



“Triggering autonomy, not just automation”: Design implications for LLM-based home automation assistants for people with intellectual disability

Diego Morra^a , Margherita Andrao^b , Qi Ai^c , Marco Mores^d , Maristella Matera^{c,*} , Barbara Treccani^e , Massimo Zancanaro^e

^a Massachusetts Institute of Technology, Cambridge, MA, USA

^b Fondazione Bruno Kessler, Trento, Italy

^c Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

^d Faculty of Engineering, Free University of Bozen-Bolzano, NOI Techpark, Bolzano, Italy

^e Department of Psychology and Cognitive Science, University of Trento, Rovereto, Italy

HIGHLIGHTS

- Smart homes promise autonomy for people with Intellectual Disability (ID), but configuration remains challenging.
- First empirical study on ChatGPT-4o interpreting home automation for users with ID.
- 3-step clarification framework explored with 5 educators and 4 user-caregiver dyads.
- 12 design implications for accessible LLM-based EUD interfaces for users with ID.

ARTICLE INFO

Keywords:

End-user development
Intellectual disability
Large language models
ChatGPT
Task-automation systems

ABSTRACT

Smart home technologies offer considerable potential to enhance autonomy, safety, and quality of life for individuals with intellectual disabilities (ID). Yet, configuring and personalizing such systems remain challenging due to complex interfaces and the technical expertise required. Traditionally, End-User Development (EUD) is often proposed as a way for non-experts to autonomously configure home devices. This work investigates the extent to which Large Language Models (LLMs) can provide effective support to people with ID in applying EUD. It explores whether an LLM-powered conversational interface can support individuals with ID in shaping smart home behaviors through natural conversations. Insights from exploratory sessions involving users with ID, caregivers, and stakeholders demonstrate a positive attitude towards this approach and suggest that conversational interaction can foster accessibility of automation programming through iterative clarifications. We conclude by outlining design opportunities and broader implications that can inform future research towards LLM-based EUD interfaces for people with ID.

1. Introduction

Individuals with intellectual disabilities (ID) often face challenges in developing autonomy and managing daily activities independently, due to difficulties in problem-solving, reasoning, and verbal communication (American Psychiatric Association, 2013). Supporting their independence is a central goal in care and rehabilitation practice, with growing interest in how technology – particularly smart home and Internet of Things (IoT) systems – can contribute to this objective (Gervais et al.,

2023; Landuran et al., 2023; Lussier-Desrochers et al., 2012; Smith et al., 2023; Morra et al., 2024). Home automation can promote assistance in performing daily activities, in engaging more actively with the surroundings, and in contributing to personal safety (Gentry, 2009; Ahmed et al., 2016; Cleland et al., 2024a; Dewsbury et al., 2003; Semary et al., 2024; Dobkin and Dorsch, 2017; Lewis et al., 2023). Additionally, these systems can support caregivers by enabling remote monitoring, optimizing intervention management, and supporting controlled transitions toward independent living (Cleland et al., 2024b).

* Corresponding author.

Email address: maristella.matera@polimi.it (M. Matera).

<https://doi.org/10.1016/j.ijhcs.2026.103938>

Received 9 December 2025; Received in revised form 20 August 2026; Accepted 3 September 2026

Available online 10 September 2026

1071-5819/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Yet, these benefits are maximized when users can control their devices directly [Lieberman et al. \(2006\)](#), rather than relying solely on pre-configured setups. End-User Development (EUD) explores how novel approaches enable non-programmers to create, modify, or extend software systems [Paternò \(2013\)](#); [Ardito et al. \(2019\)](#); [Lieberman et al. \(2006\)](#); [Barricelli et al. \(2019\)](#); [Fischer \(2013\)](#); [Paternò and Santoro \(2019\)](#). In particular, trigger-action programming offers users a relatively easy approach to personalize their smart-home environments according to their specific needs and preferences by creating rules in which triggers, associated with specific events, activate automatic actions [Blase et al. \(2016, 2014\)](#); [Ghani et al. \(2017\)](#). However, despite these promises and advancements in technology, there is still a significant gap in making EUD systems accessible to a wider range of users, specifically those with ID [Ellis et al. \(2023\)](#). This gap stems primarily from the formal-language requirement of conventional EUD platforms. Even with relatively simple paradigms such as trigger-action programming, studies have documented that non-programmers struggle to reason correctly about how rules will behave [Brackenbury et al. \(2019\)](#); [Huang and Cakmak \(2015\)](#).

Natural Language Processing (NLP) has been proposed as a means to reduce the cognitive effort involved in learning how to define a system's functioning [Blackwell \(2002\)](#). Conversational interfaces allow users to interact with systems using natural language, reducing the complexity of EUD tasks [Gallo et al. \(2021\)](#). However, recent developments in Large Language Models (LLMs) have opened new opportunities [Tetteroo \(2023\)](#). LLMs such as GPT-4 are revolutionizing software development through code generation, enabling even non-programmers to generate ready-to-deploy code from natural language instructions [Andrao et al. \(2024\)](#). By accepting natural language as the input medium, LLM-based EUD could remove the formal-language requirement. This is particularly relevant as it encourages end-users, such as people with ID, to engage with technology in a similar fashion as they do with people, fostering their involvement in EUD. Nevertheless, LLM interaction still presents challenges [Perera et al. \(2015\)](#), including limitations in understanding contexts and specifying user intents [Scoccia \(2023\)](#); [Roffarello et al. \(2023\)](#). For this reason, whether the LLM promise will translate into a meaningful interaction for individuals with ID remains an open question.

1.1. Contribution

Recognizing the potential of LLMs to democratize users' involvement in EUD and aiming to maximize impact on home automation accessibility this paper investigates challenges and opportunities of LLM-based EUD systems in serving as an intent interpreter and code generator for trigger-action routines for people with ID. We organized the research into two multi-study phases that involved a total of 28 participants, including 18 individuals with ID, four caregivers, and six educators. The goal was to (i) assess the capabilities and limitations of the LLM in extracting intents – and generate the corresponding trigger-action code – from instructions written in natural language by individuals with ID; (ii) collect user and stakeholder insights from hands-on sessions with a prototype of an LLM-based EUD interface. This work addresses the following research questions:

- **RQ1:** Can an LLM-based EUD interface support users with ID in personalizing smart home automation through natural language interaction?
- **RQ2:** How do individuals with ID and caregivers perceive and engage with an LLM-based home automation assistant?

This work contributes to the design of inclusive interactive systems for home automation in three ways:

- It documents that individuals with ID primarily struggle with structural completeness and situational specificity when formulating automation rules, a pattern that existing EUD systems could fail to accommodate.

- It shows that LLM-mediated iterative clarification can effectively scaffold rule definition for people with ID, enabling them to refine ambiguous instructions into functional automation rules through guided conversational interaction.
- It provides design principles that extend prior EUD and assistive-technology research, contributing towards a scenario where LLMs can function as adaptive, conversational mediators rather than static rule editors.

These contributions address a documented gap between existing EUD approaches, whose logic assumptions often exceed the cognitive profiles of many users, and the emerging opportunity of LLM-assisted, multimodal, co-adaptive tooling for people with ID.

1.2. Paper organization

Section 2 reviews the related work on the adoption of home automation technologies and LLMs with an impact on people with ID. Section 3 illustrates the design of our multi-study process, while Sections 4 and Section 5 detail the organization and the key results of respectively the first phase, dedicated to evaluating ChatGPT prompt understanding, and the second phase addressing the design and evaluation of an LLM-based conversational agent. Section 6 discusses the results, highlighting emerging design implications and study limitations. Finally, Section 7 draws our conclusions and outlines our future work.

2. Background & related work

Intellectual disability (ID) is a neurodevelopmental disorder characterized by significant limitations in intellectual functioning and adaptive behavior ([American Psychiatric Association, 2013](#)). Individuals with ID often experience challenges in their daily lives, including difficulties in developing autonomy and managing cognitive tasks such as problem-solving, reasoning, and decision-making. Moreover, many individuals with ID experience physical and developmental challenges, including limited fine motor skills, hyperactivity, attention deficits, and impairments in both verbal and non-verbal communication ([American Psychiatric Association, 2013](#)). According to the *Diagnostic and Statistical Manual of Mental Disorders – 5th Edition* (DSM-5; [American Psychiatric Association \(2013\)](#)), ID is frequently comorbid with other conditions, including other neurodevelopmental disorders, genetic syndromes, and medical conditions. Consequently, ID encompasses a wide and highly variable sample with specific individual needs, varying according to the severity of the ID, the level of autonomy, and adaptive behavior.

2.1. Home automation impact on people with ID

Literature reports on IoT systems promoting health and well-being (for a review, see [Tian et al. \(2024\)](#); [Lui et al. \(2025\)](#)). In a systematic review of 163 studies on home automation systems designed to support older adults, Tian et al. [Tian et al. \(2024\)](#) identify several key areas of benefit, including improved mental and emotional well-being, support for healthier lifestyles, and an improvement in perceived well-being. For example, IoT-based wearable or environmental devices are instrumental in rehabilitation, ultimately leading to more personalized and effective care [Dobkin and Dorsch \(2017\)](#). A recent study explored the effectiveness of integrating IoT technologies in co-housing environments through the DOMHO project [Bacchin et al. \(2021\)](#), with professional caregivers ($n = 7$) viewing these systems as valuable tools for aiding their work, improving the autonomy and quality of life of users with disabilities. More recently, the PETAL project (*PERSONALIZABLE assistive Ambient monitoring and Lighting*) deployed smart home systems in the homes of older adults with mild cognitive impairments, enabling end-user created automations and investigating their remote monitoring in field trials [Manca et al. \(2022\)](#).

Other studies have investigated the impact of home-automation solutions on people with ID [Landuran et al. \(2023\)](#). Lussier et al. [Lussier-Desrochers et al. \(2012\)](#) evaluated home automation for assisting

individuals with ID in meal preparation. The study involved 12 participants who prepared recipes with and without technological assistance. Results showed that the use of home automation significantly decreased the required caregiver support, particularly in terms of verbal and gestural prompting, without affecting the time needed to complete the tasks. Other studies explore the impact of mainstream Intelligent Personal Assistants (IPAs), like *Amazon Echo* and *Google Home*, on individuals with ID [Smith et al. \(2023\)](#). For example, in a semi-randomized controlled clinical trial involving individuals with ID, [Smith et al. \(2023\)](#) provided IPAs to an intervention group ($n = 22$), but not to the control group ($n = 22$). Their findings illustrate that while participants generally reported positive experiences, improved agency, and enjoyment, IPAs did not show significant changes in well-being or agency at the group level, with reported challenges in speech recognition. However, while the studies described above report a positive impact of home automation on individuals with ID, this evidence focuses on the use of pre-configured systems rather than on how people with ID can themselves configure or personalize such systems.

2.2. LLMs for EUD in home automation

With the increased availability of inexpensive sensors and the accessible approaches promoted by *Do-It-Yourself* practices, there has been a growing interest among HCI researchers in involving individuals with ID in technology creation and customization [Rajapakse et al. \(2015\)](#); [Meissner et al. \(2017\)](#); [Sas and Neustaedter \(2017\)](#); [Morra et al. \(2024\)](#); [Senaratne et al. \(2022\)](#). However, while the availability of user-friendly interfaces for commercial microcontrollers for home-automation increases, specific programming skills are still required to define their functioning [Barricelli et al. \(2019\)](#).

EUD platforms that employ a trigger-action paradigm present a compelling approach due to the streamlined structure of the configuration rules, which link events and conditions to the intended outcomes [Blase et al. \(2014\)](#). Platforms like IFTTT¹ assist users in defining trigger-action rules, and contribute to narrowing the gap between non-programmers and professional developers [Paternò \(2013\)](#); [Rahmati et al. \(2017\)](#); [Paternò and Santoro \(2019\)](#); [Barricelli et al. \(2019\)](#). These services typically provide users with logical descriptions of events and conditions, abstracting low-level details from the raw data generated by sensors [Ghiani et al. \(2017\)](#). Research has also investigated the effectiveness of innovative visual paradigms and conversation-based approaches, thereby broadening the scope of user interaction with smart-home technologies [Andrao et al. \(2023\)](#); [Desolda et al. \(2017\)](#); [Roffarello et al. \(2023\)](#); [Gallo et al. \(2021\)](#); [Corno et al. \(2019\)](#).

Despite the progress in this area, the existing body of work on EUD for home automation specifically involving people with ID remains limited [Paternò and Santoro \(2019\)](#); [Barricelli et al. \(2019\)](#); [Carmien and Fischer \(2008\)](#), with available studies mainly focusing on IoT making and co-design activities with people with ID [Morra et al. \(2024\)](#); [Ellis et al. \(2023\)](#). Designing EUD solutions for individuals with ID presents specific challenges, mainly due to the documented difficulties these individuals face in reasoning and verbal communication [Coppens-Hofman et al. \(2017\)](#); [American Psychiatric Association \(2013\)](#).

With the advent of LLMs, tools with capabilities for natural language interpretation and code generation represent an innovative approach to ensuring that individuals with ID have an active role in creating and personalizing home-automation solutions. In addition to supporting the creation and management of code [Liu et al. \(2023\)](#); [Fakhoury et al. \(2024\)](#), embedding LLM capabilities into EUD systems could enhance the interpretation of user intents expressed in natural language. Accessibility of EUD systems can be increased by eliminating the need for users to adapt their language to the system's capabilities [Blackwell](#)

(2002). With this promise, LLM-based EUD tools could fundamentally address the limitations connected to reasoning and verbal communication associated with ID: rather than requiring users to conform to a system-specific formal language, they could accept and understand natural-language input and shift the burden away from learning a new notation toward refining and building on what users already express in their own words. However, the exploration of LLMs for improving interactions with smart-home systems has only recently sparked research interest [Rey-Jouanchicot et al. \(2025\)](#); [Rivkin et al. \(2024b\)](#). Despite the potential of these models, crucial aspects remain unexplored, such as assessing their capabilities to recognize user intents or ensuring effective and safe user interaction. Non-experts' requests may contain intrinsic ambiguities or under-specifications that these models might struggle to comprehend, which may even lead to their misinterpretation. These, in turn, may generate outcomes that do not align with the user's intent. Only recent studies have attempted to address these concerns by focusing on user input disambiguation [Evan et al. \(2024\)](#); [Rivkin et al. \(2024a\)](#). Another work explored a multi-modal approach that does not rely solely on textual input [Calò and De Russis \(2024\)](#), investigating the use of LLMs to improve interaction with smart-home systems by disambiguating commands using visual and textual inputs.

While these works represent an initial step in exploring how LLMs can handle ambiguities, they remain preliminary with respect to generalizing the models' abilities. This limitation hinders researchers' ability to assess the impact of integrating LLMs into EUD systems for users with ID.

2.3. Synthesis and research gap

Across prior work on home automation, assistive technologies, and EUD, a gap becomes evident: current systems assume that users can express conditions, events, and logical dependencies with sufficient completeness to generate executable rules. However, the documented challenges that people with ID face in reasoning and verbal communication ([American Psychiatric Association, 2013](#)) suggest that they may face challenges in satisfying these assumptions, producing instructions that may be incomplete, ambiguous, or pragmatically under-specified. To our knowledge, no prior work has investigated LLM-mediated clarification as a mechanism for enabling people with ID to author home-automation rules. To date, no study has investigated how people with ID directly interact with LLM-based conversational interfaces in applied settings. Adjacent work has only explored multi-modal conversational search with adults with ID in a Wizard-of-Oz setting [Roomkham et al. \(2022\)](#), and the use of GPT-4 to interpret image-based self-expression by people with ID [Saffar et al. \(2024\)](#). This paper addresses this gap through an empirical two-phase investigation that examines both the capabilities and the impact on interaction of LLM-based scaffolding for this population.

3. Study design

As illustrated in [Fig. 1](#), the study followed a sequential, two-phase structure designed so that each phase informed the next. Phase 1 evaluated ChatGPT's ability to understand ambiguous prompts generated by people with ID. Building on insights from Phase 1, in Phase 2, a functional prototype of an LLM-based EUD interface for home automation was designed and evaluated through two user studies: a workshop with expert educators and a hands-on study with dyads of users with ID and their caregivers. Together, the two phases informed a set of design implications for accessible LLM-based EUD systems for home automation tailored to people with ID. Specific procedural details, participant numbers, and the rationale for the choice of model are reported in the corresponding phase sections ([Sections 4 and 5](#)). The study received approval from the ethical committee of Politecnico di Milano, and was conducted with the permission and under the supervision of the participants' caregivers.

