Invoking Nemo: Designing and Evaluating a Custom Voice Assistant for People with Intellectual and Developmental Disabilities

Madhuka Nadeeshani
Department of Human-Centred Computing
Monash University
Melbourne, Victoria, Australia
madhuka.koralalagedona@monash.edu

Jacqueline Johnstone
Department of Human-Centred Computing
Monash University
Melbourne, Victoria, Australia
jacqueline.johnstone@monash.edu

Kirsten Ellis
Department of Human-Centred Computing
Monash University
Melbourne, Vic, Australia
kirsten.ellis@monash.edu

Swamy Ananthanarayan
Department of Human-Centred Computing
Monash University
Melbourne, Victoria, Australia
swamy.ananthanarayan@monash.edu

Abstract

Voice assistants (VAs) have the potential to support more independent access to information and everyday tasks for people with intellectual and developmental disabilities (IDD) by enabling more natural and flexible forms of interaction. However, people with IDD encounter persistent barriers when using off-the-shelf VAs, including difficulties recalling wake words, understanding linguistically complex responses, fixed listening windows, and recovering from interaction errors. While prior customised VAs have shown promise in supporting specific tasks and domains for this population, less attention has been given to how such systems can support broader, unstructured interaction in practice. We present a customised tablet-based VA designed for adults with IDD to support open-ended interaction through multimodal access, adaptive listening, response complexity control, and interaction breakdown recovery. The VA was developed through 20 user-centred design sessions involving 11 adults with IDD, with input from support coaches, and evaluated through a three-week deployment across four community sites. Participants engaged with the VA across 161 logged queries spanning diverse topics, alongside use of multimodal interaction pathways including voice, typing, gestures, and button-based controls, demonstrating growing independence in use and supporting one another in sustaining interaction. Our findings highlight that accessible and sustained VA use was supported by a relatable character-based design, flexible interaction pathways, adaptive support for individual capabilities, and peer involvement during use.

CCS Concepts

• **Human-centered computing** → **Accessibility systems and tools**.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
ASSETS '26, Vila Nova de Gaia, Portugal
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2521-0/26/10
https://doi.org/10.1145/3797867.3828995

Keywords

intellectual and developmental disability, intellectual disability, voice assistants, conversational agents, user-centred design, accessibility

ACM Reference Format:
Madhuka Nadeeshani, Jacqueline Johnstone, Kirsten Ellis, and Swamy Ananthanarayan. 2026. Invoking Nemo: Designing and Evaluating a Custom Voice Assistant for People with Intellectual and Developmental Disabilities. In *The 28th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '26), October 25–28, 2026, Vila Nova de Gaia, Portugal.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3797867.3828995

1 Introduction

Voice assistants (VAs) have the potential to support greater independence for people with intellectual and developmental disabilities (IDD) by providing voice based access to information and everyday services without requiring complex navigation, extensive literacy, or precise fine motor interaction [11, 15, 30]. However, most off-the-shelf VAs are designed primarily for mainstream users, and their interaction mechanisms often fail to support the communication needs of people with IDD [11, 50]. Systems such as Amazon Alexa [2], Google Assistant [9], and Apple Siri [41] often struggle to recognise the atypical speech patterns [15, 32, 34] and slower speech rates [11, 34] that are common among people with IDD. Their responses can also be lengthy or linguistically complex, making them difficult for some users to follow [11, 34]. When user speech is misrecognised, these systems typically offer limited support for recovery, often generating responses based on the incorrect interpretation rather than supporting users to correct the query [11]. Importantly, these challenges have been observed even after extended periods of use [34, 43, 50], suggesting that they reflect limitations in current VA design rather than a lack of user experience.

These limitations have motivated growing research interest in developing customised VAs tailored to the needs of people with IDD. Some studies extend off-the-shelf VA platforms through custom skills such as Alexa Skills [7] or Google Dialogflow [25] within existing ecosystems [5, 26, 37]. However, such approaches remain

constrained by the underlying platform architecture, limiting control over core interaction behaviours such as wake word activation, listening window duration, and breakdown handling [6, 53]. To address these constraints, recent works have explored bespoke VAs that provide greater flexibility over interaction design [3, 14, 47]. Many of these VAs focus on specific domains such as health care [18, 47], educational support [4, 10], and communication training [3, 33], supporting task specific interactions rather than the open-ended information seeking.

Despite this progress, important gaps remain in the design and evaluation of VAs for people with IDD. Many bespoke VAs have been shaped primarily by researchers [10, 36], experts [3], or caregivers [37], with limited direct involvement of people with IDD in design decisions. Evaluations have also often been short term, sometimes limited to a single session, or conducted with proxy participants rather than people with IDD [10, 47], providing limited insight into how these systems support sustained use. These gaps highlight the need for VA design grounded in the direct input of people with IDD and evaluated through a real-world setting. To better understand these challenges, this paper is guided by the following research questions:

- **RQ1:** What interaction needs and preferences for VA emerge from people with IDD and their support coaches?
- **RQ2:** How can these needs and preferences inform the design of a custom VA for people with IDD?
- **RQ3:** How do people with IDD engage with a custom VA during technology, social and creative focused activities, and what barriers and supports influence interaction?

To address these questions, we developed a customised tablet-based VA for adults with IDD through 20 user-centred design (UCD) sessions [40] involving 11 adults with IDD and interviews with their support coaches. The custom VA was designed to be recognisable and relatable, easy to activate through the participant-selected wake word, and able to provide responses tailored to individual comprehension levels through a backend based on a large language model [1]. This backend allows the system to support open-ended, user-directed queries across diverse topics while adapting response complexity to users' comprehension needs. The system supports multiple interaction pathways, including voice, typing, gesture, and a physical button, allowing users to engage in ways that suit their abilities and context. We evaluated the prototype through a three-week deployment across four sites (96 hours) in a community-based technology introduction program in which adults with IDD engaged with emerging technologies as part of everyday skill-building and independence-focused activities. Following the deployment, we conducted semi-structured interviews with 10 adults with IDD and support coaches to examine their experiences of using the system.

This paper makes three contributions. First, we identify VA interaction needs and preferences through a user-centred design process involving people with IDD and their support coaches. Second, we present a custom VA designed in response to these findings, incorporating participant-selected wake word, adjustable response complexity and speech rate, multimodal interaction pathways, breakdown recovery support, and motivational feedback. Third, through a three-week deployment, we provide empirical insights into how people with IDD engage with a custom VA, highlighting the role of multimodality, social support, and adaptive interaction in sustaining use over time.

2 Related Work

In this section, we review prior work on VAs use by people with IDD, focusing on studies examining interaction with off-the-shelf VAs and research on customised VAs designed to address the barriers these systems present.

2.1 Off-the-Shelf VAs for People with IDD

Off-the-shelf VAs such as Amazon Alexa, Google Home, and Apple Siri have attracted growing research interest as accessible interaction technologies for underserved populations, including older adults [17, 28], people with visual impairments [16, 39], and people with physical disabilities [20, 35]. They typically operate through a wake word followed by a spoken query processed via automatic speech recognition and answered in real time. For people with IDD, VAs have been explored across everyday activities including information seeking [11, 34, 50], entertainment [11, 42], routine management [50], and smart home control [30]. Across prior studies, people with IDD have often reported preferring voice input over text based alternatives [11, 50]. This preference has been linked to the reduced spelling and literacy demands of voice interaction, with Balasuriya et al. [11] reporting that 72% of participants preferred voice input over typing. Longer term deployment studies [34, 43, 50] further suggest that these systems can support perceived independence, enjoyment of device use, and the maintenance of routines and autonomy in everyday contexts. Nadeeshani et al. [34] analysed interaction logs from two months of everyday VA use, showing that off-the-shelf VAs could support increasing independence, but persistent accessibility barriers remained in practice. Beyond task support, Smith et al. [42] demonstrated that sustained, naturalistic VA use produced measurable improvements in speech intelligibility among adults with IDD, with gains extending beyond device specific phrases to unrelated words.

However, off-the-shelf VAs interaction models are built around assumptions that often do not align with the communication needs of many people with IDD. Prior work shows that users may struggle to recall or pronounce wake words reliably [11, 32, 34], require repeated activation attempts [15, 30, 34], and need to formulate requests in ways the VA can interpret, rather than in ways that reflect their natural communication styles [11, 32, 34]. Off-the-shelf VAs often perform poorly when users' speech diverges from the clearer and more fluent patterns on which these systems typically rely [11, 32, 34]. Limited listening windows compound these difficulties further, giving users only a short window to complete their query, a factor that has itself been identified as an independent predictor of interaction failure [11, 32].

Accessibility challenges also extend beyond query initiation, as prior work shows that responses generated by off-the-shelf VAs can be difficult for some people with IDD to process and follow, particularly when they are lengthy or linguistically complex [11, 30, 34]. This suggests that accessibility challenges extend across both initiating interaction and making sense of system responses. Difficulties also persist when interactions breakdown, as off-the-shelf VAs typically require users to repeat or restart queries rather than offering

alternative ways to proceed [15, 30, 34]. As a result, users may abandon interactions, thereby limiting opportunities for independent use [30]. Repeated failures may also have broader emotional consequences where Cha et al. [15] found that adolescents with autism reported frustration and reduced confidence following unsuccessful interactions.

These findings suggest that interaction breakdowns arise not only from speech recognition limitations, but also from a mismatch between the interaction demands of off-the-shelf VAs and the communication needs of many people with IDD. Prior deployments suggest that these barriers persist even with extended use, indicating that they are structural rather than simply a matter of user familiarity [34, 43]. However, limited research has examined how VA interaction can be adapted to better support initiation, response accessibility, and recovery from breakdowns in everyday use.

2.2 Towards Customised VAs for People with IDD

In response to these limitations, researchers have explored customised VAs for people with IDD across domains including communication training [3, 14], education [4, 10], games [26, 52], and daily routine support [37]. Broadly, this work has followed two directions, extending off-the-shelf VA platforms through domain specific skills, and developing fully bespoke VAs.

