr 21-25, 2025; Al Ain, United Arab Emirates. [doi: [10.1109/ABC64332.2025.11118559](https://doi.org/10.1109/ABC64332.2025.11118559)]
45. Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA*. Jan 28, 2025;333(4):319-328. [doi: [10.1001/jama.2024.21700](https://doi.org/10.1001/jama.2024.21700)] [Medline: [39405325](https://pubmed.ncbi.nlm.nih.gov/39405325/)]
46. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digit Med*. Dec 19, 2023;6(1):236. [doi: [10.1038/s41746-023-00979-5](https://doi.org/10.1038/s41746-023-00979-5)] [Medline: [38114588](https://pubmed.ncbi.nlm.nih.gov/38114588/)]
47. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc*. Sep 1, 2018;25(9):1248-1258. [doi: [10.1093/jamia/ocy072](https://doi.org/10.1093/jamia/ocy072)] [Medline: [30010941](https://pubmed.ncbi.nlm.nih.gov/30010941/)]
48. Rizzo A, Lucas G, Gratch J, et al. Automatic behavior analysis during a clinical interview with a virtual human. *Stud Health Technol Inform*. 2016;220:316-322. [Medline: [27046598](https://pubmed.ncbi.nlm.nih.gov/27046598/)]
49. Cohen J, Richter V, Neumann M, et al. A multimodal dialog approach to mental state characterization in clinically depressed, anxious, and suicidal populations. *Front Psychol*. 2023;14:1135469. [doi: [10.3389/fpsyg.2023.1135469](https://doi.org/10.3389/fpsyg.2023.1135469)] [Medline: [37767217](https://pubmed.ncbi.nlm.nih.gov/37767217/)]
50. Jiang Z, Seyedi S, Griner E, et al. Multimodal mental health digital biomarker analysis from remote interviews using facial, vocal, linguistic, and cardiovascular patterns. *IEEE J Biomed Health Inform*. Mar 2024;28(3):1680-1691. [doi: [10.1109/JBHI.2024.3352075](https://doi.org/10.1109/JBHI.2024.3352075)] [Medline: [38198249](https://pubmed.ncbi.nlm.nih.gov/38198249/)]
51. Milintsevich K, Sirts K, Dias G. Towards automatic text-based estimation of depression through symptom prediction. *Brain Inform*. Feb 13, 2023;10(1):4. [doi: [10.1186/s40708-023-00185-9](https://doi.org/10.1186/s40708-023-00185-9)] [Medline: [36780049](https://pubmed.ncbi.nlm.nih.gov/36780049/)]
52. He Y, Yang L, Qian C, et al. Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. *J Med Internet Res*. Apr 28, 2023;25:e43862. [doi: [10.2196/43862](https://doi.org/10.2196/43862)] [Medline: [37115595](https://pubmed.ncbi.nlm.nih.gov/37115595/)]
53. Ćosić K, Popović S, Wiederhold BK. Enhancing aviation safety through AI-driven mental health management for pilots and air traffic controllers. *Cyberpsychol Behav Soc Netw*. Aug 2024;27(8):588-598. [doi: [10.1089/cyber.2023.0737](https://doi.org/10.1089/cyber.2023.0737)] [Medline: [38916063](https://pubmed.ncbi.nlm.nih.gov/38916063/)]
54. Wiederhold BK. From cyberspace to outer space: how technology will preserve mental health in the final frontier. *Cyberpsychol Behav Soc Netw*. Jan 2025;28(1):1-3. [doi: [10.1089/cyber.2024.0609](https://doi.org/10.1089/cyber.2024.0609)] [Medline: [39823387](https://pubmed.ncbi.nlm.nih.gov/39823387/)]
55. Gatti M, Palumbo R, Di Domenico A, Mammarella N. Affective health and countermeasures in long-duration space exploration. *Heliyon*. May 13, 2022;8(5):e09414. [doi: [10.1016/j.heliyon.2022.e09414](https://doi.org/10.1016/j.heliyon.2022.e09414)] [Medline: [35607498](https://pubmed.ncbi.nlm.nih.gov/35607498/)]
56. Yin Y, Liu J, Fan Q, et al. Long-term spaceflight composite stress induces depression and cognitive impairment in astronauts-insights from neuroplasticity. *Transl Psychiatry*. Nov 8, 2023;13(1):342. [doi: [10.1038/s41398-023-02638-5](https://doi.org/10.1038/s41398-023-02638-5)] [Medline: [37938258](https://pubmed.ncbi.nlm.nih.gov/37938258/)]
57. So M, Sobolev M, Menvielle G. Sanora: a conversational AI agent for multimodal digital biomarkers of mental health. *TechRxiv*. Preprint posted online on Feb 10, 2026. URL: <https://www.techrxiv.org/users/1027431/articles/1387167-sanora-a-conversational-ai-agent-for-multimodal-digital-biomarkers-of-mental-health> [Accessed 2026-03-22]
58. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. Sep 2001;16(9):606-613. [doi: [10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
59. Okuyama T, Akechi T, Kugaya A, et al. Development and validation of the Cancer Fatigue Scale. *J Pain Symptom Manage*. Jan 2000;19(1):5-14. [doi: [10.1016/S0885-3924\(99\)00138-4](https://doi.org/10.1016/S0885-3924(99)00138-4)] [Medline: [10687321](https://pubmed.ncbi.nlm.nih.gov/10687321/)]
60. Dahmen J, Cook D, Fellows R, Schmitter-Edgecombe M. An analysis of a digital variant of the Trail Making Test using machine learning techniques. *Technol Health Care*. 2017;25(2):251-264. [doi: [10.3233/THC-161274](https://doi.org/10.3233/THC-161274)] [Medline: [27886019](https://pubmed.ncbi.nlm.nih.gov/27886019/)]
61. Girard JM, Cohn JF, Mahoor MH, Mavadati SM, Hammal Z, Rosenwald DP. Nonverbal social withdrawal in depression: evidence from manual and automatic analysis. *Image Vis Comput*. Oct 2014;32(10):641-647. [doi: [10.1016/j.imavis.2013.12.007](https://doi.org/10.1016/j.imavis.2013.12.007)] [Medline: [25378765](https://pubmed.ncbi.nlm.nih.gov/25378765/)]
62. Scherer S, Stratou G, Mahmoud M, et al. Automatic behavior descriptors for psychological disorder analysis. Presented at: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; Apr 22-26, 2013; Shanghai, China. [doi: [10.1109/FG.2013.6553789](https://doi.org/10.1109/FG.2013.6553789)]
63. Girard JM, Cohn JF, Mahoor MH, Mavadati S, Rosenwald DP. Social risk and depression: evidence from manual and automatic facial expression analysis. Presented at: International Conference on Automatic Face and Gesture Recognition 2013; Apr 22-26, 2013; Shanghai, China. [doi: [10.1109/FG.2013.6553748](https://doi.org/10.1109/FG.2013.6553748)]

