International Journal of Methodology

Volume 4, Issue 1, pp. 16-31, 2026
DOI: https://doi.org/10.21467/ijm.4.1.10215

The Vilans Observation Tool for Human-Technology Interaction: A Blended Observation System for Interactions with Socially Assistive Technologies

Enikő Ágotai[1,2†], Bob M. Hofstede[1,3†*]

[1] Vilans, Centre of Expertise for Long-Term Care, Utrecht, The Netherlands
[2] Department of Informatics and Media, Uppsala University, Uppsala, Sweden
[3] Human-Technology Interaction group, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

† Authors contributed equally
*Corresponding author's email: b.hofstede@vilans.nl
Received: 20 August 2025 / Revised: 22 January 2026 / Accepted: 26 January 2026 / Published: 29 January 2026

ABSTRACT

The increasing development and adoption of gerontechnology, which are assistive technologies designed to support older adults' independence, well-being, and social connection, underscore the critical need for robust methods to assess their real-world impact. While observational tools are essential for this purpose, existing methods often present a trade-off between efficiency and depth. Rating scales are simple to use but lack the temporal and behavioural granularity needed to analyse interaction processes, while detailed behavioural coding schemes are resource-intensive and inflexible. This paper addresses a significant gap by introducing the Vilans Observation Tool for Human-Technology Interaction (VOTHI) framework, a novel, low-threshold, hybrid observational tool. Developed through a literature analysis of existing instruments, VOTHI combines the structural clarity of rating scales with the temporal sensitivity of coding schemes, organized in a modular structure to adapt to various research contexts and technologies. The framework's design principles emphasise the holistic capture of interactional dynamics, including verbal, non-verbal, affective, and contextual cues, focusing on the process of interaction rather than just the outcome. Pilot-testing demonstrates VOTHI's capacity to provide a comprehensive, adaptable, and practical method for researchers to analyse human-technology interactions. The tool serves as a valuable solution for capturing subtle, dynamic, and process-oriented aspects of engagement, particularly in gerontechnological contexts.

Keywords: Human-Technology Interaction, Observational tools, Gerontechnology

1 Introduction

With the ageing population pressing on healthcare systems worldwide [1], the role of assistive technologies is becoming increasingly important [2]. In particular, Socially Assistive Technologies (SATs) are being used to enhance the capabilities of people living with dementia (PWD), supporting their independence, autonomy in daily life, and social participation [3]. Despite growing interest, the adoption of these technologies is still hindered by several implementation factors [4]. To better understand these challenges, research suggests investigating contextual factors through in-situ testing with potential end-users, such as older adults [4], [5], [6], [7]. Beyond informing insights for more effective technologies [8], in-situ testing provides the empirical depth required for the increasingly prominent Health Technology Assessment (HTA) processes [9], [10], [11]. Consequently, direct observation of a technology's use in real-world settings offers a methodologically sound way to enrich HTA processes with insights into latent user needs and interactional dynamics [12].

A common paradigm to in-situ testing is ethnographic research, where researchers immerse themselves in specific groups or environments to understand their culture, behaviours and social interactions [13]. Within this method, structured schemes or systems are frequently used for ensuring systematic data collection. However, as we prepared for an ethnographic study on SAT use among older adults, a significant methodological challenge emerged. While various observation systems offer diverse approaches to documenting human responses and technology behaviour within Human-Technology Interaction (HTI) research, few have the flexibility and specificity required for the nuanced nature of gerontechnological interactions. Gerontechnological interactions can be explained as the interaction between older adults and technology, as the concept of gerontechnology focuses on designing technology that supports the independence and quality of life of older adults [14]. This methodological gap is further evidenced by a systematic review of twenty-two observational instruments for people with dementia [15], which found that few tools have been rigorously tested or validated for interactive, multimodal contexts. Existing instruments generally fall into two categories: observational rating scales and behavioural coding schemes [16], [17], [18]. Within these categories, the current literature offers several established instruments to assess human behaviour in care contexts [16], [17], [18]. However, their application to diverse gerontechnological settings remains limited.

Observational rating scales, such as the Observational Measurement of Engagement (OME) [19], [20], the Engagement of a Person with Dementia Scale (EPWDS) [21], the Menorah Park Engagement Scale (MPES) [22], and the Observed Emotion Rating Scale (OERS) [23] are widely cited and validated [16]. These instruments were primarily designed for use in naturalistic residential care and clinical settings, considering vulnerable populations like PWD. In practice, they typically conceptualise user response through high-level constructs, focusing mainly on dimensions of engagement, social-emotional initiation, and affective states. While these tools are highly usable and suitable for rapid, low-resource environments, they share a lack of temporal sensitivity and modularity. By relying on session-level or post-session impressions, they focus on aggregate outcomes rather than processual data. Consequently, they may omit critical contextual nuances and fail to account for micro-level dynamics or the way in which engagement fluctuates on a moment-to-moment basis in response to technological stimuli.

In contrast, relevant behavioural coding systems and ethograms are typically adapted or developed in direct relation to technological use, aiming to document more detailed and temporally sensitive sequences of user behaviour [16]. These methods involve the structured observation of embodied actions, encompassing both non-verbal and verbal modalities. Such schemes are often organised around predefined ethograms and coding categories, though some adopt a more flexible or layered approach to behavioural description. They vary wildly in their technological requirements (live or video annotation, specific software) and the units of analysis (the primary focus of the observation: group or individual). A leading example of a behavioural coding system is the Ethographic and Laban-Inspired Coding System of Engagement (ELICSE), which captures individual-level directional attentional behaviours and affective modifiers across specific body modalities [16]. This framework has been further refined through a context-specific adaptation by [24] to include conversational modalities, broadening its applicability to interactions involving speech and multimodal augmented displays. Other field-oriented approaches prioritise field-readiness and collective dynamics through varied units of analysis, such as the group-level coding protocol by [25] used for robot-assisted group activities, or the live coding protocol by [26] developed for systematic time-sampling with the robot PARO where video recording was ethically restricted. More comprehensive approaches like the structured observational framework by [27] employs behavioural grammars to interpret how human proximity and interaction attempts change in response to specific robot stimuli. Similarly, the generalised observational framework [28] provides a comprehensive taxonomy for HRI, categorising the situational context, robot capabilities, and user responses into a shared vocabulary for researchers.

While these behavioural coding systems offer high granularity and event-level data, their practical utility is often hampered by high resource demands, such as the need for specialised software and extensive coder training that limit scalability in the field. Furthermore, many tools are tied to specific robot types or institutional contexts, resulting in a lack of standardised protocols and modularity. Consequently, existing tools lack the necessary balance between granular detail and field-ready flexibility required for gerontechnological research. This methodological gap limits the rigour with which emerging technologies can be evaluated within HTA frameworks, particularly preventing a holistic understanding of HTI.

To address this gap, we developed a new observation system, the Vilans Observation Tool for Human-Technology Interaction (VOTHI), by blending elements of multiple reviewed schemes. The aim of this paper is to introduce VOTHI as a robust, adaptable tool designed to contribute to HTA processes and inform technology design to increase the successful integration of technology in care. This work presents the methodological rationale for the tool, provides formal definitions for its core concepts, and offers an illustrative application of its use by presenting our initial experiences while using VOTHI in a case study.

2 Methods

In this section, we present a synthesis of existing observation schemes and systems, identifying the specific elements that serve as the conceptual building blocks for our proposed tool (Section 2.1). Building on these blocks, we detail our iterative development process of the tool (Section 2.2) and describe the design of a first-use case study (Section 2.3). Consequently, the Results section presents the VOTHI tool itself as the primary research outcome, followed by first experiences of using the tool in our case study.

