2026 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR) | 979-8-3315-4967-1/26/$31.00 ©2026 IEEE | DOI: 10.1109/AIXVR67263.2026.00060

Towards Augmented Reality Authoring Tools that Ensure Cognitive Accessibility of Instructions

Valentin Knoben*, Jonas Blattgerste†, Thies Pfeiffer†, Björn Hein‡ and Christian Wurll‡
*Karlsruhe Institute of Technology
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
Email: valentin.knoben@kit.edu
†University of Applied Sciences Emden/Leer
Emden, Germany
Email: jonas.blattgerste@hs-emden-leer.de, thies.pfeiffer@hs-emden-leer.de
‡Karlsruhe University of Applied Sciences
Karlsruhe, Germany
Email: bjoern.hein@h-ka.de, christian.wurll@h-ka.de

***Abstract*—Current Augmented Reality (AR) authoring tool abstractions, while simplifying content creation, force trade-offs that often neglect cognitive accessibility as a result. To address this gap, we conducted a Hierarchical Task Analysis (HTA) of an exemplary AR-guided task created in Microsoft Dynamics 365 Guides and mapped it to the WHO's International Classification of Functioning, Disability and Health (ICF) to identify specific cognitive barriers. Based on these findings, we introduce a framework where multimodal AI shifts the accessibility burden from the author to the authoring tool through a combination of authoring-time guidance and adaptive runtime support. We discuss specific leverage points for AI-driven auditing and outline a research agenda of design opportunities. This approach could empower authors to produce inclusive instructions by default without needing specialized expertise.**

***Index Terms*—Augmented Reality, Cognitive Accessibility, Authoring Tools, Artificial Intelligence**

I. Introduction

Mixed Reality (MR), as conceptualized in 1994 by Milgram and Kishino [1], describes a continuum of technologies that merge real and virtual environments. The proposed "reality-virtuality continuum" ranges from the physical world to fully immersive virtual settings. Augmented Reality (AR) is a key technology on this spectrum, characterized by displaying computer-generated content in situ, where it is needed. This method of displaying information directly within the user's context has been shown to aid in transferring instructions from an abstract representation to a real-world task, proving especially useful for untrained workers learning a new process [2]. Because of this, AR has shown noteworthy potential across broad assistance and educational training use cases. This is especially true for people with cognitive impairments, who often require this form of assistance the most and struggle with exactly this cognitive transfer process [3].

While research has increasingly explored these potentials in various contexts, creating this AR content remains of ongoing interest for the research community [4]. To address this challenge, a significant research trend focuses on making AR content creation more accessible by developing authoring tools that simplify the process of building AR instructions for domain experts, even without programming knowledge [5].

As Hampshire et al. [6] already discussed 20 years ago, these AR authoring tools always inherently introduce a challenge: by increasing interface abstraction to make them easier to use, they also necessarily limit the expressive power and control authors have over the created content. One consequence of this abstraction is that some aspects are not considered; for example, and of particular interest for us in this paper, the resulting instructions are not necessarily cognitively accessible. This specific gap is then also compounded by several peripheral factors: a general lack of awareness among developers, a shortage of guiding frameworks, as accessibility research has historically prioritized sensory over cognitive impairments [7], and the subjectively high effort of implementation [8]. Even when awareness and frameworks exist, this additional step is often not taken as an economic decision [8], a pattern also observed in web accessibility [9]. Consequently, in the pursuit of simplified AR authoring capabilities, cognitive accessibility has not been adequately addressed to date.

This problem cannot, however, be solved by abandoning the simplified authoring paradigm. The level of abstraction and simplification must be maintained, as it is unrealistic to expect authors to be experts in both the nuances of AR and the complex requirements of cognitive accessibility. Even if they were, the manual creation of differentiated instructions of the same authored content for multiple target groups or individuals with differing requirements would negate the very efficiency gains these authoring tools are meant to provide.

Therefore, we propose that automated systems should ensure cognitive accessibility of AR instructions in AR authoring tools. We believe this support must occur at two distinct stages: partly during the authoring process, through automated analysis and human-in-the-loop guidance for the author, and partly

This work was supported by the Ministry of Science, Research and the Arts of Baden-Württemberg (MWK) (grant BW6_03) and zukunft.niedersachsen.

2771-7453/26/$31.00 ©2026 IEEE
DOI 10.1109/AIxVR67263.2026.00060

at runtime, through adaptive features that are automatically delivered with the authored AR instructions. We anticipate that recent advances in multimodal Artificial Intelligence (AI), specifically AI that can both process and generate multimodal input and output, now make such a sophisticated, dual-stage approach practically feasible for the first time and of particular interest to us as applied researchers. While this technological shift is promising, the usage of these models for cognitive accessibility is not trivial and creates a need to define clear requirements. This leads to the core question of this paper:

How can the capabilities of multimodal AI be leveraged within AR authoring tools to automatically analyze, adapt, and validate instructional AR content so that created instructions are cognitively accessible?