¹ <https://ifttt.com/>.

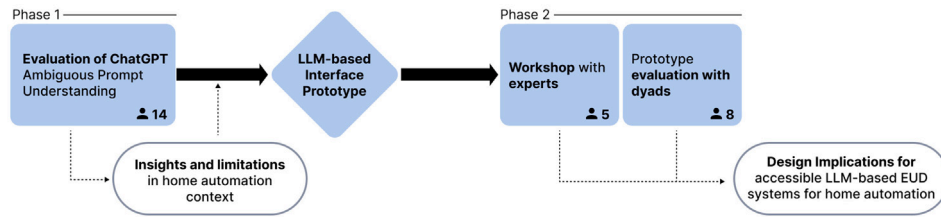


Fig. 1. Study design and phases.

Table 1

Phase 1 participants' demographic data, diagnoses, speaking and typing skills, and familiarity with home automation systems, reported as frequency-of-use categories (Never, rarely, sometimes, often).

id	age	sex	diagnosis	speaking & typing ability	smart-home use
P1	32	F	Mild Intellectual Disability	Difficulty writing on paper, but types easily on digital keyboards. Fluent speech, no phonological issues.	Rarely
P2	28	F	Moderate Intellectual Disability, Obsessive-Compulsive Disorder, Behavioral Disorders	—	Often
P3	38	F	Moderate Intellectual Disability	—	Sometimes
P4	41	F	Moderate Intellectual Disability	—	Never
P5	38	F	Moderate Intellectual Disability	Simple but effective language; initiates conversations.	Sometimes
P6	56	F	Borderline Intellectual Functioning	—	Never
P7	41	M	Mild Intellectual Disability	Speaks and writes effectively; fluent with strangers.	Often
P8	24	M	Moderate Intellectual Disability, Drug-Resistant Epilepsy, Congenital Encephalopathy	—	Sometimes
P9	53	M	Moderate Intellectual Disability, Developmental Coordination Disorder, Encephalopathy	Complex sentences, but phono-articulatory difficulties affect comprehension.	Never
P10	27	M	Moderate Intellectual Disability, Psychotic Symptomatology, Emotional Regulation Difficulties in Childhood, Epilepsy	—	Sometimes
P11	28	M	Kabuki Syndrome	Good conversational skills, but phono-articulatory issues may affect understanding.	Never
P12	34	M	Moderate Intellectual Disability, Spastic Tetraparesis, Hyperkinesia in the Upper Limbs, Post-Anoxic Encephalopathy, Sensorineural Deafness	Speaking and typing are slow but accurate; severe phono-articulatory difficulties hinder comprehension.	Sometimes
P13	29	M	Borderline Intellectual Functioning, Autism Spectrum Disorder Level 1, Unspecified Chromosomal Condition	—	Sometimes
P14	37	M	Moderate Intellectual Disability, Behavioral Disorders, Epilepsy	No typing difficulties; minor phono-articulatory problems do not affect comprehension.	Never

4. Phase 1: Evaluation of ChatGPT ambiguous prompt understanding

The study was conducted between June and July 2024 using ChatGPT-4o (the latest version available at that time). ChatGPT-4o was adopted given the conversational fluidity it could offer also compared to other models (Chiang et al., 2024). This phase involved 14 participants aged between 23 and 56 years ($M = 35.86$, $SD = 9.67$) – six females ($M_f = 38.67$, $SD_f = 9.67$ years) and eight males ($M_m = 33.75$, $SD_m = 9.75$ years). All the participants were recruited through the *Fraternità & Amicizia* daycare center in Milan. They had a diagnosis of ID and, during the study, were assisted by a caregiver (Table 1 presents Phase 1 participants' information). This phase was structured in two consecutive steps described as follows.

4.1. Collection of participant-generated instructions

In the first step, participants were invited to generate natural language instructions based on home automation scenarios. To support this task, they were shown eight short videos illustrating various smart home situations² and were asked to describe, in their own words, how they would instruct a system to reproduce each scenario (see Fig. 2). This task design was adapted from the methodology proposed by Andrao et al.

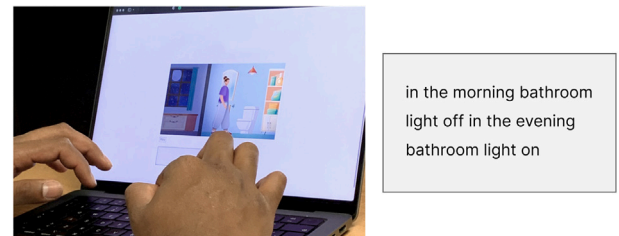


Fig. 2. On the left: a participant writing a rule after observing a video scenario during the study. On the right: an example of one of the rules expressed by a participant.

Andrao et al. (2024). The scenarios used for this step are described in Appendix A.

To ensure a familiar environment for the participants, the study was conducted within the daycare center that participants regularly attended. During the task, participants were supervised by an expert educator who assisted them as needed, ensuring that the instructions were clear and within their scope of understanding. Participants were encouraged, though not explicitly instructed, to verbalize their reasoning while composing instructions. Because individuals with ID vary widely in their expressive abilities, we adopted a soft think-aloud approach: caregivers and the expert educator gently prompted participants to articulate their reasoning when possible, without imposing a formal protocol.

² The use of an illustrative visual style for the video scenarios was a design choice suggested by the expert educator, and this approach ensured neutrality by avoiding potential biases associated with verbal descriptions.

Before the day of the session, the legal guardians signed the informed consent to participate in the study, and assent was also obtained from each participant with an ID. Data were collected using the online platform *Qualtrics*³ and sessions (30 minutes duration) were audio-recorded and later transcribed verbatim to collect the verbal reasoning of participants.

4.1.1. Analysis of participant-generated instructions

A total of 127 instructions were gathered (112 written, 15 transcribed from audio recordings⁴). These were analyzed independently by three researchers. Each researcher received a clean copy of the instruction dataset and independently classified each sentence based on the type of instruction:

- A **“rule”** if it contains a general guideline or condition, often including introductory conjunctions (e.g., *“if”*, *“when”*, or *“while”*), a sequence of conditions/events (e.g., *“during the night close all the windows”*), or if it contains some elements of a rule formulation (e.g., *“the person when enters the house turns on the lights”*),
- A **“command”** if it only contains the request for an immediate action (e.g., *“turn on light”*);
- A **“description”** if it merely describes the content of the video, the context of the scene, or in general the situation illustrated by the video-scenario (e.g., *“in this scene there is a person that enters the kitchen”* or *“in this scene, two people who are in two different rooms can be seen.”*)

Afterward, annotations from the three researchers were compared, disagreements were discussed, and when possible were resolved. The analysis focused only on the instructions classified as **“rule”** ($n = 56$) by minimum of two annotators. The decision to exclude the **“command”** ($n = 14$) category was made to ensure consistency with the usual structures of home-automation systems, which primarily rely on trigger-action programming. This paradigm is incompatible with direct commands, as they do not involve any conditional logic behind the execution of actions. The **“description”** ($n = 57$) category was discarded from the study because its inclusion could have introduced bias into the data, not representing an instruction from the user but merely an objective description of the scenario.

4.1.2. Rules intent analysis from external judges

To evaluate the degree of ambiguity of the instructions classified as rules, three external evaluators were involved.⁵ They had diverse academic backgrounds: a computer science student, a law student, and an architect. None of them had prior experience with projects related to ID or specific knowledge of the study. They were also unaware of both the origin of the rules and the characteristics of the videos from which participants had generated them. This selection allowed us to establish a non-domain-expert baseline for interpreting rule intent. Our goal was to assess the inherent interpretability of participants' instructions by lay adults, providing a reference point for comparison with LLM performance.

In individual online sessions with a researcher, each judge evaluated the rules presented in a pseudo-randomized order, ensuring that rules of the same scenario were never presented consecutively. The intents described by the judges for each rule were then compared to assess agreement among them and to determine whether their

³ <https://www.qualtrics.com>.

⁴ In some cases, participants provided only minimal written responses (e.g., a single word); thus, verbal explanations recorded during the session were treated as a second version of the instruction.

⁵ External judges were recruited to avoid bias, as researchers involved in the study might have been influenced by their familiarity with the video design and the data collection.

Table 2

Prompts used to query ChatGPT. The context table refers to the apartment features shown in Table 3.

prompt id	prompt text
Code-from-rule	Given the following rule {row['rule']} and the following context {table}, generate a pseudo-code. Just output the code without any further comments.
Check-rule	Given the following rule {row['rule']}, context {table}, and definitions {definitions}', identify if there are possible ambiguities or errors in the way the rule was written. Just output the list without any further comments.
From-rule-quest	Given the following rule {row['rule']} make a list of what clarification questions you would ask to reduce the ambiguity level of the received instruction. Just output the list without any further comments.

* Definitions: "1. Being ambiguous: the possibility of being interpreted in various ways. 2. Erroneous: contains one or more elements that, either logically or due to the presence of other conflicting elements, prevent the production of something functional."

interpretations aligned with the original scenario depicted in the video (see Appendix B). Based on the level of agreement among judges:

- 17 rules were classified **“easy”** (30.36%), as all judges agreed on the same intent interpretation (3/3);
- 27 were considered **“moderately difficult”** (48.21%), as one interpretation differed from the others (2/3);
- 12 rules were categorized as **“difficult”** (21.43%), as each judge provided a different interpretation (1/3).

4.2. Generation of ChatGPT outputs

In the second step of Phase 1, using the instruction classified as **“rule”** as input, the LLM⁶ was tasked with the three main prompts specified in Table 2:

1. *Generate pseudo-code based on the natural language rule and the context provided* – The context specified in the prompts was a detailed list of smart devices and their specific locations across various rooms and a basic floor plan. This approach (similar to Ghiani et al. (2017) and Evan et al. (2024)) was used to simulate a real-world usage scenario. The details of the simulated house setting are reported in Table 3;
2. *Identify possible ambiguities or errors within the user-generated instructions;*
3. *Generate a series of questions to disambiguate the user-generated instruction.*

4.2.1. Analysis of ChatGPT outputs

The responses generated by the LLM were organized and evaluated by two researchers (for examples of the LLM output see Table B.11 in Appendix B). This evaluation involved checking the pseudo-code to ensure it adhered to the users' prompts (prompt ID: *Code-from-rule* in Table 2).

Furthermore, the two annotators analyzed the list of ambiguities and errors detected by ChatGPT in the user-generated instructions and the questions suggested by the LLM to disambiguate them (respectively, prompt IDs *Check-rule* and *From-rule-quest* in Table 2). A thematic analysis Braun and Clarke (2006) was conducted to identify emerging themes in the LLM outputs. A deductive approach was employed for the content

⁶ GPT-4o was accessed through the OpenAI API. Decoding hyperparameters were left at their default API configuration to assess the model's behavior in a standard, non-tuned setup.

Table 3
Context table with room characteristics, installed sensors, and smart devices.

room	elements	installed sensors	smart devices
Living room/kitchen	1 entrance, 2 doors, 1 window	Temperature/Humidity, Light, Noise, Entrance (doors/windows), Presence (movement), Smoke/fire	Door, TV, Window, Coffee Machine, Fridge, Ceiling lamp, Oven, Floor lamp, Thermostat, Smart plugs, Heater, Curtains
Bedroom	1 door, 1 window	Temperature/Humidity, Light, Noise, Entrance (doors/windows), Presence (movement)	Door, Smart plugs, Window, Curtains, Heater, Desk light, TV, Alarm, Ceiling lamp
Hallway	2 doors	Temperature/Humidity, Entrance (doors/windows), Presence (movement)	Door, Ceiling lamp, Window, Heater
Bathroom	1 door, 1 window	Temperature/Humidity, Light, Entrance (doors/windows), Presence (movement)	Door, Smart plug, Window, Toilet, Heater, Mirror, Ceiling lamp, Curtains

Table 4

The table reports descriptive statistics and non-parametric tests performed on ambiguities detected (“*ambiguity_n*”) and questions generated (“*question_n*”) by ChatGPT, considering the recognition of rule intent (“yes”/“no”) and clarity of intent. To examine differences between rule types, we adopted non-parametric tests – *Mann-Whitney U tests* and *Welch’s ANOVA (F)* – since the test of homogeneity of variances detected unequal variances. All assumption checks are reported in [Appendix C](#).

	intent	<i>M</i>	<i>SD</i>	min	max	<i>test</i>	<i>p</i>
ambiguity _n	no	6.14	1.64	4.00	11.00	<i>U</i> (372.50)	0.76
	yes	5.93	1.66	3.00	11.00		
	easy	6.18	1.59	4.00	11.00	<i>F</i> (4.77)	0.02
	moderate	6.41	1.65	3.00	11.00		
	difficult	5.00	1.21	3.00	7.00		
question _n	no	12.52	3.10	10.00	20.00	<i>U</i> (500.00)	0.07
	yes	13.52	2.49	10.00	20.00		
	easy	13.06	2.33	10.00	18.00	<i>F</i> (1.60)	0.23
	moderate	12.33	2.24	10.00	16.00		
	difficult	14.48	4.06	10.00	20.00		

analysis of ambiguities while an inductive thematic analysis was conducted for the generated questions [Braun and Clarke \(2006\)](#); [Clarke and Braun \(2017\)](#). In both cases, following the initial coding, the themes were iteratively refined through dialog, and divergences were resolved through further discussion between the two researchers until agreement was reached [Braun and Clarke \(2021, 2019\)](#). In particular, the analysis assessed whether the questions generated by ChatGPT aimed to address ambiguities or other elements that hindered intent recognition.

4.2.2. ChatGPT intent recognition

To determine whether ChatGPT understood the intent behind the rules, two researchers independently assessed whether the pseudo-code, if executed, would reproduce the original scenario depicted in the video. Out of the 56 evaluated codes, ChatGPT successfully recognized the intent in 27 instances, corresponding to 48.21% of the total number of rules. Performance was further analyzed based on interpretation difficulty:

- For “easy” rules ($n = 17$), the LLM correctly understood the intent in 6 cases (35.29%).
- For “moderately difficult” rules ($n = 27$), intent was recognized in 17 cases (62.96%).
- For “difficult” rules ($n = 12$), intent was identified in 4 cases (33.33%).

4.2.3. Ambiguities and errors detected by ChatGPT

The second prompt used to query ChatGPT (prompt ID: *Check-rule*; [Table 2](#)) aimed to obtain a list of ambiguities and errors identified by the model in the user-generated instructions. The LLM returned 338 items, consisting of 178 ambiguities and 160 errors. The number of errors and ambiguities identified per instruction ranged from 3 to 11, with most rules presenting between 5 and 7 issues (see [Table 4](#)). Mann-Whitney *U tests* revealed no significant difference in the number of ambiguities detected between rules where intent was recognized and those where

it was not recognized ($U = 372.50$, $p = .76$), indicating that ChatGPT’s ambiguity detection operates independently of its intent recognition accuracy. However, a Welch’s ANOVA showed a significant effect of clarity level on detected ambiguities ($F = 4.77$, $p = .02$): “difficult” rules yielded fewer flagged ambiguities ($M = 5.00$) than “moderate” ones ($M = 6.41$), suggesting highly ambiguous instructions lack sufficient structure for the model to pinpoint specific issues.

All identified ambiguities and errors fall into the ambiguity categories identified in the studies from which the Phase 1 study procedure was derived [Andrao et al. \(2024\)](#). No new ambiguity categories emerged from the rules written by people with ID.

4.2.4. Disambiguation questions generated by ChatGPT

With the third prompt (prompt ID: *From-rule-quest*; [Table 2](#)), ChatGPT generated a total of 730 questions to clarify the 56 ambiguous rules. We focus our thematic analysis on the 363 questions generated for the 29 rules with a failed intent identification, to examine how the LLM would address the situations in which the pseudo-code generated – if executed – would not align with the participants’ intentions. From this analysis, we classify questions into four main categories:

1. **Questions related to the necessity of context extension (186 instances).** This included 77 questions related to “Specifications on States” (e.g., “Does the rule apply every morning or just under certain conditions?” [Rule: First it was morning and was in the bathroom then turned on the bathroom light]) and 43 questions related to “Details on the Actions” to be performed. Furthermore, 30 questions addressed possible “Exceptions and Alternative Cases”, 23 requesting further “Specifications on Sensors”, and 13 questioning the need for a “Delay on Actions”.
2. **Questions related to the outcome of the rule and its variables (95 instances).** On 43 occasions, ChatGPT asked for more details on the general “Motivations and Intentions” in defining the rule. For example, a question in this category was “What is the purpose of opening the window in both the morning and at night?” [Rule: in the morning a girl opens the window, and at night she opens the window and the alarm sounds]. In 14 questions, ChatGPT asked for information about the “End of the Action” and in 13 about “Error Handling”. Eleven (11) questions were related to “Safety and Security” aspects. Eight (8) questions regarding the presence of “Automation” questioned whether the action was supposed to be manual or automated; Six (6) questions related to specifications of “Temporal Aspects”, for example, the sequence of events.
3. **Questions related to the language used (55 instances).** Twenty-nine (29) questions were related to clarifications on the “Structure” of the instructions and “Words” used, and 26 requested to specify values when “Figurative Values” were used. For example, a question in this category was “What does ‘had to go to the toilet’ precisely mean in terms of actions to be taken?” [Rule: The girl had to go to the toilet; the first time, the door opened. The second time the light was triggered because there was another person inside].