64. Menne F, Dörr F, Schröder J, et al. The voice of depression: speech features as biomarkers for major depressive disorder. *BMC Psychiatry*. Nov 12, 2024;24(1):794. [doi: [10.1186/s12888-024-06253-6](https://doi.org/10.1186/s12888-024-06253-6)] [Medline: [39533239](https://pubmed.ncbi.nlm.nih.gov/39533239/)]
65. Wang Y, Liang L, Zhang Z, et al. Fast and accurate assessment of depression based on voice acoustic features: a cross-sectional and longitudinal study. *Front Psychiatry*. 2023;14:1195276. [doi: [10.3389/fpsy.2023.1195276](https://doi.org/10.3389/fpsy.2023.1195276)] [Medline: [37415683](https://pubmed.ncbi.nlm.nih.gov/37415683/)]
66. Cummins N, Sethu V, Epps J, Krajewski J. Probabilistic acoustic volume analysis for speech affected by depression. Presented at: Interspeech 2014; Sep 14-18, 2014; Singapore, Singapore. [doi: [10.21437/Interspeech.2014-311](https://doi.org/10.21437/Interspeech.2014-311)]
67. Shinohara S, Toda H, Nakamura M, et al. Evaluation of emotional arousal level and depression severity using voice-derived sound pressure change acceleration. *Sci Rep*. Jun 30, 2021;11(1):13615. [doi: [10.1038/s41598-021-92982-7](https://doi.org/10.1038/s41598-021-92982-7)] [Medline: [34193915](https://pubmed.ncbi.nlm.nih.gov/34193915/)]
68. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. *Biol Psychiatry*. Oct 1, 2012;72(7):580-587. [doi: [10.1016/j.biopsych.2012.03.015](https://doi.org/10.1016/j.biopsych.2012.03.015)] [Medline: [22541039](https://pubmed.ncbi.nlm.nih.gov/22541039/)]
69. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geraltz DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics*. Jan 2007;20(1):50-64. [doi: [10.1016/j.jneuroling.2006.04.001](https://doi.org/10.1016/j.jneuroling.2006.04.001)] [Medline: [21253440](https://pubmed.ncbi.nlm.nih.gov/21253440/)]
70. Yamamoto M, Takamiya A, Sawada K, et al. Using speech recognition technology to investigate the association between timing-related speech features and depression severity. *PLoS One*. 2020;15(9):e0238726. [doi: [10.1371/journal.pone.0238726](https://doi.org/10.1371/journal.pone.0238726)] [Medline: [32915846](https://pubmed.ncbi.nlm.nih.gov/32915846/)]
71. Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. *Brain Cogn*. Oct 2004;56(1):30-35. [doi: [10.1016/j.bandc.2004.05.003](https://doi.org/10.1016/j.bandc.2004.05.003)] [Medline: [15380873](https://pubmed.ncbi.nlm.nih.gov/15380873/)]
72. Sundelin T, Lekander M, Kecklund G, Van Someren EJ, Olsson A, Axelsson J. Cues of fatigue: effects of sleep deprivation on facial appearance. *Sleep*. Sep 1, 2013;36(9):1355-1360. [doi: [10.5665/sleep.2964](https://doi.org/10.5665/sleep.2964)] [Medline: [23997369](https://pubmed.ncbi.nlm.nih.gov/23997369/)]
73. Zargari Marandi R, Madeleine P, Omland Ø, Vuillerme N, Samani A. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Sci Rep*. Sep 3, 2018;8(1):13148. [doi: [10.1038/s41598-018-31577-1](https://doi.org/10.1038/s41598-018-31577-1)] [Medline: [30177693](https://pubmed.ncbi.nlm.nih.gov/30177693/)]
74. Stern JA, Boyer D, Schroeder D. Blink rate: a possible measure of fatigue. *Hum Factors*. Jun 1994;36(2):285-297. [doi: [10.1177/001872089403600209](https://doi.org/10.1177/001872089403600209)] [Medline: [8070793](https://pubmed.ncbi.nlm.nih.gov/8070793/)]
