Challenges in accessing generative AI for users with cognitive disabilities: an exploratory case study

Virginia Francisco[1,2], Raquel Hervás[1,2] and Ricardo García-Mata[3]

[1]Facultad de Informática, Universidad Complutense de Madrid, 28040 Madrid, Spain
[2]Instituto de Tecnología del Conocimiento, Universidad Complutense de Madrid, 28233 Pozuelo de Alarcón, Spain
[3]Computing Services, Research Support, Universidad Complutense de Madrid, 28040 Madrid, Spain

Abstract

Generative Artificial Intelligences (GAIs) have great potential for people with cognitive disabilities, since they can increase their autonomy, offer them personalized learning and break down the barriers that this group traditionally has to access information. To address this gap, this paper investigates the specific difficulties experienced by users with cognitive disabilities when interacting with GAIs.

We present an exploratory case study involving 16 students with cognitive disabilities enrolled in a university-based educational program. ChatGPT was selected as the representative GAI tool due to its popularity and the similarity of its interface and interaction paradigms to those of other widely used Generative AI systems. The intervention was integrated into the students' academic curriculum and involved completing practical academic tasks using ChatGPT. A mixed methodological approach was employed, combining pre- and post-intervention questionnaires completed by students and tutors/experts, along with on-site observations during the sessions.

Results of this exploratory study corroborate the existence of remarkable barriers in basic interaction areas: question formulation (69% of students required support), response comprehension (56% reported understanding answers only occasionally), and information verification (only 19% were able to verify the system responses independently). Suggestions for improvement, from the perspective of the users and their tutor, include more guided interfaces with predefined question templates and contextual suggestions, customizable output formats with language simplification, enhanced spelling correction tools, interactive explanations of complex terms and built-in verification aids.

This study highlights the need to design GAI interfaces with a strong focus on cognitive accessibility. Future development should prioritize multimodal and adaptive interaction models, support tools for comprehension and verification, and interaction designs that promote greater autonomy for users with cognitive disabilities.

Keywords

cognitive disabilities, cognitive accessibility, Generative Artificial Intelligence (GAI), ChatGPT, Human-Computer Interaction (HCI), inclusive design

1. Introduction

The advancement of Large Language Models (LLMs), powered by deep neural networks, has resulted in human-like conversational assistants, including chatbots and voice bots. Leading technology companies, such as OpenAI, Google and Meta, among others, are developing chatbots that can generate text and images in response to user-defined prompts. These are commonly referred to as Generative Artificial Intelligences (GAIs). Their outputs often appear remarkably human, and the use of chatbots has become widespread, with both the general public and professionals utilizing this technology to tackle a wide range of tasks.

However, the benefits of this technological progress may not be reaching all segments of the population equally. For individuals with cognitive impairments, accessing and effectively using GAIs might continue being a significant challenge. Despite their broad capabilities, current GAI systems are often not designed for cognitive accessibility, creating barriers that limit this user group's ability to interact with these tools on equal terms. Yet, GAIs hold enormous potential for people with cognitive disabilities: they can promote autonomy, offer personalized learning, and help break down traditional barriers to

Interacción'25: XXV CONGRESO INTERNACIONAL DE INTERACCIÓN PERSONA-ORDENADOR, September 03–05, 2025, Valladolid, Spain

virginia@fdi.ucm.es (V. Francisco); raquelhb@fdi.ucm.es (R. Hervás); rgarcia@ucm.es (R. García-Mata)
0000-0002-4492-5633 (V. Francisco); 0000-0003-2900-9992 (R. Hervás)

accessing information, communication, and knowledge. To ensure that this group can take advantage of this potential, it is necessary to research and identify the specific difficulties these users face in accessing GAI. In this context, equitable access to emerging technologies must explicitly consider and address the needs of individuals with cognitive impairments.

Generative AI interfaces may often appear as counterintuitive and potentially lack necessary adaptations for cognitive accessibility, usually featuring complex menus, confusing icons, and limited customization options. Additionally, these tools may employ complex and abstract language, potentially making it difficult for people with cognitive impairments to understand and interact effectively with them. These challenges might include difficulties in using interfaces, creating clear prompts, interpreting results, and customizing options. All of this could potentially have a negative impact on the user experience and the quality of the responses generated for this specific group. To date, there is a significant gap in research on the adoption, impact, and cognitive accessibility needs associated with the use of GAI tools by individuals with cognitive disabilities. It is essential to identify functional, linguistic, and design adaptations that enable effective, understandable, and safe use of these tools.

Addressing this gap, this paper poses the following research question: What are the challenges faced by people with cognitive disabilities in accessing and using generative AI? The exploratory case study presented in this paper aims to identify the specific difficulties and barriers that users with cognitive disabilities face when interacting with a representative Generative AI chatbot, seeking to corroborate our initial assumption of unequal access.

It was essential for the purpose of testing our hypothesis and achieving our research goals to have the individuals with cognitive disabilites to interact with a GAI. ChatGPT was selected as the representative tool for this study due to its status as the most widely known text-based generative AI system. Developed by OpenAI and released for public use in 2022 with the launch of version 3.5, it quickly became the fastest-growing application, reaching 100 million users [1]. Subsequent versions (ChatGPT-4, ChatGPT-4o, and ChatGPT-4o-mini) further increased its robustness and efficiency, establishing it as the most widely used generative AI tool in 2023 and 2024 [2]. Furthermore, ChatGPT was also chosen because its interface and interaction paradigms share numerous common elements with other prominent GAIs such as Claude, Mistral or Gemini. These commonalities include reliance on text-based prompts and responses, similar structures for input fields and output displays, and comparable approaches to generating and presenting information. Therefore, the challenges encountered by the participants of our study with ChatGPT are likely to offer valuable insights into the broader accessibility barriers present across various GAIs.

The study presented in this work was conducted with students from the ACCEDE program, a university-specific certificate offered by Universidad Complutense de Madrid (UCM) for individuals with intellectual and developmental disabilities. The activity was intentionally integrated into the students' academic curriculum to offer a tangible benefit to the participants, ensuring that the research provided a meaningful and relevant learning experience for them.

This study pursues the following specific objectives:

- To evaluate the satisfaction and perceived usefulness of ChatGPT by students with cognitive disabilities.
- To identify the main difficulties encountered during the interaction, including asking questions, understanding answers, and verifying information.
- To explore improvements suggested by participants to make ChatGPT, in particular, and all GAIs in general, more accessible.
- To analyze students' perceptions of their future ability to use this tool independently.

The expected outcomes of the study include the collection of relevant information to facilitate more equitable access and more effective use of GAIs, the identification of specific areas for improvement in cognitive accessibility in tools such as ChatGPT, and the generation of useful knowledge to guide future research on the inclusive design of GAIs for people with cognitive disabilities.

The rest of the paper is organized as follows. Section 2 presents the state of the art on generative AI technologies and their use and utility for users with cognitive disabilities. The methodology of this

research is described in Section 3. Then, the obtained results are presented in Section 4, and discussed in Section 5. Finally, we acknowledge the limitations of the study in Section 6 and the conclusions and future work are addressed in Section 7.

2. State of the art

The rapid growth of Generative AI (GAI) technologies has led to increased research on their use and utility for end users [3, 4, 5]. Although Generative AI interfaces could be considered end user friendly, in fact their successful use depends on how to design prompts that get the best result, which is not always an intuitive process. Nielsen [6] recently identified this form of interaction as *intent-based outcome specification* and argued that it is the first new UI interaction paradigm in 60 years. In this paradigm, users specify what they want, often using natural language, but not how it should be produced.

Even though the usability of GAI tools is currently understudied, some research is emerging on the design principles to consider for Generative AI User Experience and their implications in the field of Human-Computer Interaction (HCI) [7, 8]. However, recent works are mostly focused on the interaction with generative AIs from the point of view of different domains, like education [9], chatbot design for non-programmers [10] or art [11], often overlooking the accessibility challenges faced by specific user populations.