Work extending off-the-shelf VA platforms has primarily focused on adding simplified interaction mechanisms tailored to the needs of people with IDD, such as step-by-step prompting, personalised routines, and structured feedback within existing environments such as Amazon Alexa Skills [10, 26, 37]. For instance, Alexism uses step-by-step prompting and personalised feedback to support oral care [37], MotivateMe uses a structured study support interaction informed by visual support principles to scaffold task completion [10], and Social Adventure uses interactive, narrative based design to support social skills development [26]. While this work demonstrates the value of tailoring prompts, routines, and feedback to scaffold specific activities, these systems remain constrained by the underlying platform. As a result, they inherit core interaction limitations such as defined wake word formats, fixed listening windows, limited support for recovery, and speech recognition challenges for users whose communication styles diverge from system expectations [10, 26, 37].

To address these constraints, other work has explored fully bespoke VAs that provide greater flexibility over interaction design. These systems allow researchers to move beyond platform imposed constraints and introduce a wider range of interaction and support mechanisms, including multi-topic conversational flows for communication practice [3, 14], immersive and motivating training environments [22], integrated support functions such as health monitoring and configurable interfaces [18, 36], and multimodal interaction designs that combine speech with visual and interactive elements [3, 24]. These approaches demonstrate that bespoke systems can support more flexible, multimodal, and engaging forms of interaction compared to off-the-shelf platforms. However, despite this increased flexibility, many bespoke VAs remain centred on predefined use cases and structured dialogue flows, limiting interaction to narrow, researcher defined domains [3, 4]. As a result, they provide limited insight into how people with IDD engage with broader, open-ended interaction in everyday contexts. Supporting such use requires systems that can respond flexibly to diverse queries while ensuring that responses remain accessible and understandable. More broadly, prior work has largely customised how support is delivered within specific tasks, rather than reconfiguring the underlying VA interaction model to better support initiation, response accessibility, and breakdown recovery in real-world use.

Building on this work, our study extends prior research by directly involving people with IDD throughout the design process, in contrast to many systems shaped primarily by researchers, clinicians, or caregivers [3, 37]. It also moves beyond fixed use cases and scripted interaction flows by designing for open-ended, user directed interaction. In addition, it integrates accessibility considerations identified in prior work, including response complexity, adaptive listening, multimodal interaction, and breakdown recovery, within a unified system evaluated through sustained real-world deployment with people with IDD.

3 Methods

We employed a two-stage study design to identify interaction preferences, inform the design of a customised VA, and examine its use among people with IDD (Figure 1). The first stage involved interviews with support coaches and UCD card-based activities with people with IDD, which guided the development of the prototype. The second stage involved deploying and evaluating the prototype within the program to identify factors that enabled and constrained interaction in practice, drawing on deployment observations, system logs, and interviews with people with IDD and their support coaches.

Study context: The study was conducted in partnership with a local disability support organisation that delivers group programs for adults with IDD across 20 sites within a single geographic region, supporting over 500 participants. The research team has an ongoing collaboration with the organisation through its community based technology introduction program, in which participants engage with technologies such as Bee-Bots [49] and Spheros [46] alongside other activities that support everyday skills and independence building. This collaboration includes contributing activity materials and volunteering with the program, with the authors contributing over 400 hours of volunteer time before the study.

Across four sites, support coaches facilitated sessions lasting six hours per day, four days per week. Off-the-shelf VAs, including the Google Nest Hub and Amazon Echo Show, were already available within the program before the study began, and participants had used them as part of regular program activities. Within this setting, participants used the customised VA across a range of self-directed questions and activities, enabling us to examine VA use in practice across diverse interests and everyday interactions.

Participants: Participants were adults with IDD enrolled in the disability support organisation's technology introduction program. Eleven participants (5 male, 6 female; mean age = 25.82 years) contributed to the UCD design sessions, and 10 participants (5 male, 5 female; mean age = 25.80 years) took part in the deployment and evaluation phase (Table 1). Of the 10 participants in the evaluation phase, eight had previously participated in the UCD sessions,

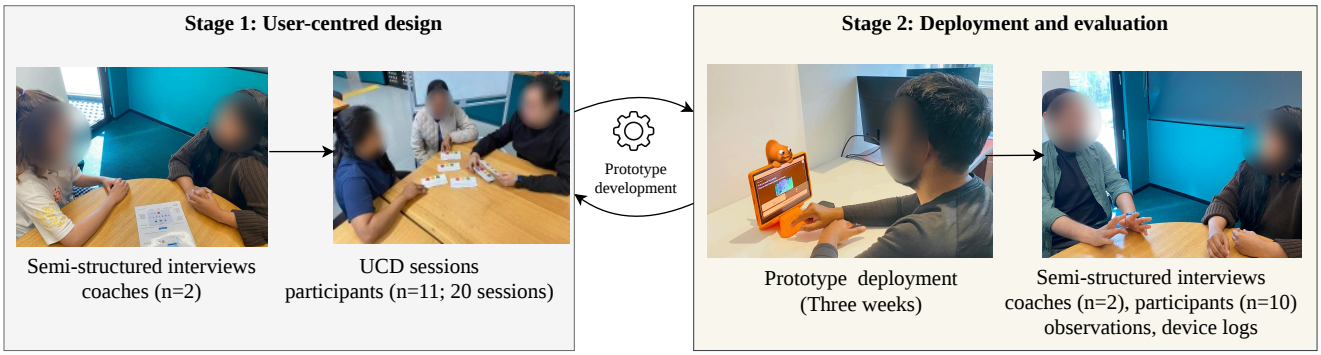

Figure 1: Overview of the two-stage study. Stage 1 involved semi-structured interviews with coaches and UCD sessions with participants, informing the development of the custom VA. In Stage 2, the prototype was deployed for three weeks and evaluated through observations, device logs, and post-study interviews with participants and coaches.

while two were newly enrolled and had not been involved in the design process. This difference in group composition reflects the organisation's rolling enrolment model, with some participants transitioning to other programmes or leaving during the study period. In our study, all participants were classified by the organisation as having lower support needs, indicating they could engage with minimal assistance, follow simple instructions, and complete short tasks with occasional reminders. All participants used English as the primary language of communication within the program, and all UCD sessions and deployment interactions with the custom VA were conducted in English.

Coaches: Coaches were professional support workers employed by the disability support organisation who facilitate the technology introduction program across four sites (Table 2). The coaches rotated between sites throughout the week to oversee program activities and participated in both UCD sessions and the deployment study.

Ethical considerations: Ethics approval for the study was obtained from the university human research ethics committee and disability support organisation. To support informed consent, study information was provided in accessible formats, including a plain language information sheet with images, a step-by-step verbal explanation, and a short video overview. All participants were able to provide their own informed consent to take part in the study. Families and carers were informed about the study to support transparency, communication, and participants' decision-making, but they did not provide consent on behalf of participants. Prior to each session, support coaches conducted brief comprehension checks to confirm participants' understanding of the study and their willingness to participate. These checks involved asking simple questions about the study purpose and participants' roles, ensuring that consent remained informed and voluntary throughout the study.

3.1 Stage 1: UCD Process (RQ1)

3.1.1 Coach Interviews. We began with semi-structured interviews with two coaches to understand the interaction challenges and contextual factors shaping VA use in practice within the program, based on their prior experience supporting participants' use of off-the-shelf VAs to address RQ1. Coach interviews lasted approximately 13 to 24 minutes. In line with prior research, coaches also identified recurring barriers related to interaction initiation and accessibility, including difficulties recalling and pronouncing wake words [11, 30, 34], responses that were too long or linguistically complex to follow [32, 34], and challenges associated with limited listening windows [11, 34] and diverse speech patterns, including stammering [11, 34, 42]. C1 and C2 also highlighted the importance of alternative interaction pathways beyond voice, such as typing and gesture based input, noting that some participants experienced difficulties with voice interaction in different contexts. In addition, coaches highlighted the role of motivational support [10, 34] and suggested that a familiar, recognisable character design could help sustain engagement over time [44]. Beyond these accessibility concerns, coaches emphasised that the VA should support queries across a wide range of topics rather than limiting users to fixed options. Coach interview data were also analysed using the inductive codebook thematic analysis approach described in Section 3.2.2, with insights informing the development of the UCD session materials.

3.1.2 UCD Sessions. We conducted 20 UCD sessions across four sites, with five sessions at each site over six weeks. The design activities were informed by both the coach interviews and our ongoing engagement with participants in the broader program context, which helped us identify interaction preferences and shape the accessible materials used in the sessions. Prior work on UCD with people with IDD emphasises the value of visual, low-text artefacts for supporting communication and reducing cognitive load [12, 21]. In line with this, we developed cards combining images and minimal text to support accessible preference expression and exploration of design options.

The information, images, and examples included on the cards were informed by coach interviews, prior literature, and the research team's extended engagement with the program before the UCD sessions, including over 400 hours of volunteer time. This engagement helped researchers become familiar with participants' routines, communication preferences, interests, and topics they commonly enjoyed, which informed the development of accessible

Table 1: Participant Demographics and Study Participation

Participant ID	Site	Age	Gender	Self-Reported Disability	UCD Sessions	Evaluation
P1	Site 1	26	M	ASD	1,2,3,4,5	N
P2	Site 1	20	F	IDD	1,2,3,4,5	N
P3	Site 1	21	M	IDD	1,2	N
P4	Site 1	23	M	IDD, Cerebral Palsy	3,4,5	Y
P5	Site 2	31	F	IDD, ASD	1,2,3,4,5	Y
P6	Site 2	24	M	IDD, Down syndrome	1,2,3,4,5	Y
P7	Site 2	32	M	IDD, ASD	1,2,3,4,5	Y
P8	Site 3	28	F	ASD	1,2,3,4,5	Y
P9	Site 3	26	F	IDD	1,2,3,4,5	Y
P10	Site 3	25	F	ASD, Partial Trisomy 9p	1,2,3,4,5	Y
P11	Site 4	28	F	ASD	1,2,3,4,5	Y
P12	Site 1	22	M	ASD	–	Y
P13	Site 2	19	M	ASD, IDD	–	Y

UCD sessions indicate participation across sessions 1–5; evaluation indicates deployment participation (Y = yes, N = no)

Table 2: Coach Demographic and Study Participation

Coach ID	Site	Gender	Disability Experience	UCD Sessions	Evaluation
C1	Site 1–4	F	5+ years	1,2,3,4,5	Y
C2	Site 1–4	F	3 years	1,2,3,4,5	Y

UCD sessions indicate participation across sessions 1–5; evaluation indicates deployment participation (Y = yes, N = no)

and relevant card content for discussion. The cards were used as discussion prompts rather than fixed design decisions, and participants could adapt the options or add their own ideas during the sessions. Sessions were typically held in small groups of around three participants, depending on attendance, and each structured activity lasted approximately 10–30 minutes.