4. **Questions related to specifications on system functioning (27 instances).** Questions related to monitoring the system status, with reference, for example, to the necessity of installing or connecting additional devices, and to details on the system functioning in general (e.g., “Who or what entity is responsible for following or enforcing this rule?” [Rule: opened the window and there was the alarm]).

The themes that emerged from the qualitative analysis of the generated questions are similar to those identified in the analysis of the ChatGPT-detected ambiguities. In fact, 91 questions (25.07%) were intended to address the ambiguities detected by the LLM. For each rule, at least 1 (with a maximum of 7) question addressed an ambiguity ($M = 3.14$, $SD = 1.55$ questions).

4.3. Key results of phase 1

Ambiguity due to lack of specification

The most frequent ambiguity category was related to incomplete details in the rules, specifically regarding conditions, alternative states, or triggers (“Specification on States”). In addition to language, which was an anticipated critical factor American Psychiatric Association (2013); Coppens-Hofman et al. (2017), the lack of specification or omission of essential information appears to be a prominent issue, associated with the logical processes required to define trigger-action rules with sufficient specificity in actions or conditions. Notably, the LLM was less accurate in identifying the intent of rules classified as “easy”. This could be linked to implicit contextual or pragmatic assumptions that were not explicitly encoded in the rule itself. Prior work comparing humans and LLMs on pragmatic tasks shows that models tend to default to literal interpretations of utterances rather than recovering the pragmatically intended meaning (Jennifer et al., 2023), thereby limiting the model’s ability to recognize intents that humans find easy to infer. For example, rules that contain “switch on the light” are unanimously interpreted by judges as implying a dark environment, even though this condition is not explicitly stated in the participant’s rule, thereby matching the original scenario. Human judges, in these cases, may have relied on implicit contextual or pragmatic assumptions to recover the intended meaning. A specific example is a rule classified as “easy”: “at night, the fire lights up in the fireplace”. When reading the rule, a human spontaneously mentally represents a picture of the situation as related to darkness, thus recognizing the lighting of the fireplace as the action, with night as the triggering condition. Failing to make this situational inference, ChatGPT appears to have resolved the rule’s ambiguity by relying on the statistically most plausible interpretation of a fire-related scenario – namely, the outbreak of a fire.

Overall, findings suggest that ChatGPT’s current abilities in intent recognition, and thus its effectiveness in assisting individuals with ID in home automation, are promising but limited. Aligned with previous studies that have explored the impact of non-determinism on LLM’s code-generation tasks Ouyang et al. (2023), this result appears to be influenced by the LLM’s unpredictability, and no consistent patterns in its code-generation capability could be identified.

ChatGPT’s capability to disambiguate instructions over iterations

When prompted for clarification, ChatGPT can formulate questions that align with the identified categories of ambiguity. In line with the iterative reasoning process suggested by other studies Evan et al. (2024), this result suggests that the model can propose questions to guide the user in reviewing ambiguous instructions. These questions require the user to extend the original instruction to include the requested information, thus augmenting the prompt’s syntactic clarity and length. Nevertheless, the LLM frequently generates more questions than the number of ambiguities in the code. Indeed, while the model can formulate ambiguity-relevant questions, all the remaining questions might be redundant or unnecessary for instruction disambiguation.

5. Phase 2: user studies with preliminary prototype

After designing an early LLM-based prototype (Section 5.1) to conduct hands-on activities involving stakeholders and users with ID, Phase 2 focused on evaluating strengths and weaknesses of this solution, through two complementary activities:

- **Workshop with experts.** Five expert educators participated in a structured workshop aimed at gathering domain-specific insights on the interface design, interaction, and overall user experience. The workshop included demonstrations, prototype and scenario evaluations, and a co-design activity consisting of collaborative sketching sessions to elicit reflection on challenges and solutions Dorst (2011). The co-design activity was intentionally conducted after the hands-on experience with the early prototype, following established practice in evaluative co-design Sanders and Stappers (2014). The early prototype functioned as an instrument of inquiry, supporting focused discussion of interaction-specific issues and giving experts a tangible reference point that abstract conversation alone could not provide.
- **Prototype evaluation with dyads of a person with ID and their caregiver.** Four dyads were involved in hands-on sessions where they interacted with the prototype. The sessions were structured to explore autonomous use by participants, with caregivers offering support as needed. Through observation and post-session interviews, we assessed usability, learnability, and perceived challenges.

5.1. Material

Based on insights from Phase 1 – in particular, the observed variability in the LLM’s responses and its tendency to over-generate clarification questions – a three-step framework was defined and subsequently adopted to guide the design of the LLM-based system prototype. As shown in Fig. 3, the framework consists of an iterative and guided process that aims to maximize the model performance while addressing observed limitations:

1. **Step 1: Onboarding.** A first step consists of an onboarding process to assist users in formulating the basic prompt. The interface awaits the user’s input. If the input is unrelated to home automation, the interface specifies that it is necessary to write an instruction, even if it is incomplete. Such design choices can help reduce cognitive load, ensuring a smoother initial interaction with the system. Step 1 addresses the high frequency of *descriptions* and *commands* observed in Phase 1 by prompting users to formulate at least a *partial rule*.
2. **Step 2: Iterative refinement.** This step operationalizes LLM-mediated ambiguity reduction while preventing the cognitive overload caused by ChatGPT’s tendency to over-generate questions. Upon receiving an initial instruction or an idea for a home automation task, the interface identifies ambiguities and poses up to five sequential questions to refine the instruction. Questions are presented one at a time to avoid overwhelming the user. This ensures consistency with the user’s intent, thereby reducing misinterpretations and supporting iterative refinement of the prompt. User’s responses to all clarification questions are then used by the LLM to output a refined version of the instruction that addresses the identified ambiguities while preserving the user’s original intent.
3. **Step 3: Scenario visualization.** Step 3 introduces narrative-based simulation to support the mental modeling of automation outcomes, shown in prior work to reduce logical reasoning load for users with cognitive impairments Huang and Cakmak (2015). This high-level abstraction of code functionality aims to facilitate

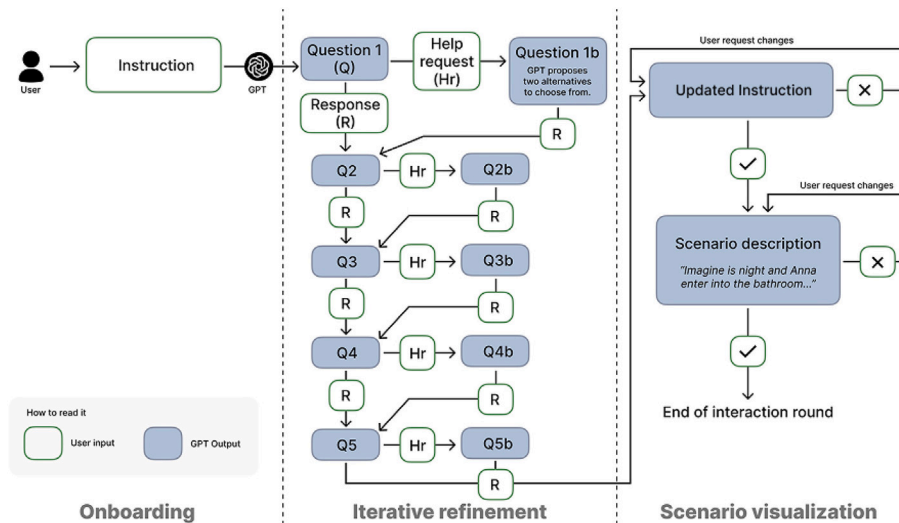


Fig. 3. Interaction flow in a round of interaction with the interface.

user comprehension of the generated code, reducing the knowledge burden required to evaluate whether their instructions are logically sound and aligned with their original intent. This is achieved by generating a hypothetical scenario illustrating how the instruction would function in practice, presented as a form of storytelling:

"Imagine it's a cold winter day. The temperature in the house drops to 17 °C. Without anyone having to do anything, the smart fireplace automatically turns on, warming up the room. After a while, the temperature rises and exceeds 22 °C. At this point, the fireplace turns off by itself, preventing the house from overheating".

After describing the scenario, the interface awaits the user's feedback. If the user is satisfied, the round of interaction concludes. Otherwise, the interface prompts the user to specify modifications they would like to make.

A chat-based prototype tool was developed based on the three-step framework (see Fig. C.6). It uses a custom GPT configuration with ChatGPT-4o as the inference model and applies structured prompting to ensure that system behavior strictly aligns with the guidance, refinement, and visualization phases (see Appendix C.1). Hyperparameters were left at their default API values, consistent with Phase 1. A rule-based governance mechanism constrains the model to predefined operations at each phase (e.g., clarification questions, instruction rewriting, contextual narrative generation) and allows state transitions only upon explicit user confirmation. This multi-layer organization forms a repeatable, finite-state interaction pattern for LLM-assisted EUD. The prototype was iteratively calibrated by two researchers, primarily by refining prompts to reduce failures, until it consistently reproduced expected interaction flows across predefined trigger-action scenarios. Evaluation centered on adherence to phase boundaries, correctness of returned rules, and narrative coherence.

5.2. Workshop with experts

5.2.1. Participants

Five individuals ($F = 3$; $M = 2$) were recruited for the workshop (see Table 5), all affiliated with DOWN DADI, a non-profit association acting in the Padua area (Italy), dedicated to improving the quality of life for individuals with Down Syndrome, Autism Spectrum Disorder, Intellectual Disabilities, and other behavioral or relational difficulties.

Table 5

Workshop participant demographics. Indicated with "exp." are the years of experience as educators and with "AI exp." participants' previous experience with AI ("yes/no").

id	gen.	exp.	role	age	AI exp
E1	F	+20	coordinator	> 40	no
E2	F	19	coordinator/educator	> 40	no
E3	M	3	educator	< 30	yes
E4	F	5	educator	< 30	yes
E5	M	—	educator	< 30	yes

The participants worked as educators and coordinators within the association, with professional experience ranging from early career to over two decades.

5.2.2. Tasks and procedure

The workshop was conducted remotely via Zoom and facilitated by two researchers. All five participants were physically located in a shared room and joined the session through a single computer. During the workshop, one researcher shared his screen to display the slides and guide the session flow. Participants submitted their responses using personal mobile devices through the interactive presentation platform *Mentimeter*.⁷ The workshop activities (duration of 2.5 hours) were video-audio recorded and followed this structure:

- 1. Introduction and ice-breaking.** A brief ice-breaking activity with the *Mentimeter* functionalities was conducted to encourage participation and reduce initial hesitation. Participants were invited to respond to two questions regarding their roles within the association and prior experience with AI.
- 2. Hands-on and evaluation.** One of the researchers administered two hands-on moments. The first introduced ChatGPT and its capabilities, while the second focused on the interface prototype. In this second hands-on session, the researcher introduced participants to the prototype's aim and functionalities, showcasing a full interaction from instruction to code generation, passing through the 5-question interactive disambiguation process. All the data were collected through the *Mentimeter* platform. Following the

⁷ <https://www.mentimeter.com/>.

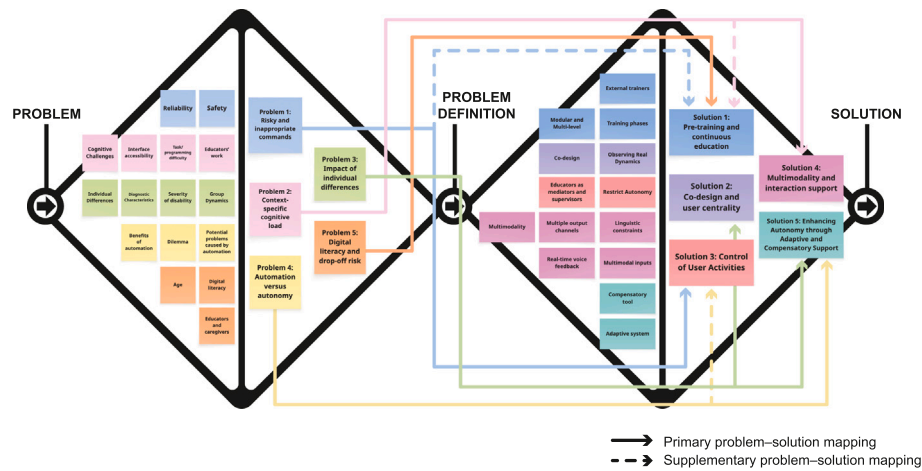


Fig. 4. Problem-solution mapping across the double diamond process (Ball, 2019). The left diamond summarizes five problem themes, each with its associated code; the right diamond summarizes five solution themes, each with its corresponding code. Connector colors correspond to the colors assigned to the originating *problem theme* (P1–p5). The central “Problem definition” column marks the transition between discovery and solution development.

hands-on session, we asked experts to complete a User-Experience Questionnaire (UEQ-S) – a standardized questionnaire designed to assess the user experience (Schrepp et al., 2017) – that we administered to collect quantitative and standardized usability dimensions. This short form consists of eight items, each represented by a pair of opposing adjectives (e.g., “confusing” – “clear”), positioned at the extremes of a seven-point scale ranging from “−3” to “+3”. The questionnaire comprises two scales that measure the following aspects:

- (1) *Pragmatic Quality Scale*. Evaluates the effectiveness and efficiency of the product (four items, e.g., “useful” – “useless”).
- (2) *Hedonic Quality Scale*. Assesses the appeal and attractiveness of the design (four items, e.g., “engaging” – “boring”).

The UEQ-S questionnaire responses were collected using *Mentimeter*, with each expert submitting their response via smartphone. The results of the responses were displayed to the experts only after the completion of the UEQ-S sessions to mitigate potential bias or influence.

3. **Co-design through Design Thinking activities.** Following the structure of the Double Diamond Ball (2019), an initial divergent phase (problem space) prompted participants to generate and share individual reflections using *Mentimeter* on perceived challenges related to the interface prototype and overall framework. This was followed by a collective discussion. In the subsequent convergent phase, thematic areas were identified and ranked using *Mentimeter*. In the final divergent phase (solution space), participants ideated potential solutions to the identified thematic.

5.2.3. Data analysis

Workshop audio-video recordings were transcribed verbatim. Two researchers independently coded the transcripts following an inductive thematic analysis approach Braun and Clarke (2006); Clarke and Braun (2017). Researchers individually generated and assigned codes. After individual coding, the two coders compared their codebooks and resolved discrepancies through discussion. Codes were then collated into broader themes through discussion (Clarke and Braun, 2017).

Experts’ responses to the UEQ-S questionnaire were analyzed in accordance with the UEQ handbook guidelines (Schrepp et al., 2017) to compute dimension scores and generate benchmark comparisons.

5.2.4. Results

Ten (10) themes related to problems and solutions (see Fig. 4) emerged from the inductive thematic analysis (Clarke and Braun, 2017) of the Design Thinking session (Dorst, 2011). The complete list of themes and codes is provided as Supplemental material.