One of these overlooked groups is people with cognitive disabilities. These individuals face significant challenges when accessing GAIs, such as cognitively inaccessible interfaces and language not adapted to their needs. Glazko et al. [12] conducted a three-month auto-ethnography study of the use of GAI to meet personal and professional needs in a team of researchers with and without disabilities, documenting experiences independently in a shared document. This documentation detailed their motivations for GAI use, the outcomes (both successful and unsuccessful in addressing access needs), and any issues related to ableism or representation. Findings were then discussed in weekly meetings. Their findings demonstrate a wide variety of potential accessibility-related uses for GAI but also highlight concerns around verifiability, training data, ableism, and false promises.

Recent evaluations using established accessibility standards reveal systematic barriers in current GAI tools. Acosta-Vargas et al. [13] assessed 50 Generative AI applications using both manual and automated (WAVE tool) review, aligned with WCAG 2.2. The study identified many accessibility issues, highlighting the barriers for the adoption of GAIs by people with cognitive disabilities. Similarly, a study assessing 20 GAI applications [14], such as ChatGPT and DALL-E, also employed automated tools and manual review to find that 79% of the evaluated tools failed to meet the WCAG 2.2 “Perceptible” principle, followed by deficiencies in the “Operable”, “Understandable” and “Robust” principles. Common issues included inadequate image descriptions, a lack of semantic structures, and challenges with keyboard navigation.

These findings are consistent with concerns raised by Alshaigy and Grande [15] who, through conceptual analysis informed by professional accessibility insights, point out the inequity in the development of GAI tools, as these tools were predominantly developed by people without disabilities, along with a notable absence of design guidelines specifically tailored for the inclusion of people with disabilities. Despite evidence emphasizing the importance of integrating accessibility early into the design process, this crucial aspect continues to be overlooked in the development of new GAI technologies. Among the significant barriers faced by people with disabilities, these authors point to dynamic interfaces, unlabeled buttons, and inaccuracies in automated transcription services. These authors call for people with disabilities to raise their voices and share their experiences to ensure their needs are prioritized and addressed, thus closing the existing gap and encouraging a more inclusive approach to Generative AI.

Despite these challenges, there is growing research exploring the potential benefits of GAI for people with cognitive disabilities, particularly in educational environments. Liu et al. [16] explored how students from special schools in Hong Kong engage with text-to-image GAI tools in their design processes, gathering data through collected student designs, five-point Likert scale questionnaires, and

informal conversations focusing on student attitudes and future intentions to use GAI for design. Their findings reveal a strong interest in AI learning among students with special education needs, and how incorporating AI into design demonstrates the substantial potential for enhancing students' skills and literacy, positively influencing their future career development and life experiences. Similarly, Mitre and Zeneli [17] conducted a systematic literature review on AI-driven solutions, like assistive technologies, adaptive-learning systems, and generative AI chatbots and virtual assistants, and results show AI-driven solutions' potential to transform the learning process of people with disabilities by creating personalized learning paths, increasing access to educational resources, and supporting real-time communication. However, they also raised ethical concerns about the integration of AI in education, emphasizing the need for the participation of disabled individuals in the development process.

Several studies have examined the real-world use of generative tools by students with disabilities. Pierrés et al. [18] explored the role of GAI, centered on ChatGPT, in higher education for students with disabilities, gathering data through semi-structured interviews with 33 students. This detailed instrument focused on current GAI use, specific ChatGPT experiences, identified opportunities/benefits, limitations/challenges, general concerns, and desired future applications. The study revealed that ChatGPT offers significant opportunities as an assistant in teaching, writing, reading, research, and self-organization. The results suggest that ChatGPT can facilitate written communication for students with communication disabilities, improve reading comprehension for neurodiverse students, and help to establish routines for individuals with TDAH. However, the study also identified limitations and challenges, including accessibility issues, concerns about information accuracy, and the need to develop prompting and critical thinking skills.

Other recent efforts have explored GAI tools specifically designed for or tested with neurodivergent users. TwIPS [19] is a LLM-powered texting application to simplify conversational nuances for autistic users. TwIPS can assist users with deciphering tone and meaning of incoming messages, ensuring the emotional tone of their message is in line with their intent, and coming up with alternate phrasing for messages that could be misconstrued and received negatively by others. TwIPS was evaluated through a user study in which participants took part in semi-structured interviews and completed a follow-up survey. The interviews explored perceptions of the tool's usefulness and gathered suggestions for improvement. The survey consisted of 19 items rated on a 7-point Likert scale, with an additional option available when standard responses did not apply. Participants' audio and screen activity were recorded throughout the study for later analysis. The results of the evaluation highlights the importance of AI interfaces that balance personalization and privacy, proposing adaptive adjustments that respect user autonomy without cognitively overloading them. They also highlight the need to promote critical trust in AI through interfaces that explicitly communicate the uncertainty of its responses and evolve with the user's experience. Mullen et al. [20] present a study that seeks to investigate real-life interactions between people with disabilities and LLM-based chatbots, primarily through interviews, complemented by a 7-point Likert post-survey evaluating perceived utility, understanding, likelihood of future use, and reliability of chatbots, along with open-ended questions on chatbot behavior feedback and concerns. Roomkham and Sitbon [21] present a multimodal and collaborative search systems for people with intellectual disability. They performed an ethnographic study conducted in a collaborative setting with twenty participants across four sessions and follow-up interviews. The results suggest that multimodal and conversational interaction can play a crucial role in social support, peer awareness, and personal interests. Jang et al. [22] investigate the phenomenon of LLM use by autistic adults at work and explore opportunities and risks of LLMs as a source of social communication advice. Data was collected via semi-structured interviews, chatbot interaction logs, and post-interaction surveys with 7-point Likert scales for utility, understanding, future use, and reliability, along with open-ended written responses. Their evaluation shows that participants strongly preferred LLM over confederate interactions. However, a coach specializing in supporting autistic job-seekers raised concerns that the LLM was dispensing questionable advice. This divergence in participant and practitioner attitudes reflects existing schisms in HCI on the relative privileging of end-user wants versus normative good and proposes design considerations for LLMs to center autistic experiences.

In summary, while GAI technologies offer great potential, especially for people with cognitive

disabilities, current design practices often fail to meet their needs as they do not take them into account. The literature highlights both significant opportunities, and significant accessibility gaps resulting from non-inclusive design. Most existing studies focus on specific use cases or individual tools, often without exploring how these tools perform in real contexts involving people with cognitive disabilities. Our study addresses this gap by evaluating its accessibility in authentic tasks, with the direct participation of users with cognitive disabilities. With this, we seek to provide practical design ideas that go beyond compliance with accessibility standards and move toward genuinely inclusive interaction paradigms.

3. Methodology

This research takes the form of an exploratory case study that investigates the access, perceived usefulness, difficulties, and future potential of using ChatGPT by students with cognitive disabilities within the context of a specific educational activity at university. Our main goal is not to assess the effectiveness of ChatGPT as a learning tool, but shed light on the underlying challenges related to cognitive accessibility, HCI, and the comprehension of information generated by GAIs, from the perspective of users with cognitive disabilities. Identifying these difficulties is essential for designing more inclusive GAIs and for developing effective support strategies.

Students of the ACCEDE program were selected for this study. The ACCEDE program[1] is a university-specific certificate offered by the Complutense University of Madrid (Universidad Complutense de Madrid, UCM) for individuals with intellectual and developmental disabilities. Its goal is to train participants to become assistant technicians in the evaluation of inclusive environments, promoting the social and labor inclusion of young people with cognitive disabilities through a training program for employment and university inclusion within the environment of the UCM.

Next sections present a thorough description of the research design, the participants involved in the study, the used data collection instruments and the tools employed for data analysis.

3.1. Research design

To explore the interaction of students with cognitive disabilities with ChatGPT, a mixed methodological approach was adopted, combining quantitative and qualitative data collection and analysis. Data collection was conducted through questionnaires given to students before and after the intervention, as well as parallel questionnaires completed by their tutor at the ACCEDE program and the support experts that participated in the intervention. Figure 1 presents the structure and stages of the study.