2.1 Literature synthesis to inform design criteria

To establish an empirical foundation for our observation tool (i.e., VOTHI), we conducted a targeted review of existing observational instruments and frameworks [16], [17], [18], [19], [21], [22], [23], [20], [23], [25], [26], [27], [28], and relevant literature [29]. Given the scarcity of tools specifically validated for gerontechnological settings, we expanded our scope to include instruments relevant to human-human [28] and human-robot interaction assessment, provided they offered methodological utility for documenting complex behaviours in HTI care settings. We specifically selected instruments that either demonstrated rigorous validation in (dementia) care or addressed the multimodal complexities of interactive technologies, regardless of their original context or location of use.

Our analysis took the form of a technical deconstruction of these instruments' design principles, coding logics, and operational limitations. We categorised each instrument based on its unit of analysis (the primary entity being observed), the nature of its indicators (e.g., Likert vs. binary event coding of an observable event, cue, or action that can be consistently recognised and tracked during an interaction), and its adaptability to the changing roles of participants in gerontechnological settings. Each instrument was ultimately analysed through four primary criteria: granularity (temporal resolution), modularity (adaptability of the framework), resource requirements (feasibility in the field), and observational focus (outcome vs. process). The synthesis is based on an interpretation of the reported information found in literature, as these systems were not operationally tested for this study. Still, the synthesis reveals several persistent technical patterns that hinder their immediate application to gerontechnological research, as presented in Table 1. Therefore, it can be used to construct a new tool that is intended to overcome these barriers in care settings.

Table 1: *synthesis of existing observation tools for HTI.*

Tool	Type	Focus	Granularity	Modularity*	Resources needed	Observed component
OME [19]	Rating scale	Overall engagement & attitude	Session-level	Low—fixed dimensions	Pen and paper	Outcome(s)
MPES [22]	Rating scale (categorical)	Engagement type classification	Session-level	Low—rigid categories	Pen and paper	Outcome(s)
OERS [23]	Rating scale	Affective states	Session-level	Low—fixed categories	Pen and paper	Outcome(s)
EPWDS [21]	Rating scale	Social & emotional engagement	Session-level	Low—fixed items	Pen and paper	Outcome(s)
ELICSE [16]	Behavioural coding scheme	Detailed, non-verbal engagement	Event/frame-level	High—codes/modalities adaptable	Video, software, training of observers	Process
Group-level protocol [25]	Behavioural coding scheme	Group engagement & affect	Moment-level (short clips)	Moderate—dimensions adjustable	Video (retrospective coding)	Outcome(s)
Live coding protocol [26]	Behavioural coding scheme	Detailed, social mediation	Moment-level (1-min intervals)	Moderate—behaviours list editable	Digital form	Process
Structured observational framework [27]	Behavioural coding scheme	Engagement & physical interaction	Event/moment-level	High—behaviours/grammar adaptable	Video and software	Process
Generalised observational framework [28]	Conceptual framework for behavioural coding scheme	Structured vocabulary for describing interaction	Depends on adaptation	High—structure adaptable	Theoretical	Process

*Low modularity: the tool's items/categories are fixed and changing them would essentially create a new instrument requiring new validation. Moderate modularity: the tool has some fixed structure, but individual dimensions or behaviours can be swapped, added, or removed with minor adjustments. Not fully open-ended, but still practical to adapt. High modularity: the tool's structure is inherently open: categories, behaviours, and coding rules are meant to be modified or expanded without breaking the system. It's designed for adaptation.

2.1.1 Granularity and modularity

Regarding granularity and modularity, a clear technical divide exists between rating-scale instruments and behavioural coding schemes. The rating scales analysed are all characterised by session-level granularity and low modularity. These tools rely on post-session aggregate impressions or the selection of a single dominant behaviour [22] during an activity, which provides a high-level summary but lacks the temporal sensitivity to capture interactional sequences or micro-level dynamics. Because their items are fixed and validated as a set (rated on Likert-scale [19], [21], [23] or categorically [22]), they offer limited flexibility for adaptation to new technologies or participant roles. Changing the fixed items in these scales can even be considered as creating a new instrument, and hence, would requiring re-validation. Eventually, this would limit their flexibility for use across diverse technologies or evolving participant roles. In contrast, behavioural coding schemes and frameworks offer moment- or event-level resolution. These systems allow for the tracking of specific sequences, such as directional attentional behaviours across body modalities [16] or physical proximity changes [27]. They typically demonstrate moderate-to-high modularity, as seen in the ability to swap or add conversational modalities to accommodate speech or multimodal displays [23].

2.1.2 Resource intensiveness

The assessment of resource requirements and reproducibility highlights a tension between the depth of data collected and the ease of using the tool in practice. While the rating scales are highly practical for low-resource, "pen and paper" environments [19], [22] they lack the detail found in more granular coding schemes. On the other hand, highly granular systems [16], [27] are resource-intensive, requiring high-quality video, specialised annotation software, and extensive coder training, which can limit scalability in naturalistic field research. Middle-ground approaches attempt to bridge this gap, for example, by using short video clips [25] or digital time-sampling [26] to capture process-oriented data more feasibly. However, video-based schemes, though methodologically rich, may encounter barriers to implementation in field settings due to ethical, technical, or logistical constraints. A reproducibility gap remains, as many pragmatic, field-oriented protocols lack the published coding manuals or standardised validation metrics necessary for scientific replication.

2.1.3 Behavioural scope

The behavioural scope of these tools is often fragmented. Some schemes emphasise affect (e.g., OERS) and others focus on engagement (e.g., OME, EPWDS) or physical or non-verbal behaviours (e.g., ELICSE). Few integrate all relevant dimensions for gerontechnological interactions (verbal and non-verbal behaviour, engagement, affect, and context) into a coherent and structured framework. This fragmentation constrains their applicability for studies not only focused on engagement but also on the broader interactional dynamics of HTI, such as user confusion, adaptation, or social mediation. The group-level coding protocol [25] provides a useful benchmark for assessing the scope, granularity, and ecological fit of observational tools.

2.1.4 A trade-off between granularity, flexibility, and resource intensiveness

In our sample of reviewed tools, there is a near-even split between instruments that primarily summarise outcomes and those that document the process of interaction. Outcome-focused tools, while efficient and often easier to deploy, generally offer less granularity and flexibility, which can limit their usefulness in exploratory studies seeking to understand how users respond to and make sense of technologies in-situ. This limits their use in exploratory studies that seek to understand how users respond to or make sense of technologies in context. Understanding such interactional dynamics requires tools that can sensitively capture aspects that are often under-represented in current observational schemes, like user agency, confusion, adaptation, and facilitation. Similarly, in longitudinal studies, where technologies evolve over time, observational tools must support dynamic, multi-session tracking, rather than static or session-bound summaries. Ethogram-based approaches can deliver this level of behavioural resolution, yet they are typically resource-intensive and difficult to adapt quickly for iterative or field-based work.

2.2 Instrument development process and design rationale

Building on the synthesis of existing schemes we found that VOTHI should be balancing on the trade-off between granularity, flexibility, and resource intensiveness. Therefore, VOTHI is conceptually grounded in a multimodal and process-oriented view of interaction. Its development followed an iterative process in which the technical deconstruction of existing observational instruments (Section 2.1), combined with the practical requirements identified during our ethnographic field studies, served as the primary decision criteria for its design principles. We followed three developmental steps. First, we translated the design criteria derived from the synthesis into candidate building blocks for an observation system. Second, we reviewed our conceptual system with experienced colleagues with expertise in ethnographic research. Third, we used VOTHI in two cases of our multi-case study (Section 2.3). Doing so, we could reflect on our experience while using VOTHI. Each round of probing resulted in further refinement of definitions, reduction of ambiguity, and simplification of the recording logic, while we continuously assessed whether the emerging observation logic remained low-threshold for use in naturalistic care environments, retained high modularity to accommodate diverse socially assistive technologies, and prioritised the interaction process over static outcomes. Crucially, our design work explicitly addressed the need for a structural format capable of capturing the co-temporality of human and system behaviours.