Perceived challenges with an LLM-based interface. Five themes encapsulate expert-perceived challenges in the interaction with the prototype by people with ID:

- **Problem 1: Risky and inappropriate commands (21 instances).** While the system was viewed positively, experts expressed concerns about “Safety” ($n = 18$) and “Reliability” ($n = 3$). Three of them (E3, E4, and E5) were particularly concerned that users with ID might create frivolous or dangerous rules, and that unfiltered use could encourage childish or inappropriate actions. This encompasses, for example, turning on the stove in uncontrolled situations, as stated by E5. Furthermore, experts emphasized the importance of predefined filters and blocks to automatically restrict dangerous requests. They also discussed possible issues related to the user over-reliance on the system, especially in contexts lacking technical support. Because habits formed through automation are not always transferable to unsupported environments, malfunctions could lead to frustration or unsafe behaviour. Educator E3 also feared that unfiltered use could encourage childish or inappropriate actions, exemplified by “turning on the sink at 3 A.M.”, “switching on the lights at unreasonable hours”, or requesting “a party at 2 A.M.”.
- **Problem 2: Context-specific cognitive load (40 instances).** Experts (E3, E4, E5) pointed out that long explanations and complex texts produced cognitive overload (“Cognitive Challenges”; $n = 22$), leading to loss of attention and abandonment. Excessive technical language, unclear instructions, and limited opportunities to ask for clarification could hinder comprehension. Other barriers were linked to “Interface accessibility” ($n = 6$). As E3 noted, this involved both “interaction with the interface and, for some, even the comprehension of the language”, adding difficulties in spatial orientation: “not only text but also space-related [issues], like ‘where should I place the answer?’”. He also mentioned issues related to device choice (mobile vs. computer), voice interaction (perceived as useful but difficult to execute), and doubts about autonomous use by users with ID. Educator E3 further emphasized “Task/programming difficulty” ($n = 8$), arguing that programming is burdensome and often

incomprehensible for users with ID, leading to negative impacts on time loss and social life. Such challenges also increase educators' workload, raising economic and time costs for facilities (*"Educators' work"*, $n = 4$).

- **Problem 3: Impact of individual differences (66 instances).** Experts highlighted that *"Diagnostic Characteristics"* ($n = 25$), and more general cognitive and behavioral *"Individual Differences"* ($n = 30$) deeply shaped interaction modalities. For example, expert E4 observed that while long, complex written explanations could overwhelm some users, others can find them manageable. E3 further highlighted the importance of considering *"Severity of disability"* ($n = 9$) in prototype use, noting that *"for medium-high disability it is necessary to assess how limiting it is and what the person could do without home automation, while it may be very useful in medium-severe cases"*. Expert E5 also pointed out the issues related to the *"Group Dynamics"* ($n = 2$), as divergent personal preferences in a co-living group could lead to conflicting rules that contradict each other.
- **Problem 4: Automation versus autonomy ($n = 35$).** The group grappled with the tension between the *"Benefits of automation"* ($n = 8$) and the need to preserve autonomy (*"Dilemma"*, $n = 16$), challenged by the presence of automation (*"Potential problems caused by automation"*, $n = 11$). On the one hand, automation can provide safety, as seen in the case of automatic smoke detection systems, simplifying tasks and serving as compensatory support. On the other hand, experts are concerned that this could negatively impact user autonomy, potentially fostering dependency on system automation. The perceived negative impact varied by type and *"Severity of disability"*, the role of the educator, and the chosen supervision strategy. Excessive external control or rigid rules can undermine freedom, while the absence of explanation or mediation can diminish responsibility and choice.
- **Problem 5: Digital literacy and drop-off risk (17 instances).** Experts E2, E3, and E5 argue that motivation to use the prototype can be influenced by *"Age"* ($n = 6$) and *"Digital literacy"* ($n = 7$). Older users are often less inclined towards technology, whereas younger users – though more accustomed to digital tools – could quickly become bored or disengage if the interface seems complex or reading-intensive. The same challenges were also associated with *"Educators and caregivers"* ($n = 4$), as E3 stated, *"the result is a drop-off in wanting to use the tool, because if it becomes too time-consuming, even for the educator, you end up losing interest"*. Overall, lack of training and familiarity with digital tools is seen as a cause of prejudice against the technology and hinders the ability to convey its potential.

Experts envisioned solutions to ensure LLM-based systems' usability and accessibility. Five themes highlight solutions suggested by the experts:

- **Solution 1: Pre-training and continuous education (23 instances).** All experts stressed the requirement of training, which must be structured, multi-level, and ongoing to ensure safe and effective use by both users and educators. *"External trainers"* ($n = 5$) – psychologists, experienced educators, or technical specialists – were seen as crucial for filling competence gaps and offering unbiased perspectives, but their involvement should complement rather than replace educators' work. Training should be *"Modular and Multi-level"* ($n = 13$), combining individual, group, association, and co-living activities. E3 also pointed out that involving parents should be calibrated in a way to avoid creating what they called *"protective bubbles"*, which can reduce autonomy. Importantly, training should involve different *"Training phases"* ($n = 5$): an initial pre-training and some follow-up that continue over time to support adaptation, strengthen metacognition, and foster progressive autonomy, while maintaining an active supervisory role.
- **Solution 2: Co-design and user centrality (9 instances).** A participatory and collaborative approach is viewed as essential. Experts

envision a *"Co-design"* ($n = 6$) approach by involving users, educators, and other stakeholders in decision-making, defining rules and limits together, avoiding protective bubbles. The expert underscored that the prototype should be both effective for professionals and enjoyable for users, justifying economic investment in inclusive design. Experts highlighted that it is essential *"Observing Real Dynamics"* ($n = 3$) to effectively co-design, involving participants with ID to ensure authentic dynamics and issues.

- **Solution 3: Control of User Activities (39 instances).** This theme emerged from comments on the dual role of *"Educators as mediators and supervisors"* ($n = 28$). While oversight is necessary to prevent dangerous rules, guarantee safety, and support users with severe disabilities, constant control or externally imposed rules can *"Restrict Autonomy"* ($n = 11$). Experts advocated for a co-participatory approach, where limits are explained and negotiated between educators and users, thereby enhancing their understanding and sense of responsibility.
- **Solution 4: Multimodality and interaction support (34 instances).** *"Multimodality"* was discussed as a key design principle ($n = 13$), not just a technical solution. Integrating *"Multiple output channels"* ($n = 2$), such as visual, auditory, and textual, increases system adaptation to diverse communicative and cognitive needs. Experts also emphasized the importance of *"Linguistic constraints"* ($n = 6$): language should remain clear, simple, and concise, avoid unfamiliar terminology, and rely on short sentences with few words. Providing continuous, step-by-step instructions (e.g., 1. "click here", 2. "enter your name") rather than lengthy explanations can further enhance accessibility and clarity of understanding. In addition, images and photos help identify objects and explain situations; videos are particularly powerful for scaffolding steps and supporting decisions in multiple-choice situations. *"Real-time voice feedback"* ($n = 7$), as reminders or guidance, can be a valuable support but must be carefully calibrated to avoid confusion. Expert also discussed the importance of *"Multimodal inputs"* ($n = 6$): non-verbal audio signals, images, and video inputs can further enrich communication and increase system flexibility.
- **Solution 5: Enhancing Autonomy through Adaptive and Compensatory Support (42 instances).** Participants reflected on the role of the prototype as a *"Compensatory tool"* ($n = 27$) that assists individuals with ID in problem-solving without substituting their abilities, thereby fostering learning, metacognitive reflection, and skill development. Rather than encouraging dependency, the tool provides contextual guidance through reminders, confirmations, and feedback. At the same time, participants highlighted the importance of designing an *"Adaptive system"* ($n = 15$), emphasizing its capacity to respond to individuals' needs and features and to enable educators to personalize according to individual differences, and to serve as a resource for educators. In this way, the tool also functions as a monitoring and observation aid, enabling real-time analysis of behaviors and facilitating tailored interventions that promote autonomy and responsibility.

UEQ-s results. Analyses of results from the UEQ-S questionnaire show an average user experience score of 0.75 ($SD = 0.29$), placing it in the "below average" range.⁸ Appendix D summarizes the mean ratings for each UEQ-S item.

Items 5–8, which assess Hedonic Quality such as excitement and novelty, received the highest averages falling in the "excellent" range ($M = 1.85$; $SD = 0.49$). Specifically, the level of originality (item 7) and novelty (item 8) attributed to the system were 2.00 and 2.40 on average, respectively. The Pragmatic Quality Scale (Items 1–4) falls in the "bad"

⁸ Above +1.50 is considered "excellent", between +0.80 and +1.50 "good", between 0.00 and +0.80 "below average", between 0.00 and -0.80 "poor", and below -0.08 "very bad" (Schrepp et al., 2017).

Table 6

Demographic and diagnostic information of the four participants with ID who took part in the prototype evaluation with dyads, including ID level and verbal production (“verbal”) capacity.

id	age	gen.	diagnosis	ID	IQ	verbal
P1	23	M	Partial fetal alcohol syndrome, symptoms of ADHD	mild	< 60	good
P2	24	M	Autism spectrum (level 2)	mild	< 61	good
P3	19	F	Down syndrome, oppositional defiant behaviors	severe	< 35	low
P4	18	F	Partial trisomy of chromosome 18	severe	< 35	moderate

range of the UEQ benchmark ($M = -0.35$, $SD = 0.67$). Particularly, the issues in the practical usability of the prototype were related to ease of use (item 2) and clarity (item 4).

5.3. Prototype evaluation with dyads

5.3.1. Participants

We recruited four dyads, each composed of one participant with ID and one caregiver. The recruiting occurred through snowball sampling, facilitated by the *DOWN DADI* association in Padua. The four participants with ID – two males and two females – were aged between 18 and 24 ($M = 21.00$, $SD = 2.94$ years). All participants reported a diagnosis of a specific genetic or medical syndrome associated with ID (see Table 6). The caregivers involved were the parents of the participants (referred to in the text, as C1, C2, C3, and C4).

5.3.2. Task and procedure

The study was organized as semi-structured interviews conducted remotely over two weeks using the interface prototype via *Zoom*. In addition to the dyad, two researchers participated in each session: one acting as a moderator and the other as a supervisor responsible for data collection. Each session lasted approximately 90 minutes, with participants given the freedom to take necessary breaks. Before starting, the procedure was carefully explained to each dyad. Consent to participate in the study were obtained from the caregivers, who in this case was also the legal guardians of the participants with ID. Assent to take part in the study was also collected from participants with ID. They were also informed that their participation was entirely voluntary and that they could withdraw their assent at any time.

At the start of each session, the moderator read the instructions that detailed the process and tasks to be performed by the participants. Participants were asked to create a home automation instruction following the same methodology adopted in Phase 1: after watching a short video, participants were asked to compose an instruction for a home automation system to execute the scenario presented. Since the study sessions were conducted remotely, the moderator shared their screen to present the interface prototype. Participants could interact with the system in one of two ways: by typing their instructions in the *Zoom* chat, which the moderator copied and pasted into the prototype, or by dictating the instruction aloud for the moderator to transcribe verbatim into the interface input field. Neither the moderator nor the caregiver independently reformulated participants’ input. Before submitting any instruction, the moderator asked participants to confirm that the text displayed in the interface accurately reflected what they intended to express. After submitting the instructions, participants were invited to continue interacting with the prototype, providing any additional prompts or responses required by the interface. Upon reaching the final step of the interaction, participants were given the option to start another interaction round by watching another video scenario, self-defined instruction to input into the interface, or ending the study.

After completing each interaction round, participants were asked to provide feedback, responding to three pre-defined questions concerning the usefulness, clarity, and number of questions generated by the

Table 7

The frequency table shows the distribution of the 39 generated questions among participants and interaction rounds.

id	round 1	round 2	round 3	round 4
P1	2	3	5	–
P2	2	5	5	5
P3	5	3	–	–
P4	4	–	–	–

interface. At the end of the entire interaction session, participants were asked to provide some demographic information, experience with home automation systems, and programming languages. Answers were collected through *Qualtrics*. Next, we asked caregivers to complete the UEQ-S questionnaire (Schrepp et al., 2017) and answer some open-ended questions about the prototypical interface. Questions covered both positive and negative aspects of the interface, as well as future expectations about LLM-based systems and their possible impact on home automation and support for people with ID.

5.3.3. Data analysis

In addition to collecting the user-generated chats from the prototype, the sessions were audio-recorded. Two researchers independently analyzed the transcribed recordings using a qualitative thematic analysis. A deductive approach was employed to analyze the user-generated chats, building on the results from Phase 1. The caregiver UEQ-S responses were analyzed using descriptive statistics and benchmark comparisons (Schrepp et al., 2017). Engagement levels were assessed based on the number of interaction rounds that participants completed with the interface after the first video scenario.

5.3.4. Results

This section presents the empirical findings of the prototype evaluation along six complementary dimensions: (i) the LLM-based interface’s performance in generating clarification questions; (ii) participants’ engagement with the iterative interaction; (iii) participants’ ability to understand the questions posed by the interface; (iv) the role taken by caregivers; (v) qualitative themes that emerged from their open-ended feedback; and (vi) the standardized usability assessment via UEQ-S. Together, these findings address RQ2 by characterizing how individuals with ID and their caregivers experience, understand, and evaluate an LLM-based home automation assistant.

LLM-based interface performance. All participants completed at least 1 round (maximum 4 rounds), obtaining a refined version of the original instruction (see Table E.17 in Appendix). Due to technical problems, round 4 of P1 was not completed and was excluded from the analysis. All initial and revised instructions are provided in Appendix E. Among all participants, we analyzed the performance of the LLM-based prototype in 11 rounds of interaction. For each interaction round, the interface generated a minimum of 2 and a maximum of 5 clarification questions ($M = 3.90$, $SD = 1.29$) for a total of 39 questions (see Table 7).

Following the classification that emerged in Phase 1, among those 39 questions, 24 fall within the **Content Extension** category (see Table 8). These questions asked the user to specify the “*Sensor*” to be used, the “*Action*” to perform, or “*Delay*” times. Ten (10) questions were categorized under **Outcome** ambiguity, where the system seeks clarification on “*End of Actions*” to determine whether the action should be repeated indefinitely or interrupted when specific conditions are met. Four (4) questions were related to **Language** ambiguity, where the interface requests clarification of “*Figurative values*”. Finally, 1 question falls into the **System specification** category, with the interface asking about the specific brand of smart plug available in the household. No new ambiguity-related categories emerged from any of the generated questions.

Table 8

Examples of questions generated for each theme are presented, with the corresponding participant and interaction round indicated.

theme	sub-theme	id	round	question	class
Content extension	Conditions	P3	2	<i>Do you want the window to close automatically only when it is nighttime and the light is off (i.e., it is dark), or is it enough that it is simply nighttime to close the window?</i>	two-options
Language	Figurative value	P2	3	<i>When you say 'the house is cold', at what temperature do you want the fire to turn on? For example, below 18 °C?</i>	open-ended
Outcome	End of Action	P4	1	<i>Here are two alternatives to choose from: 1. Automatic shutdown: The bathroom light turns off automatically when the sensor detects that it is no longer dark. This option is useful for saving energy. 2. Manual shutdown: The light stays on even if it is no longer dark and can only be turned off manually (e.g., via a switch or voice command).</i>	two-options
System specification	Functioning	P2	3	<i>The external device that controls the stove is a specific smart plug (e.g., from a brand such as Sonoff, TP-Link, etc.), or do you want the system to be generic and compatible with any plug?</i>	two-options

User engagement and feedback. Engagement levels were assessed based on the number of rounds of interaction. Different levels of engagement with the interface were observed. Both P1 and P2 engaged fully without hesitation after completing their first round of interaction. P1 proposed self-defined instructions from the second round, inspired by personal experience. For example, “*When I get up to go to the center, Alexa should warm the milk*”. This idea, discussed with the caregiver, prompted P1 to reflect on how the system could optimize daily routines and address personal needs.

P3 and P4 decided to interrupt the interaction after the second and the first rounds, respectively. P3 expressed a lack of interest in continuing the activity as she was tired from the task, saying that we could proceed with other instructions the following day. P4 was not interested in the interaction, so she immediately interrupted the activity after the first round.

After the first interaction round, participants were asked about their opinions regarding the interface's questions. P1, P2, and P3 found the questions understandable and valuable. P2 noted that the questions were very easy and expected a higher number of questions. P4 preferred not to answer.

Questions understanding. Analyzing the verbal reports, we observed different levels of understanding of clarification questions generated by the interface. P1, P2, and P4, while requiring active support from caregivers to fully understand some of the questions, generally grasped the overall requests posed by the interface.

In contrast, P3 required substantial assistance from the caregiver to comprehend the questions and formulate responses that often consisted of single words or concepts that closely echoed the caregiver's input. The caregiver later explained that tasks that encompass the definition of time duration are challenging for P3 and reported that she typically responds to questions by choosing the last option provided.