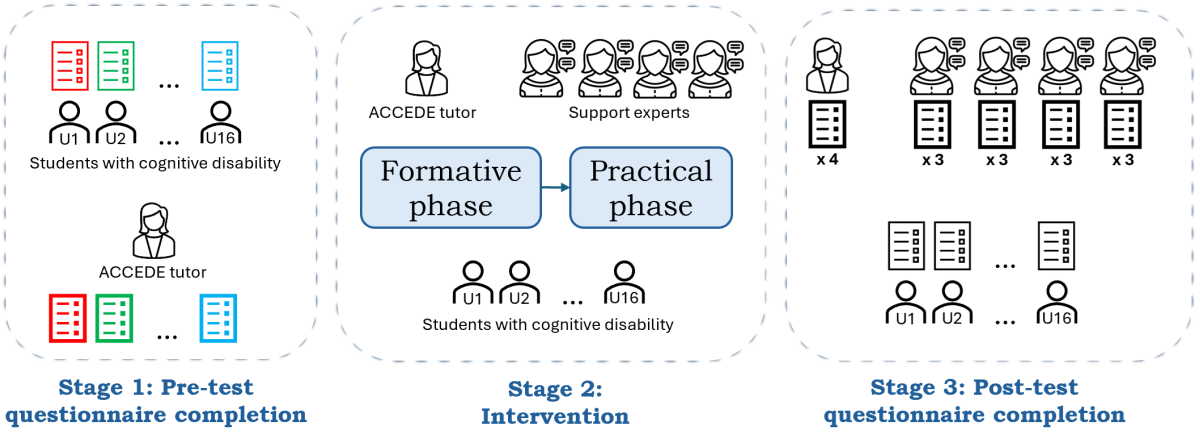

Figure 1: Structure and stages of the study. Different questionnaires were administered to students, tutor and support experts before and after the intervention

[1] https://ucm.es/programa-accede

Stage 1 comprised the completion of pre-test questionnaires by the participants (students with cognitive disability) and their tutor in the ACCEDE program. Each student individually completed their own questionnaire, while the tutor filled out a separate questionnaire for each participant in the study. More details about the questionnaires can be found later in Section 3.3.

Stage 2, which involved the main intervention, was structured into two main phases:

- Formative Phase: The first phase consisted of a training session designed to introduce the fundamental concepts of Artificial Intelligence and Generative Artificial Intelligence (GAI), with a specific focus on ChatGPT. The contents addressed included:
  - Conceptual introduction to AI and its various applications.
  - Explanation of the distinctive characteristics of AIs and their ability to generate content.
  - Presentation of ChatGPT as an example of conversational GAI, explaining its basic functions.
  - Exploration of practical examples of how ChatGPT can be used as a support tool.
  - Emphasis on the importance of being aware of the possible limitations and errors of ChatGPT, promoting a critical attitude towards the information generated.
  - Recommendations for a responsible and safe use of the tool.
  - Basic strategies for the formulation of effective questions and for the verification of the information provided by ChatGPT, emphasizing the importance of being specific in the input questions, asking for clarifications and contrasting the information with other sources.
- Practical Phase: In the second phase, students participated in practical activities in a computer laboratory where they used ChatGPT to perform specific academic tasks. To ensure the tasks' practical relevance and the pedagogical value of the activities, they were selected in collaboration with the students' tutors, prioritizing those they had already worked in class. From various tutor-suggested ideas (e.g., preparing interview questions, creating a university leaflet on cultural activities, researching the university's coat of arms, evaluating building/poster accessibility, classifying expenses,...), three were chosen. Our selection criteria focused on tasks representing distinct AI interaction purposes and varied forms of engagement to ensure a broad exploration of challenges. Feasibility, quick completion, minimal student input, and clear ChatGPT applicability were also key factors (e.g., complex accessibility evaluations were excluded). Each selected task was designed to elicit different cognitive demands and interaction patterns, enabling a comprehensive analysis of potential barriers and facilitators relevant to individuals with cognitive disabilities. The following three academic tasks were finally selected to investigate interaction challenges with ChatGPT:
  - Research about the origin of some elements on the university coat of arms. This activity focused on ChatGPT's ability to provide factual information and required the formulation of specific search questions. Students formulated direct questions (e.g., "Why is the swan the symbol of the Complutense University of Madrid?"), read and attempted to comprehend the information provided, identifying key facts, and finally, put their findings in common in a group discussion.
  - Classifying different expenses into needs and wants. This task aimed to assess how students could use ChatGPT for conceptual reasoning and classification tasks, requiring understanding definitions and applying diverse criteria. Students first asked ChatGPT to define "need-based expense" vs. "want-based expense" (e.g., "What is the difference between a need and a want regarding expenses?"). Subsequently, they classified a provided list of expenses (e.g., Trousers, Education, Soft drinks, Holidays, Housing, etc.) individually without ChatGPT, then prompted ChatGPT to classify the same list, and finally compared their personal classifications with ChatGPT's, reflecting on any discrepancies.
  - Preparation of interview questions for the Vice-Dean of Quality of their school. This task focused on the use of ChatGPT as a tool for planning and generating content for a formal academic interaction, involving the identification of relevant topics and the formulation of relevant and well-structured questions. Students informed ChatGPT of the interviewee's

role and interview purpose, asked it to generate relevant questions, and then reviewed and selected the most appropriate and professional ones. Finally, students put their questions in common in a group discussion.

During the practical phase, the students were assisted by their ACCEDE tutor and four intervention support experts. In order to focus their attention on specific participants, each of the four experts attended three specific students, and the ACCEDE tutor took care of the remaining four.

After the intervention, Stage 3 consisted on the completion of post-test questionnaires by all the actors involved. Each participant filled out a personal questionnaire, while each support expert and the tutor completed one questionnaire for each of the 3-4 students they had assisted during the intervention. More details about these materials are included in Section 3.3.

3.2. Participants

Sixteen students (n=16) with cognitive disabilities from the ACCEDE program participated in the study. Participants were between 18 and 27 years old (M = 21, SD = 2.50), with an equal gender distribution (50% women, 50% men). Only one student reported having an additional motor impairment (6.25%), and another student reported a visual impairment (6.25%).

Regarding the level of cognitive disability, 75% of the students reported having a mild level, while 18.75% had a borderline level, and 6.25% had a moderate level. All participants have difficulties with logical reasoning, 56.25% have alternating attention difficulties and 25% have cognitive flexibility difficulties. Less frequent difficulties included selective attention (18.75%), inhibition (18.75%), and problems with working memory (18.75%), short-term memory (12.5%), written expression (12.5%), sustained attention (6.25%), and long-term memory (6.25%).

In terms of technological autonomy, the majority of participants (81.25%) reported being very autonomous or needing little help when using technology. All participants reported using technology on a regular basis, with 87.5% using it daily and only 12.50% using it occasionally or rarely. Half of the students (50%) felt very comfortable using new applications, 43.75% felt somewhat comfortable, and only one participant (6.25%) reported low comfort levels.

3.3. Data collection instruments

The review of relevant studies shown in Section 2 provided an overview of various data collection instruments such as semi-structured interviews, Likert-scale surveys, qualitative data collection, and observational methods. However, there remains limited availability of standardized or directly applicable instruments specifically designed to evaluate the nuanced interaction challenges posed by Generative AI technologies like ChatGPT, particularly within our academic context and for this specific population. Existing measures often focus on general usability, different types of AI, or populations without cognitive disabilities, and many are designed for a broader scope than our focused inquiry. Therefore, to align with the exploratory nature and specific objectives of this case study, new questionnaires were specifically developed. These custom-designed instruments allowed us to directly assess the specific interaction aspects most relevant to our research questions and target population in a contextually appropriate manner.

To achieve this comprehensive assessment, data was collected primarily through these newly developed questionnaires, administered to students and their ACCEDE tutor before and after the intervention, and to support experts after the intervention. The questionnaires had closed-ended questions (with dichotomous categorical answers -Yes/No- and 3- and 4-category ordinal answers) as well as open-ended questions. All questionnaires were administered in Spanish, the native language of both the students with cognitive disabilities and their ACCEDE tutors/support experts.