Its concrete construction was informed by both rating-based and ethogram-based traditions. From the former, it borrows the clarity and usability of categorical observation items, such as those found in the OME or the OERS. From the latter, particularly the ELICSE, it inherits the attention to situated, embodied action and hierarchical structure. VOTHI's conceptual scaffolding is further grounded in the layered taxonomic approaches emphasising the layered nature of HTI [28], ensuring that the tool covers the situational, capability, and response layers required for a holistic assessment. Inspiration was also drawn from the RAISE tool [29], which provided a valuable model for adapting affective and contextual cues to gerontechnological scenarios. Internal conceptual validation was achieved through iterative mapping of these theoretical inspirations against the specific interactional nuances observed in the field, ensuring the tool's categories remained both theoretically grounded and practically relevant. Furthermore, one of our primary objectives was to establish a conceptual vocabulary to ensure consistency and methodological rigour during the observation process. To achieve this, we defined three core methodological units that form the internal logic of the tool: the interaction episode, the observation block, and the indicator.

The interaction episode is defined as a continuous sequence of engagement or activity between the user and the technology that is bounded by a clear start and end point (e.g., the activation of a robot until the user moves away). This serves as the primary temporal unit of analysis. The observation block represents a modular conceptual category used to structure and guide the observation. These blocks act as configurable lenses focusing on specific dimensions of the interaction, such as user behaviour or system feedback, allowing researchers to focus on specific dimensions of the interaction without losing sight of the whole. Within these blocks, indicators are the most granular units of the tool. While the blocks provide the structure, the indicators are specific, observable, and recordable units of actions or states, such as cues, actions, or events, that are predefined by the researchers to be consistently recognised and tracked. Researchers select or adapt the indicators to suit the specific technology being studied, captured as occurrence (binary), frequency (countable), or quality (graded intensity). We established coding rules where each indicator can be accompanied by descriptive notes to capture literal language and to preserve ethnographic nuance.

To operationalise these units, we adopted a matrix-based logic as the structural backbone of the tool. In this framework, observation blocks and their indicators are arranged along the vertical axis, while interaction episodes are mapped along the horizontal axis. This layout was chosen specifically to allow the observer to document simultaneous occurrences within a single temporal column, effectively capturing the co-occurrence of different behaviours as they unfold. The final validation was performed through a qualitative cross-check between the two authors and an expert consultation with senior researchers in HTI. This panel reviewed the tool for intuitiveness, clarity, comprehensiveness of the blocks, and practical feasibility. By confirming that the tool's parameters directly reflect the observable realities of the field, we ensured that the instrument is methodologically fit for its intended purpose and could address the identified research gap without significant data loss.

2.3 A multi-case study to experience VOTHI's first use

VOTHI was applied within a multi-case ethnographic study investigating the latent interactional needs of older adults when engaging socially assistive robots (SARs), focusing on observing both verbal and non-

Agotai & Hofstede, *Int. J. Methodol.;* Vol. 4, Issue 1, pp. 16-31, 2026

verbal interactions. This pilot served to assess the usability, the clarity of the tool's conceptual units, and its overall applicability in the unpredictable environment of naturalistic care settings. Specifically, it functioned as a performance test for VOTHI in environments where high-resource requirements (such as extensive video recording) were ethically or logistically restricted. By deploying the tool in these diverse contexts, we evaluated how effectively the matrix-based logic and modular observation blocks functioned under the time constraints of live field research, and whether the predefined indicators were sufficiently exhaustive to capture interactional nuances without overwhelming the observer.

The study spanned both intramural (residential care) and extramural (home care) settings in the Netherlands, involving older adults (N=17) interacting with Tinybots Tessa, PARO, or SARA Robotics across four distinct case studies. These robots provided a rigorous testbed for the tool's modularity, representing a diverse range of embodiments, from functional, voice-based assistants to zoomorphic and anthropomorphic systems. Furthermore, the pilot was designed to account for the heterogeneous nature of the target population. This included a wide spectrum of physical and cognitive states among the older adults, ranging from high levels of independence and cognitive clarity to advanced stages of dementia and reduced verbal capacity. Locations also varied significantly, from private homes to busy, communal care-home corridors. While the older adults were the primary focus, the presence of informal caregivers (N=2) in specific sessions provided an additional opportunity to test the tool's ability to document social mediation and multi-party dynamics, leading to a total sample size of *N=19*. To ensure rigour and consistency during this first-use case, a two-observer setup was frequently employed. The framework's modularity allowed us to focus on different lenses during observation sessions, which were then cross-validated. While the technologies and participants observed were shared, observer A documented verbal cues while observer B focused on non-verbal body language of participants.

It should be noted that while this application of VOTHI generated significant empirical data, participant-specific demographics (such as age and gender) and individual case outcomes are excluded here. Within the scope of this paper, the first-use case study serves exclusively as a methodological context to illustrate the performance of VOTHI as an instrument. The following sections report on the practical insights gained from this application, focusing on the tool's performance in documenting real-world interactional sequences.

3 Results

3.1 Structure and functioning of VOTHI

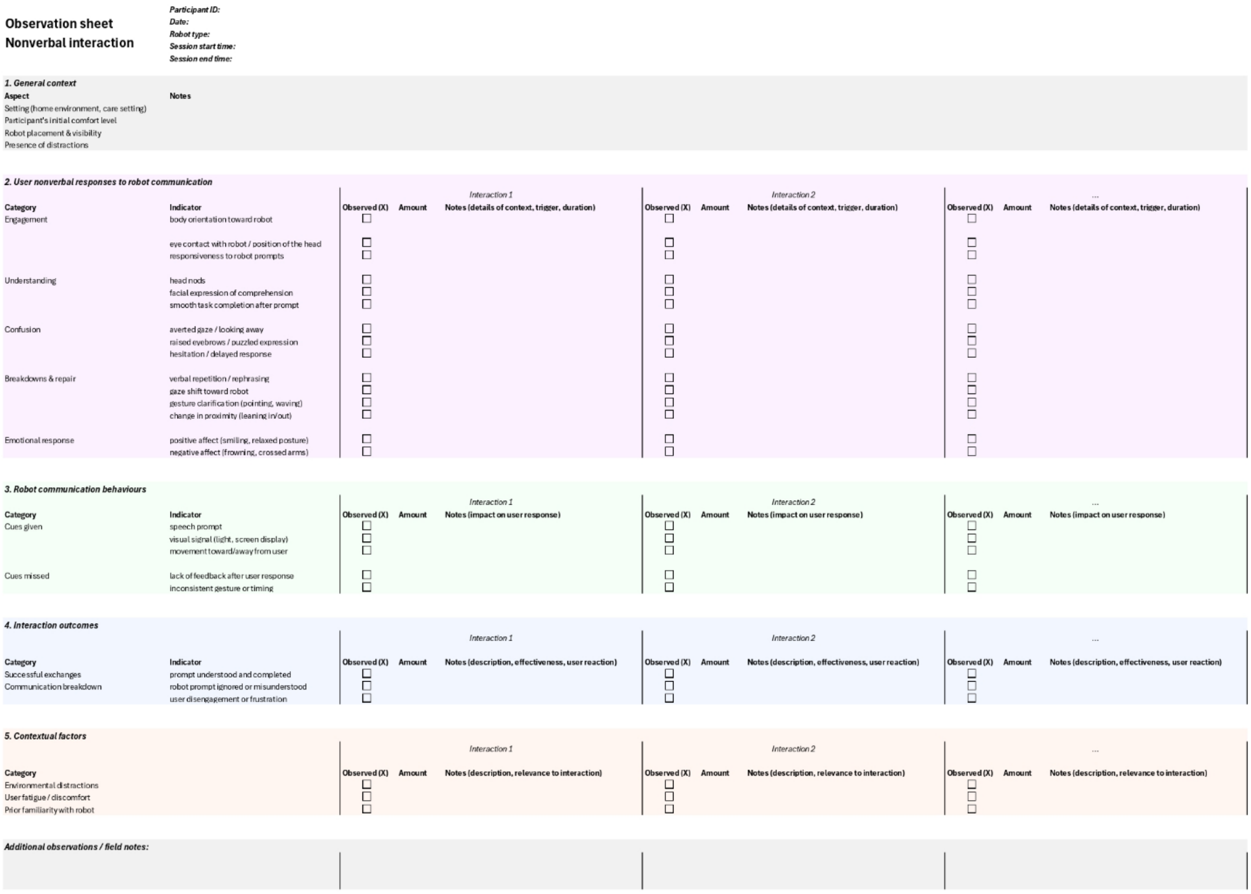
As the final output of the development, VOTHI is manifested as a visual matrix template designed for both in-situ and retrospective annotation. It functions as a modular framework that allows researchers to construct context-specific protocols while maintaining a standardised structural backbone. VOTHI is structured around six core observation blocks: (1) general context, (2) user behaviour, (3) technology/system behaviour, (4) interaction outcomes, (5) contextual factors, and (6) additional notes. Each block focuses on a particular dimension of the interaction for structuring and guiding observations and are intended to be populated with indicators specific to the technology or care context being studied. This modularity allows the tool to remain domain-agnostic. As detailed in Table 2, these dimensions are analytically partitioned to capture the triangulation of an interaction, as, by separating "user behaviour" from "technology behaviour", the tool forces the observer to look for causality and response patterns. Similarly, the "general context" and "contextual factors" blocks provide the necessary ecological grounding, ensuring that user reactions are interpreted within their physical and social surroundings. This categorical separation allows for a more granular identification of where outstanding moments occur, and whether they stem from the system's output, the user's cognitive state, or external environmental interference.