Caregivers' role in the interaction. Caregivers played an active, supportive role during activities, often reading the interface's questions aloud to prompt participants' reflection. An exception was observed with participant P4, where C4 took a more substitutional role, directly providing answers on the participant's behalf or even pushing the participant to change the response (e.g., “*No, no, let's do the first one it suggested instead*”). It was observed that all caregivers used or recalled concrete examples related to the participants' personal experiences to encourage reflection or simplify the understanding of the tasks. In particular, caregivers leveraged participants' prior experiences with systems like *Alexa*, referencing these interactions to help them understand the type of instructions expected by the interface.

Qualitative themes from caregivers' open-ended feedback.

- **Automation vs. Autonomy.** The LLM-based system was recognized as a promising tool for improving autonomy, particularly in co-living scenarios or if caregivers can no longer provide direct care. Despite its potential, caregivers expressed concern that excessive

automation may undermine the development of essential problem-solving skills. C2 brought up the example of using a calculator in school to support students, which occurs only after they have developed basic numeracy skills. They emphasized the need for a balanced approach that empowers users without fostering over-reliance on technology.

- **Individual Differences.** Caregivers of participants with mild ID confirmed that the questions, vocabulary, and system interactions were accessible and valid despite some difficulties. Caregivers of participants with severe ID (P3 and P4) reported that the tasks and questions were too complex for independent use. They suggested redesigning the system as a collaborative tool in which caregivers and participants work together or as a moderate setup in which caregivers assist in defining instructions, while the LLM-based system supports participants in performing daily routines.
- **Input-output Modality.** Users' interactions with the interface could be conducted by writing instructions or expressing them vocally. It was observed that, in all cases, participants chose to interact verbally. This preference was driven by the convenience of speaking, as noted by C4, who stated “*It is easier to do it verbally*”. C2 stated that for P2 interaction through writing is normally easier than verbal interaction. Unfortunately, the setting of the online study posed some difficulties. Although participants could type their responses, P2 struggled to type them because he was used to a traditional keyboard while he was using a tablet during the study.
- **Visual and Concrete Cues.** C3 and C4 emphasized the importance of visual aids and limited concrete options to support decision-making. For example, C4 noted that open-ended questions posed significant challenges, while offering two simple alternatives for decision-making was more manageable and understandable. C2, C3, and C4 also noted that incorporating visual cues alongside the questions could significantly improve the system's accessibility, enhance comprehension, and support independent responses.

UEQ-s results. Analyses of results from the UEQ-S questionnaire show an average user experience score of 1.41 ($SD = 1.53$), placing it in the “good” range (see [Appendix D](#)). The Pragmatic Quality Scales fall in the “below average” range of the UEQ benchmark ($M = 0.81$, $SD = 2.18$), highlighting some issues in the practical usability of the prototype, specifically related to efficiency (*item 3*) and ease of use (*item 2*). On the other hand, the Hedonic Quality Scale scored in the “excellent” range ($M = 2.00$, $SD = 0.94$). Specifically, the level of interest (*item 6*) and novelty (*item 8*) attributed to the system were both 2.50 on average.

6. Discussion

Our findings extend prior HCI research on EUD by showing that individuals with ID primarily struggle with structural completeness and situational specificity, rather than the syntax of rule creation. LLM-mediated scaffolding can compensate for these gaps through iterative clarification, shifting the design space from “teach users to express rules correctly” toward “design systems that refine rules with them”. This

reconceptualizes EUD for assistive contexts as a co-constructive process, and aligns with contemporary models of shared agency in assistive technology [Cynthia et al. \(2018\)](#).

In the following, we discuss our findings in relation to our research questions.

6.1. User with ID refined and disambiguated prompt guided by the LLM-based home automation assistant (RQ1)

Our results suggest that ChatGPT's capabilities can be leveraged to disambiguate prompts interactively: when prompted with incomplete instructions, the model is able to articulate disambiguation questions that align with the ambiguity categories identified by previous research ([Andrao et al., 2024](#)). This convergence with established ambiguity taxonomies indicates that LLMs can serve as a viable foundation for accessible EUD: rather than requiring users to learn formal rule languages, the burden of disambiguation can be transferred to the model, opening a design space in which the system actively supports users instead of merely interpreting them.

Furthermore, the evaluation of our chat-based prototype provides evidence that an iterative process, guided by LLM-generated disambiguation questions (as in [Evan et al. \(2024\)](#)), can be effective in supporting users with ID in generating home automation rules. The iterative process resulted in refined prompts that were improved compared to the initial ones, which were often overly abstract and ambiguous (see [Appendix E](#) for initial vs. refined prompts). The interactive process successfully addressed one of the main ambiguities that emerged in Phase 1: the lack of specificity in instructions. This finding provides a positive, but nuanced, answer to RQ1: an LLM-based EUD interface can support users with ID in personalizing smart home automation through natural language, provided that the interaction is structured through controlled clarification dialogues that compensate for language ambiguity and users' communication difficulties.

The framework we used to guide users in interacting with the prototype also addresses the LLM's tendency to generate more questions than necessary. This high productivity, which emerged from Phase 1, could have overwhelmed users if left unchecked [Finlay and Antaki \(2012\)](#). Our framework limits clarification questions to a maximum of five sequential questions. This modification proved effective, and participants generally found the questions understandable. A key takeaway is that the effectiveness of LLM-based EUD for users with ID depends not only on the model's linguistic comprehension capabilities, but also on how the interaction itself is structured.

However, both caregivers and experts emphasized that controlling the number of questions must be balanced with consideration of their type and formulation. Open-ended questions generated by the interface might be overly complex, while multiple-choice or yes/no questions are generally considered as more accessible to users with ID (in line with research in other fields [Finlay and Antaki \(2012\)](#); [Hollomotz \(2018\)](#)). Moreover, questions requiring specification of temporal or numerical values need to be supported by concrete examples or predefined options. These insights suggest the need to control the topic and the number of questions, as well as to adapt the format to align with individuals' capabilities.

During our study we observed challenges and highlighted key improvements needed to make the experience accessible to a larger number of people. The first way to achieve this, as experts suggest, is to develop an adaptable system. This system should respond to individual needs and be personalized by the caregiver [Bacchin et al. \(2021\)](#); [Cleland et al. \(2024b\)](#); [Carmien and Fischer \(2008\)](#). This points toward a design model in which a single deployable solution must accommodate heterogeneous cognitive profiles through caregiver-mediated configuration rather than one-size-fits-all defaults.

In line with ongoing work on multimodal interaction and output ([Calò and De Russis, 2024](#); [Barricelli et al., 2022](#); [Roffarello et al., 2023](#); [Artizzu et al., 2024](#)), our findings further suggest that multimodality

is essential for accessibility. Experts in the workshop emphasized that multimodal outputs—such as images, pictograms, videos, and voice prompts—can better accommodate diverse communicative and cognitive needs. Similarly, caregivers in the dyadic study highlighted that some individuals with ID communicate more effectively verbally than in writing, and they recommended offering concrete visual options instead of open text fields. Taken together, these observations indicate that multimodality should not be treated as an enhancement layer but as a precondition for accessibility in this user group: input/output channels need to be matched to expressive abilities (verbal vs. textual) and to cognitive load constraints.

6.2. Balancing autonomy and automation in a co-participative future with LLM-based assistants (RQ2)

Both experts and caregivers, as well as individuals with ID themselves, were broadly positive about the capabilities and potential of the proposed LLM-based home automation system, as was also made evident by the UEQ-S responses. Caregivers likewise described the interface as interesting and regarded it as a tool that could encourage reflection and self-expression. Notably, no comments or behaviors showing inhibition towards the tool were observed, and the main reasons for disengagement were tiredness or disinterest.

Safety considerations emerged from experts' comments concerning unsupervised usage of the system by individuals with ID who might inadvertently set dangerous behavior, such as turning on the stove when not needed [Palekar et al. \(2019\)](#); [Lewis et al. \(2023\)](#); [Chan et al. \(2025\)](#). While neither generation of dangerous rules nor hallucinations – i.e., outputs referring to non-existent devices, sensors, or scenarios – occurred in our study, the experts' comments highlighted a mandatory safety requirement for LLM-based systems. Therefore, a human-machine safety control approach can be envisioned (e.g., [Rahmati et al. \(2018\)](#)), with caregivers and educators empowered on set limits to user rule-generation freedom, which can be matched with LLM capability to recognize dangerous or uncommon outputs from the user-generated instructions. Even if the model's capability to manage safety control requires future investigation, LLM capabilities to recognize and address ambiguities suggest that this represents a promising approach.

However, our results also suggest that autonomous use remains out of reach for many. While this might be due to usability limitations of our prototype, it also aligns with caregivers' and experts' emerging perceptions towards the future direction of LLM-based assistants in the home automation context. They stressed that cognitive and behavioral differences, as well as the severity of the disability, make a universal solution unrealistic and repeatedly emphasized that co-participation with users with ID – and not complete autonomy – should be the design goal.

Along these lines, it is worth noting how caregivers' presence often transforms a frustrating interaction into a meaningful one. Rather than seeking to replace caregivers, we should envision shared EUD tools that enrich this human connection. In practice, this means designing for two simultaneous users with distinct but interleaved roles: the user with ID as the primary actor and the caregiver as a meta-designer who shapes, monitors, and intervenes in the interaction.

This vision is also linked to the tension between automation benefits and the risk of undermining autonomy. This aspect is particularly sensitive, as literature shows that autonomy is essential for the development and well-being of individuals with ID [American Psychiatric Association \(2013\)](#); [Savulescu \(2009\)](#); [Sands and Doll \(1996\)](#). This includes teaching life skills that enable individuals to make informed decisions and manage daily tasks without requiring constant oversight (e.g., [Carey et al., 2023](#)). In particular, autonomous living initiatives dedicated to individuals with ID are part of modern therapeutic approaches designed to enhance independence and self-determination ([Bacchin et al., 2021](#)). An emerging comment on the balance between autonomy and automation could encapsulate expert concern regarding this topic and the possible impact on established current practices and therapies that

aim to increase independence. In fact, involved experts noted that automation can improve safety and simplify routines, but cautioned that excessive automation may foster dependency. Caregivers echoed this concern by likening the LLM-based interface to a calculator, useful for reducing task effort but only after core skills have already been learned. They advocated for designs that reduce cognitive load for non-essential tasks at home while leaving room for users to practice essential skills or learn basic behaviors. This points to a fundamental design principle for LLM-based assistive automation: compensation should be selective. Automation must scaffold rather than substitute, distinguishing between tasks where it lowers barriers (cognitive offloading) and tasks where it would erode skill acquisition.

Based on this, it emerges that current user and expert perception of LLM-based assistants for home automation holds promise and inspires curiosity rather than fear. This evidence marks one of the first steps towards understanding current user and expert perception on the matter. Possible emerging open questions that build on our results and could guide future research are: *What role should caregivers/educators assume in interactions between users with ID and LLM-based tools? What is the ideal balance between autonomy and automation for individuals with ID in the home automation context? Can LLMs effectively recognize and address safety issues in home automation instructions? Which LLM better responds to the needs of users with ID in home automation?*

6.3. Design implications

Leveraging the combined findings and themes that emerged from Phase 2, we present design implications to inform the development and use of future LLM-based EUD interfaces for people with ID. In particular, these implications can guide prompt engineering, shaping how LLMs are instructed to generate accessible and effective dialogues. We organize them around the three temporal phases of system interaction: before usage (BU), during usage (DU), and after usage (AU).

- **BU1: Establish caregiver-supervised safety limits.** This approach entails two levels of control. First, limits can be jointly negotiated by caregivers and users, applying the principle of scaffolding to turn rule setting into opportunities for learning and responsibility (Pea, 2018). Second, for high-risk situations, caregivers should be able to set non-negotiable hard limits. The system can leverage LLM capabilities to explain these blocks and suggest safe alternatives (e.g., proposing allowed actions when a restricted rule is triggered). This layered approach also serves as a safeguard against unpredictable LLM behaviors — such as hallucinations — which, although not observed in our study, represent a limitation of LLM-production.
- **BU2: Enable progressive personalization.** The system should act as a highly configurable compensatory tool, providing adaptive personalization to individual needs (Morra et al., 2024). As emerged from our analysis, an initial profiling phase can capture users' daily routines (e.g., waking times or activity sequences), enabling the system to adjust interactions while maintaining safety. Personalization should be guided by educators or caregivers and dynamically adapted over time to match users' skill development to foster progressive autonomy.
- **DU1: Design for user-caregiver joint-use.** Educators and caregivers must be enabled to act as facilitators. The mediation and autonomy support they provide should be matched by the system, calling for designs that explicitly support dual use, where both the user and the educator can intervene in the interaction. This is in line with the interdependence perspective on assistive technology, which frames access as a collaborative accomplishment in which people with disabilities retain and exercise agency (Cynthia et al., 2018). It is also consistent with empirical work on in-home conversational agents, where co-design with children and their parents has highlighted how interactions may extend beyond the primary user, with household members playing an active role in guiding engagement with the system (Garg and Sengupta, 2020). Such a design perspective leaves room for educators' contributions in a meta-design role (Fischer et al., 2004), enabling them to shape, adapt, and guide the interaction as needed.
- **DU2: Ensure multimodal and adaptive interaction.** Multimodal interaction can reduce cognitive overload while improving user engagement (Calò and De Russis, 2024) and supporting comprehension and adaptation to individual differences (Chen et al., 2013). When designed adaptively, it accommodates the diverse needs of users, particularly those with ID Standen et al. (2020). Our analysis highlights three key directions: (i) enabling input through both text and speech to match users' expressive abilities; (ii) combining text, audio, and visual output to enhance understanding; and (iii) using image or video sequences to clarify cause-and-effect relations in home automation scenarios.
- **DU3: Provide clear and concrete language.** Interfaces should employ plain, concrete, and sequential language, following cognitive accessibility design principles Cosentino et al. (2021). For instance, posing “How many minutes should it last?” instead of “How much time should it last?”, or offering structured options such as “Do you want to watch TV or go for a walk?” rather than open-ended prompts, aligns with guidelines for enhancing comprehension of users with ID Sieghart et al. (2024); Finlay and Antaki (2012). These guidelines should be instilled into the LLM prompts governing the dialog.
- **Decomposing cognitively demanding tasks into smaller, sequential steps can also reduce overload and foster understanding Sieghart et al. (2024). Based on our analysis, complex requests should therefore be reformulated as progressive prompts (e.g., first “Do you want a specific duration?” followed by “Should it be longer than 5 min?”).**
- **DU4: Recall familiar contexts.** The system should prioritize real-life and concrete examples (e.g., familiar names, scenarios, and routines) to make interactions more familiar, increase engagement, and reduce caregiver intervention (Sieghart et al., 2024; Hollomotz, 2018; Morra et al., 2024). This can be achieved through adequate pre-training of the LLM model handling the dialog (BU3).
- **AU1: Provide continuous and multi-level training.** Given the rapid evolution of LLM technologies, systems should provide structured, multi-level, and ongoing training for users and caregivers. Training should progress from onboarding to periodic refreshers, involving the broader support network (educators, families, co-living groups) to sustain adaptation. Systems can also foster Communities of Practice (CoP) that enable stakeholders to share strategies and build collective capacity Lave and Wenger (1991).
- **AU2: Support flexible reconfiguration.** As user behaviors and contexts change, caregivers should be able to easily adjust system limits and blocks to maintain adaptation to individual needs, in line with EUD goals (Lieberman et al., 2006). The system should support this through intuitive management tools (e.g., dashboards or guided conversations) that enable caregivers to modify rules, set temporal conditions, and monitor user-generated patterns over time.
- **AU3: Foster autonomy through adaptive automation.** Automation should follow principles of adaptive assistive technologies, scaffolding rather than substituting user skills, to foster independence and ensure long-term autonomy Pea (2018); Gildenpennig et al. (2019). Our analysis suggests that LLM-based tools should act as compensatory systems, providing contextual guidance (e.g., reminders, confirmations, feedback) while preserving space for decision-making. This requires automation to dynamically adjust, gradually reducing assistance as users

grow competence, and enabling caregivers to deliver tailored interventions that promote progressive autonomy.

6.3.1. Limitations

Some limitations must be acknowledged to frame the results of this work in a broader perspective. In Phase 1, defining instructions from video scenarios represented a challenge for some of the involved users, drawing on specific cognitive and linguistic skills. While this aspect was partially addressed through the educator's assistance, it could have impacted the written production of some participants. In addition, prompt design (see Table 2) could have impacted Phase 1's results: the decision to prompt pseudo-code generation without specifying syntax or a reference framework might have affected the variability observed in the generated code.