The inclusion of the ACCEDE tutor as informant was due to her in-depth knowledge of the students' individual abilities, difficulties and learning styles. In addition, the responses of the tutor and support experts to the post-test questionnaires were based not only on their prior knowledge of the students, but also on direct observations made during the intervention. This on-site observation of students'

interaction with ChatGPT offered a privileged perspective and allowed cross-validation of students' self-reported data, helping to identify possible discrepancies between self-perception and behavioral manifestation during GAI use, especially in areas where students might have difficulties to self-assess accurately their skills or understanding (e.g., difficulty admitting to not understanding something or specifying the cognitive areas they have affected).

The researchers and the ACCEDE tutor collaborated actively in the development of the data collection instruments. Based on her direct experience with students in the ACCEDE program, the tutor provided valuable input regarding the language and forms of expression most accessible to the participants. The joint review made it possible to identify and modify possible ambiguities or terms that could be confusing for the students, ensuring that the questions were interpreted in the manner intended by the researchers. The collaboration also ensured that the Likert range used for each question was appropriate for the participants' ability to discriminate and that the response options were mutually exclusive and exhaustive.

Prior to the intervention, pre-test questionnaires, which can be consulted in Table 1[2], were administered to both students and their tutor:

- The student questionnaire collected information about their general familiarity with technology, comfort using new applications, confidence in searching for information online, prior knowledge on Artificial Intelligence and specifically about ChatGPT, etc.
- The tutor questionnaire collected student demographics (age, gender, type and level of disability, main areas of cognitive difficulty), their level of autonomy with technology, and any other considerations relevant to the student's participation in the study. The data of participants described in Section 3.2 were obtained from the pre-test questionnaire filled out by her tutor.

Table 1
Pre-test questionnaries for students and tutor

PRE-TEST QUESTIONNAIRE FOR STUDENTS	
PRE-STU_1	Do you like to ask for help or do you prefer to do things on your own?
PRE-STU_2	Do you usually ask questions if something isn't clear?
PRE-STU_3	What do you do when you don't understand the answer to your question?
PRE-STU_4	How often do you use technology (computer, phone, tablet)?
PRE-STU_5	Are you comfortable using new apps or programs?
PRE-STU_6	What technology do you use the most?
PRE-STU_7	Is it easy for you to search for information on the internet?
PRE-STU_8	Do you always trust the information you find on the internet?
PRE-STU_9	Who would you ask for help if you need to know if something on the internet is true?
PRE-STU_10	Have you heard of Artificial Intelligence? Could you describe it in your own words?
PRE-STU_11	Do you know any AI tools? Which ones? Have you ever used them? What have you used them for?
PRE-STU_12	Do you know ChatGPT? What is ChatGPT and what is it used for?
PRE-STU_13	What do you hope to learn from this activity?
PRE-STU_14	Is there anything specific that worries you or seems difficult about this activity?
PRE-STU_15	Do you think technology can help you complete college assignments?
PRE-TEST QUESTIONNAIRE FOR TUTOR	
PRE-EXP_1	Age
PRE-EXP_2	Sex
PRE-EXP_3	Type of disability
PRE-EXP_4	Level of disability
PRE-EXP_5	Main areas of cognitive difficulty
PRE-EXP_6	Level of autonomy with technology
PRE-EXP_7	Other issues that should be taken into account

[2]It is important to note that the questionnaires presented are English translations of the original Spanish instruments used during the study.

During the practical phase of the intervention, the tutor and four support experts observed the students' interaction with ChatGPT and took note of everything that happened. In order to focus their attention on specific participants, each of the four experts attended three specific students, and the ACCEDE tutor took care of the remaining four. Once the intervention was completed, post-test questionnaires were administered to both the students and the tutor and support experts. These questionnaires explored participants' experiences with ChatGPT, including difficulties encountered in asking questions and understanding answers, their ability to verify information, their perception of the usefulness of the tool, their willingness to use it in the future, and improvements they would suggest to make it more accessible. The questionnaire for the tutor/experts also included questions about their perception of the students' learning and autonomy during the activity, as well as general recommendations for improving the accessibility of ChatGPT for students with cognitive disabilities. Post-test questionnaires can be consulted in Table 2.

Table 2
Post-test questionnaries for students, tutor and support experts

POST-TEST QUESTIONNAIRE FOR STUDENTS	
POST-STU_1	Did you enjoy using ChatGPT?
POST-STU_2	Did you find ChatGPT useful for the activities?
POST-STU_3	What did you like the most about ChatGPT?
POST-STU_4	What did you like the least about ChatGPT?
POST-STU_5	Did you have any difficulties when asking ChatGPT questions? What kind of difficulties?
POST-STU_6	Were you able to check if the information was correct? How did you try to verify it?
POST-STU_7	Did you understand the answers ChatGPT gave you?
POST-STU_8	Were you able to use the information ChatGPT gave you to complete the tasks?
POST-STU_9	Was there anything in the answers that you didn't understand? How did you solve it?
POST-STU_10	Do you feel you could use ChatGPT for schoolwork on your own in the future?
POST-STU_11	Do you think ChatGPT can give you useful answers?
POST-STU_12	Would you like to keep using ChatGPT to learn?
POST-STU_13	Do you feel more confident using technology after this activity?
POST-STU_14	Do you think ChatGPT needs improvements to be easier to understand? If yes, what would you suggest?
POST-TEST QUESTIONNAIRE FOR TUTOR/EXPERTS	
POST-EXP_1	Did the student need a lot of help to formulate appropriate questions?
POST-EXP_2	Did the student understand ChatGPT's answers?
POST-EXP_3	Was the student able to follow the steps to verify the information?
POST-EXP_4	Did the student show frustration or lack of motivation at any point?
POST-EXP_5	Did the student seem to enjoy the activity?
POST-EXP_6	Did the student seem to be learning during the activity?
POST-EXP_7	Did you observe any improvements in the student's autonomy when using ChatGPT during the activity?
POST-EXP_8	What kind of difficulties did the student have?
POST-EXP_9	Do you think ChatGPT would be useful in future activities for the student?
POST-EXP_10	What changes would you recommend to make ChatGPT more accessible?

All questionnaires were completed in physical format and later digitized and anonymized for subsequent analysis.

3.4. Data analysis

The quantitative data were analyzed using descriptive and inferential statistics with SAS 9.4 statistical software. Variables were summarized using absolute frequencies and percentages. To examine associations between variables, Pearson's Chi-square test was used when the expected cell counts were sufficient, and Fisher's Exact Test was applied when Chi-square assumptions were not met. For compar-

isons involving ordinal variables, the Wilcoxon rank-sum test was employed. In addition, Spearman's rank correlation coefficient was used to assess the strength and direction of monotonic relationships between ordinal variables that did not meet normality assumptions.

The qualitative data, derived from open-ended responses, were analyzed using thematic analysis, following the procedures outlined by Green and Thorogood [23] and Braun and Clarke [24]. A classical content analysis approach was used to identify, analyze, and report patterns within the data [25]. First, each sentence or meaningful segment was assigned one or more codes that summarized its core idea. These codes were then grouped into categories based on semantic similarity. From these categories, key descriptive themes emerged. Representative textual quotations were selected to illustrate each theme, and all quotes were translated into English for reporting.

4. Results

This section presents the results of the study, structured around four main themes that emerged from the data analysis. These themes reflect main aspects of the participants' experience before, during, and after their interaction with ChatGPT. First, we describe the technological profile of participants and their prior expectations and beliefs about Artificial Intelligence (Section 4.1). We then analyze the difficulties encountered during the interaction with ChatGPT, focusing on question formulation, comprehension of responses, and information verification (Section 4.2). Next, we explore students' attitudes, emotional experiences, and perceptions of usefulness and learning after using the tool (Section 4.3). Subsequently, we summarize improvement suggestions provided by both students and tutor/experts to enhance the accessibility and effectiveness of ChatGPT (Section 4.4).

4.1. Technological profile and expectations for AI

The results presented in this subsection are based on the pre-test questionnaires completed by the students. Regarding their help-seeking behaviors, a large majority of participants (87.5%) stated that they ask for help (to other people or computer tools) when something is unclear at least occasionally.