Beyond these structured sessions, the authors spent extended time within the program throughout the six weeks (150 hours), attending full-day sessions and engaging with participants and coaches in everyday activities. This time helped the researchers understand participants' routines, communication practices, and support contexts, which informed how the UCD activities were facilitated, interpreted, and refined across sessions. These observations were used to support the design process and were not analysed as a separate field-note dataset.

Across the UCD sessions, we explored key aspects of VA interaction design, including wake-word use, response complexity, motivational preferences, alternative input methods, and visual appearance. We first elicited participants' experiences with existing VAs through open-ended discussions, focusing on interaction challenges and preferences. These insights then informed accessible, card-based design activities that supported more structured exploration of potential design directions. The full card set represented candidate features such as wake words, response complexity levels, feedback mechanisms, multimodal inputs, and appearance designs. Figure 3 shows examples of the wake-word and response-complexity cards used during these activities. To support participants in understanding the proposed VA features, researchers introduced each card using interaction scenarios, spoken examples, and demonstrations of how the feature would work on the tablet. For activities related to speech input, response complexity, and motivational feedback, researchers read example questions and responses aloud, compared different response styles, and explained how the custom VA would listen, reply, or provide encouragement. For gesture-based activities, researchers demonstrated each gesture, described its intended function, and asked participants what they thought the custom VA would do in response. This allowed researchers to clarify misunderstandings before asking for participants' preferences. Blank cards were also provided during the activities to allow participants to contribute additional ideas beyond the prepared options. The outcomes of these sessions informed key design directions in the development of the custom VA (section 4).

3.2 Stage 2: Deployment and Evaluation (RQ3)

3.2.1 Deployment Procedure. The system was implemented as a working prototype that supported core interaction functions, including wake-word detection, speech recognition, query processing, response generation, response simplification, and multimodal input and output. The prototype was then deployed within the program to examine its use in a community program setting. The deployment took place across four program sites for three weeks. Each site participated in three deployment days, resulting in a total of 12

deployment sessions. Sessions took place within the broader six-hour program day, although participants engaged with the custom VA intermittently rather than continuously throughout this period. Participants interacted with the prototype either individually or in small groups of two to three (Figure 2). The device was placed centrally on a table, and participants were free to ask questions of their own choosing, enabling exploratory use. A researcher was present during sessions to provide assistance when needed.

3.2.2 Data Collection and Data Analysis. We collected data through a combination of device usage logs, researcher notes, and semi-structured interviews. Device usage data captured interaction activity across deployment sessions, while researcher notes documented relevant contextual observations during use. Following the deployment, semi-structured interviews were conducted with participants and support coaches to gather reflections on their experiences using the custom VA.

Device Usage Data: The backend logged system usage data throughout the deployment, including user queries, system responses, input modality, response latency, listening duration, speech confidence, utterance characteristics, and interaction settings such as reading level and speech speed. These logs were analysed to examine interaction patterns, modality use, speech production characteristics, and system behaviour across participants and sessions.

Participant queries were further analysed using topic-based content categorisation to examine the range of information participants sought over time. Two authors reviewed all logged queries to determine an appropriate level of categorisation. Given the specificity of many requests, a simplified scheme was adopted to capture the general intent of each query. For example, *"Show me an image of Dora"* was classified as Cartoon; *"Show me a video of a Mazda car"* was classified as Vehicles, resulting in 13 categories. Backend log measures, including query counts, input modality, speech confidence, disfluency scores, listening-window duration, and interaction settings, were analysed descriptively to examine usage patterns across participants and sessions.

Observational Data. Field notes were recorded during each deployment session to capture verbal and non-verbal behaviours, including gestures, facial expressions, and vocalisations, as well as contextual details such as peer collaboration and instances of interaction support. These notes were reviewed alongside usage logs and interview transcripts to provide contextual support for interpreting interaction patterns, including instances of assisted use, peer involvement, and how participants engaged with the system in practice.

Interview Data. Semi-structured interviews with participants and support coaches were conducted at the end of the deployment to capture reflections on usability, accessibility, and interaction over time. Participant interviews lasted approximately 10–12 minutes, and coach interviews 10–18 minutes; all interviews were audio recorded.

Interview data were analysed using an inductive codebook thematic analysis approach [13]. This approach supported data-driven coding while enabling coding consistency and coordinated interpretation across the research team. To develop the initial coding framework, two authors independently coded one participant interview transcript and one coach interview transcript, then discussed their coding decisions and refined the framework. The first author then conducted open coding across the remaining transcripts, generating descriptive codes closely grounded in the data. Two authors reviewed these codes and developed candidate higher-level themes, which were discussed and refined with the research team. During these meetings, the team reviewed coded excerpts, discussed coding differences, and clarified, renamed, or consolidated codes as the analysis progressed. When new codes were identified during later stages of analysis, the first author returned to earlier transcripts to annotate additional relevant instances, supporting consistent application across the dataset. This process moved from descriptive codes to higher-level thematic groupings that captured broader patterns across participants. For example, low-level codes such as *"wake-word not recognised," "gesture recognition failure,"* and *"slow system response"* were grouped under the higher-level code *"system reliability barriers."*

4 Stage 1: UCD Findings and Design Outcomes (RQ1, RQ2)

This section presents the interaction needs and preferences identified through the UCD process, and the design outcomes that followed from them. Drawing on data from adults with IDD and support coaches, and connecting these findings to challenges identified in prior work, we describe how participants' experiences, preferences, and support contexts shaped key design decisions for the custom VA, including its wake word, response complexity, motivational feedback, multimodal interaction pathways, and visual appearance.

4.1 Designing an Accessible Wake Word

Participants (P1, P5, P9, P11) and coaches (C1, C2) consistently described unfamiliar wake words as effortful and difficult to recall. These findings suggest that interaction initiation should be simpler and more memorable for people with IDD, consistent with prior work identifying wake word recall and pronunciation as recurring barriers [11, 15]. Participants and coaches expressed a preference for single-word wake words, as these reduced recall and pronunciation demands in practice. Familiar cartoon character names were particularly favoured during early discussions and were therefore explored further in the design activities. From these suggestions, the most commonly raised options were selected and turned into wake word cards for use across sites.

Each card featured a cartoon image on the front and the corresponding name on the back (*Max*, *Nemo*, *Dora*, *Peppa*, and *Lunar*). During the wake-word activity, participants were also given the option to use blank cards to propose character ideas beyond the prepared options. For example, P12 suggested *"stitch"* by showing an image of the character, and this suggestion was later shown to other participants in subsequent sessions to explore whether it reflected broader preferences. Across the 11 participants, Nemo emerged as the most frequently preferred wake word, selected by eight participants as their first choice. Participants' explanations suggested that wake word preferences were shaped primarily by familiarity, ease of pronunciation, and personal connection to the

Figure 2: Two participants interacting with the customised VA during a deployment session (A) collaboratively typing a question, and (B) viewing VA response.

character. For example, P1 selected *"Nemo"* because he liked fish and recognised the familiar clownfish character associated with the name, describing it as *"easy to pronounce."* P9 likewise linked her preference for *"Nemo"* to prior familiarity with the character, noting that she had seen the clownfish character at her nephew's home. These findings suggest that accessible wake words should not only be short and easy to pronounce, but also familiar and personally meaningful to participants.

4.2 Designing Accessible Responses

Lengthy responses were described by participants (P3, P6, P7) and coaches (C1, C2) as difficult to follow, particularly when they included too much information or unfamiliar wording, echoing challenges commonly encountered in responses from commercial VAs. These accounts suggested that responses needed to be shorter and easier to process, consistent with prior work showing that responses from commercial VAs can be difficult for some people with IDD to understand [11, 32, 34]. To examine how response complexity shaped comprehension, we prepared response cards using a Lexile-inspired readability approach [29]. Each card responded to the same prompt, *"What is an apple?"*, but varied in linguistic complexity across 11 levels, with Level 1 using the shortest and simplest wording and Level 11 using longer sentences and more complex vocabulary. Participants also compared text-only answers with versions that presented the same answers alongside accompanying images.

All participants consistently preferred responses that combined answers with images, suggesting that visual support aided understanding. They also showed a clear preference for shorter and easier-to-understand responses. Seven participants selected Level 1, while one participant each selected Levels 2, 3, and 4. No participant selected cards above Level 4, and these higher-level responses were generally treated as too complex because they introduced longer explanations and less familiar vocabulary. This shaped how response complexity was translated into the prototype, with the custom VA using five response-complexity settings rather than the full 11-level card set, prioritising simpler responses while still allowing adjustment for participants who preferred more detailed answers. Participants consistently justified these choices in terms of comprehension. For example, P1 preferred a simpler response because *"I can understand that."* More complex responses were often rejected when they introduced unfamiliar terms. Some participants explicitly identified words such as *"vitamins"* and *"antioxidants"* as difficult, with P8 noting, *"I don't know about vitamins."* These findings suggest that responses should be calibrated to the comprehension level of the individual, with simpler language and shorter sentences as the default for most participants.