Table 2: *Overview of VOTHI's core observation blocks and their application.*

Block	Purpose	Application across care technology contexts
General context	▪ Records the situational and environmental conditions of the observation. ▪ This includes baseline info about the type setting (e.g. at home, care facility, day centre), timing, who is present (e.g. caregiver, family, professional), and whether the tech/device is visible or active.	▪ This applies no matter the topic. Helps situate the interaction.
User behaviour	▪ Captures how the user (e.g., client, resident, informal caregiver, or professional) responds to the system or technology. ▪ Can include actions, reactions, verbal responses, emotional signs, and behaviour over time.	▪ It allows insight into usability, comfort, or adoption. ▪ Example use cases: tracking acceptance, resistance, confusion, learning, trust, empowerment, or distress in interaction with technologies.
Technology/ system behaviour	▪ Logs what the technology or intervention does: how it gives feedback, communicates, signals, prompts, adapts, or fails. ▪ Includes any form of interface output or system action: sounds, notifications, prompts, movements, visual displays, responses, or lack of responses.	▪ Helps assess usability, timing, relevance, and responsiveness. Can be used for any tool or system from apps, robots, wearables, interfaces to hybrid human-tech workflows. ▪ Example use cases: smart systems, monitoring tools, decision support tools, virtual assistants, telehealth platforms, care robots.
Interaction outcomes	▪ Tracks the result of the interaction attempt between the user and the technology/system. ▪ Did it lead to task completion, misunderstanding, avoidance, success.	▪ Supports evaluation of effectiveness, ease of use, user satisfaction. ▪ Example use cases: identifying breakdowns or noting how and when users complete or abandon a task.
Contextual factors	▪ Notes any external, environmental, or personal conditions that could influence behaviour or outcomes, even if they're not part of the technology itself. ▪ This could include noise levels, time of day, the person's mood or fatigue, technical issues, or social dynamics.	▪ Helps interpret data more accurately across settings. ▪ Example use cases: identifying factors like fatigue, interruptions, prior experience, sensory limitations, staff support, or emotional state that may affect use.
Additional notes	▪ Provides space for free-text observations, open-ended reflections, or anything not captured by structured fields.	▪ Adds nuance and depth. ▪ Useful for documenting surprises, user quotes, emotional reactions, or unexpected events that could inform future design or analysis.

The functioning of the tool relies on a horizontal-vertical matrix logic that prioritises the chronological flow of events. Indicators are arranged horizontally within their respective blocks, while interaction episodes are listed sequentially across columns. Its layout is critical for analytical rigour for two reasons. First, it enables the observer to document the co-temporality of an interaction, as, by looking across a single column, what the technology did and how the user responded at that specific moment can be followed. Second, the horizontal progression allows for sequential analysis, making it possible to track how interactions evolve, escalate, or break down over time. In other words, observers can track events interaction by interaction, making it suitable for capturing both brief interaction episodes and longer sequences.

A basic digital Matrix Template has been developed for public use (Attachment A) and can also be printed for paper-based observation. For visual illustration purposes, Figure 1 reproduces the VOTHI Matrix Template. The layout is hierarchically structured and colour-coded to distinguish observation blocks, which helps manage the high cognitive workload of real-time field observations by supporting quick scanning across blocks and reducing loss of focus. The template is implemented as an Excel workbook, leveraging the multi-sheet functionality to keep related materials together while separating perspectives (e.g., different observers, sessions, or configuration variants). This format preserves the instrument's low-threshold character (no specialised software required), supports multi-observer workflows in a single file, and remains print-friendly for settings where digital devices or video are not feasible.



Observation sheet
Nonverbal interaction

Participant ID:
Date:
Robot type:
Session start time:
Session end time:

1. General context

Aspect	Notes
Setting (home environment, care setting)	
Participant's initial comfort level	
Robot placement & visibility	
Presence of distractions	

2. User nonverbal responses to robot communication

		Interaction 1			*Interaction 2*			*...*		
Category	**Indicator**	**Observed (X)**	**Amount**	**Notes (details of context, trigger, duration)**	**Observed (X)**	**Amount**	**Notes (details of context, trigger, duration)**	**Observed (X)**	**Amount**	**Notes (details of context, trigger, duration)**
Engagement	body orientation toward robot	☐			☐			☐		
	eye contact with robot / position of the head	☐			☐			☐		
	responsiveness to robot prompts	☐			☐			☐		
Understanding	head nods	☐			☐			☐		
	facial expression of comprehension	☐			☐			☐		
	smooth task completion after prompt	☐			☐			☐		
Confusion	averted gaze / looking away	☐			☐			☐		
	raised eyebrows / puzzled expression	☐			☐			☐		
	hesitation / delayed response	☐			☐			☐		
Breakdowns & repair	verbal repetition / rephrasing	☐			☐			☐		
	gaze shift toward robot	☐			☐			☐		
	gesture clarification (pointing, waving)	☐			☐			☐		
	change in proximity (leaning in/out)	☐			☐			☐		
Emotional response	positive affect (smiling, relaxed posture)	☐			☐			☐		
	negative affect (frowning, crossed arms)	☐			☐			☐		

3. Robot communication behaviours

		Interaction 1			*Interaction 2*			*...*		
Category	**Indicator**	**Observed (X)**	**Amount**	**Notes (impact on user response)**	**Observed (X)**	**Amount**	**Notes (impact on user response)**	**Observed (X)**	**Amount**	**Notes (impact on user response)**
Cues given	speech prompt	☐			☐			☐		
	visual signal (light, screen display)	☐			☐			☐		
	movement toward/away from user	☐			☐			☐		
Cues missed	lack of feedback after user response	☐			☐			☐		
	inconsistent gesture or timing	☐			☐			☐		