Addressing this question, we adopt a perspective based on functional barriers, focusing on the interplay between a user's mental functions and task-relevant factors rather than on specific diagnoses. In our view, this functional perspective offers a more actionable basis for system design than medical diagnoses, as it accounts for the fact that distinct conditions often share overlapping barriers, while individuals with the same diagnosis may not necessarily face the same functional limitations. This view encompasses cognitive barriers related to memory, attention, and executive functions [10], as well as psychological barriers like motivation, confidence, and stress [11]. Based on this functional view, we systematically analyze the problem using the International Classification of Functioning, Disability and Health (ICF) [12] by the World Health Organization (WHO) to identify how specific functional impairments create barriers at different stages of an exemplary AR-guided task. We then map these challenges to the current capabilities of multimodal AI, outlining a framework of leverage points for intervention. Finally, we discuss the path forward, distinguishing between immediate opportunities and remaining complex, long-term research challenges.

II. Related Work

Existing research related to our work spans across three active fields. The first is research on AR authoring, which aids non-programmers in creating content [4], [5]. While recent work incorporates AI into AR authoring [13], it lacks an accessibility focus. The second is MR accessibility, which identifies fundamental user barriers. The third is AI-driven cognitive accessibility, an active topic in web and software design that utilizes AI for adaptation. Together, this motivates our work, which specifically addresses AI-ensured cognitive accessibility within AR authoring tools.

A. Mixed Reality Accessibility Efforts

The research community has argued that for MR technologies to be truly inclusive, accessibility must be considered a fundamental component from the start, not as a post-hoc patch [14]. In line with this, Creed et al. [15] published a comprehensive research agenda for inclusive AR and VR, derived from multidisciplinary workshops with stakeholders. They identified unaddressed challenges, noting that barriers are not just technical but also structural and ethical. Key areas requiring further investigation include the need for better disability representation, the development of accessible authoring platforms to empower disabled users as creators, and the necessity of embedding users with disabilities in all stages of the research and design process. Dudley et al. [7] introduced the concept of "inclusive immersion", synthesizing research and commercial efforts to improve accessibility by maximizing both access and enjoyment. Their work maps existing approaches into key design strategies and highlights unaddressed challenges, including the difficulty of designing for the vast diversity of user needs, a lack of developer guidance and tools, and the ethical and logistical difficulties in conducting empirical research with disabled participants.

However, this broad coverage can be misleading, as the focus of prior research has been unevenly distributed. A systematic literature review by Gerling et al. [16] demonstrates that the majority of work targets motor-physical or sensory impairments, while only a single paper focused on cognitive impairments. In line with the other efforts, the authors argue that the research tends to be "barrier-centric" and often neglects the user's experience.

This skewed focus is also reflected in the development of practitioner guidelines and technical standards. Work such as Heilemann et al. [17], which synthesizes accessibility guidelines for VR games, and reports from bodies like the IEEE [18], predominantly emphasize adaptable input and output modalities (e.g., captions, controller remapping, contrast modes) that primarily address sensory and motor barriers. When cognitive accessibility is addressed, the recommendations are frequently limited to general simplifications, such as skipping tasks or reducing interaction speed [18]. At the level of formal technical standardization, as surveyed by Makamara and Adolph [19], efforts like the W3C's "XR Accessibility User Requirements" are broad but do not define mechanisms for cognitive accessibility beyond simplifications [20].

B. Cognitive Accessibility through Artificial Intelligence

Parallel to the accessibility efforts within MR, a separate research field concerns the use of AI to automate accessibility in general. A comprehensive systematic review by Bhavana et al. [21] states that AI is "transforming" traditional assistive technologies into "intelligent systems" that enhance accessibility across visual, auditory, motor, and cognitive domains. A significant focus has emerged on advancing cognitive accessibility specifically, and researchers describe the move from static, "one-size-fits-all solutions" to dynamic, AI-driven systems as a "paradigm shift" towards fewer learning and communication barriers [22].

Much of this work leverages Large Language Models (LLMs) to address the high cost and time of manual content adaptation. For instance, Ledoyen et al. [23] investigate multi-task LLM approaches to automate the generation of easy-to-read textual content, a format specifically designed for individuals with cognitive impairments. This technological shift is not limited to content generation alone but is also influ-

encing the methods for designing accessible systems. Moreno et al. [24] present cognitive accessibility design patterns for user interfaces that provide content simplification. Similarly, Pascual et al. [25] explore the use of generative AI to create and refine usability and accessibility heuristics focused on cognitive diversity. Notably, this trend aligns with the broader vision for the next decade of accessibility research described by Gerling et al. [26], who envision generative AI being embedded directly into design and development processes to "address accessibility from the start," similarly to Mott et al. [14] which described these visions for MR accessibility.