Beyond linguistic complexity, coaches identified response presentation as a key design consideration. Participants compared text-only answers with versions paired with images, and all consistently preferred the latter. They reported that images supported understanding, with P8 noting *"pictures and sounds help to learn."* Coaches emphasised visually simple presentation, recommending one sentence per line and cautioning that too many images or overly dense descriptions could increase confusion and cognitive load. These findings suggest that accessible responses should combine spoken, visual, and textual elements rather than rely on a single modality.

4.3 Designing for Engagement and Motivation

Participants (P1, P5, P8, P9) and coaches (C1, C2) preferred feedback that acknowledged successful interaction, rather than responses that simply provided information. They noted that its absence in previously used VAs made interaction feel less engaging and sometimes discouraged further questioning. These responses suggest that feedback and encouragement were important for sustaining engagement, consistent with prior work incorporating feedback mechanisms in bespoke VAs [3, 10, 26].

To explore these preferences further, we presented participants with a range of feedback options, including celebration animations, text-based praise such as *"Well done,"* *"Great job,"* and *"Awesome,"* and star or badge-based rewards. Participants consistently described these forms of feedback as enjoyable, motivating, and affirming, and many expressed a preference for multiple types rather than a single fixed option. For example, P4 selected the *"Well done"* celebration card, explaining that *"I feel it's like an appreciation and I like any of them"* and noting that she would like to receive any positive feedback. P9 preferred a celebration style response due to its familiarity from computer games, stating, *"That's cool. I get this kind of message when I'm playing computer games too."* P6 similarly

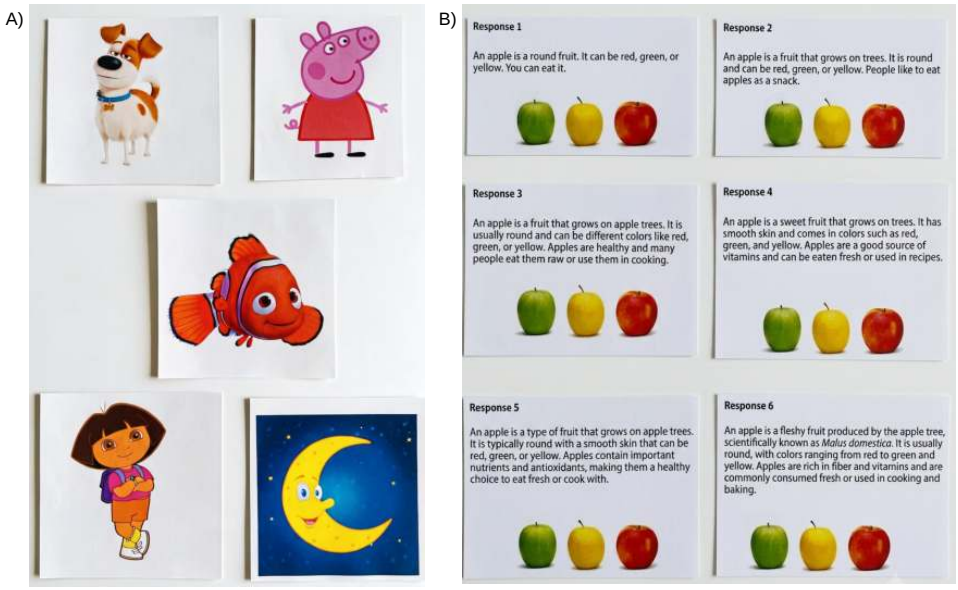

Figure 3: Cards used with participants during UCD sessions. (A) Wake-word cards featuring familiar characters are used to explore preferences for interaction initiation. (B) Response cards presenting answers at different levels of linguistic complexity are used to examine comprehension and preference.

associated such feedback with recognition, noting, *"I like different types of motivations"* and *"I like it when people notice what I've done."* These findings indicate that participants valued feedback that was positive, affirming, and varied. In response to these preferences, the prototype incorporated a range of feedback and motivational cues.

4.4 Supporting Multimodal Interactions

A voice-only interface was considered insufficient to support the range of communication abilities, preferences, and situational contexts across participants. This concern was raised by both participants (P5, P8, P11) and coaches (C1, C2) during the design sessions, consistent with prior work highlighting the importance of multimodal access [3, 26, 34]. While some participants were comfortable asking questions verbally, others described situations in which speaking was less convenient, comfortable, or accessible. C1 and C2 also noted that even participants who were comfortable using voice did not always prefer it. This points to the need for systems that accommodate changing preferences and abilities rather than assuming a fixed preferred modality.

Participants and coaches identified several alternatives to speech, including gestures, typing, and a physical button. These alternatives were discussed as complementary interaction pathways that could support users when voice input was not preferred. Gesture-based interaction was seen as particularly useful for participants whose speech was inconsistent or who needed a non-verbal way to initiate or control interaction. C1 described gestures as a useful option for participants whose speech was inconsistent, while P1 responded enthusiastically to waving, stating, *"Yes, I do like waving, it's like saying hi."* P9 likewise highlighted the accessibility value of gesture input, noting that *"for fully non-verbal people this is useful."*

Before finalising the gesture cards, the research team considered additional options, including a pointing gesture and clap. These options were not presented in the UCD activity because coaches advised that they may be difficult for some participants with limited motor control to perform reliably. We therefore presented three candidate gestures through printed cards: a wave gesture for activation, a palm-facing gesture towards the camera to stop the assistant, and a thumbs-up gesture to request a joke. These gestures were selected for their relative simplicity, low motor demand, and visual distinctiveness. Participants responded positively to gesture-based input, often describing the gestures as easy to understand, engaging, and familiar. Preferences were broadly consistent across participants, with all participants preferring a wave gesture for activation, eight preferring a palm-facing gesture to stop the assistant, and seven preferring a thumbs-up gesture to request a joke. These responses suggest that gestures were valued not as a replacement for speech, but as an additional interaction pathway that could support more flexible access.

Typing and button-based input were also identified as important alternatives to speech and gesture interaction. C1 proposed a physical button and typing option as simple and, reliable input methods, particularly in situations where gestures might not be recognised or where users preferred a more direct and controllable form of interaction. P5 similarly noted, *"I'm fine asking questions, but this could be good for some people who might not be able to speak."*

In response to these preferences, the prototype incorporated gesture input, a physical button, and a typing option to support more flexible interaction.

4.5 Designing the Prototype's Visual Appearance

During the design sessions, participants (P7–P10) and coaches (C1, C2) suggested that the tablet design should visually align with the participant-selected wake word and its associated clownfish character. Participants searched for and showed the research team character images on their own devices, indicating specific visual cues they wanted the custom VA to reflect, such as a bright orange colour scheme and cartoon-like appearance. This indicated that visual appearance was valued not only as a decorative feature, but also as a way of making the custom VA more familiar, recognisable, and approachable, consistent with prior work on character-based and engaging designs [44].

To explore these preferences, we presented three cards showing different ways of incorporating the character into the tablet design, positioned at the top of the tablet, along the sides, or as a more minimal image on the back. The design with the character positioned at the top of the tablet was preferred by seven of the ten participants, suggesting that participants valued a design in which the character was clearly visible during interaction. Participant responses further indicated that recognisability and character visibility shaped these choices. For example, P6 reacted positively to the design, stating, *"Oh, that's interesting."* P7 responded positively to the orange colouring because it looked like Nemo, while P8 explained, *"I like the one like the Nemo cartoon."*

5 System Overview (RQ2)

The custom VA is a tablet-based system designed to support open-ended, everyday interaction for people with IDD. Building on needs and preferences identified through the UCD process, the prototype was developed with a familiar wake word, adjustable response complexity, multimodal interaction pathways, engagement and motivational cues, a recognisable character-based design, and adaptive listening support. The system was deployed on a Samsung Galaxy Tab S9, selected based on coach feedback that participants were familiar with larger screens and that these supported readability and ease of interaction. The physical prototype included a 3D-printed cover and stand incorporating a clownfish-inspired character design, reflecting participants' visual appearance preferences from the UCD sessions.

The custom VA supports multiple complementary interaction pathways, including voice, typing, gestures, and a physical button, allowing users to switch modalities depending on context, preference, and ability. Its typing-based options include prompted queries, content retrieval such as YouTube search, and open-ended input through a *"type your own"* function. Users can activate the assistant by saying the participant-selected wake word *"Nemo"*, performing a wave gesture, or pressing an external button connected via Bluetooth, and can then ask questions by speaking to the VA. Responses are delivered through synchronised speech and on-screen text, with optional visual support through images. For example, a participant might activate the assistant, ask *"What is a rainbow?"*, and receive a spoken and written response tailored to their selected readability level, optionally accompanied by a relevant image (Figure 4).

The custom VA follows a client-server architecture that supports open-ended, user-directed interaction across diverse topics. To move beyond predefined intents and scripted dialogue flows, it uses a backend based on a large language model for response generation, enabling flexible queries while maintaining accessibility through a readability control pipeline that regulates response complexity. A React Native frontend on Android manages multimodal input, interface rendering, and text-to-speech output, while a Python FastAPI backend handles streaming speech recognition, response generation, readability control, and session management. Wake-word activation is supported by the Picovoice Porcupine engine [38], configured with the participant-selected custom keyword *"Nemo"*. A physical button provides an additional option for explicit interaction initiation.

Gesture interaction is implemented through a native Android module built on MediaPipe [31], which classifies hand landmark configurations in real time using the device's front-facing camera. The system supports two static gestures, an open palm that returns the interface to the home screen and provides a consistent exit mechanism, and a thumbs-up gesture that triggers a predefined joke request to encourage engagement. A third gesture, wave, is detected using a custom temporal algorithm that tracks lateral wrist displacement across consecutive frames and initiates voice recording. These gestures were selected based on their simplicity and recognisability during the UCD activities.