4. Interaction outcomes

		Interaction 1			*Interaction 2*			*...*		
Category	**Indicator**	**Observed (X)**	**Amount**	**Notes (description, effectiveness, user reaction)**	**Observed (X)**	**Amount**	**Notes (description, effectiveness, user reaction)**	**Observed (X)**	**Amount**	**Notes (description, effectiveness, user reaction)**
Successful exchanges	prompt understood and completed	☐			☐			☐		
Communication breakdown	robot prompt ignored or misunderstood	☐			☐			☐		
	user disengagement or frustration	☐			☐			☐		

5. Contextual factors

	Interaction 1			*Interaction 2*			*...*		
Category	**Observed (X)**	**Amount**	**Notes (description, relevance to interaction)**	**Observed (X)**	**Amount**	**Notes (description, relevance to interaction)**	**Observed (X)**	**Amount**	**Notes (description, relevance to interaction)**
Environmental distractions	☐			☐			☐		
User fatigue / discomfort	☐			☐			☐		
Prior familiarity with robot	☐			☐			☐		

Additional observations / field notes:

Figure 1: Example observation sheet of VOTHI.

VOTHI can be independently replicated and used in observation sessions following a four-stage workflow: (1) pre-defining indicators, (2) configuring the matrix, (3) in-situ or retrospective documentation, and (4) data synthesis.

The workflow begins with preparation, where the researcher establishes the specific parameters of the observation. During this phase, the researcher identifies the specific behaviours and system actions and selects or develops a set of indicators relevant to the study's objectives. Crucial to this stage is the creation of formal operational definitions for each indicator. This ensures that if multiple observers are involved, they maintain a shared understanding of what constitutes a specific action, thereby increasing inter-rater reliability.

Once the indicators are defined, the process moves into configuration. Here, the researcher populates the observational blocks of the VOTHI matrix with the chosen indicators and determines the coding rules for the session. A decision is made for each indicator regarding whether it will be recorded as a binary occurrence (present/absent), a frequency count (e.g., number of physical touches), or a graded intensity (e.g., a 3-point scale or description for affective state). The researcher also determines the expected scope of an interaction episode for that session (e.g., a single conversational turn or a complete multi-step task sequence).

During active observation, the tool functions as a real-time interface between the observer and the interaction. As events unfold the observer documents the process sequentially, as shown in Figure 2. The operational procedure for the annotation process is designed to be fast yet descriptive. When a predefined indicator occurs, observers simply tick a checkbox. Depending on the observational aim, they can optionally record frequency, duration, or intensity. If an indicator does not occur but the moment is still meaningful, the "notes" field can be used to capture verbatims or unexpected reactions. Otherwise, the field remains blank. For every new interaction episode, the observer moves to the next column on the matrix, scanning for indicator occurrence in the rows. This design is intended to accelerate the observational process while preserving space for ethnographic nuance. Descriptive notes are encouraged to be literal, capturing observed language, reactions, or context-specific details that enrich interpretation.

Finally, the workflow concludes with the data synthesis and analysis. We recommend that the observer reviews the matrix to ensure all marks are clear and expands any abbreviated notes while the interaction is still fresh in their memory. This produces a structured, chronological log that serves as the primary data source. These logs are particularly well-suited for sequential analysis or thematic coding, as they provide a visualised map of how interactions emerge, flourish, or break down over time.

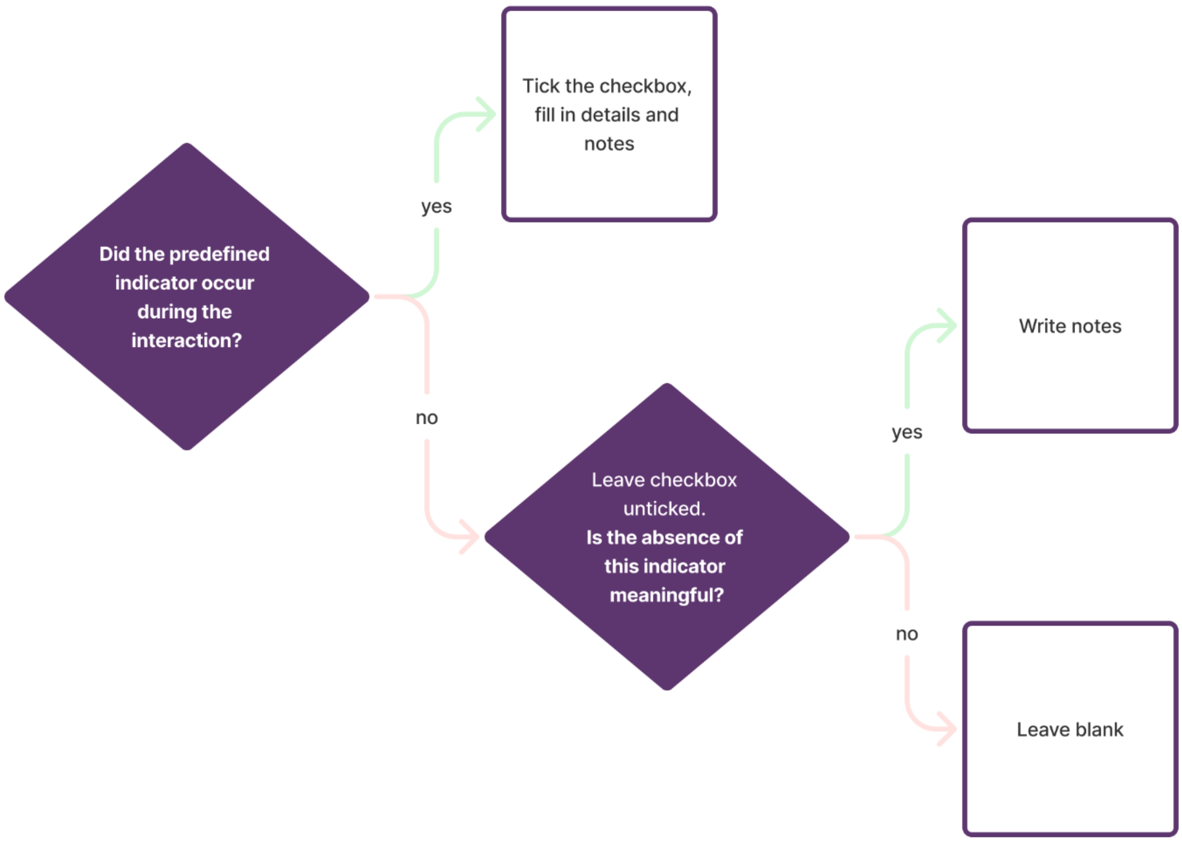


Figure 2: Step-by-step flow-chart for using VOTHI.

Depending on the structure of the observation and the defined indicators, VOTHI typically yields structured, descriptive logs that can later be used for thematic analysis, design iteration, or cross-case comparison. Its format supports both brief and extended observations and is adaptable to a variety of

DOI: 10.2147/ijm
Available online at Journals.aijr.org

settings. Essentially, it serves as an analytical map that forces the researcher to consider the relationship between different dimensions of interaction simultaneously.

3.2 Instrument performance results

The application of VOTHI in our multi-case ethnographic study provided critical insights into its performance as a field instrument. These findings highlight the practical feasibility of the framework and the operational adjustments required for real-time field use.

3.2.1 Initial operational successes

The initial field-testing confirmed that the VOTHI framework is a feasible and effective alternative to existing schemes. A key strength observed was the tool's capacity to facilitate multi-observer collaboration. By clearly defining our roles, we could independently configure the modular blocks to focus on complementary aspects of the interaction. Because the two observers focused on non-overlapping lenses (observer A: verbal interaction; observer B: non-verbal cues), formal inter-rater reliability on identical units was not applicable. Instead, immediate post-session cross-checks were conducted: after each session, we jointly reviewed each other's sheets to verify temporal alignment, resolve any interpretive ambiguities, and confirm that the combined notes formed a coherent account of the episode. This consensus-based convergence check served to guard against omissions (e.g., missed co-occurrences), to harmonise terminology, and to document clarification where needed.