Phase 2 involved four dyads, a size aligned with research involving people with ID, where ethical, logistical, and cognitive constraints limit recruitment. Given the high heterogeneity of ID profiles, our findings should be interpreted as *analytical generalizations* that illustrate what types of interactions are possible, rather than statistical generalizations. The goal of Phase 2 is thus exploratory: to surface interaction patterns, breakdowns, and opportunities that can guide the design of future LLM-based systems and larger studies.

Additionally, the online setting of the Phase 2 studies and the absence of physical feedback on real devices may have influenced how users perceived and engaged with the prototype, potentially affecting its real-world applicability and generalization.

Finally, while the latest available version of the ChatGPT model was employed in this study, an LLM's capabilities are rapidly evolving, and results related to intent understanding and code and question generation may require re-evaluation with future model releases. While our findings necessarily reflect the behavior of a proprietary and evolving LLM, future systems can however mitigate this volatility by constraining generation through local, model-independent mechanisms. Retrieval-Augmented Generation (RAG) approaches can ground the model on a stable and auditable knowledge base, reducing reliance on undocumented prompt preprocessing and postprocessing and increasing the reproducibility of intent clarification and rule construction.

Our results should be interpreted as interaction patterns that can be preserved even within more controllable architectures. Nevertheless, it is essential to account for the fact that technological advancements do not automatically lead to improvements in accessibility or usability. Investigating current accessibility with a user-centric perspective serves as a benchmark for assessing future model improvements.

7. Conclusion

This work investigates the potential of an LLM-based home automation assistant for people with ID. The first part of our study shows that while current ChatGPT capabilities for recognizing and addressing ambiguities in user-generated prompts are promising, they remain limited due to the model's unpredictability. To address this limitation, we developed a prototype of an LLM-based EUD interface and evaluated it through two comprehensive user studies involving participants with ID, caregivers, and educators. By leveraging a scaffolded 3-step framework that guides users from ideating instructions to generating ready-to-deploy code, we demonstrate that LLM-based EUD systems hold promise for enabling people with ID to personalize and control their domestic environment through natural language interaction. We thus derive a set of design implications to advance accessibility and usability in this emerging context. Beyond informing system design, these implications highlight the need to situate such technologies in a socio-technical framework where caregivers, individuals with ID, and technology collaborate towards enhancing autonomy without replacing the ability and responsibility of human actors (Pea, 2018). Future work will pursue this direction by developing a platform where principles from EUD and meta-design Fischer et al. (2004); Fischer (2013) help promote human roles while safeguarding technology.

More generally, our research is positioned within the broader context of intent-based interfaces, which are rapidly gaining traction through the diffusion of LLMs and the promise of natural language as a universal medium of interaction Min et al. (2025). Yet, our findings highlight that equating "natural" with "inclusive" risks overlooking the needs of people with disabilities Pucci et al. (2024). This work takes one of the first steps toward addressing this gap by investigating what "natural" interaction actually means for individuals with ID, and by deriving design implications that move beyond the natural language alone. By reframing intent-based interaction and the related LLM-prompt engineering through inclusivity, we aim to spark a broader research agenda within HCI that ensures future LLM-based interfaces are not only powerful but also meaningfully accessible to all.

CRedit authorship contribution statement

Diego Morra: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Margherita Andrao:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Qi Ai:** Writing – original draft, Software, Data curation. **Marco Mores:** Investigation. **Maristella Matera:** Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Barbara Treccani:** Supervision. **Massimo Zancanaro:** Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to the daycare center "Fraternità & Amicizia" (Milan) and to the non-profit association "Down DADI" (Padua) for their contribution to the study presented in this article. We also thank the study participants, their family caregivers, and the educators.

This work is funded by "Future Artificial Intelligence Research" (FAIR – Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 – Partnerships extended to universities, research centers, companies, and research D.D. MUR n. 341 del 15.03.2022, PE0000013, CUP: H97G22000210007), and the project "Direct Manipulation in Multimodal Intent-Based User Interfaces", funded by the Google's "Society-Centered AI" award program.

Appendix A. Phase 1: materials

Supplementary materials used in Phase 1, including the home-automation scenarios presented to participants during the study.

Table A.9

Home-automation scenarios illustrated in the videos. Each video depicted a *state–event–action* sequence (Huang and Cakmak, 2015; Ghiani et al., 2017) to demonstrate the functioning of interconnected smart home devices. The scenarios were designed to represent a range of common home-automation tasks and were distributed evenly across four themes, identified in consultation with an expert educator. Video durations ranged from 13 to 16 seconds ($M = 15.00$, $SD = 1.41$).

id	state	event	action	theme	s
s1	night	person enters	light on	bathroom	16
s2	person inside	handle is lowered	red light on	bathroom	16
s3	during the night	door opens	red light on	home security	16
s4	during the night	window opens	red light on	home security	16
s5	it is cold	person enters	fireplace on	temperature	16
s6	it is cold	it becomes night	fireplace on	temperature	13
s7	dog inside	smoke	door opens	personal safety	14
s8	smoke	temperature rises	sprinklers on	personal safety	13

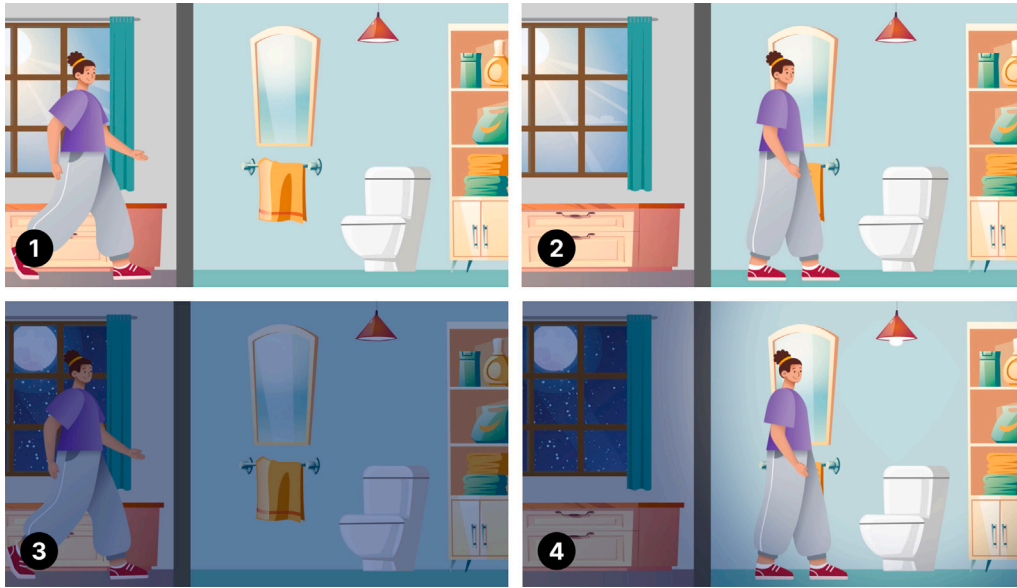


Fig. A.5. Example frames extracted from scenario *s1*. The top-left frame (1) shows the event “a person enters” while the state “during the night” is false (i.e., “during the day”); the top-right frame (2) shows that no action is executed; the bottom-left frame (3) shows the event “a person enters” while the state “during the night” is true; and the bottom-right frame (4) shows the execution of the action “light turns on”. In each video, the first part depicts the situation with the state “false” and no action is triggered, followed by the one where the state is true and the action is performed. To mitigate potential biases related to spatial positioning, an alternative mirrored version of each video was also created. Each participant was shown half of the videos in the original orientation and half in the mirrored version.

Appendix B. Phase 1: instruction analysis supplementary material

Supplementary material for the Phase 1 instruction analysis, including examples of judges’ intent interpretations and LLM-generated outputs.

Table B.10

Examples of intent descriptions provided by judges (*J1*, *J2*, *J3*) for participant-generated instructions. The “*compl.*” column indicates whether the rule was complete (“yes”) or incomplete (“no”). The “*intent*” column specifies the interpretation difficulty level. Two scenarios (“*sc.*”) were selected: *s4* (“night, window opens, red light”) and *s1* (“night, person enters bathroom, light”).

id	sc.	rule	compl.	J1	J2	J3	intent
P4	<i>s4</i>	<i>opened the window and there was the alarm</i>	no	if alarm is on and window is opened, it should trigger	wants the alarm to sound when the window opens	wants the alarm to be triggered when someone opens the window	easy
P8	<i>s4</i>	<i>the girl opens the window when it is light the alarm did not flash when it is dark the alarm was triggered</i>	yes	in daytime, alarm is off, allowing windows to be opened. in evening, alarm is on, so if someone opens the windows, the alarm will be triggered	wants the alarm to sound if one tries to open the window in the dark	wants to activate alarm at sunset and deactivate it at sunrise	mod. diff.
P10	<i>s1</i>	<i>first it was morning and was in the bathroom then turned on the bathroom light</i>	no	when residents wake up in the morning, the lights must be on	wants bathroom light to turn on in the evening when entering the bathroom	wants the bathroom light to stay on during night and to turn off in the morning	difficult
P6	<i>s1</i>	<i>we would need emergency lights when it's all dark. Because [name] sometimes falls, once I found her here on my legs.</i>	yes	when it is dark, user would like to have some lights on	wants to have emergency lights on at night with the lights off	wants emergency lights to be on during the night	easy

Table B.11

Two examples of GPT-4o outputs when prompted with participants' rules, both classified as "easy" to interpret. For p12, the generated pseudo-code did not correctly reflect the rule's intent, whereas for p11, it did.

id	rule [scen.]	Code-from-rule	Check-rule	From-rule-request
P12	<i>in the morning a girl opens the window and at night she opens the window and sets off the alarm [s4]</i>	<pre> morning_routine(): if time == 'morning': if room == 'Bedroom': open(Bedroom_window) night_routine(): if time == 'night': if room == 'Bedroom': open(Bedroom_window) trigger(Bedroom_alarm) run_routines(): while True: current_time = get_current_time() if current_time == 'morning': morning_routine() elif current_time == 'night': night_routine() sleep(60) Check every minute </pre>	1. Ambiguity in Timing: The rule states "in the morning" and "at night". These terms are subjective and can vary depending on personal or system-defined times. For instance, when exactly does 'morning' start and 'night' begin? Is 'night' after a specific time (e.g., 10 PM)? 2. Unclear Subject: The rule mentions "a girl" opens the window. However, this real person or entity is not linked to the smart system as described in the table. The rule should either refer to actions taken by users via the smart devices or some form of automation. 3. Lack of Specific Room: The rule specifies actions without referencing a specific room. Since the table lists multiple rooms with windows and alarms in the Bedroom, it should clearly state which room's window is being opened and in which room the alarm is being activated. 4. Action Sequence Issue (Error): The rule has two actions linked to "at night": opening the window and sounding the alarm. These actions might be considered conflicting. If a window is expected to be opened at night, setting off an alarm could be counterproductive unless the alarm is meant to sound only when the window is opened unexpectedly. 5. Noise and Alarm Sensors: There is no direct mention of an 'alarm' being a sensor or a device in any room other than Bedroom. It's not clear whether the alarm sound is generated by the "Noise" sensor or a specific "Alarm" smart device. 6. Logical Coherence and Device Availability: The rule execution depends on installed sensors or devices. Ensure that the necessary sensors and devices required to implement the rule are appropriately mentioned in the context and table.	1. What specific time in the morning does the girl open the window? 2. Does the girl open the same window in both the morning and at night? 3. What time period constitutes "night" in this context? 4. Is the alarm only on at night or also in the morning? 5. What type of alarm is being referred to? 6. Is the alarm triggered automatically when the window is opened at night or does she manually activate it? 7. What is the purpose of opening the window in both the morning and at night? 8. Should the process of opening the window be done daily, or only under certain conditions? 9. Is there any specific sequence to the actions taken at night (opening the window first then sounding the alarm)? 10. Are there any exceptions to when the window should not be opened or the alarm not sounded? 11. Are there different actions to take during weekends or holidays? 12. Is there a different protocol if the girl is not present at home? 13. Are there any safety precautions to be followed while opening the window or sounding the alarm?
P11	<i>set the alarm at night [s3]</i>	<pre> BEGIN FUNCTION setNightAlarm() ROOMS = ["Living room/Open kitchen", "Bedroom", "Halway", "Bathroom"] FOR EACH ROOM IN ROOMS IF ROOM.contains("Alarm") ROOM.Alarm.set("ON") ELSE PRINT "No alarm found in " + ROOM END IF END FOR END FUNCTION FUNCTION main() setNightAlarm() END FUNCTION END </pre>	1. Time Frame: The term "at night" is vague and could be interpreted in various ways. Does "night" start at sunset, a specific hour like 8 PM, or at bedtime? This ambiguity could lead to different interpretations and implementations. 2. Definition of "Alarm": It is unclear what is meant by "alarm". Is it a burglar alarm, fire alarm, or a general security system? Different rooms have sensors like smoke/fire detectors, presence detectors, and entrance sensors that could trigger different types of alarms. 3. Scope of Rooms: The rule does not specify if the alarm should be set for all rooms or only specific rooms. Each room has different sensors and smart devices that could influence how and where the alarm should be configured. 4. Sensor and Device Interaction: If "alarm" means a security system, there is a logical issue if certain sensors or smart devices required to set the alarm are not present in all rooms. For instance, if no presence detection (movement sensor) is in some rooms, the alarm system may not function as intended. 5. Specific Smart Device Control: If the rule intends to involve smart devices to set the alarm, it is not clear which devices should be used or activated to set the alarm. For example, should the "door" smart device lock as part of the alarm setting process?	1. What time should the alarm be set for at night? 2. Is the alarm supposed to recur every night or just once? 3. Should the alarm be set for the same time every night? 4. Is there a specific sound or volume that the alarm should use? 5. Should the alarm also be set for weekends? 6. Are there any specific conditions under which the alarm should not be set? 7. Is there more than one alarm that needs to be set? 8. Should the alarm be connected to any other devices or systems? 9. Is it acceptable to use a smartphone alarm, or is there a preferred device? 10. Are there any specific instructions for how to disable the alarm in the morning? 11. Should the alarm only be set on certain days of the week? 12. Are there any other settings or configurations required for the alarm?

Table B.12

Results of the normality test (*Shapiro-wilk*) and homoscedasticity tests (*Brown-forsythe* and *Levene*) performed on the number of detected ambiguities ("ambiguity_n") and generated questions ("question_n"), considering intent recognition ("intent_yes", "intent_no") and clarity levels ("easy", "moderate", "difficult") of incomplete rules. Significant shapiro-wilk tests ($p < .05$) indicate deviations from normality, justifying the use of non-parametric tests. Non-significant brown-forsythe and levene tests ($p > .05$) confirm that group variances are sufficiently homogeneous.

	test	data	statistic	value	p
ambiguity_n	<i>Shapiro-Wilk</i>	intent_yes	<i>W</i>	0.90	0.01
		intent_no	<i>W</i>	0.91	0.01
	<i>Brown-Forsythe</i>		<i>F</i> (1, 54)	0.45	0.51
		easy	<i>W</i>	0.83	.001
		moderate	<i>W</i>	0.95	0.19
	<i>Shapiro-Wilk</i>	difficult	<i>W</i>	0.92	0.25
			<i>F</i> (2, 53)	0.38	0.69
question_n	<i>Shapiro-Wilk</i>	intent_yes	<i>W</i>	0.90	0.01
		intent_no	<i>W</i>	0.79	0.01
	<i>Brown-Forsythe</i>		<i>F</i> (1, 54)	0.97	0.33
		easy	<i>W</i>	0.91	0.09
		moderate	<i>W</i>	0.81	.001
	<i>Shapiro-Wilk</i>	difficult	<i>W</i>	0.85	0.03
			<i>F</i> (2, 53)	5.195	0.01

Appendix C. Phase 2: chat-based prototype interface

Additional details on the chat-based prototype used in Phase 2, including the interface and the system prompt implementing the three-step interaction framework.

C.1. System prompt of phase 2 prototype

Phase 2 chat-based prototype was configured through the system prompt reported below, which operationalizes the three-step framework of [Section 5.1](#):

1. Onboarding phase

Users will provide an initial prompt based on a short video that illustrates a home-automation scenario using trigger/action rules. The video is shown on a separate screen, and the user summarizes the observed scenario in their own words. It is very likely that, at this stage, the prompt will be incomplete, ambiguous, or hard to understand. If you recognize that the prompt received is unrelated to home automation, respond by specifying this and prompting the user to write a home-automation instruction, even in incomplete or draft form.