Concerning information literacy, half of the participants reported that they always find it easy to search for information on the internet, while 43.75% indicated that they sometimes do. Only a small percentage (6.25%) reported difficulty with this task. However, a considerable percentage (43.75%) admitted to trusting the information they find online without verifying it.

Regarding their familiarity with Artificial Intelligence (AI), 62.5% of the students had heard of AI, although only 37.5% were able to name a specific AI tool. Among this latter group, only three students (23.1% of those responding affirmatively) had actually used an AI tool before. Specific knowledge of ChatGPT was even more limited: 37.5% of the group knew about it, and another 18.75% had heard of it but did not know exactly what it was. Notably, no students with moderate or borderline intellectual disabilities were familiar with ChatGPT, suggesting a potential relationship between the level of disability and the degree of familiarity with this type of AI tool (*Fisher's exact test,* $p = 0.0346$).

Finally, the majority of participants expressed a positive belief in the potential of technology to support their academic tasks (75%), with 18.75% expressing uncertainty and only one student reporting a negative view.

4.2. Interaction with ChatGPT

The analysis of students' interaction with ChatGPT is structured around three dimensions that reflect the main stages of a typical user experience with a generative AI tool: asking questions, understanding the answers, and verifying the accuracy of the information provided. These findings are based on the analysis of the post-test questionnaires filled out by the students, as well as the tutor and support experts involved in the intervention.

4.2.1. Asking questions

Regarding the students' experience in asking questions to ChatGPT, the results indicate a low perceived level of difficulty, with only 19% of participants reporting having encountered challenges when formulating questions. The tutor/experts' perspective offered a more nuanced view regarding the need for help (students could only answer Yes or No, while the tutor/experts had four response categories): 19% required a lot of help, 25% moderate help, another 25% little help, and 31.25% did not need any help.

A significant relationship was found between difficulties in writing and the need for help formulating questions (*Fisher's exact test,* $p = 0.0250$). Specifically, 100% of the students with writing difficulties were rated as needing a lot of help, while only 7.1% of students without such difficulties received that same rating. Among students without writing difficulties, the remaining distribution was: 28.6% moderate help, 28.6% little help, and 35.7% no help. These differences were also confirmed by a Wilcoxon rank-sum test ($W = 4.00$, $p = 0.0401$).

An analysis of the open-ended responses provided by students and the tutor/experts led to the identification of several main categories of difficulties that students experienced when asking questions to ChatGPT:

- Difficulties related to the question's conceptualization and clarity. It was observed that some participants struggled to formulate questions that were relevant to the context of the task or required a specific focus. One student exemplified this by commenting: *"I phrased the expense question poorly because I asked about unnecessary expenses instead of necessary ones"*. Similarly, the lack of clarity manifested itself in the formulation of broad and poorly defined questions, as evidenced by the following tutor's observation: *"She asks about the vice dean in general and not which specific vice dean she wants to interview"*.
- Difficulties in question generation and articulation: In addition to conceptualization, some participants faced challenges when it came to actually generating the questions needed for the task. Some participants expressed uncertainty about how to start asking the questions they needed to complete the activity, as one student noted: *"I didn't know how to start asking the questions I needed for the activity"*. Furthermore, difficulty identifying necessary information and how to request it to ChatGPT was a common obstacle, as one expert noted: *"She struggled with formulating questions (she didn't know how to ask the chat what she needed, and we had to help her)"*.
- Linguistic formulation difficulties: The presence of spelling errors, as noted in the observation *"I made many spelling mistakes when asking questions"* could have affected the GAI's ability to understand the user's intent and provide relevant answers.
- Difficulties when requesting refinement and clarification: Some difficulties related to formulating questions aimed at refining the information provided by ChatGPT or requesting clarification were identified. Comments such as *"I wasn't able to ask it to summarize or simplify"* and *"It's difficult for them to ask for simplifications or to express that there are things they don't understand"* illustrate a lack of awareness or skill in using questions as a tool to deepen understanding or adapt information to their needs.

4.2.2. Understanding answers

Regarding the understanding of the answers generated by ChatGPT, students' perceptions varied: 43.75% stated they always understood the answers and 56.25% indicated that they understood them only sometimes. Going deeper into this issue, half of the students reported having understood all of the answers, while the other half reported finding some aspects incomprehensible. From the tutor/experts' perspective, answer comprehension was evaluated more positively: 75% of the students understood the answers, 18.75% had difficulty understanding them, and only 6.25% did not understand them.

Exploring the potential influence of specific areas of cognitive difficulty on understanding ChatGPT answers, the results suggest a statistically significant relationship between cognitive flexibility difficulties and the tutor/experts' assessment of answer comprehension (*Fisher's exact test,* $p = 0.0198$). A Wilcoxon

rank-sum test also showed that students with cognitive flexibility difficulties tended to experience more difficulties processing ChatGPT responses (W = 50.50, p = 0.0104). Descriptively, 91.67% of students without difficulties in cognitive flexibility were rated as understanding the responses well. In contrast, among the students with such difficulties, only 25% was rated as understanding well, while 50% were rated as understanding "with difficulty", and 25% as not understanding at all. These findings suggest that the ability to adapt to the information provided by ChatGPT, which often requires some flexibility in processing, is compromised in students with this cognitive difficulty. Similarly, a significant relationship was identified between difficulties in written expression and tutor/experts' assessment of answer comprehension (*Fisher's exact test,* p = 0.0500). The Wilcoxon rank-sum test supported this association, indicating that students with written expression difficulties required more effort to understand the answers (W = 2.20, p = 0.0277). In fact, all students with difficulties in written expression were rated as understanding the answers with difficulty, whereas among the students without such difficulties, 85.7% were assessed as understanding well, 7.1% as understanding with difficulty, and 7.1% as not understanding.

Analyzing the responses to the open-ended questions of tutor/experts, several main categories of difficulties that students experienced when understanding answers were established to summarize the qualitative results:

- Difficulties related to format and amount of information: Participants experienced difficulty when the response consisted of long, unstructured blocks of text. This was evident in comments such as: *"He found it difficult to make progress when ChatGPT gave him all the text at once without bullet points or when it returned a lot of text"*.
- Difficulty requesting information adaptation: Another significant difficulty was students' inability or unwillingness to ask ChatGPT to adapt the information to their needs. This was reflected in statements such as *"He couldn't ask it to summarize or simplify it"*.
- Difficulties understanding specific terms: Some participants encountered words or phrases within the responses that were unfamiliar to them or whose meaning they did not fully understand. This is evidenced by the following comment from this participant: *"There were some words in ChatGPT's responses that he did not understand"*.

Analyzing the responses to the open-ended question about the strategies to try to solve the difficulties faced understanding the answers, several main categories of strategies were established:

- Seeking external (human) help: A common strategy used by participants was to seek help from tutor/experts. This is evidenced by direct comments such as *"Ask the teacher"* and *"To understand the answers, he needed a lot of help from me"*.
- Self-correction and question refinement strategies: In some cases, participants attempted to obtain more understandable answers by detailing or specifying their initial question, as described in *"Further detailing and making the question more specific"*.
- Requesting adaptation and clarification from ChatGPT: A strategy directly related to interacting with the AI was asking it to adapt or clarify the information. Participants attempted to overcome their lack of understanding by asking for summaries, as described in *"Asking for a summary"*. Furthermore, when faced with unfamiliar terms, some chose to directly ask ChatGPT for their meaning or explanation, as illustrated by comments such as *"Asking ChatGPT for words I didn't understand in the answer (e.g., etymology)"*.

4.2.3. Verifying the accuracy of the information provided

A critical aspect of interacting with GAI tools is verifying the accuracy of the information provided. In this study, students' perceptions of their ability to verify ChatGPT information was mostly negative, with 81.25% reporting that they were unable to perform this verification. From the tutor/experts' perspective, a similar situation was observed, although with more variability: 53.33% of students were unable to verify the information, 26.67% did so with assistance, and only 20% were able to verify it

independently. Regarding the methods used to attempt to verify information, a small percentage of students mentioned asking ChatGPT directly about the validity of their answer (6.25%), while another minority group appealed to internet searches (12.5%).