The matrix-based layout was particularly successful in capturing the co-temporality of events. For instance, in interaction with Tinybots Tessa, observer A documented a participant's verbal response of "Yes, I've had coffee", while observer B simultaneously recorded subtle non-verbal cues, such as a slight head nod and a cheerful smile. This combination provided a richer understanding of the participant's reaction, confirming that her verbal response was supported by positive non-verbal signals of comprehension. Additionally, this layout captured behaviours that might otherwise be missed: for example, a participant's finger raised in anticipation of a robot's prompt was recorded as a clear sign of engagement that was only visible through the synchronised lens of the matrix. This provided a more precise understanding of the interactional sequence than would be possible with a session-level rating scale.

The combined data from the observation blocks also gave concrete evidence of communication breakdowns and social referencing. For example, when a participant asked her daughter, "am I not allowed to say this?", observer B simultaneously noted a gesture of mild confusion and the participant looking to her daughter for reassurance. In similar cases, the tool's design facilitated the later analysis of causality and discord, by documenting these social chain reactions across the user behaviour and contextual factor blocks.

3.2.2 Practical challenges

Despite these successes, the field-testing revealed two primary practical challenges inherent in structured observation. Firstly, in multi-user scenarios, such as the PARO case study (N=9), the need to use separate sheets for each participant complicated the observation of shared social dynamics. While the design maintained clarity at the individual level, capturing how different individuals in a room reacted simultaneously to the robot required significant effort to reconstruct a cohesive group-level picture. For example, while one participant might be laughing and touching PARO, another might be watching with curiosity from a distance.

Secondly, particularly in fast-paced interactions, we occasionally struggled to maintain the chronological flow while capturing complex dialogue due to the high cognitive load required to document multiple verbal and non-verbal reactions in real-time. This was most notable in the sessions with SARA Robotics and PARO, where the robot sometimes was moved quickly between participants. The difficulty in capturing multi-user dynamics and the cognitive load experienced during real-time observation are not unique to this tool but rather highlight known methodological trade-offs in structured observation studies [30], [31]. This is because capturing complex, emergent group behaviours require a different methodological lens than individual interactions, a challenge widely discussed in human-robot interaction literature [32]. Similarly, the cognitive load experienced is a well-documented limitation of live coding schemes, which often struggle to balance the need for high-fidelity data with the real-time demands of complex, naturalistic settings [33]. Addressing these challenges, several iterative refinements were made to the observation workflow. On the one hand, to manage the cognitive load during rapid interactions, we developed a workaround where real-time documentation was supplemented with rapid ethnographic notes taken in the "additional notes" block of the matrix, which were then transferred to the formal indicators immediately after the session. On the other hand, to ensure the accuracy of these transfers and maintain observer alignment, we refined the workflow to include a formal post-session debriefing. During these meetings, we compared our respective sheets to cross-validate recorded indicators and ensure that the modular blocks were being applied uniformly. This reflexive approach ensured that the final dataset remained grounded in literal, descriptive observation notes of raw behaviour before moving to interpretation.

4 Discussion

4.1 Contributions

This paper introduced the Vilans Observation Tool for Human-Technology Interaction (VOTHI), a hybrid observation system developed in response to the lack of observational tools, found during our own ethnographic research, suitable for gerontechnological assessment [15]. Our comprehensive review of established observational instruments and frameworks confirmed the holistic shortcomings and a general lack of adaptability for nuanced, process-oriented interactions in care settings. VOTHI fills this gap by offering both a flexible, conceptual scaffolding and a base template that facilitates detailed, reproducible data collection while maintaining the necessary ethnographic openness to capture emergent social behaviours [16], [34]. The application of VOTHI in the first-use case study demonstrates its practical feasibility and relevance across varying robotic embodiments and social configurations. By providing a standardised structural backbone that remains sensitive to the situated nuances of care environments, the tool enables researchers to document interactional micro-cues that are often overlooked by more rigid methodologies. These initial findings, while exploratory, suggest that VOTHI serves as a productive framework for capturing the complexity of HTI in-situ, offering a level of operational detail that supports both local qualitative insight and broader methodological rigour.

4.2 Comparison with existing tools

VOTHI represents a departure from existing observational instruments by specifically addressing the unique interactional requirements of gerontechnology. Its design acknowledges the heterogeneous nature of ageing populations, including those living with dementia, sensory impairments, or reduced verbal capacity. Most existing tools do not account for the asymmetrical interactional patterns often found in ageing populations, such as delayed processing times, sensory impairments, or non-verbal micro-cues of engagement. VOTHI improves upon these by shifting the focus from documenting static outcomes to mapping the sequential trajectory of the interaction, a vital approach that offers insights into user behaviour that existing tools cannot capture [35]. This way, VOTHI addresses the process-blindness noted in traditional gerontological assessments, where the reasons behind a user's actions are often left unrecorded. The primary technical distinction between VOTHI and established instruments lies in its visual matrix logic. By organising observations into six configurable blocks that represent the layered nature of HTI, VOTHI enables researchers to create tailored protocols, adapting specific indicators to the nuances of the technology and setting without sacrificing a standardised structural backbone. This modular architecture is central to its adaptability, allowing for a swappable indicator system that maintains the methodological integrity of the tool. Moreover, unlike tools that record variables as discrete data points, the matrix facilitates the explicit co-registration of user behaviour, system behaviour, and contextual factors in a unified simultaneous frame. This columnar, episode-by-episode layout maintains a chronological trace of interactions and preserves co-temporal alignments, making phenomena such as emergence, breakdown,

Agotai & Hofstede, *Int. J. Methodol.;* Vol. 4, Issue 1, pp. 16-31, 2026

adaptation, and social mediation visible. Consequently, this allows researchers to move beyond asking "what happened" to work toward an exploratory understanding of "how" the interaction unfolded, specifically identifying potential interactional breakdowns and social referencing patterns [35].

A significant procedural advantage over traditional instruments is its capacity to facilitate multi-observer triangulation. As demonstrated in our first-use case, the modularity of the blocks allows independent observers to focus on distinct modalities of the same interaction which can then be synchronised and cross-validated. This also addresses a common limitation in field research where a single observer's cognitive load limits the depth of data captured [30], [31]. This approach not only enhances the reliability of the findings but also allows for a level of comparative internal analysis that session-level rating scales or single-observer coding schemes cannot achieve. Furthermore, VOTHI provides a low-threshold alternative to high-fidelity schemes, as it accommodates both in-situ and retrospective annotation but does so through a multi-dimensional lens that allows for the simultaneous analysis of individual and system-level perspectives without the excessive resource demands and ethical hurdles associated with high-resolution video infrastructure.

4.3 Strengths and limitations

A significant strength of VOTHI is its systematic development, which remains firmly grounded in both a comprehensive literature synthesis and established HTI theory. By mapping theoretical dimensions (e.g., situational and response layers) [28] directly onto the tool's observation blocks, we ensured that the instrument remains methodologically robust. Furthermore, its modularity ensures that it is not tied to a specific device, technology, or observational objects, allowing it to evolve alongside the rapidly changing landscape of SATs. While the tool emerged from gerontechnology research, its domain-agnostic structure is designed to support HTI observation across various fields. This flexibility is particularly valuable for longitudinal gerontechnology research, where the technology may be updated or replaced, yet the observational backbone remains consistent.