Speech input is processed through a streaming pipeline and transmitted incrementally via WebSocket to Google Cloud Speech-to-Text. UCD findings indicated that participants often paused mid-utterance, repeated words, or exhibited irregular speech patterns, which can lead to premature termination in standard voice interfaces. To address this, the system implements a dynamic listening window that adapts silence thresholds based on a lexical repetition score derived from interim transcripts and a timing-based score computed from word-level speech recognition timestamps. These signals are combined into a composite disfluency score, extending the silence threshold by up to 2500 ms and allowing users additional time to complete their input. The initial listening-window thresholds were informed by prior work on VA interaction barriers for people with IDD and coach input about participants' speech patterns, and were then examined in practice through backend logs capturing listening-window duration, speech confidence, utterance timing, and disfluency scores. Post-processing further improves transcription robustness by collapsing repeated tokens and removing filler words. When speech confidence is low, the system presents alternative interpretations as selectable options, allowing users to correct recognition errors without restarting the interaction.

For response generation, the system uses GPT-4o-mini through the OpenAI API. When the system could not confidently interpret a participant's request, it generated a clarification prompt rather than producing an answer based on uncertain input. This design aimed to support interaction repair and reduce the likelihood of generating responses from misrecognised or ambiguous queries. Design sessions also highlighted substantial variation in reading and comprehension abilities, motivating a five-level readability control pipeline. Each level defines constraints over sentence length,

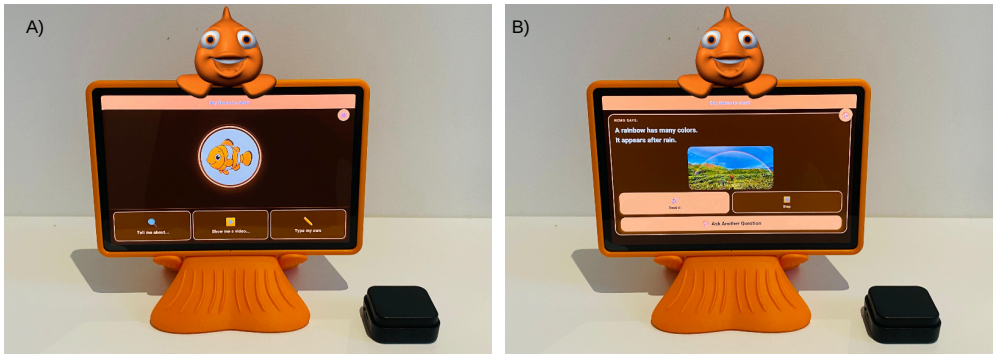

Figure 4: Custom VA interface and physical prototype. (A) Home screen supporting multiple input pathways, including prompted queries, video retrieval, open-ended text input, voice activation, and a physical button. (B) Response screen showing a simplified text response, accompanying image, synchronised text-to-speech, and controls to continue or end the interaction.

vocabulary frequency, and response structure, enabling systematic variation in linguistic complexity. Lower levels prioritise short, direct responses using high-frequency vocabulary, while higher levels allow more detailed explanations and varied phrasing. Generated responses are evaluated using a similar Lexile proxy metric based on mean sentence length and word frequency scores derived from the wordfreq library [45], and are iteratively adjusted to align with the selected level.

For example, the same query could produce different conversational outputs depending on the selected readability level. For the query *"What is a rainbow?"*, a lower readability level could produce a short response such as *"A rainbow is colours in the sky. It can happen when sunlight shines through rain."* A higher readability level could produce a more detailed response, such as *"A rainbow forms when sunlight passes through raindrops in the air. The light bends and separates into different colours, which is why we see a colourful arc in the sky."* The prototype also supported short back-and-forth repair interactions when speech recognition confidence was low. For example, after a low-confidence transcription of a short query, the system could ask, *"did you mean Dora?" or "did you mean door?"* and display the two clarification options on the interface for the participant to select. The participant could also repeat the query or type the intended question, allowing the interaction to continue rather than ending after a recognition failure.

Response complexity and speech rate could be manually selected through the interface by participants, researchers, or support coaches, allowing responses to be adjusted to individual comprehension and pacing needs during use. Other supports operated automatically, with the listening window adapting to detected pauses and disfluency, while clarification prompts were triggered when recognition confidence was low. The base listening window was also linked to the selected speech-speed setting, so slower speakers received proportionally extended interaction time. To enrich responses and support comprehension, topic-relevant images are retrieved via the Wikipedia REST API and displayed alongside responses with automatically extracted captions.

6 Stage 2: Deployment and Evaluation Findings (RQ3)

Analysis of the deployment data for RQ3 identified four themes describing how participants used the custom VA, multimodal flexibility as an adaptive communication strategy, familiar character design as a support for engagement, scaffolding independence through support and adaptive use, and resilience and openness in interacting with emerging technologies.

6.1 Multimodal Flexibility as an Adaptive Communication Strategy

Participants interacted with the custom VA in flexible ways, shifting between voice, typing, gestures, and button-based input depending on what was most accessible, efficient, or comfortable in the moment. P5 noted, *"Sometimes voice, sometimes typing... it depends on the environment"*, and further explained, *"if it's noisy I prefer typing"*. P8 similarly described how fatigue influenced this choice, stating, *"When I'm tired, I like typing. But I also like talking to them"*. Usage logs demonstrated this pattern across the three-week deployment: text input accounted for 68% of queries in Week 1 and 54.5% in Week 3, while speech use increased from 32% to 45.5%, suggesting that participants became progressively more confident with voice interaction over time.

Multimodal interaction extended beyond input preference to include how participants initiated, navigated, and concluded interactions. Participants (P4, P6, P8, P10) did not rely on a single activation method. Some activated the system by speaking the participant-selected wake word *"Nemo"*, others waved to activate the system, and others pressed a physical button before voicing their query. In several instances, modalities were combined across different stages of the same interaction rather than used in isolation. For example, participants (P5, P9, P11) sometimes used a gesture or button press to initiate interaction, and then continued through voice to complete their query. Gestures were also used beyond activation, such as using a thumbs-up gesture to request jokes or a stop gesture to end an interaction. This shows that gestures operated not only as alternatives to speech, but also as complementary controls within a broader multimodal interaction structure.

Preferences further varied according to participants' communication abilities, motor capabilities, and support needs. Some participants expressed greater confidence in typing, while others preferred speaking. P5 noted, *"Both are good. It depends on what someone is comfortable with and what they need help with."* Coaches further highlighted how communication ability and motor constraints shaped what was feasible in practice. For instance, coach C1 explained, *"P10 can speak, but recently her speech has not been very clear. In such cases, she could use typing and gestures instead of voice."* Coaches also noted that not all participants could reliably perform specific gestures due to motor limitations, meaning that simpler or more accessible alternatives were often required. Importantly, the value of multimodality extended beyond individual preference to support inclusion in group settings. Coaches emphasised that *"relying on a single interaction mode could exclude some users"* particularly non-verbal participants or those with differing communication abilities. Providing multiple modalities allowed participants to engage in ways that suited their abilities, helping ensure that interaction remained accessible to a wider range of users within shared activities.

6.2 Familiar Character Design as a Support for Engagement

Participants engaged with the custom VA in ways that reflected familiarity, enjoyment, and sustained interest throughout the deployment. Their accounts suggest that the assistant was experienced not only as a source of information, but also as an enjoyable and encouraging presence during interaction. For example, P8 described the assistant as *"awesome"* and *"cool,"* adding that *"he tells me a lot of things."* P7 similarly responded enthusiastically to the system's affirming feedback, stating, *"Yes, a lot, he says I'm a superstar."* Coaches C1 and C2 also observed strong positive reactions to the system, noting that participants *"love Nemo."* These accounts suggest that participants engaged with the custom VA not only as a functional tool, but as a familiar and approachable character.

This familiarity was also reflected in participants' physical and embodied interactions with the system. Some participants, such as P2 and P7, gently touched or patted the interface and physical cover while referring to it by name, suggesting that the character was not only recognised, but also approached with a sense of comfort and familiarity. Coaches further noted that using a familiar character name made the system easier to recognise, remember, and relate to in practice. These observations suggest that the character-based design contributed to how approachable the system felt during use.

Beyond affective responses, engagement was also evident in patterns of use over time. As shown in Table 3, 10 participants used the custom VA across the deployment, generating 161 logged queries in total. Of these, 135 queries were successful (83.9%), and 26 were unsuccessful (16.1%). Because some interactions involved multimodal input, peer support, or support from coaches and researchers, not every logged query could be reliably attributed to a single participant. We therefore report weekly active participants together with aggregate query counts, rather than participant-level query totals.

Query volume rose from 41 queries in Week 1 to 76 in Week 2, before decreasing to 44 queries in Week 3. No participant discontinued use of the custom VA during the deployment, and differences in weekly active participation reflected program attendance rather than withdrawal from the study. Importantly, the proportion of successful queries remained high across all three weeks, increasing from 80.5% in Week 1 to 84.2% in Week 2 and 86.4% in Week 3. This suggests that participants continued to use the system successfully over time, even though the number of active participants and total queries varied across weeks. As shown in Table 4, participants asked about a wide range of topics, with jokes and animals being the most frequent categories across the deployment. This suggests that engagement was exploratory and interest-driven rather than limited to a narrow set of predefined tasks.

The system's affirming feedback appeared to reinforce this engagement by making interaction feel encouraging rather than purely task-oriented. Participants (P5–P11) responded positively to praise and affirmation during interaction, and C1 contrasted this with existing communication tools, which were described as repetitive and less engaging. C1 explained, *"Right now, he uses a communication pack where he presses the same buttons repeatedly, which is not very encouraging or engaging. A system like this could encourage more interaction and exploration."* These findings suggest that engagement with the custom VA was shaped not only by access to information, but also by how the system was presented and experienced. A familiar character design, encouraging feedback, and opportunities for open-ended exploration together supported continued and self-directed interaction across the deployment.

6.3 Scaffolding Independence Through Support and Adaptive Use

Participants' independent use of the custom VA developed gradually through repeated use and was supported by both system features and the surrounding social context. Both participants and support coaches emphasised that initial guidance was important when participants were first introduced to the system. As coach C1 explained, *"you cannot just give it to them and expect them to use it immediately... with proper guidance and regular use, they will be able to learn and use it independently."* Participants also described increasing confidence and ease of use as the deployment progressed. Several participants noted, *"I can use it by myself"* (P5–P9, P12, P13), while also reflecting that *"it's a lot easier now"* (P5, P8–P13). By the end of the second week, P10 was independently searching for topics of interest without support from coaches or peers. Support also emerged between participants, with some helping peers continue interactions by assisting with typing or spelling words, for example, entering terms such as *"butterfly"*.