When analyzing the relationship between verification ability and specific areas of cognitive difficulty, several statistically significant relationships were identified. A significant relationship was found between the ability to follow steps to verify information and difficulties with alternating attention (*Fisher's exact test,* $p = 0.0174$). A Wilcoxon rank-sum test further revealed that students with alternating attention difficulties exhibited a greater ability to verify information compared to those without these difficulties ($W = 69.00$, $p = 0.0078$). Specifically, 33.33% students without alternating attention difficulties were able to verify the information on their own, 44.44% needed help, and 22.22% were unable to verify the information. In contrast, 100% students with alternating attention difficulties were able to verify the information on their own. Similarly, a significant relationship was found between verification ability and short-term memory difficulties (*Fisher's exact test,* $p = 0.0286$). The Wilcoxon rank-sum test revealed that none of the students with short-term memory difficulties had trouble verifying information, as all of them were able to verify it independently ($W = 4.00$, $p = 0.0315$). In contrast, the majority of students without these difficulties were unable to verify information (61.54%). Finally, a significant relationship was found between verification ability and working memory difficulties (*Fisher's exact test,* $p = 0.0374$): none of the students with working memory difficulties had difficulty verifying information, and they were able to do so alone or with help, while the majority of students without these difficulties (61.54%) were unable to do so.

4.3. Perceptions and attitudes toward ChatGPT

As in the previous subsection, the findings to be shown in this subsection are based on the analysis of the post-test questionnaires completed by the students, as well as by the tutor and support experts involved in the intervention.

The majority of students expressed a positive attitude toward their future use of ChatGPT: 68.70% expressed a desire to continue using ChatGPT for learning, while 18.25% were unsure about it and 12.5% indicated they would not like to continue using it. Regarding perceived autonomy in future use, opinions were more divided: 37.5% were optimistic and believed they could use the tool without help, while a slightly larger proportion, 43.75%, anticipated the need for some assistance and 18.75% expressed a less autonomous view, believing they would require constant help to use ChatGPT. When analyzing the relationship of these results with specific areas of cognitive difficulty, a significant relationship was found between short-term memory difficulties and the tutor/experts' assessment of whether the student had difficulties using ChatGPT (*Fisher's exact test,* $p = 0.0500$). Tutor/experts mostly rated students with short-term memory difficulties as not experiencing difficulties when interacting with ChatGPT, with 100% rated as not having difficulty. In contrast, 85.71% of students without short-term memory difficulties were rated by tutor/experts as having difficulties interacting with ChatGPT.

Overall, most students perceived ChatGPT as a useful tool for the activities they carried out: 56.25% considered it very useful and 43.75% considered it somewhat useful. Regarding the usefulness of the answers provided by ChatGPT for completing academic activities, the majority of students indicated that the experience was positive: 43.75% indicated that the answers were always useful, while 37.50% indicated that they were useful in almost all situations. Only 18.75% indicated that the answers were only useful in some situations.

According to the tutor/experts' observations, most students seemed to experience learning during the activity: 66.67% of students learned during the session, 20% experienced moderate learning, and 13.33% did not seem to learn at all. It is important to note that a significant relationship was found between the students' disability level and the tutor/experts' perception of learning (*Fisher's exact test,* $p = 0.0220$). Tutor/experts rated 83.33% with a mild disability level as having learned during the session, while none with moderate or borderline disabilities were perceived as having learned. In addition, a Wilcoxon rank-sum test revealed that students with mild disabilities were perceived as those who learned more during the activity compared to students with borderline or moderate disabilities ($W =$

38.50, $p = 0.0155$).

Regarding the emotional experience during the session, most of the students reported positive feelings: 75% showed no signs of frustration or demotivation at any point, while 12.5% experienced these feelings occasionally and another 12.5% consistently. Regarding enjoyment, 62.50% of students seemed to enjoy the session, 18.75% did so moderately, and another 18.75% did not seem to enjoy it at all. It is significant to note that a statistically significant association was found between enjoyment of the activity and the desire to continue using ChatGPT for learning (*Fisher's exact test,* $p = 0.0357$; *Spearman's rank correlation,* $r_s = 0.532$, $p = 0.0339$), suggesting that a more pleasant experience with the tool is directly related to a greater willingness to use it in the future for learning purposes. 90% of the students who reported enjoying the activity expressed a desire to continue using ChatGPT for learning, while only 33.33% of the students who reported moderate enjoyment or not enjoyment shared this desire.

Regarding the question about the most appreciated aspects of ChatGPT posed by students, several main categories were established to summarize the qualitative results:

- Perceived usefulness and benefits of the answers: Comments such as *"The useful answers it gave me"* underline the practical value they found in the AI's responses. ChatGPT's ability to offer concrete examples, as mentioned in *"How it gave me examples of expenses based on need and expenses based on desire"* was also appreciated for facilitating understanding. Furthermore, the usefulness of the information for learning and completing academic tasks was highlighted in statements such as *"Learning things with it"*. The AI's ability to provide information and help to find answers was also a positive aspect, as indicated by *"If you don't know the answer to something, you can ask ChatGPT"*.
- Responsiveness and functionality: ChatGPT's ability to answer a variety of questions was highly valued. This includes the ability to answer conceptual questions, as exemplified by *"Asking what a need expense and a want expense mean"* as well as the general function of answering questions, as indicated by *"Being able to ask questions"*.
- Positive User Experience: The speed with which ChatGPT provided responses was a notable factor in the positive user experience, as mentioned in *"How quickly it answers things"*.

On the other hand, after analyzing the aspects that students liked least about ChatGPT, we obtain the following categories that summarize the qualitative responses given:

- Problems with response quality and format: Several participants expressed frustration with the way the information was presented. Comments such as *"the information was very long and I had to read a lot"* reflected difficulty processing lengthy responses. The *"lack of concise formatting"* and the fact that the responses were not "outlined" were also noted as negative aspects, with a clear preference for brevity and outlines to facilitate understanding.
- Questioning the reliability and accuracy of information: A significant concern among participants was the reliability of the information provided by ChatGPT. The *"generation of incorrect information"* was mentioned multiple times, generating a *"distrust of the information"*.
- Interaction and functionality issues: Some participants experienced difficulties interacting with the AI. It was mentioned that *"sometimes it doesn't understand the questions"*.

Regarding the perceived impact on students' confidence in using the technology, a vast majority of the participants (81.25%) reported feeling more confident using the technology after the session, 6.25% were undecided about the decision, and 12.5% indicated they did not feel more confident after the experience.

4.4. Suggestions for improvement

Participants also offered valuable suggestions for improving ChatGPT's functionality and user experience in post-test questionnaires. These suggestions were grouped into three main categories:

- Improvements to the quality and timeliness of information: Participants expressed a desire for the tool to be *"more up-to-date and to provide recent information correctly"*. They also emphasized the need for *"greater accuracy in responses"* and to *"avoid errors in information"* even suggesting that incorrect responses should be eliminated altogether to increase reliability.
- Improvements to the clarity and format of responses: Several participants indicated a preference for *"shorter, less text-heavy responses"* advocating for *"more concise responses"* and *"less text-heavy responses"*. The idea of *"making responses simpler from the outset without having to be prompted"* was also a recurring point to facilitate understanding. Additionally, it was suggested that *"quick access to adaptation functions (simplify, summarize)"* must be provided.
- Improvements to user interaction and assistance: The ability to *"correct questions if they are incorrect so they are answered correctly"* was a key suggestion. Also mentioned was the need to *"help me make questions more specific and detect general questions and ask for them to be more specific"*. Furthermore, the implementation of a *"guide to provide initial information"* and *"instructions to contextualize ChatGPT"* from the outset was suggested, making it easier to provide user characteristics to obtain more personalized and relevant answers. Finally, the inclusion of a *"list of predefined help commands"* and *"useful command suggestions"* could empower users to more easily tailor answers to their needs.

5. Discussion

Overall, the results highlight the central role that specific cognitive difficulties play in interacting with AI-based tools, affecting both question formulation and the understanding and verification of answers. These findings allow us to identify key areas for improving the accessibility of systems like ChatGPT.