2. Rule correction phase

You need to analyze the user's input to identify any ambiguities that could cause errors in code generation. You can ask up to five clarification questions, but always one at a time, moving to the next only after the user has answered. You always have to specify that, if the user does not know how to answer, they may say so and you will offer two alternatives to choose from. Once all necessary information has been collected, you should present an updated version of the user's original prompt, refined on the basis of the information gathered.

3. Scenario visualization phase

Always use the updated version of the user's prompt, generate pseudo-code that reflects the corresponding trigger-action rule. You have to provide only a detailed description of the code's functionality to confirm it matches the user's intent, without showing the code or any technical details. To aid comprehension you may offer a brief hypothetical scenario, possibly featuring a human figure (e.g., "it's night, Anna enters the kitchen, and..."). Ask the user to confirm that the described functionality matches the desired one. Only upon the user's confirmation can you inform the user that the code has been generated successfully, otherwise, prompt the user to tell you what changes need to be made. Repeat this confirmation step until the user confirms that the functionality matches the desired one.

In your responses, you should emphasize clarity, simplicity, and accessibility for non-expert users while maintaining technical rigor in code generation. Responses should also be structured to guide the user, minimizing technical jargon and ensuring that all questions and generated outputs are accurate and relevant to the described scenarios.

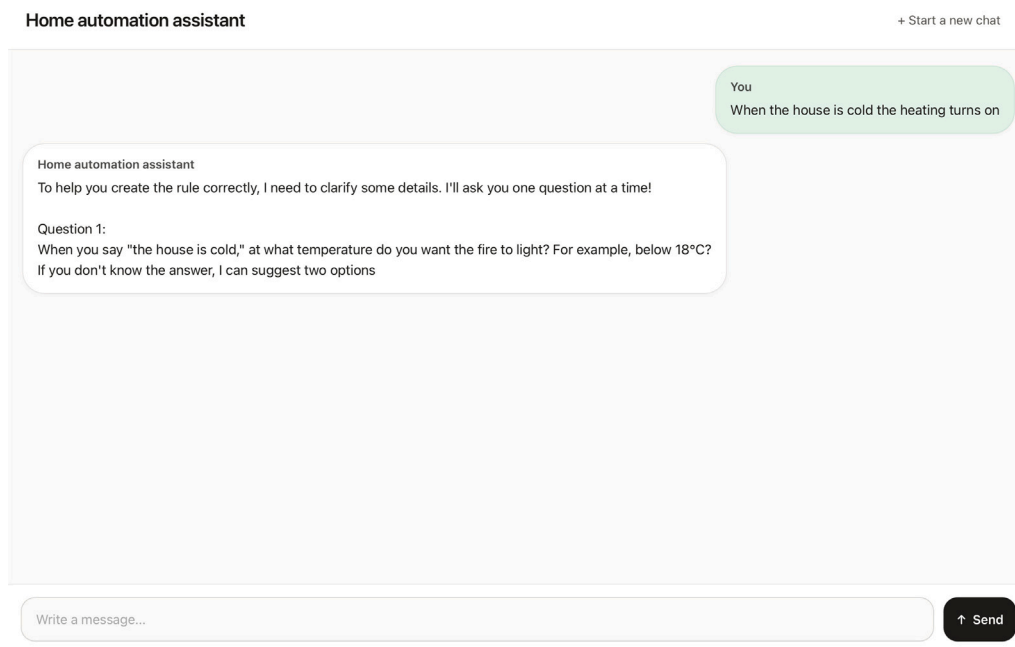


Fig. C.6. Screenshot of the chat-based prototype tool used in phase 2.

Appendix D. Phase 2: results UEQ-s

Supplementary UEQ-S results from the Phase 2 evaluations, covering both the workshop with experts and the prototype evaluation with user-caregiver dyads.

D.1. Workshop with experts

Table D.13

Summary of UEQ-s scores (divided by scales and the global score) collected from the experts ($N = 5$).

UEQ-S scales	<i>M</i>	<i>SD</i>	<i>N</i>
<i>Pragmatic Quality</i>	−0.35	0.68	5
<i>Hedonic Quality</i>	1.85	0.49	5
Overall	0.75	0.29	5

Table D.14

Descriptive statistics of UEQ-s score per item (it.). Interpretation guidelines (Schrepp et al., 2017): values between −0.80 and 0.80 represent a neutral evaluation of the corresponding scale, values > 0.8 represent a positive evaluation, and values < −0.8 represent a negative evaluation. The range of the scales was between −3 (“negative pole”) and +3 (“positive pole”).

it.	<i>M</i>	<i>SD</i>	<i>N</i>	negative p.	positive p.	scale
1	1.00	1.00	5	obstructive	supportive	pragmatic
2	−1.40	0.90	5	complicated	easy	pragmatic
3	0.20	0.40	5	inefficient	efficient	pragmatic
4	−1.20	0.80	5	confusing	clear	pragmatic
5	1.80	1.30	5	boring	exciting	hedonic
6	1.20	0.40	5	not interesting	interesting	hedonic
7	2.00	0.70	5	conventional	inventive	hedonic
8	2.40	0.50	5	usual	leading edge	hedonic

D.2. Prototype evaluation with dyads

Table D.15

Summary of UEQ-s scores (divided by scales and the global score) collected from the caregivers ($N = 4$).

UEQ-S scales	<i>M</i>	<i>SD</i>	<i>N</i>
<i>Pragmatic Quality</i>	0.81	2.18	4
<i>Hedonic Quality</i>	2.00	0.94	4
Overall	1.41	1.53	4

Table D.16

Descriptive statistics of UEQ-s scores per item (it.). Interpretation guidelines (Schrepp et al., 2017).

it.	<i>M</i>	<i>SD</i>	<i>N</i>	negative p.	positive p.	scale
1	1.75	2.50	4	obstructive	supportive	pragmatic
2	0.25	3.20	4	complicated	easy	pragmatic
3	0.00	2.16	4	inefficient	efficient	pragmatic
4	1.25	2.06	4	confusing	clear	pragmatic
5	1.25	2.06	4	boring	exciting	hedonic
6	2.50	1.00	4	not interesting	interesting	hedonic
7	1.75	1.50	4	conventional	inventive	hedonic
8	2.50	0.58	4	usual	leading edge	hedonic

Appendix E. Phase 2: user-generated instructions

Instructions generated by participants during the Phase 2 prototype evaluation, including their initial formulations and the versions refined through interaction with the system.

Table E.17

User-generated instructions before and after rounds (“r”) of interaction with the interface.

id	r	scenario	user-generated instruction	refined instruction
P1	1	video1	<i>Turn on the light when you enter the bathroom and turn it off when you leave the bathroom and flush the toilet.</i>	Light activation: When the bathroom door is opened (entry). Light deactivation and flushing: When the door is closed (exit), without delay.
P1	2	self-defined	<i>When I wake up to go to the daily center warm the milk.</i>	Every weekday, at 7:30 AM, when the alarm rings, the system will turn on the stove to heat the milk.
P1	3	self-defined	<i>Alexa turn on the washing machine and puts the 30-minute short.</i>	Alexa prepares the washing machine: Turns it on and asks which program to set (e.g., 30-minute short cycle). This happens before you leave for athletics. Start the wash: When you return, tell Alexa, “I’m back, start the washing machine”, and it starts the chosen program. Automatic shutdown: The washing machine turns off automatically after the wash.
P2	1	video1	<i>The girl that is going to the bathroom needs to turn on the light.</i>	When the motion sensor detects the girl entering the bathroom, the light turns on automatically. The light stays on until the girl manually turns it off using the switch.
P2	2	video2	<i>The window doesn’t open from outside, call the police.</i>	If the window opening sensor detects an attempt to open it: 1. An automatic call to the police is made immediately. 2. A notification is sent to the iPad and computer. 3. A sound alarm is activated in the house to scare the intruder.
P2	3	video3	<i>When the house is cold the fire turns on.</i>	The smart fireplace automatically turns on when the temperature drops below 18°C. The smart fireplace automatically turns off when the temperature exceeds 22°C. This rule is always active, at any time of the day.
P2	4	video4	<i>When the pot overheats the fire detector activates.</i>	When the pot reaches 200°C, the fire detector activates, and the system automatically turns off the stove using a smart plug.
P3	1	video1	<i>Is night, she goes to the bathroom, the light turns on, and she brushes her teeth.</i>	At night, when a person enters the bathroom, the light turns on automatically. After 5 min, a scent diffuser activates to make the environment more pleasant. When the person leaves the bathroom, the light turns off.
P3	2	video2	<i>Is night dark, close the window.</i>	If it's night and the room is dark, close the window only if it's open. Send a notification message when the window is closed.
P4	1	video1	<i>Turn on when it's dark.</i>	The bathroom light turns on automatically when a light sensor detects it's dark and turns off automatically when the brightness returns to sufficient levels. This rule is active at any time of the day.

Appendix F. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.ijhcs.2026.103938.

Data availability

Data will be made available on request.

References

- Ahmed, E., Ashraf Islam, F.S., Huda, M.N., Abdullah-Al-Mamun, K., 2016. A road to independent living with smart homes for people with disabilities. In: 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 472–477. <https://doi.org/10.1109/ICIEV.2016.7760048>
- American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Publishing. 9780890425572 <https://books.google.it/books?id=-JivBAAAQBAJ>
- Andrao, M., Balducci, F., Breve, B., Cena, F., Desolda, G., Deufemia, V., Gena, C., Matera, M., Mattioli, A., Paternò, F., Santoro, C., Treccani, B., Vernerio, F., Zancanaro, M., 2023. Understanding concepts, methods and tools for end-user control of automations in ecosystems of smart objects and services. In: End-User Development: 9th International Symposium, IS-EUD 2023, Cagliari, Italy, June 6–8, 2023 Cagliari, Italy. Springer-Verlag, Berlin, Heidelberg, pp. 104–124. https://doi.org/10.1007/978-3-031-34433-6_7. 978-3-031-34433-9.
- Andrao, M., Morra, D., Paccosi, T., Matera, M., Treccani, B., Zancanaro, M., 2024. This sounds unclear: evaluating ChatGPT capability in translating end-user prompts into ready-to-deploy Python code. In: Proceedings of the 2024 International Conference on Advanced Visual Interfaces (Arenzano, Genoa, Italy) (AVI '24). Association for Computing Machinery, New York, NY, USA, pp. 4. <https://doi.org/10.1145/3656650.3656693>, Article 4 9798400717642.
- Ardito, C., Costabile, M.F., Desolda, G., Manca, M., Matera, M., Paternò, F., Santoro, C., 2019. Improving tools that allow end users to configure smart environments. In: End-User Development: 7th International Symposium, IS-EUD 2019, (10-12 July 2019), vol. 7, Springer, Hatfield, UK, pp. 244–248.
- Artizzu, V., Carcangiu, A., Manca, M., Mattioli, A., Mereu, J., Paternò, F., Santoro, C., Simeoli, L., Spano, L.D., 2024. End-user development for eXtended reality using a multimodal intelligent conversational agent. In: Proceedings of the Workshop on Prototyping and Developing Real-World Applications of Extended Reality at the 17th International Conference on Advanced Visual Interfaces (Genoa, Italy).
- Bacchin, D., Pluchino, P., Grippaldi, A.Z., Mapelli, D., Spagnoli, A., Zanella, A., Gamberini, L., 2021. Smart co-housing for people with disabilities: a preliminary assessment of caregivers' interaction with the DOMHO system. *Front. Psychol.* 12, 734180.
- Ball, J., 2019. The double diamond: a universally accepted depiction of the design process. <https://www.designcouncil.org.uk/our-resources/the-double-diamond/>.
- Barricelli, B.R., Cassano, F., Fogli, D., Piccinno, A., 2019. End-user development, end-user programming and end-user software engineering: a systematic mapping study. *J. Syst. Softw.* 149, 101–137, 0164-1212. <https://doi.org/10.1016/j.jss.2018.11.041>
- Barricelli, B.R., Fogli, D., Iemmolo, L., Locoro, A., 2022. A multi-modal approach to creating routines for smart speakers. In: Proceedings of the 2022 International Conference on Advanced Visual Interfaces (Frascati, Rome, Italy) (AVI '22). Association for Computing Machinery, New York, NY, USA, pp. 5. <https://doi.org/10.1145/3531073.3531168>, Article 37. 9781450397193.
- Blackwell, A.F., 2002. First steps in programming: a rationale for attention investment models. In: Proceedings of the IEEE 2002 Symposium on Human Centric Computing Languages and Environments (HCC'02), vol. 2, IEEE Computer Society, USA, <https://doi.org/10.1109/HCC.2002.1046334>. 0769516440.
- Blase, U., McManus, E., Ho, M.P.Y., Littman, M.L., 2014. Practical trigger-action programming in the smart home. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, pp. 803–812. <https://doi.org/10.1145/2556288.2557420>. 9781450324731 26 April 2014–1 May 2014.
- Blase, U., Ho, M.P.Y., Brawner, S., Lee, J., Mennicken, S., Picard, N., Schulze, D., Littman, M.L., 2016. Trigger-action programming in the wild: an analysis of 200,000 iftt recipes. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, pp. 3227–3231. <https://doi.org/10.1145/2858036.2858556>. 9781450333627.
- Brackenbury, W., Deora, A., Ritchey, J., Vallee, J., Weijia, H., Wang, G., Littman, M.L., Blase, U., 2019. How users interpret bugs in trigger-action programming. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, pp. 1–12. <https://doi.org/10.1145/3290605.3300782>. 9781450359702.
- Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3 (2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Braun, V., Clarke, V., 2019. Reflecting on reflexive thematic analysis. *Qual. Res. Sport Exerc. Health* 11 (4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>