Mean listening-window duration decreased across the deployment, from 17.4 seconds in Week 1 to 14.0 seconds in Week 2 and 12.8 seconds in Week 3. As the system employs an adaptive listening window that extends in response to pauses and disfluency, this measure reflects not only participants' interaction timing but also the extent to which the system needed to prolong the listening period. The decline over time therefore suggests that fewer such extensions were required as the deployment progressed. This interpretation is supported by the concurrent decrease in mean

Table 3: Deployment usage summary across the three-week deployment.

Week	Active participants	Total queries	Successful queries, n (%)	Unsuccessful queries, n (%)
Week 1	6	41	33 (80.5%)	8 (19.5%)
Week 2	7	76	64 (84.2%)	12 (15.8%)
Week 3	6	44	38 (86.4%)	6 (13.6%)

Table 4: Distribution of query categories across the three-week deployment.

Category	Week 1	Week 2	Week 3	Total
Jokes	14	26	7	47
Animals	12	26	6	44
Sports	3	1	5	9
Games	–	2	6	8
Vehicles	–	2	5	7
Food	–	–	5	5
Books	2	2	–	4
Weather	–	3	–	3
Cartoon	2	–	–	2
Science	–	2	–	2
Festive Activities	6	5	4	15
General Knowledge	–	3	2	5
Other	2	4	3	9
Total	**41**	**76**	**44**	**161**

disfluency scores (0.458, 0.397, 0.381), indicating that participants produced fewer pauses or disruptions during interaction. While this does not imply broader changes in speech ability, it suggests that participants' spoken interaction with the system became more efficient or stable with repeated use.

Adaptive features also supported this progression by allowing interaction to be tailored to different abilities and preferences. Participants and coaches highlighted the value of adjustable response complexity, speech rate, and readability, noting that *"everyone has different reading levels"* (P5) and *"people process information at different speeds"* (P9). Use of settings also reflected different levels of independence and support. For example, P8 consistently selected higher response complexity levels, indicating a preference for more detailed answers, whereas P4–P6 and P11–P13 relied on coaches to configure settings on their behalf. C1 similarly explained that configurable settings allowed some participants to adjust the system independently, while coaches could support others by configuring settings when needed.

This use of configurable settings was also reflected in the logs. Across the three-week deployment, 87 setting updates were recorded. Of these, 45 were made independently by participants, while the remaining 42 were made with support from coaches or researchers. This suggests that some participants were able to personalise the system independently, while others benefited from supported configuration during use. All five response complexity levels were used across participants, showing that different levels were needed in practice rather than simply being available as options. Lower levels were selected more often overall, with Level 2 emerging as the most frequently used setting because it provided simple responses, while Levels 4 and 5 were consistently used by participants who preferred more complex responses. Speech rate showed a similar pattern of calibration. Although the slower rate was used most often overall, some participants gradually shifted toward standard or faster rates over time as they became more comfortable with the system. These patterns show that the system's adaptive features supported personalised interaction in ways that reflected diverse comprehension needs and changing levels of confidence.

Breakdown recovery features also helped sustain interaction without requiring participants to restart. The system incorporated an automatic error correction mechanism that resolved misrecognised or misspelled queries by inferring the intended query and producing a corrected response. Rather than returning a generic error, this allowed the interaction to continue while reducing the burden on the participant. P8 responded to a successful auto-correction with evident satisfaction, noting *"I like that and suddenly it worked"*. This account suggests that automatic error correction can not only resolve interaction breakdowns, but also support a sense of accomplishment that reinforces motivation and encourages continued engagement.

6.4 Resilience and Openness Toward Emerging Technologies

Participants encountered several forms of interaction breakdown during the deployment, including inconsistent wake-word recognition, speech misrecognition, gesture detection failures, and response latency. Of the 26 unsuccessful interactions, pronunciation-related issues were the most frequent source of breakdown (n=11),

followed by ambiguous queries (n=9), with fewer cases of unwanted activations (n=3), listening-window-related issues (n=2), and stammering-related difficulties (n=1). Participants also encountered occasions where wake-word attempts did not successfully trigger recording, further disrupting interaction flow. P8 noted that wake-word activation was not always reliable, *"sometimes it's not working"*, and also commented on response delays, *"it's just slow"*. P11 similarly described gesture recognition as inconsistent, stating, *"I like wave but it's not working sometimes"*. Coaches likewise observed that wake-word detection and gesture commands did not always perform reliably across sessions. Despite these difficulties, participants continued to engage with the custom VA. When asked whether slow or failed responses caused frustration, P8 replied, *"I like Nemo"*. Participants consistently demonstrated persistence in their interactions, even when the system failed to respond as expected.

This persistence was also reflected in observable interaction behaviour. When gestures were not recognised, participants repeated them rather than abandoning the attempt, and when the wake word failed, they repeated it until the system responded. P5 made this explicit, stating, *"I will tell this until I get this thing worked"*, highlighting a willingness to persist through system failure rather than disengage. Query volume nearly doubled from Week 1 to Week 2 despite these early difficulties, suggesting that breakdowns did not discourage continued use. Rather than withdrawing, participants continued returning to the custom VA across sessions, and several also expressed interest in using it beyond the deployment. Some participants also compared the custom VA positively with their prior experiences with Google Home. For example, P5 stated *"Nemo is better than Google."* P11 linked this preference to the prototype's alternative input options, noting, *"Nemo is better than Google because I can type if it does not hear me."* These comparisons reflect participants' perceived experiences rather than a direct comparative evaluation.

7 Discussion

Our findings suggest that accessible VA use for people with IDD depends not only on system performance, but also on how interaction is supported and sustained in practice. This section addresses RQ3 by examining participants' use of the custom VA and the barriers, supports, and opportunities that shaped sustained interaction over time.

7.1 Resilience and Social Support Reframe Accessible VA Use

A key finding from our deployment is that accessible VA use for people with IDD depended not only on technical reliability, but also on whether interaction could continue through moments of breakdown. Participants remained willing to engage with the custom VA even when interaction did not proceed smoothly, often retrying, switching modalities, or continuing with support from coaches and peers when wake-word activation failed, gestures were not recognised, or responses were delayed. This contrasts with prior work showing that interaction failures can lead to frustration, disengagement, and reduced confidence for people with IDD [15], and has also been observed in other accessibility-focused contexts, including work with older adults [39, 48] and blind and low vision people [19]. In our deployment, interaction was sustained over time despite such difficulties.

This points to the importance of recoverability in accessible VA use, referring to whether interaction can continue when breakdowns occur. In our deployment, participants were often able to continue through modality switching, correction prompts, and alternative response pathways rather than being forced to restart or abandon the interaction. They moved between voice, typing, gestures, and button-based input when one mode became difficult, while the system also supported recovery by presenting alternative interpretations when speech recognition confidence was low. Prior work on accessible VA design has often focused on improving recognition accuracy and reducing breakdowns [5, 34]. Our findings extend this work by showing that the effects of breakdowns depend not only on whether they occur, but also on whether participants are supported with meaningful pathways for continuation and repair.

At the same time, resilience in our deployment was closely tied to the social context in which the custom VA was used. Independent use did not emerge in isolation, but through sustained support from coaches and peers who gradually introduced features, adjusted settings when needed, and helped participants continue interacting through moments of difficulty. Participants also supported one another, for example by helping with spelling, pronunciation, or feature use. These patterns suggest that, in this setting, continued VA use often benefited from social support not only as a stepping stone to independence, but also as an ongoing and complementary part of sustained use. This aligns with prior work showing that technology use for people with IDD and related populations is shaped by peers, supporters, and broader social contexts rather than by isolated individual interaction alone [34, 50]. More broadly, this suggests that how VAs are introduced may matter alongside how they are designed, with deployment in group settings and support from peers and coaches potentially contributing to confidence and sustained use.

7.2 Adaptive Configuration for Accessible VA Use

Our findings suggest that accessible VA use for people with IDD could not be achieved through a single fixed interaction configuration. Rather than relying on a one-size-fits-all interaction model [51], accessible use depended on making interaction adjustable in practice to accommodate differences in how participants provided input, received responses, and paced interaction. This aligns with prior work showing that generic VA interaction models often do not accommodate the diversity of communication, literacy, and comprehension needs among people with IDD [11, 34]. In the custom VA, this was operationalised through adjustable response complexity and listening-window settings that could be adapted to different speaking styles and interaction needs.

These adjustments mattered in practice because no single configuration was appropriate for all participants or all situations. Participants and coaches used different response complexity levels during deployment, indicating that preferences for readability

varied across participants and contexts. Similarly, changes to listening duration supported participants who needed more time to formulate queries. Settings logs further supported this pattern of active personalisation, with 87 setting updates recorded across the three-week deployment and all five response complexity levels used across participants. Prior work has similarly highlighted the value of personalised and supporter-mediated interaction for people with IDD, showing that supporters can play an important role in tailoring systems to users' needs, for example by adding routines, reminders, or navigation support [8, 36]. However, such configuration is often handled through separate settings panels or caregiver-controlled interfaces [8, 36]. The custom VA addressed this by making settings such as response complexity and speech rate adjustable during use, enabling personalisation to occur within the interaction itself, either by participants directly or with support from coaches when needed.

Personalisation was therefore valuable not because it produced a single ideal configuration for each user, but because it allowed the system to remain usable across changing needs, contexts, and levels of support. While this was feasible in the supported deployment context, it may be more difficult in less structured settings. Future systems could move toward more automatic forms of personalisation by recognising users' voice and interaction patterns and adjusting settings such as response complexity and listening duration accordingly, while still allowing participants and supporters to review and control those changes.