The results of the presented exploratory study support our initial hypothesis that people with cognitive disabilities face significant difficulties when interacting with ChatGPT, a general-purpose GAI. While the tool demonstrated potential and generated positive feedback in terms of perceived usefulness and enjoyment (81% of participants liked using ChatGPT and 100% found it useful), the difficulties encountered in formulating questions, understanding answers, and verifying information highlights the need to redesign GAI interfaces for this group.

The difficulty in asking questions, as evidenced both by students' self-perceptions (19% of students reported needing assistance formulating questions) and by tutor/experts' observations (69% of students required some assistance), highlights the need for adapted interfaces. Incorporating predefined question templates for common tasks (e.g. defining a concept, explaining how to do something, etc.) or providing contextual suggestions could help users express their needs more effectively, especially for those with writing difficulties, who were the ones most frequently needing support formulating questions. Voice input, a feature suggested by students themselves, appears to be a valuable alternative for those with typing difficulties. Additionally, improved spelling error detection and correction mechanisms, beyond basic fixes to offer meaningful rephrasings, could mitigate problems caused by spelling issues, which tutor/experts identified as a notable obstacle during participant observation.

Regarding answer understanding, the fact that a significant percentage of students (56%) reported only occasionally understanding the responses suggests a mismatch between the standard way ChatGPT presents information and the specific presentation needs of this user group. To address this, the interface should offer customizable output formatting options such as concise responses or summary layout, which many students preferred. Incorporating language simplification features (as recommended by the ACCEDE tutor) and enabling interactive explanations of complex terms or concepts directly within the interface could significantly improve information accessibility. Additionally, tutor/experts observed that students find it difficult to admit that they do not understand something, making them less likely to ask ChatGPT for clarification, highlighting the need for more proactive and supportive interaction design.

The most critical finding of this experiment is the majority's inability to verify information (only 19% were able to verify the information provided by ChatGPT for the activities), suggesting a limited

understanding of effective verification strategies by most participants. Future GAI interfaces should integrate source reliability indicators in a visual and understandable manner. Furthermore, direct links to sources should be provided and predefined prompts for verifying information should be included, guiding users through the information verification process. The interface could also explicitly warn about the probabilistic nature of some answers, encouraging a more cautious attitude toward the information generated.

Suggestions for improvement from students and tutor/experts highlight the need for more intuitive and adaptable interfaces. Incorporating interactive tutorials and contextual assistance directly within the interface could reduce reliance on external help and encourage greater user autonomy (particularly relevant given that only 37.5% of students currently feel confident using ChatGPT independently). The observed increase in confidence in using the technology after the session indicates a positive short-term impact of the intervention, although further research is needed to determine the long-term sustainability of this effect. Interestingly, preliminary findings suggest a potential association between certain cognitive difficulties (alternating attention, short-term memory and working memory) and a greater capacity for verifying information. This unexpected result requires deeper investigation to better understand the mechanisms behind this relationship.

Finally, the observation that no students with moderate or limited disabilities were familiar with ChatGPT suggests a potential digital gap and a need to facilitate access and familiarization with these technologies for students with greater support needs.

These results are consistent with findings from previous work. Like Acosta-Vargas et al. [13, 14], we identified systematic accessibility barriers in GAI tools, especially related to understanding long, unstructured responses. In line with Pierrés et al. [18] our participants highlighted the difficulties faced by inexperienced AI users in formulating effective questions, a difficulty that in our case was exacerbated by written expression problems. Furthermore, our results reinforce the concerns about the veracity of information expressed by Glazko et al. [12], Alshaigy and Grande [15], Pierrés et al. [18] and Haroon & Doga [19] , showing a limited capacity for autonomous verification by students. We also observed, like Liu et al. [16], a generally positive attitude towards the use of GAI, although with important differences in the perception of usefulness and autonomy according to the cognitive profile of the user. In addition, our conclusions support the need for personalized interaction mechanisms, as proposed by Haroon & Dogar [19], especially those that provide adaptive guidance without overwhelming users with excessive customization complexity. However, unlike most previous studies, our work offers first-hand observational and self-reported data from users with cognitive disabilities performing real academic tasks with ChatGPT, providing grounded insights into the specific challenges and support needs of this underrepresented population.

6. Threats to validity

This study has certain threats to the validity of our findings that should be considered when interpreting the results:

- A significant threat to the external validity stems from the small sample size, which consisted of only 16 participants. Furthermore, the homogeneity of the sample (all students from a specific university-based educational program for individuals with cognitive disabilities) inherently limits the generalization of our findings. While providing rich, in-depth insights for this exploratory case study, these factors mean that the observed patterns of interaction, difficulties, and perceptions regarding ChatGPT's utility may not be directly extrapolate to broader populations of users with diverse cognitive disabilities, or to different educational and social contexts. The limited sample size also constrained the statistical power of certain analyses, impacting the robustness of inferential findings.
- Threats to statistical conclusion validity arose from the analysis of categorical data. Specifically, in some analyses, 75% of cells had expected frequencies less than 5. This condition can compromise the validity of Chi-square test. To address this limitation, Fisher's exact test was used, although

its sensitivity is also limited by sample size, potentially affecting the precision of the statistical inferences drawn from these specific analyses.

- A potential threat to validity relates to the reliance on self-perception data obtained from student-completed questionnaires. Such questionnaires could introduce bias in the perception and reporting of difficulties. Although this limitation is generally attenuated by the parallel questionnaires completed by the tutor/experts, and high overall concordance was observed between their perceptions and those of the students (thereby enhancing the reliability of our findings), instances of discordance highlight a potential divergence in perspectives. This divergence suggests that 'difficulty' or 'understanding' can be perceived differently by the user and the observer, which impacts the precise interpretation of these constructs and the full scope of faced barriers.

Addressing these threats to validity will be crucial for future research. Expanding the sample size and incorporating a broader range of objective measures to complement users' perceptions of their own difficulties will be essential to enhance the generalization and provide more conclusive evidence regarding Generative AI accessibility for individuals with cognitive disabilities.

7. Conclusions and future work

This exploratory study has revealed the significant difficulties faced by people with cognitive disabilities when interacting with ChatGPT, despite the perceived potential and generally positive attitude toward the tool. Although the presented study focused on ChatGPT, the difficulties identified in asking questions, understanding answers, and verifying information can be extrapolated to other GAIs that employ text-based conversational interfaces such as Gemini, DeepSeek, Claude, or Mistral. Considerations regarding the need for guided interfaces, customizable formatting options, language simplification, and verification support tools are applicable to the inclusive design of any conversational generative AI system intended for users with cognitive disabilities.

These results highlight the need for a design approach focused on cognitive accessibility for GAIs. Future research and development of these technologies should emphasize the creation of interfaces that not only enable interaction but also facilitate understanding, critical evaluation, and autonomy for users with cognitive disabilities. The exploration of multimodal interfaces, advanced customization of information output, and the integration of verification support tools are presented as promising solutions for achieving more inclusive human-computer interaction in the field of Generative AI.

Future work should focus on developing and implementing more guided interfaces, incorporating predefined question templates and contextual suggestions to facilitate query formulation. Implementing voice input and improving spelling error detection and correction with understandable rephrasing are also key areas for development. To improve response comprehension, it is recommended to explore customizable output formatting options, integrate language simplification features, and request interactive explanations. In addition, interfaces could integrate proactive comprehension support mechanisms such as:

- Automatic detection of potential comprehension difficulties: Analyze user interaction (e.g., repeated questions, rephrasing of the same question, prolonged periods of inactivity after a complex answer) to infer potential comprehension difficulties.
- Automatically offer simplifications: After an initial response, the interface could discreetly offer options like "Do you need a simpler version?" or "Would you like a summary of this?"
- Automatic breakdown of complex answers: Divide long answers into smaller, more manageable sections, with the option to explore each section in detail.
- Key concept highlighting: Identify and highlight key terms or ideas in the answer, offering the ability to access definitions or explanations with only a click.
- Use of analogies or examples: When AI detects a potential difficulty, it may automatically offer analogies or simple examples to illustrate abstract concepts.