- Braun, V., Clarke, V., 2021. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and psychotherapy research* 21 (1), 37–47. <https://doi.org/10.1002/capr.12360>
- Calò, T., De Russis, L., 2024. Enhancing smart home interaction through multimodal command disambiguation. *Pers. Ubiquitous Comput.* 1–16.
- Carey, E., Ryan, R., Sheikh, A., Dore, L., 2023. Exercising autonomy—the effectiveness and meaningfulness of autonomy support interventions engaged by adults with intellectual disability. A mixed-methods review. *British Journal of Learning Disabilities* 51 (3), 307–323. <https://doi.org/10.1111/bld.12464>
- Carmien, S.P., Fischer, G., 2008. Design, adoption, and assessment of a socio-technical environment supporting independence for persons with cognitive disabilities. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, pp. 597–606. <https://doi.org/10.1145/1357054.1357151>. 9781605580111.
- Chan, S., Jiachen, L., Siman, A., Wang, Y., Bilau, I., Jones, B.D., Yang, E., Mynatt, E.D., Tan, X.Z., 2025. Insights from designing context-aware meal preparation assistance for older adults with mild cognitive impairment (MCI) and their care partners. In: *Proceedings of the 2025 ACM Designing Interactive Systems Conference* (DIS '25). Association for Computing Machinery, New York, NY, USA, pp. 3263–3279. <https://doi.org/10.1145/3715336.3735817>. 9798400714856.
- Chen, F., Ruiz, N., Choi, E., Julien Epps, M.A.K., Taib, R., Yin, B., Wang, Y., Jan 2013. Multimodal behavior and interaction as indicators of cognitive load. *ACM Trans. Interact. Intell. Syst.* 2 4, Article 22 36, 2160–6455. <https://doi.org/10.1145/2395123.2395127>
- Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A.N., Tianle, L., Dacheng, L., Zhu, B., Zhang, H., Jordan, M., Gonzalez, J.E., Stoica, I., 2024. Chatbot arena: an open platform for evaluating LLMs by human preference. In: *Forty-First International Conference on Machine Learning*. <https://openreview.net/forum?id=3MW8GKNyZl>.
- Clarke, V., Braun, V., 2017. Thematic analysis. *The journal of positive psychology* 12 (3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Cleland, J., Hutchinson, C., Williams, P.A.H., Manuel, K., Laver, K., 2024a. The experience of using home automation by individuals with disability. *Disabil. Rehabil. Assist. Technol.* 19 (6), 2389–2396. <https://doi.org/10.1080/17483107.2023.2288391>
- Cleland, J., Hutchinson, C., Patricia, A.H., Kisan, Manuel, W., Laver, K., 2024b. A scoping review to explore the health, social and economic outcomes of home automation for people with disability. *Disabil. Rehabil. Assist. Technol.* 19 (4), 1446–1453. <https://doi.org/10.1080/17483107.2023.2196308> PMID: 37012673.
- Coppens-Hofman, M.C., Terband, H., Snik, A.F.M., Maassen, B.A.M., 2017. Speech characteristics and intelligibility in adults with mild and moderate intellectual disabilities. *Folia Phoniatrica et Logopaedica* 68 (4), 175–182.
- Corno, F., De Russis, L., Roffarello, A.M., 2019. Empowering end users in debugging trigger-action rules. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, pp. 1–13. <https://doi.org/10.1145/3290605.3300618>. 9781450359702.
- Cosentino, G., Morra, D., Gelsomini, M., Matera, M., Mores, M., 2021. COBO: a card-based toolkit for co-designing smart outdoor experiences with people with intellectual disability. In: *Ardito, C., Lanzilotti, R., Malizia, A., Petrie, H., Piccinno, A., Desolda, G., Inkpen, K. (Eds.), Human-Computer Interaction – INTERACT 2021*. Springer International Publishing, Cham, pp. 149–169. 978-3-030-85623-6.
- Cynthia, L., Erin Brady, B., Branham, S.M., 2018. Interdependence as a frame for assistive technology research and design. In: *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) (ASSETS '18). Association for Computing Machinery, New York, NY, USA, pp. 161–173. <https://doi.org/10.1145/3234695.3236348>. 9781450356503.
- Desolda, G., Ardito, C., Matera, M., Apr 2017. Empowering end users to customize their smart environments: model, composition paradigms, and domain-specific tools. *ACM Trans. Comput.-Hum. Interact.* 24 2, Article 12 52, 1073–0516. <https://doi.org/10.1145/3057859>
- Dewsbury, G., Clarke, K., Hughes, J., Rouncefield, M., Sommerville, I., 2003. Designing dependable digital domestic environments. In: *Proceedings of HOIT*.
- Dobkin, B., Dorsch, A., 2017. The evolution of personalized behavioral intervention technology: will it change how we measure or deliver rehabilitation? *Stroke* 48, 2329–2334. <https://doi.org/10.1161/STROKEAHA.117.016620>
- Dorst, K., 2011. The core of design thinking and its application. *Des. Stud.* 32 (6), 521–532. <https://doi.org/10.1016/j.destud.2011.07.006>
- Ellis, K., Kruesi, L., Ananthanarayan, S., Senaratne, H., Lindsay, S., 2023. Piece it together: insights from one year of engagement with electronics and programming for people with intellectual disabilities. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17.
- Evan, K., Haoxiang, Y., Lee, S., Julien, C., Mar 2024. Sasha: creative goal-oriented reasoning in smart homes with large language models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8 (1), Article 12 38. <https://doi.org/10.1145/3643505>
- Fakhoury, S., Naik, A., Sakkas, G., Chakraborty, S., Lahiri, S.K., 2024. LLM-based test-driven interactive code generation: user study and empirical evaluation. *IEEE Trans. Softw. Eng.* 50 (9), 2254–2268.
- Finlay, W.M.L., Antaki, C., 2012. How staff pursue questions to adults with intellectual disabilities. *J. Intellect. Disabil. Res.* 56 (4), 361–370.
- Fischer, G., 2013. End-user development: from creating technologies to transforming cultures. In: *End-User Development: 4th International Symposium, IS-EUD 2013, Copenhagen, Denmark, (10-13 June 2013), vol. 4*. Springer, pp. 217–222.
- Fischer, G., Giacardi, E., Ye, Y., Sutcliffe, A.G., Mehndjeev, N., Sept 2004. Meta-design: a manifesto for end-user development. *Commun. ACM* 47 (9), 33–37, 0001-0782. <https://doi.org/10.1145/1015864.1015884>
- Gallo, S., Manca, M., Mattioli, A., Paternò, F., Santoro, C., 2021. Comparative analysis of composition paradigms for personalization rules in IoT settings. In: *Fogli, D., Tetteroo, D., Barricelli, B.R., Borsci, S., Markopoulos, P., Papadopoulos, G.A. (Eds.), End-User Development*. Springer International Publishing, Cham, pp. 53–70. 978-3-030-79840-6. https://doi.org/10.1007/978-3-030-79840-6_4
- Garg, R., Sengupta, S., 2020. Conversational technologies for in-home learning: using co-design to understand children's and parents' perspectives. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20). Association for Computing Machinery, New York, NY, USA, pp. 1–13. <https://doi.org/10.1145/3313831.3376631>
- Gentry, T., 2009. Smart homes for people with neurological disability: state of the art. *NeuroRehabilitation* 25 (3), 209–217. <https://doi.org/10.3233/NRE-2009-0517> PMID: 19893189.
- Gervais, M., Bouchard, B., Labreque, L., Tremblay, V., Bouchard, J., Chouinard, M.-C., Dionne, C., Bouchard, K., Gaboury, S., 2023. Technological tools for assisting people with autism spectrum disorder (ASD), intellectual development disorder (IDD) and physical disabilities (PD): developing a technological tool that assists people with different disabilities or disorders to help their personal autonomy. In: *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments* (Corfu, Greece) (PETRA '23). Association for Computing Machinery, New York, NY, USA, pp. 171–176. <https://doi.org/10.1145/3594806.3594857>. 9798400700699.
- Ghani, G., Manca, M., Paternò, F., Santoro, C., Apr 2017. Personalization of context-dependent applications through trigger-action rules. *ACM Trans. Comput.-Hum. Interact.* 24 2, Article 14 33, 1073–0516. <https://doi.org/10.1145/3057861>
- Güldenpennig, F., Mayer, P., Panek, P., Fitzpatrick, G., 2019. An autonomy-perspective on the design of assistive technology experiences of people with multiple sclerosis. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.
- Hollomotz, A., 2018. Successful interviews with people with intellectual disability. *Qual. Res.* 18 (2), 153–170.
- Huang, J., Cakmak, M., 2015. Supporting mental model accuracy in trigger-action programming. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) (UbiComp '15), (7-11 September 2015). Association for Computing Machinery, New York, NY, USA, pp. 215–225. <https://doi.org/10.1145/2750858.2805830>. 9781450335744.
- Jennifer, H., Floyd, S., Jouravlev, O., Fedorenko, E., Gibson, E., 2023. A fine-grained comparison of pragmatic language understanding in humans and language models. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), pp. 4194–4213.
- Landuran, A., Sauzeon, H., Consel, C., N'kaoua, B., 2023. Evaluation of a smart home platform for adults with down syndrome. *Assist. Technol.* 35 (4), 347–357.
- Lave, J., Wenger, E., 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Lewis, B., Ranalli, T.-M., Gourley, A., Kirupaharan, P., Venkatasubramanian, K., 2023. "I...caught a person casing my house...and scared him off." the use of security-focused smart home devices by people with disabilities. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, pp. 16. <https://doi.org/10.1145/3544548.3581007>, Article 846. 9781450394215.
- Lieberman, H., Paternò, F., Klann, M., Wulf, V., 2006. End-user development: an emerging paradigm. In: *Lieberman, H., Paternò, F., Wulf, V. (Eds.), End User Development*. Springer, Dordrecht, pp. 1–8. 978-1-4020-5386-3. https://doi.org/10.1007/1-4020-5386-X_1
- Liu, M.X., Sarkar, A., Negreanu, C., Zorn, B., Williams, J., Toronto, N., Gordon, A.D., 2023. What it wants me to say": bridging the abstraction gap between end-user programmers and code-generating large language models. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–31.
- Lui, C.X.Y., Yang, N., Tang, A., Tam, W.W.S., 2025. Effectiveness evaluation of smart home technology in preventing and detecting falls in community and residential care settings for older adults: a systematic review and meta-analysis. *J. Am. Med. Dir. Assoc.* 26 (1), 105347.
- Lussier-Desrochers, D., Lachapelle, Y., Leclerc, N., Pigot, H., Bauchet, J., Giroux, S., 2012. Assessing the effect of domotics used as an assistant to meal preparation with people with an intellectual disability. In: *The Fifth International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services*, pp. 1–6.
- Manca, M., Paternò, F., Santoro, C., 2022. Remote monitoring of end-user created automations in field trials. *J. Ambient Intell. Humaniz. Comput.* 13 (12), 5669–5697.
- Meissner, J.L., Vines, J., McLaughlin, J., Nappey, T., Maksimova, J., Wright, P., 2017. Do-it-yourself empowerment as experienced by novice makers with disabilities. In: *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) (DIS '17). Association for Computing Machinery, New York, NY, USA, pp. 1053–1065. <https://doi.org/10.1145/3064663.3064674>. 9781450349222.
- Min, B., Chen, A., Cao, Y., Xia, H., 2025. Malleable overview-detail interfaces. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, pp. 25. <https://doi.org/10.1145/3706598.3714164>, Article 688. 9798400713941.
- Morra, D., Caslini, G., Mores, M., Garzotto, F., Matera, M., 2024. MakeNodes: opening connected-IoT making to people with intellectual disability. *Int. J. Hum.-Comput. Stud.* 190, 103325. 1071-5819. <https://doi.org/10.1016/j.ijhcs.2024.103325>
- Ouyang, S., Jie, M.Z., Harman, M., Wang, M., 2023. LLM is like a box of chocolates: the non-determinism of ChatGPT in code generation. [cs.SE]. *arXiv:2308.02828*.
- Palekar, M., Fernandes, E., Roesner, F., 2019. Analysis of the susceptibility of smart home programming interfaces to end user error. In: *2019 IEEE Security and Privacy Workshops* (SPW), pp. 138–143. <https://doi.org/10.1109/SPW.2019.00034>

- Paternò, F., 2013. End user development: survey of an emerging field for empowering people. *Int. Sch. Res. Not.* 1, 532659. <https://doi.org/10.1155/2013/532659>
- Paternò, F., Santoro, C., 2019. End-user development for personalizing applications, things, and robots. *Int. J. Hum.-Comput. Stud.* 131, 120–130, 1071–5819. <https://doi.org/10.1016/j.ijhcs.2019.06.002>
- Pea, R.D., 2018. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. In: *Scaffolding*. Psychology Press, pp. 423–451.
- Perera, C., Aghae, S., Blackwell, A., Díaz, P., Pipek, V., Ardito, C., Jensen, C., Ignacio, A., 2015. Natural notation for the domestic internet of things. In: Boden, A. (Ed.), *End-User Development: 5th International Symposium, IS-EUD 2015 (26–29 May 2015)*, Springer International Publishing, Madrid, Spain, Cham, pp. 25–41. https://doi.org/10.1007/978-3-319-18425-8_3
- Pucci, E., Piro, L., Andolina, S., Matera, M., 2024. From conversational web to inclusive conversations with LLMs. In: *Proceedings of the 2024 International Conference on Advanced Visual Interfaces (Arenzano, Genoa, Italy) (AVI '24)*. Association for Computing Machinery, New York, NY, USA, pp. 3. <https://doi.org/10.1145/3656650.3656739>, Article 87 9798400717642.
- Rahmati, A., Fernandes, E., Jung, J., Prakash, A., 2017. IFTTT vs. zapier: a comparative study of trigger-action programming frameworks. *arXiv preprint [cs.CR]*: <https://doi.org/10.48550/arXiv.1709.02788> arXiv:1709.02788.
- Rahmati, A., Fernandes, E., Eykholt, K., Prakash, A., 2018. Tyche: a risk-based permission model for smart homes. In: *2018 IEEE Cybersecurity Development (SecDev)*, pp. 29–36. <https://doi.org/10.1109/SecDev.2018.00012>
- Rajapakse, R., Brereton, M., Sitbon, L., Roe, P., 2015. A collaborative approach to design individualized technologies with people with a disability. In: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, <https://doi.org/10.1145/2838739.2838824>
- Rey-Jouanchicot, J., Bottaro, A., Campo, E., Bouraoui, J.-L., Vigouroux, N., Vella, F., 2025. A user personalized and adaptive smart home using large language models. In: *2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC)*. IEEE, pp. 1–4.
- Rivkin, D., Hogan, F., Feriani, A., Konar, A., Sigal, A., Liu, S., Dudek, G., 2024a. SAGE: smart home agent with grounded execution. [cs.AI]. [arxiv:2311.00772](https://arxiv.org/abs/2311.00772)
- Rivkin, D., Hogan, F., Feriani, A., Konar, A., Sigal, A., Liu, X., Dudek, G., 2024b. Aiot smart home via autonomous LLM agents. *IEEE Internet Things J.* 12 (3), 2458–2472.
- Roffarello, A.M., De Russis, L., 6–8 Jun 2023. Defining trigger-action rules via voice: a novel approach for end-user development in the IoT. In: Cagliari, I., Spano, L.D., Schmidt, A., Santoro, C., Stumpf, S. (Eds.), *International Symposium on End User Development*. Springer, Springer, Nature Switzerland, Cham, pp. 65–83. 978-3-031-34433-6. https://doi.org/10.1007/978-3-031-34433-6_5
- Roomkham, S., Terris, S., Sitbon, L., 2022. Multi-modal conversational search for people with intellectual disability: an exploratory study. In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, pp. 6. <https://doi.org/10.1145/3491101.3519821>, Article 304 9781450391566.
- Saffar, A.H., Sitbon, L., Hoogstrate, M., Abbas, A., Roomkham, S., Miller, D., 2024. Human and large language model intent detection in image-based self-expression of people with intellectual disability. In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval (Sheffield, United Kingdom) (CHIIR '24)*. Association for Computing Machinery, New York, NY, USA, pp. 199–208. <https://doi.org/10.1145/3627508.3638347>. 9798400704345.
- Sanders, E.B.-N., Stappers, P.J., 2014. Probes, toolkits and prototypes: three approaches to making in codesigning. *Codesign* 10 (1), 5–14. <https://doi.org/10.1080/15710882.2014.888183>
- Sands, D.J., Doll, B., 1996. Fostering self-determination is a developmental task. *The Journal of Special Education* 30 (1), 58–76. <https://doi.org/10.1177/002246699603000104>
- Sas, C., Neustaedt, C., Apr 2017. Exploring DIY practices of complex home technologies. *ACM Trans. Comput.-Hum. Interact.* 24 2, Article 16 29, 1073–0516. <https://doi.org/10.1145/3057863>
- Savulescu, J., 2009. Autonomy, well-being, disease, and disability. *Philosophy, Psychiatry, & Psychology* 16 (1), 59–65. <https://doi.org/10.1353/ppp.0.0217>
- Schrepp, M., Hinderks, A., Thomaschewski, J., 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *Int. J. Interact. Multimed. Artif. Intell.* 4 (6), 103–108. <https://doi.org/10.9781/ijimai.2017.09.001>
- Scoccia, G.L., 2023. Exploring early adopters' perceptions of ChatGPT as a code generation tool. In: *2023 38th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, pp. 88–93. <https://doi.org/10.1109/ASEW60602.2023.00016>
- Semary, H.E., Al-Karawi, K.A., Abdelwahab, M.M., Elshabrawy, A.M., 2024. A review on internet of things (IoT)-related disabilities and their implications. *Journal of Disability Research* 3 (2), 20240012. <https://doi.org/10.57197/JDR-2024-0012>
- Senaratne, H., Ananthanarayan, S., Ellis, K., 2022. Tronicboards: an accessible electronics toolkit for people with intellectual disabilities. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.
- Sieghart, S., Rohles, B., Bessemans, A., 2024. Empowering independence through design: investigating standard digital design patterns for easy-to-read users. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, pp. 18. <https://doi.org/10.1145/3613904.3641911>, Article 911. 9798400703300.
- Smith, E., Sumner, P., Hedge, C., Powell, G., 2023. Smart-speaker technology and intellectual disabilities: agency and wellbeing. *Disabil. Rehabil. Assist. Technol.* 18 (4), 432–442. <https://doi.org/10.1080/17483107.2020.1864670> PMID: 33378247.
- Standen, P.J., Brown, D.J., Taheri, M.H., et al., 2020. An evaluation of an adaptive learning system based on multimodal affect recognition for learners with intellectual disabilities. *Br. J. Educ. Technol.* 51, 1748–1765. <https://doi.org/10.1111/bjett.13010>
- Tetteroo, D., 2023. Challenges of enabling end-users to develop systems with AI. In: *End-User Development: 9th International Symposium, IS-EUD 2023, (6-8 June 2023)*. Springer-Verlag, Berlin, Heidelberg, Cagliari, Italy, pp. 241–249. https://doi.org/10.1007/978-3-031-34433-6_16. 978-3-031-34432-9.
- Tian, Y.J., Felber, N.A., Pageau, F., Schwab, D.R., Wangmo, T., 2024. Benefits and barriers associated with the use of smart home health technologies in the care of older persons: a systematic review. *BMC Geriatr.* 24 (1), 152.