7.3 Multimodal Interaction as a Necessity for Accessibility

Prior work with people with IDD has highlighted the value of voice interaction, particularly in reducing literacy and fine-motor demands [11, 30]. Prior studies have also shown that voice-only interaction can be difficult to sustain in practice, particularly when speech production, timing, or recognition becomes challenging [11, 23, 34]. Our deployment similarly suggests that voice alone was insufficient to support accessible VA use in practice. Even participants who were comfortable using voice did not rely on it exclusively, instead turning to typing, gestures, or button-based input as contexts, fatigue, and access needs changed. Importantly, this was not only a matter of different users preferring different modalities. The same participant might comfortably use voice in one moment and prefer typing later in the session when fatigued or when the environment changed. Accessibility therefore depended not only on offering multiple modalities, but on enabling flexible switching between them as individual needs evolved throughout the day.

Alternative modalities did not merely replace voice. They allowed interaction to continue while participants built confidence with voice over time. This was reflected in usage patterns across the deployment, where speech use increased progressively while typing, gestures, and button-based input remained actively used, suggesting that alternative modalities supported rather than replaced voice interaction over time. In several instances, participants also combined modalities within the same interaction, for example using a gesture or button to initiate interaction and then continuing through voice to complete their query. Our findings therefore suggest that accessible VA use depends not only on the availability of voice input, but on the ability to shift flexibly between interaction pathways.

Beyond individual use, multimodality also mattered in shared community settings, where interaction often involved peers and support coaches. In these contexts, multiple input pathways helped accommodate different communication abilities, preferences, and support needs within the same activity. Rather than competing with one another, voice, typing, gestures, and button-based input functioned as complementary ways of maintaining access across changing user needs and shared settings. At the same time, presenting all modalities at once may increase cognitive load for people with IDD. Future systems may support flexible switching while foregrounding the most appropriate options for a given user and context, enabling more intuitive and context-sensitive modality selection for people with IDD in diverse community settings.

7.4 Character Familiarity and Encouragement as Accessibility Supports

Our findings suggest that accessibility in VA use for people with IDD depends not only on functional support, but also on whether interaction feels familiar, recognisable, and encouraging. Prior work has incorporated motivational feedback into VA systems for people with IDD, for example to support study engagement [10], suggesting that these experiential qualities can play an important role in sustaining use. Prior work has largely focused on task-specific systems and on functional barriers such as speech recognition and initiation demands [32, 34], with comparatively less attention to how familiarity and encouragement function in open-ended, everyday VA use [27]. Our deployment extends this perspective by showing that character design and motivational feedback, including celebration animations and positive affirmations such as *"Well done"* and *"Great job"* can function as accessibility supports in a fully autonomous VA, rather than as peripheral aesthetic features.

Participants engaged with the custom VA not only as a functional source of information, but also as a character they could recognise and relate to. This resonates with prior work on character-based assistive systems for people with IDD, such as RoboParrot [44], which suggests that socially recognisable and responsive character forms can help support engagement. In our deployment, positive responses to the participant-selected character suggest that familiarity made the system feel more approachable and meaningful in use. Importantly, this effect cannot be explained simply by the presence of a character within the interface. Rather, its value appeared to lie in the fact that the character was recognisable, familiar, and meaningful to participants, having been selected and refined through the design process. This familiarity made interaction feel more approachable and encouraged continued use.

This familiarity, together with the system's affirming feedback, appeared to support continued engagement over time. Coaches observed that other tools participants used were often experienced as repetitive or disengaging, suggesting that sustained interaction depended not only on whether the system could provide information, but also on whether it felt rewarding and encouraging to use. Encouragement was embedded not only in explicit feedback cues, but also in the overall experience of interacting with a familiar and responsive character. Our findings therefore suggest that character

design and motivational feedback can contribute to accessibility by reinforcing successful interaction and making repeated use feel more inviting. At the same time, these elements should be designed with users rather than introduced generically, ensuring that they are familiar, meaningful, and contextually appropriate, particularly in adult and shared settings where excessive encouragement or inappropriate character choices may not support engagement in the same way. A fixed character design may also not sustain engagement indefinitely. Future systems could therefore consider allowing character personalisation, enabling users to adapt or change the character over time as their preferences evolve.

8 Limitations and Future Work

This work demonstrates the potential of a customised VA to support more flexible and accessible interaction for people with IDD, while also highlighting practical considerations that shaped its design and use in deployment settings. The UCD process relied on structured, card-based activities to support accessible participation. While these cards were developed based on participant and coach input, they may still have foregrounded a predefined set of interaction options, potentially limiting the range of preferences that could be expressed. Although participants were able to raise ideas beyond the prepared options, some participant-generated interface ideas, such as on-screen topic buttons for common interests, are not considered in the current prototype. Future work could explore how such shortcuts can support access to familiar topics while still preserving open-ended interaction.

The deployment was designed to examine how participants used and experienced the custom VA in a community setting, rather than to compare its performance directly with commercial VAs. The deployment also highlighted technical challenges in wake-word detection, speech recognition, gesture recognition, and response latency, which affected interaction flow in some situations. Future work should examine direct comparisons with existing VAs using both interaction outcomes and technical measures of reliability, latency, and breakdown recovery.

Although we did not identify hallucinated responses during deployment, LLM-generated responses can still introduce such risks. The absence of observed issues may reflect the types of queries participants asked, which were mainly everyday information or content-search questions rather than inference-based questions requiring reasoning or analysis of user-provided data. Future work should examine response accuracy and safety in contexts involving more reasoning-based, or decision-relevant queries.

Participants in this study had lower support needs, meaning our findings may not directly generalise to people with higher support needs or those with very limited verbal communication. Although the prototype supported multiple input pathways, future work should examine how customised VAs can be adapted for a broader range of communication abilities, support needs, and interaction contexts, including users who may require more extensive non-verbal or supporter-mediated access pathways.

9 Conclusion

This work presents a customised VA designed to support accessible, open-ended interaction for people with IDD. Through a UCD process with adults with IDD and their support coaches, we identified key barriers in existing VA interaction, including difficulties with wake-word recall, complex responses, limited support for recovering from errors, limited motivational feedback, and limited support for non-verbal users. These insights informed the design of the prototype, which incorporates the participant-selected wake word, adjustable response complexity, multimodal input, motivational feedback, character-based design, interaction repair mechanisms, and adaptable listening windows.

The three-week deployment demonstrated that participants engaged with the custom VA in diverse and exploratory ways, with modality preferences varying across users and contexts. Participants also demonstrated increasing independence over time, while peer support remained important in sustaining interaction when difficulties arose. Their continued engagement despite these difficulties reflects resilient interaction behaviour supported by the system's flexible design. These findings suggest that accessible VA design for people with IDD should move beyond fixed, speech-centric interaction models toward approaches that are adaptable, supportive of varied communication needs, and grounded in users' preferences. We hope this work contributes to the design of VAs that better support people with IDD in accessing information and participating more independently in everyday contexts.

Acknowledgments

We thank our participants and coaches from the disability support organisation for their time and insights. This research was funded by the Australian Government through the Australian Research Council (Project No. LP210300465). This work was also supported by the Monash University Faculty of Information Technology International Postgraduate Research Scholarship.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. (2023).

[2] Amazon Alexa. 2026. https://alexa.amazon.com/. Accessed: 2025-02-20.

[3] Mohammad Rafayet Ali, Seyedeh Zahra Razavi, Raina Langevin, Abdullah Al Mamun, Benjamin Kane, Reza Rawassizadeh, Lenhart K. Schubert, and Ehsan Hoque. 2020. A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, IVA 2020*. Association for Computing Machinery, Inc. doi:10.1145/3383652.3423900

[4] Sumayh Aljameel, James O'Shea, Keeley Crockett, Annabel Latham, and Mohammad Kaleem. 2019. LANA-I: An Arabic Conversational Intelligent Tutoring System for Children with ASD. In *Advances in Intelligent Systems and Computing*, Vol. 997. Springer Verlag, 498–516. doi:10.1007/978-3-030-22871-2_34

[5] Anna A. Allen, Howard C. Shane, and Ralf W. Schlosser. 2018. The Echo™ as a Speaker-Independent Speech Recognition Device to Support Children with Autism: an Exploratory Study. *Advances in Neurodevelopmental Disorders* 2, 1 (3 2018), 69–74. doi:10.1007/s41252-017-0041-5

[6] Amazon. 2024. Choose the Invocation Name for a Custom Skill. Alexa Skills Kit Developer Documentation. Accessed: 2025.

[7] Amazon Developer. 2024. Alexa Skills Kit. https://developer.amazon.com/en-US/alexa/alexa-skills-kit. Accessed: 2026-04-06.

[8] Lavanya Arya, Pranita Ranade, and Tanmoy Goswami. 2023. Intellectually Disabled Child Holistic Growth Management Platform: Voice Interaction and Scheduling. In *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*. 1–6. doi:10.1109/OTCON56053.2023.10113913

[9] Google Assistant. 2026. https://assistant.google.com/. Accessed: 2025-02-20.

[10] Ami Bakhai, Aurora Constantin, and Cristina Adriana Alexandru. 2020. MotivateMe!: An Alexa Skill to Support Higher Education Students with Autism. In *Proceedings of the 14th International Conference on Interfaces and Human Computer Interaction (IADIS I-HCI 2020)*. International Association for Development of the Information Society (IADIS), 187–191.

[11] Saminda Sundeepa Balasuriya, Laurianne Sitbon, Andrew A. Bayor, Maria Hoogstrate, and Margot Brereton. 2018. Use of voice activated interfaces by people with intellectual disability. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 102–112. doi:10.1145/3292147.3292161

[12] Laura Benton and Hilary Johnson. 2014. Structured approaches to participatory design for children with autism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 445–454.