Although some recent versions of ChatGPT occasionally offer some of these features, such as highlighting key concepts or providing explanations, they are not consistently triggered, nor can users easily control when or how they appear. Even if such features appeared to users during the intervention, the difficulties reported by participants and the observed misunderstandings still hold. Therefore, we consider that integrating these functionalities in a more systematic and user-driven way would reduce the cases in which the user has to explicitly acknowledge their lack of understanding, offering more a intuitive and proactive support.

The widespread inability to verify information highlights the urgent need to incorporate visual and understandable indicators of source reliability, along with direct links to original sources and predefined prompts to guide the verification process. The unexpected relationships found between certain cognitive difficulties and verification ability point to the need for future research to explore the cognitive mechanisms underlying these findings. Furthermore, more research is needed to determine the long-term sustainability of the increase in technological confidence observed after participating in the study. The limited sample size of this study is acknowledged, and future research with larger samples is necessary to confirm and generalize these findings. Finally, the relationship between disability level and familiarity with AI tools should be further investigated, given the finding that no students with moderate or borderline disabilities were familiar with ChatGPT.

These lines of future work seek to advance the development of more inclusive and accessible GAIs that can truly empower people with cognitive disabilities in their learning and digital autonomy.

Acknowledgments

This publication is part of the R&D&I project HumanAI-UI, Grant PID2023-148577OB-C22 (Human-Centered AI: User-Driven Adaptative Interfaces) funded by MICIU/AEI/10.13039/501100011033 and by FEDER/UE. We want to thank the ACCEDE program tutor and students for their collaboration, and the support experts for their help during the intervention.

Declaration on generative AI

During the preparation of this work, the authors used ChatGPT in order to: grammar and spelling check and paraphrase and reword. Further, the authors used Google Translator for text translation. After using these tools, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] P. Rao, How Long it Took for Popular Apps to Reach 100 Million Users, 2023. URL: https://www.visualcapitalist.com/threads-100-million-users/.

[2] B. Thormundsson, ChatGPT - statistics & facts, 2024. URL: https://www.statista.com/topics/10446/chatgpt/.

[3] E. Choi, A comprehensive inquiry into the use of ChatGPT examining general, educational, and disability-focused perspectives, International Journal of Arts, Humanities & Social Science 4 (2023) 1–7. doi:10.56734/ijahss.v4n11a1.

[4] M. D. Korea, A. Panagiotis, Factors arising from the utilization of artificial intelligence and large language models in special education and training, European Journal of Special Education Research 10 (2024). doi:10.46827/ejse.v10i2.5209.

[5] K. G. Barman, N. Wood, P. Pawlowski, Beyond transparency and explainability: on the need for adequate and contextualized user guidelines for LLM use, Ethics and Information Technology 26 (2024) 47. doi:10.1007/S10676-024-09778-2.

[6] J. Nielsen, AI: First New UI paradigm in 60 years, 2023. URL: https://www.nngroup.com/articles/ai-paradigm/.

[7] J. D. Weisz, J. He, M. Muller, G. Hoefer, R. Miles, W. Geyer, Design Principles for Generative AI Applications, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 378:1–378:22. doi:10.1145/3613904.3642466.

[8] E. F. Duarte, P. T. Palomino, T. P. Falcão, G. L. Proença, M. B. Porto, C. dos Santos Portela, D. F. Ribeiro, A. Nascimento, Y. P. Costa, M. Souza, A. M. Segoria, A. M. Toda, GranDIHC-BR 2025-2035 - GC6: Implications of Artificial Intelligence in HCI: A Discussion on Paradigms Ethics and Diversity Equity and Inclusion, in: Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems, IHC '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 22:1–22:19. doi:10.1145/3702038.3702059.

[9] A. Han, Z. Cai, Design implications of generative AI systems for visual storytelling for young learners, in: Proceedings of the 22nd Annual ACM Interaction Design and Children Conference, IDC 2023, Chicago, IL, USA, June 19-23, 2023, ACM, 2023, pp. 470–474. doi:10.1145/3585088.3593867.

[10] J. D. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, Q. Yang, Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023, ACM, 2023, pp. 437:1–437:21. doi:10.1145/3544548.3581388.

[11] E. Zhou, D. Lee, Generative artificial intelligence, human creativity, and art, PNAS Nexus 3 (2024) 1–8. doi:10.1093/pnasnexus/pgae052.

[12] K. S. Glazko, M. Yamagami, A. Desai, K. A. Mack, V. Potluri, X. Xu, J. Mankoff, An Autoethnographic Case Study of Generative Artificial Intelligence's Utility for Accessibility, in: Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 99:1–99:8. doi:10.1145/3597638.3614548.

[13] P. Acosta-Vargas, B. Salvador-Acosta, S. Novillo-Villegas, D. Sarantis, L. Salvador-Ullauri, Generative artificial intelligence and web accessibility: Towards an inclusive and sustainable future, Emerging Science Journal 8 (2024) 1602–1621. doi:10.28991/ESJ-2024-08-04-021.

[14] P. Acosta-Vargas, G. Acosta-Vargas, B. Salvador-Acosta, J. Jadan-Guerrero, Addressing web accessibility challenges with generative artificial intelligence tools for inclusive education, in: 10th International Conference on eDemocracy and eGovernment, ICEDEG 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1–7. doi:10.1109/ICEDEG61611.2024.10702085.

[15] B. Alshaigy, V. Grande, Forgotten again: Addressing accessibility challenges of generative ai tools for people with disabilities, in: Adjunct Proceedings of the 2024 Nordic Conference on Human-Computer Interaction, NordiCHI '24 Adjunct, Association for Computing Machinery, New York, NY, USA, 2024, pp. 68:1–68:6. doi:10.1145/3677045.3685493.

[16] X. Liu, N. Lau, A. Chuin, W. K. R. Leung, A. H. S. Ho, M. Das, M. Liu, C. L. Kwok, Understanding Students' Perspectives, Practices, and Challenges of Designing with AI in Special Schools, in: Proceedings of the Eleventh International Symposium of Chinese CHI, CHCHI '23, Association for Computing Machinery, New York, NY, USA, 2024, p. 197–209. doi:10.1145/3629606.3629625.

[17] X. Mitre, M. Zeneli, Using AI to Improve Accessibility and Inclusivity in Higher Education for Students with Disabilities, in: 2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET), 2024, pp. 1–8. doi:10.1109/ITHET61869.2024.10837607.

[18] O. Pierrès, A. Darvishy, M. Christen, Exploring the role of generative AI in higher education: Semi-structured interviews with students with disabilities, Education and Information Technologies 30 (2024) 8923–8952. doi:10.1007/s10639-024-13134-8.

[19] R. Haroon, F. Dogar, TwIPS: A Large Language Model Powered Texting Application to Simplify Conversational Nuances for Autistic Users, in: Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 24:1–24:18. doi:10.1145/3663548.3675633.

[20] K. Mullen, W. Xue, M. Kudumu, "I'm treating it kind of like a diary": Characterizing How Users

with Disabilities Use AI Chatbots, in: Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 133:1–133:7. doi:10.1145/3663548.3688549.

[21] S. Roomkham, L. Sitbon, Restarting the conversation about conversational search: exploring new possibilities for multimodal and collaborative systems with people with intellectual disability, in: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval, CHIIR '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 231–242. doi:10.1145/3627508.3638339.

[22] J. Jang, S. Moharana, P. Carrington, A. Begel, "It's the only thing I can trust": Envisioning Large Language Model Use by Autistic Workers for Communication Assistance, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 77:1–77:18. doi:10.1145/3613904.3642894.

[23] J. Green, N. Thorogood, Qualitative Methods for Health Research, Introducing Qualitative Methods series, SAGE Publications, 2009. doi:10.7748/nr.13.2.91.s14.

[24] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative Research in Psychology 3 (2006) 77–101. doi:10.1191/1478088706qp063oa.

[25] U. H. Graneheim, B. Lundman, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, Nurse Education Today 24 (2004) 105–112. doi:10.1016/j.nedt.2003.10.001.