However, several limitations must be acknowledged. Firstly, our understanding of the existing tools discussed in Section 2.1 was based solely on a review of the literature and we did not operationally test these tools. Therefore, our understanding of their features and limitations is an interpretation of the reported information we found.

Secondly, it is important to stress that VOTHI is not a statistical measurement instrument. It does not come with pre-scored items or automated templates. Instead, it is a flexible approach that prioritises rich, contextual insight over numerical data, aligning with the principles of qualitative research design where exploratory understanding is favoured over hypothesis testing or numerical benchmarking [31]. Consequently, this paper does not provide psychometric validation, such as inter-rater reliability coefficients or formal construct validity. Such validation remains a necessary step for future research before the tool can be used for benchmarking.

Thirdly, the results obtained from the two-observer setup in our example may not be representative of a single-observer application. While the two-observer setup provided a holistic view of the interaction, a single observer would likely need to prioritise specific blocks to maintain accuracy. Relatedly, the quality of the data is highly dependent on the researcher's choices, which may lead to unintentional bias or the potential to miss crucial behaviours [36]. Hence, the quality seems dependent on the number of observers as well as the observer's training, attentiveness, and consistent application of the framework.

Lastly, we acknowledge our dual role as both the developers of the VOTHI framework and the primary observers in the first-use case study. However, it is important to note that the framework was developed and refined organically during the ethnographic research process to address specific observational challenges in the field. Because the intention to publish VOTHI as a formal methodology arose only after its successful practical application, the risk of observation bias during the initial data collection was naturally mitigated, and our primary goal at the time was the accuracy of the ethnographic data rather than the validation of a new tool.

4.4 Future directions: implications for research and practice

The exploratory findings from this study suggest several avenues for the formal development and scaling of the VOTHI framework. To transition from a conceptual scaffold to a validated instrument, we propose two future research priorities.

Firstly, the most immediate requirement for future research is the establishment of psychometric robustness. We recommend future studies to employ multiple independent observers to code identical interaction sequences to calculate Inter-Rater Reliability (IRR) coefficients. Such studies are essential to determine the level of consistency achievable across different professional backgrounds (e.g., designers, healthcare providers).

Secondly, to manage the high cognitive load associated with real-time documentation and to support observer mobility, future iterations of VOTHI could focus on its UX/UI optimisation. While the current Excel-based format of VOTHI provides a practical, low-threshold starting point that supports both digital entry and paper-based printing, future iterations could benefit from a dedicated user interface designed for high-paced observation. In many naturalistic care settings, researchers must move alongside the participant and the technology, making stationary laptop entry troublesome. Therefore, we recommend future iterations to explore tablet-optimised or similar interfaces. This could potentially include features for timestamped entries, pre-populated indicator libraries, and the ability to link qualitative notes directly to specific interaction episodes. Crucially, this digital evolution must not come at the expense of VOTHI's hybrid versatility. Our first-use case confirmed that paper-based tools remain essential in environments where digital devices may be intrusive or logistically impractical. Furthermore, beyond pure academic research, there is significant potential for VOTHI to be integrated into HTA frameworks.

5 Conclusion

VOTHI offers a blended, low-threshold approach to observing HTI that keeps the process in view without demanding full ethogram infrastructures. Its modular blocks and interaction-by-interaction matrix let researchers reconfigure the same scaffold across technologies and settings while preserving the co-temporal traces of user action, system behaviour, outcomes, and context. Early application with older adults and socially assistive robots suggests the framework is feasible, adaptable, and analytically fertile, while also clarifying where procedures should be tightened. We do not present VOTHI as a finished instrument but as a practical scaffold that the community can refine through comparative studies and reliability work. Used alongside complementary methods, it can help HTI researchers and evaluators move "did it work?" to "how did it work, for whom and under what conditions?".

6 Declarations

6.1 Acknowledgements

The authors would like to thank the participants and care personnel from the four case studies (Vérian, Surplus, Accolade Zorg) for their cooperation, which made it possible to conduct this research and test the VOTHI system. We also express our gratitude to Sander Merkus, Henk Herman Nap, Sima Ipakchian Askari, Raymond Cuijpers, and Wijnand IJsselsteijn for their insightful feedback and guidance on the research design which included the design of VOTHI.

6.2 Funding source

This research was funded by an internal money from the Vilans foundation as part of its methodology development for value determination and co-design.

6.3 Ethical Approval

Ethical approval for this study was provided by the Ethical Review Board of Eindhoven University of Technology, with approval number ERB2025JADS02.

6.4 Informed Consent

Informed consent was obtained from all participants.

6.5 Competing Interests

The authors declare that they have no competing interests.

6.6 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in institutional affiliations.

How to Cite this Article:

E. Agotai and B. M. Hofstede, "The Vilans Observation Tool for Human-Technology Interaction: A Blended Observation System for Interactions with Socially Assistive Technologies", *Int. J. Methodol.*, vol. 4, no. 1, pp. 16–31. Jan. 2026. https://doi.org/10.21467/ijm.4.1.10215

References

[1] WHO, 'Ageing and health'. Accessed: Aug. 06, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[2] M. Ienca *et al.*, 'Intelligent Assistive Technology for Alzheimer's Disease and Other Dementias: A Systematic Review', *J. Alzheimer's Dis.*, vol. 56, no. 4, pp. 1301–1340, Feb. 2017, doi: 10.3233/JAD-161037.

[3] W. Q. Koh *et al.*, 'Bridging gaps in the design and implementation of socially assistive technologies for dementia care: the role of occupational therapy', *Disabil. Rehabil. Assist. Technol.*, vol. 19, no. 3, pp. 595–603, Apr. 2024, doi: 10.1080/17483107.2022.2111610.

[4] W. Q. Koh, S. A. Felding, K. B. Budak, E. Toomey, and D. Casey, 'Barriers and facilitators to the implementation of social robots for older adults and people with dementia: a scoping review', *BMC Geriatr.*, vol. 21, no. 1, p. 351, Jun. 2021, doi: 10.1186/s12877-021-02277-9.

[5] G. Cortellessa, R. De Benedictis, F. Fracasso, A. Orlandini, A. Umbrico, and A. Cesta, 'AI and robotics to help older adults: Revisiting projects in search of lessons learned', *Paladyn J. Behav. Robot.*, vol. 12, no. 1, pp. 356–378, Aug. 2021, doi: 10.1515/pjbr-2021-0025.

[6] N. Gasteiger *et al.*, 'Older adults' experiences and perceptions of living with Bomy, an assistive dailycare robot: a qualitative study', *Assist. Technol.*, vol. 34, no. 4, pp. 487–497, Jul. 2022, doi: 10.1080/10400435.2021.1877210.

[7] B. M. Hofstede *et al.*, 'A field study to explore user experiences with socially assistive robots for older adults: emphasizing the need for more interactivity and personalisation', *Front. Robot. AI*, vol. 12, p. 1537272, Mar. 2025, doi: 10.3389/frobt.2025.1537272.

[8] S. Suijkerbuijk, L. Cornelisse, S. Van Der Weegen, and H. H. Nap, 'Technologie implementeren met het honingraatmodel', *TVZ - Verpleegkd. Prakt. En Wet.*, vol. 131, no. 4, pp. 14–17, Aug. 2021, doi: 10.1007/s41184-021-1016-6.

[9] M. Drummond, R. Tarricone, and A. Torbica, 'European union regulation of health technology assessment: what is required for it to succeed?', *Eur. J. Health Econ.*, vol. 23, no. 6, pp. 913–915, Aug. 2022, doi: 10.1007/s10198-022-01458-6.

[10] H. S. M. Kort, 'Criteria and assessment of assistive technologies for ageing-in-place', *Gerontechnology*, vol. 23, no. s, pp. 1–1, Jul. 2024, doi: 10.4017/gt.2024.23.s.1131.1.sp.