[13] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3 (2021), 328–352. doi:10.1080/14780887.2020.1769238

[14] Fabio Catania and Franca Garzotto. 2023. A conversational agent for emotion expression stimulation in persons with neurodevelopmental disorders. *Multimedia Tools and Applications* 82, 9 (4 2023), 12797–12828. doi:10.1007/s11042-022-14135-w

[15] Inha Cha, Sung In Kim, Hwajung Hong, Heejeong Yoo, and Youn Kyung Lim. 2021. Exploring the use of a voice-based conversational agent to empower adolescents with autism spectrum disorder. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery. doi:10.1145/3411764.3445116

[16] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. "Nobody Speaks that Fast!" An Empirical Study of Speech Rate in Conversational Agents for People with Vision Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376569

[17] Melisa Duque, Sarah Pink, Yolande Strengers, Rex Martin, and Larissa Nicholls. 2021. Automation, wellbeing and digital voice assistants: Older people and Google devices. *Convergence* 27, 5 (2021), 1189–1206. doi:10.1177/13548565211038537

[18] Fatima Ezzahrae El Rhatassi, Btihal El Ghali, and Najima Daoudi. 2023. A Chatbot's Architecture for Customized Services for Developmental and Mental Health Disorders: Autism. In *Lecture Notes in Networks and Systems*, Vol. 669 LNNS. Springer Science and Business Media Deutschland GmbH, 134–141.

[19] Paola Esquivel, Kayden Gill, Mary Goldberg, S. Andrea Sundaram, Lindsey Morris, and Dan Ding. 2024. Voice Assistant Utilization among the Disability Community for Independent Living: A Rapid Review of Recent Evidence. *Human Behavior and Emerging Technologies* 2024, 1 (2024), 6494944. doi:10.1155/2024/6494944

[20] Anne Fota and Hanna Schramm-Klein. 2024. The Impact of Digital Voice Assistants on the Everyday Life of Physically Disabled Users: Barriers, Drivers and Potential for Improvement. (2024). https://scholarspace.manoa.hawaii.edu/items/1d981edf-b981-40fa-8dce-1f0680e205f4

[21] Christopher Frauenberger, Judith Good, Wendy Keay-Bright, and Helen Pain. 2013. Designing technology for children with special needs: Bridging perspectives through participatory design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 875–884.

[22] Brittany Garcia-Pi, Rohan Chaudhury, Miles Versaw, Jonathan Back, Dongjin Kwon, Caleb Kicklighter, Paul Taele, and Jinsil Hwaryoung Seo. 2023. AllyChat: Developing a VR Conversational AI Agent Using Few-Shot Learning to Support Individuals with Intellectual Disabilities. In *Human-Computer Interaction – INTERACT 2023: 19th IFIP TC13 International Conference, York, UK, August 28 – September 1, 2023, Proceedings, Part IV* (York, United Kingdom). Springer-Verlag, Berlin, Heidelberg, 402–407. doi:10.1007/978-3-031-42293-5_43

[23] Franca Garzotto, Mattia Gianotti, Alberto Patti, Francesca Pentimalli, and Francesco Vona. 2024. Empowering Persons with Autism Through Cross-Reality and Conversational Agents. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (May 2024), 2591–2601. doi:10.1109/TVCG.2024.3372110

[24] Mattia Gianotti, Alberto Patti, Francesco Vona, Francesca Pentimalli, Jessica Barbieri, and Franca Garzotto. 2023. Multimodal interaction for persons with autism: the 5A case study. In *International Conference on Human-Computer Interaction*. Springer, 581–600.

[25] Google. 2024. Dialogflow. https://cloud.google.com/dialogflow. Accessed: 2026-04-18.

[26] Stefan Greuter, Joanne Watson, and Susan Balandin. 2022. Social Adventure. *International Journal of Gaming and Computer-Mediated Simulations* 14, 1 (7 2022), 1–21. doi:10.4018/ijgcms.303107

[27] Syed Mahmudul Huq et al. 2024. Dialogue Agents for Artificial Intelligence-Based Conversational Systems for Cognitively Disabled: A Systematic Review. *Disability and Rehabilitation: Assistive Technology* 19, 3 (2024), 1059–1078. doi:10.1080/17483107.2022.2146768

[28] Sunyoung Kim. 2021. Exploring How Older Adults Use a Smart Speaker–Based Voice Assistant in Their First Interactions: Qualitative Study. *JMIR Mhealth Uhealth* 9, 1 (13 Jan 2021), e20427. doi:10.2196/20427

[29] Colleen Lennon and Hal Burdick. 2003. *The Lexile Framework for Reading Measurement and Success.* Technical Report. MetaMetrics, Inc., Durham, NC, USA.

[30] Lundy Lewis and Andre Vellino. 2021. Helping Persons with Cognitive Disabilities using Voice-Activated Personal Assistants. In *2021 8th International Conference on ICT and Accessibility, ICTA 2021*. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/ICTA54582.2021.9809777

[31] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. (2019).

[32] Fabio Masina, Valeria Orso, Patrik Pluchino, Giulia Dainese, Stefania Volpato, Cristian Nelini, Daniela Mapelli, Anna Spagnolli, and Luciano Gamberini. 2020. Investigating the Accessibility of Voice Assistants With Impaired Users: Mixed Methods Study. *J Med Internet Res* 22, 9 (25 Sep 2020), e18431. doi:10.2196/18431

[33] Emily Mower, Matthew Black, Elisa Floresa, Marian Willams, and Shrikanth Narayanan. 2011. Design of emotionally targeted interactive agent for children with autism. In *2011 IEEE International Conference on Multimedia and Expo*. doi:10.1109/ICME.2011.6011990

[34] Madhuka Nadeeshani, Jacqueline Johnstone, Kirsten Ellis, and Swamy Ananthanarayan. 2026. I Can Do It: Exploring Voice Assistants for Adults with Intellectual Disabilities. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, Article 987, 20 pages. doi:10.1145/3772318.3790573

[35] Kenichiro Noda. 2018. Google Home: smart speaker as environmental control unit. *Disability and rehabilitation: assistive technology* 13, 7 (2018), 674–675. doi:10.1080/17483107.2017.1369589

[36] Maria Papadogiorgaki, Nikos Grammalidis, Athina Grammatikopoulou, Konstantinos Apostolidis, Ekaterini S. Bei, Kostas Grigoriadis, Stylianos Zafeiris, George Livanos, Vasileios Mezaris, and Michalis E. Zervakis. 2023. An Integrated Support System for People with Intellectual Disability. *Electronics (Switzerland)* 12, 18 (9 2023). doi:10.3390/electronics12183803

[37] Parvaneh Parvin, Marco Manca, Caterina Senette, Maria Claudia Buzzi, Marina Buzzi, and Susanna Pelagatti. 2022. Alexism: ALEXa supporting children with autISM in their oral care at home. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 5 pages. doi:10.1145/3531073.3531157

[38] Picovoice Inc. 2024. Porcupine Wake Word Engine. https://picovoice.ai/platform/porcupine/. Accessed: 2026-03-28.

[39] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174033

[40] Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. *Interaction Design: Beyond Human-Computer Interaction.*

[41] Apple Siri. 2026. https://www.apple.com/siri/. Accessed: 2025-02-20.

[42] Elizabeth Smith, Petroc Sumner, Craig Hedge, and Georgina Powell. 2021. Smart speaker devices can improve speech intelligibility in adults with intellectual disability. *International Journal of Language and Communication Disorders* 56, 3 (5 2021), 583–593. doi:10.1111/1460-6984.12615

[43] Elizabeth Smith, Petroc Sumner, Craig Hedge, and Georgina Powell. 2023. Smart-speaker technology and intellectual disabilities: agency and wellbeing. *Disability and Rehabilitation: Assistive Technology* 18, 4 (2023), 432–442. doi:10.1080/17483107.2020.1864670

[44] Pegah Soleiman, Sohail Salehi, Maryam Mahmoudi, Morteza Ghavami, Hadi Moradi, and Hamidreza Pouretemad. 2014. RoboParrot: A robotic platform for human robot interaction, case of autistic children. In *2014 Second RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*. 711–716. doi:10.1109/ICRoM.2014.6990987

[45] Robyn Speer. 2022. wordfreq: Word Frequencies for Many Languages. https://github.com/rspeer/wordfreq. Accessed: 2026-03-28.

[46] Sphero Inc. 2024. Sphero Educational Robots. https://sphero.com/

[47] Franck Tarpin-Bernard, Joan Fruitet, Jean Philippe Vigne, Patrick Constant, Hanna Chainay, Olivier Koenig, Fabien Ringeval, Béatrice Bouchot, Gérard Bailly, François Portet, Sina Alisamir, Yongxin Zhou, Jean Serre, Vincent Delerue, Hippolyte Fournier, Kévin Berenger, Isabella Zsoldos, Olivier Perrotin, Frédéric Elisei, Martin Lenglet, Charles Puaux, Léo Pacheco, Mélodie Fouillen, and Didier Ghenassia. 2021. THERADIA: Digital Therapies Augmented by Artificial Intelligence. In *Lecture Notes in Networks and Systems*, Vol. 259. Springer Science and Business Media Deutschland GmbH, 478–485. doi:10.1007/978-3-030-80285-1_55

[48] Milka Trajkova and Aqueasha Martin-Hammond. 2020. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376760

[49] TTS Group. 2024. Bee-Bot Programmable Robot. https://www.tts-group.co.uk/bee-bot/

[50] E Van Wingerden, SV Vacaru, L Holstege, and PS Sterkenburg. 2023. Hey Google! Intelligent personal assistants and well-being in the context of disability during COVID-19. *Journal of Intellectual Disability Research* 67, 10 (2023), 973–985. doi:10.1111/jir.13064

[51] Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-Based Design: Concept, Principles and Examples. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*

(CHI '11). ACM, 243–252. doi:10.1145/1978942.1978973

[52] Lian Zhang, Ashwaq Z. Amat, Huan Zhao, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. 2021. Design of an Intelligent Agent to Measure Collaboration and Verbal-Communication Skills of Children with Autism Spectrum Disorder in Collaborative Puzzle Games. *IEEE Transactions on Learning Technologies* 14, 3 (6 2021), 338–352. doi:10.1109/TLT.2020.3029223

[53] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. 2019. Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems. In *2019 IEEE Symposium on Security and Privacy (SP)*. 1381–1396. doi:10.1109/SP.2019.00016