75. Yin Z, Liu B, Hao D, Yang L, Feng Y. Evaluation of VDT-induced visual fatigue by automatic detection of blink features. *Sensors (Basel)*. Jan 25, 2022;22(3):916. [doi: [10.3390/s22030916](https://doi.org/10.3390/s22030916)] [Medline: [35161662](https://pubmed.ncbi.nlm.nih.gov/35161662/)]
76. Hasan F, Kashevnik A. State-of-the-art analysis of modern drowsiness detection algorithms based on computer vision. Presented at: 29th Conference of Open Innovations Association (FRUCT); May 12-14, 2021; Tampere, Finland. [doi: [10.23919/FRUCT52173.2021.9435480](https://doi.org/10.23919/FRUCT52173.2021.9435480)]
77. Yamada Y, Kobayashi M. Detecting mental fatigue from eye-tracking data gathered while watching video: evaluation in younger and older adults. *Artif Intell Med*. Sep 2018;91:39-48. [doi: [10.1016/j.artmed.2018.06.005](https://doi.org/10.1016/j.artmed.2018.06.005)] [Medline: [30026049](https://pubmed.ncbi.nlm.nih.gov/30026049/)]
78. Guggisberg AG, Mathis J, Schneider A, Hess CW. Why do we yawn? *Neurosci Biobehav Rev*. Jul 2010;34(8):1267-1276. [doi: [10.1016/j.neubiorev.2010.03.008](https://doi.org/10.1016/j.neubiorev.2010.03.008)] [Medline: [20382180](https://pubmed.ncbi.nlm.nih.gov/20382180/)]
79. Krajewski J, Trutschel U, Golz M, Sommer D, Edwards D. Estimating fatigue from predetermined speech samples transmitted by operator communication systems. Presented at: Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design; Jun 22-25, 2009; Big Sky, MT. [doi: [10.17077/drivingassessment.1359](https://doi.org/10.17077/drivingassessment.1359)]
80. de Vasconcelos CA, Vieira MN, Kecklund G, Yehia HC. Speech analysis for fatigue and sleepiness detection of a pilot. *Aerosp Med Hum Perform*. Apr 1, 2019;90(4):415-418. [doi: [10.3357/AMHP.5134.2019](https://doi.org/10.3357/AMHP.5134.2019)] [Medline: [30922431](https://pubmed.ncbi.nlm.nih.gov/30922431/)]
81. Magliacano A, Fiorenza S, Estraneo A, Trojano L. Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. *Neurosci Lett*. Sep 25, 2020;736:135293. [doi: [10.1016/j.neulet.2020.135293](https://doi.org/10.1016/j.neulet.2020.135293)] [Medline: [32771601](https://pubmed.ncbi.nlm.nih.gov/32771601/)]
82. Fischer B, Breitmeyer B. Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia*. 1987;25(1A):73-83. [doi: [10.1016/0028-3932\(87\)90044-3](https://doi.org/10.1016/0028-3932(87)90044-3)] [Medline: [3574652](https://pubmed.ncbi.nlm.nih.gov/3574652/)]
83. Bloch Y, Aviram S, Neeman R, et al. Methylphenidate mediated change in prosody is specific to the performance of a cognitive task in female adult ADHD patients. *World J Biol Psychiatry*. 2015;16(8):635-639. [doi: [10.3109/15622975.2015.1036115](https://doi.org/10.3109/15622975.2015.1036115)] [Medline: [25945954](https://pubmed.ncbi.nlm.nih.gov/25945954/)]
84. Holmlund TB, Cohen AS, Cheng J, et al. Using automated speech processing for repeated measurements in a clinical setting of the behavioral variability in the stroop task. *Brain Sci*. Mar 4, 2023;13(3):442. [doi: [10.3390/brainsci13030442](https://doi.org/10.3390/brainsci13030442)] [Medline: [36979252](https://pubmed.ncbi.nlm.nih.gov/36979252/)]

85. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). *NPJ Digit Med*. Jul 8, 2024;7(1):183. [doi: [10.1038/s41746-024-01157-x](https://doi.org/10.1038/s41746-024-01157-x)] [Medline: [38977771](https://pubmed.ncbi.nlm.nih.gov/38977771/)]
86. Tartaglia J, Jaghab B, Ismail M, et al. Assessing health technology literacy and attitudes of patients in an urban outpatient psychiatry clinic: cross-sectional survey study. *JMIR Ment Health*. Dec 30, 2024;11:e63034. [doi: [10.2196/63034](https://doi.org/10.2196/63034)] [Medline: [39753220](https://pubmed.ncbi.nlm.nih.gov/39753220/)]
87. Sobolev M, Vitale R, Wen H, et al. The Digital Marshmallow Test (DMT) diagnostic and monitoring mobile health app for impulsive behavior: development and validation study. *JMIR Mhealth Uhealth*. Jan 22, 2021;9(1):e25018. [doi: [10.2196/25018](https://doi.org/10.2196/25018)] [Medline: [33480854](https://pubmed.ncbi.nlm.nih.gov/33480854/)]
88. Radanliev P, Santos O, Brandon-Jones A, Joinson A. Ethics and responsible AI deployment. *Front Artif Intell*. 2024;7:1377011. [doi: [10.3389/frai.2024.1377011](https://doi.org/10.3389/frai.2024.1377011)] [Medline: [38601110](https://pubmed.ncbi.nlm.nih.gov/38601110/)]
89. Trocin C, Mikalef P, Papamitsiou Z, Conboy K. Responsible AI for digital health: a synthesis and a research agenda. *Inf Syst Front*. Dec 2023;25(6):2139-2157. [doi: [10.1007/s10796-021-10146-4](https://doi.org/10.1007/s10796-021-10146-4)]
90. Chaturvedi RR, Angrisani M, Troxel WM, et al. American Life in Realtime: benchmark, publicly available person-generated health data for equity in precision health. *PNAS Nexus*. Oct 2025;4(10):pgaf295. [doi: [10.1093/pnasnexus/pgaf295](https://doi.org/10.1093/pnasnexus/pgaf295)] [Medline: [41064239](https://pubmed.ncbi.nlm.nih.gov/41064239/)]
91. Wen H, Sobolev M, Vitale R, et al. mPulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study. *JMIR Ment Health*. Jan 27, 2021;8(1):e25019. [doi: [10.2196/25019](https://doi.org/10.2196/25019)] [Medline: [33502330](https://pubmed.ncbi.nlm.nih.gov/33502330/)]

## Abbreviations

**AI:** artificial intelligence  
**CFS:** Cancer Fatigue Scale  
**PHQ-9:** Patient Health Questionnaire-9  
**TMT:** trail making test  
**USAF:** US Air Force

*Edited by Amy Schwartz, Matthew Balcarras; peer-reviewed by Ivan Rodriguez-Conde; submitted 03.Nov.2025; final revised version received 27.Feb.2026; accepted 27.Feb.2026; published 01.Apr.2026*

### *Please cite as:*

*So M, Sobolev M, Menvielle G*

*A Conversational Platform (Okaya) for Multimodal Digital Biomarkers of Fatigue, Cognition, and Mental Health: Feasibility Observational Study*

*JMIR Form Res 2026;10:e87054*

*URL: <https://formative.jmir.org/2026/1/e87054>*

*doi: [10.2196/87054](https://doi.org/10.2196/87054)*

© Matthew So, Michael Sobolev, Gregory Menvielle. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 01.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.