[11] WHO, 'Health technology assessment - Global'. Accessed: Aug. 06, 2025. [Online]. Available: https://www.who.int/health-topics/health-technology-assessment

[12] J. Bidonde, M. Vanstone, L. Schwartz, and J. Abelson, 'An institutional ethnographic analysis of public and patient engagement activities at a national health technology assessment agency', *Int. J. Technol. Assess. Health Care*, vol. 37, no. 1, p. e37, Jan. 2021, doi: 10.1017/S0266462321000088.

[13] S. Singh, 'What is Ethnographic Research? Methods and Examples | Researcher Life'. Accessed: Aug. 06, 2025. [Online]. Available: https://researcher.life/blog/article/what-is-ethnographic-research-methods-and-examples/

[14] S. Ilha, F. Casarin, B. Huppes, D. P. Gautério-Abreu, F. Fernandes, and O. L. Munhoz, 'Gerontechnologies to Support Family Caregivers of Elderly Individuals with Alzheimer's Disease: a strategic action research study', *Rev. Enferm.*, vol. 32, 2024, doi: 10.12957/reuerj.2024.84101.

[15] K. G. Madsø, E. Flo-Groeneboom, N. A. Pachana, and I. H. Nordhus, 'Assessing Momentary Well-Being in People Living With Dementia: A Systematic Review of Observational Instruments', *Front. Psychol.*, vol. 12, Nov. 2021, doi: 10.3389/fpsyg.2021.742510.

[16] G. Perugia, R. van Berkel, M. Díaz-Boladeras, A. Català-Mallofré, M. Rauterberg, and E. Barakova, 'Understanding Engagement in Dementia Through Behavior. The Ethographic and Laban-Inspired Coding System of Engagement (ELICSE) and the Evidence-Based Model of Engagement-Related Behavior (EMODEB)', *Front. Psychol.*, vol. 9, 2018, doi: 10.3389/fpsyg.2018.00690.

[17] L. M. Pirborj, C. Mills, R. Gorkin, and Karthick Thiyagarajan, 'Tactile Interaction with Socially Assistive Robots for Children with Physical Disabilities', *Sensors*, vol. 25, no. 13, pp. 4215–4215, 2025, doi: 10.3390/s25134215.

[18] S. Qiu, P. An, K. Kang, J. Hu, T. Han, and M. Rauterberg, 'A Review of Data Gathering Methods for Evaluating Socially Assistive Systems', *Sensors*, vol. 22, no. 1, p. 82, 2022, doi: 10.3390/s22010082.

[19] J. Cohen-Mansfield *et al.*, 'The Comprehensive Process Model of Engagement', *Am. J. Geriatr. Psychiatry*, vol. 19, no. 10, pp. 859–870, 2011, doi: 10.1097/jgp.0b013e318202bf5b.

[20] J. Cohen-Mansfield, M. Dakheel-Ali, and M. S. Marx, 'Engagement in Persons With Dementia: The Concept and Its Measurement', *Am. J. Geriatr. Psychiatry*, vol. 17, no. 4, pp. 299–307, 2009, doi: 10.1097/jgp.0b013e31818f3a52.

[21] C. Jones, B. Sung, and W. Moyle, 'Engagement of a Person with Dementia Scale: Establishing content validity and psychometric properties', *J. Adv. Nurs.*, vol. 74, no. 9, pp. 2227–2240, 2018, doi: 10.1111/jan.13717.

[22] K. S. Judge, C. J. Camp, and S. Orsulic-Jeras, 'Use of Montessori-based activities for clients with dementia in adult day care: Effects on engagement', *Am. J. Alzheimers Dis.*, vol. 15, no. 1, pp. 42–46, 2000, doi: 10.1177/153331750001500105.

[23] M. P. Lawton, K. Van Haitsma, and J. Klapper, 'Observed Affect in Nursing Home Residents with Alzheimer's Disease', *J. Gerontol. B. Psychol. Sci. Soc. Sci.*, vol. 51B, no. 1, pp. P3–P14, 1996, doi: 10.1093/geronb/51b.1.p3.

[24] Y. Feng, G. Perugia, S. Yu, E. I. Barakova, J. Hu, and W. Matthias, 'Context Enhanced Human Robot Interaction: Exploring the Role of System Interactivity and Multimodal Stimuli on the Engagement of People with Dementia', *Int. J. Soc. Robot.*, vol. 14, no. 3, pp. 807–826, 2022, doi: 10.1007/s12369021008234.

[25] A. Tanner, A. Urech, H. Schulze, and T. Manser, 'Older Adults' Engagement and Mood During Robot-Assisted Group Activities in Nursing Homes: Development and Observational Pilot Study', *JMIR Rehabil. Assist. Technol.*, vol. 10, pp. e48031–e48031, 2023, doi: 10.2196/48031.

[26] W.-L. Chang, S. Sabanovic, and L. Huber, 'Observational study of naturalistic interactions with the socially assistive robot PARO in a nursing home', *Robot Hum. Interact. Commun.*, 2014, doi: 10.1109/roman.2014.6926268.

[27] H. Yoon, G. Shim, H. Lee, M.-G. Kim, and S. Kim, 'Observation of Human–Robot Interactions at a Science Museum: A Dual-Level Analytical Approach', *Electronics*, vol. 14, no. 12, p. 2368, 2025, doi: 10.3390/electronics14122368.

[28] A. Andrés, D. E. Pardo, M. Díaz, and C. Angulo, 'New instrumentation for human robot interaction assessment based on observational methods', *J. Ambient Intell. Smart Environ.*, vol. 7, no. 4, pp. 397–413, 2015, doi: 10.3233/ais-150331.

[29] A. L. Snow *et al.*, 'Development of a New Tool for Systematic Observation of Nursing Home Resident and Staff Engagement and Relationship', *The Gerontologist*, vol. 58, no. 2, pp. e15–e24, 2018, doi: 10.1093/geront/gnw255.

[30] J. W. Creswell and C. N. Poth, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications, 2016.

[31] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2017.

[32] A. A. Abdelrahman, D. Strazdas, A. Khalifa, J. Hintz, T. Hempel, and A. Al-Hamadi, 'Multimodal Engagement Prediction in Multiperson Human–Robot Interaction', *IEEE Access*, vol. 10, pp. 61980–61991, 2022, doi: 10.1109/ACCESS.2022.3182469.

[33] R. Bakeman and J. M. Gottman, *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge University Press, 1997.

[34] S. Pink and V. Fors, 'Ethnography, Stakeholders, and Audiences: Toward Openness and Inclusivity', *Sociol. Res. Online*, vol. 22, no. 4, pp. 169–173, Dec. 2017, doi: 10.1177/1360780417746029.

[35] A. Tausch, A. Kluge, and L. Adolph, 'Psychological Effects of the Allocation Process in Human–Robot Interaction – A Model for Research on ad hoc Task Allocation', *Front. Psychol.*, vol. 11, Sep. 2020, doi: 10.3389/fpsyg.2020.564672.

[36] M. D. LeCompte and J. J. Schensul, *Designing & conducting ethnographic research: an introduction*, Second edition. in The ethnographer's toolkit, no. 1. Lanham, Md: AltaMira Press, 2010.

[37] A. Dang and T. Manjunath, 'The value of observational study data in healthcare decision making: An Indian perspective', *Int. J. Med. Public Health*, vol. 6, no. 1, p. 53, 2016, doi: 10.4103/2230-8598.179778.

[38] M. B. Happ, 'Mixed Methods in Gerontological Research: Do the Qualitative and Quantitative Data "Touch"?', *Res. Gerontol. Nurs.*, vol. 2, no. 2, pp. 122–127, Apr. 2009, doi: 10.3928/19404921-20090401-06.