III. Hierarchical Task Analysis

To improve the cognitive accessibility of authored AR instructions, we need to understand how cognitive disability impacts the use of AR-based instructions and which functional barriers can occur. As a first step towards this understanding, we conducted a Hierarchical Task Analysis (HTA) of an exemplary AR-guided task [27], focused on the mental processes involved. With the motivation to identify accessibility barriers for users with cognitive disabilities in AR-guided work scenarios, we determined involved mental functions as defined by the ICF [12]. In this, an impairment of an associated mental function points towards a potential barrier.

An exemplary practical use case from an ongoing study in a sheltered workshop served as the basis for our analysis. Here, users set up a pad printing machine using an AR Head-Mounted Display (HMD), following AR instructions provided in Microsoft Dynamics 365 Guides [28]. We used the fastening of the print plate inside the machine as a step to illustrate derived mental stages. Practically, this step required the worker to fasten the print plate using an Allen key to prevent it from sliding out. Instructions for this step consisted of a (1) video demonstrating the action, (2) 3D holographic hand indicating the location of the Allen key, and (3) textual instruction describing the action and additionally specifying how tightly the bolt should be fastened. This workflow and the resulting support concept are visualized in Fig. 1.

With the HTA, we also considered concepts of theoretical cognitive ergonomics, namely Endsley's theory of situation awareness [29] and Norman's seven stages of action [30]. In the following, we outline actions and decisions for users to make during every stage of a discrete step and map them to relevant mental functions as defined by the ICF [12]. In this ICF notation, the prefix 'b' denotes body functions, while the subsequent digits represent the hierarchy of increasingly specific sub-functions (e.g., b164 to b1646). The consolidated mapping is shown in Table I.

A. Perception

In an augmented environment, users have to maintain awareness of their own localization within the workspace, providing the spatial frame necessary to *perceive* and relate virtual and physical elements (b1141). Initially, the virtual environment has to be scanned for instructional content. This involves visual perception of, e.g., the virtual hand pointing to the location of the Allen key, as well as the holographic panels containing the textual and video instruction (b1561). Auditory perception also plays a role if an acoustic instruction, or simply text-to-speech, is active (b1560). Sustained attention is necessary to perceive instructions in their entirety (b1400), e.g., if following an instructional video. In AR environments, users must also form a clear spatial understanding of where virtual elements sit in relation to real-world objects (b1565). Having perceived text, video, and 3D instructions, their location needs to be remembered so that the user can revisit them on demand (b1440). This phase aligns with the perception level in Endsley's model of situation awareness, where detection and identification of relevant elements provide the raw inputs for later processing [29].

B. Comprehension

Following the perception of instructions, users enter the *comprehension* stage, where previously perceived instructional elements are interpreted and integrated into a coherent, actionable understanding (b117). In Endsley's terms, in this stage, meaning and relationships among perceived content are formed so that accurate decision options can be generated [29]. For that, users need to be able to shift between multiple different instruction modalities (b1401) and comprehend them altogether to derive instructional intent (b1402). In particular, for the latter, perceived instructions need to be memorized so that joint comprehension can take place (b1440). Maintaining obtained information in long-term memory can further aid the comprehension of future tasks (b1441), while recalling such context from previous steps helps with understanding the task at hand (b1442). Provided attentional focus on and memorization of perceived instructions, the user then has to abstract the idea and intent behind the conveyed information (b1640) and decide on the most plausible interpretation (b1645). To understand textual instructions, mental functions of language are required (b167). In case of our exemplary step, comprehension involves understanding that the holographic hand signals the location of the Allen key to use for this task, the video conveys the precise procedure of how to fasten the print plate in place, and the text additionally gives the user a feeling of how tight it should be fastened.

C. Preparation

In the *preparation* stage, the user has to convert interpreted instructions into a concrete, executable plan to solve the instructed task. Theoretically, this stage constitutes a hybrid between the transition from comprehension to decision as described by Endsley [29] and Norman's intention-formation and action-specification steps [30]. During this stage, the user has to sample the physical environment to localize and verify necessary parts and tools (b156) while recalling instruction goals (b1442) and storing object locations for later execution of the task (b1440). Cognitively, ideas have to be formed with the objective of accomplishing the instructed task (b1646), the task has to be decomposed into a sequence of micro-actions

TABLE I
MENTAL FUNCTIONS MAPPED ONTO ASSOCIATED ACTIONS INVOLVED IN AR-ASSISTED TASKS AND ORGANIZED BY COGNITIVE STAGES

Stage	Mental Function	Sub-Function	Task Action
Perception	Orientation (b114)	Orientation to place (b1141)	Understand one's position within the augmented environment
	Attention (b140)	Sustaining (b1400)	Concentrate on a single instruction modality
	Memory (b144)	Short-term (b1440)	Remember location of instructional holograms
	Perceptual (b156)	Auditory (b1560)	Hear acoustic instruction
		Visual (b1561)	See text, image/video, 3D instructions
		Visuospatial (b1565)	Link physical and virtual world
Comprehension	Intellectual (b117)		Overall capacity to understand and learn instructed task
	Attention (b140)	Shifting (b1401)	Shift focus between multiple instructions
		Dividing (b1402)	Jointly comprehend multiple instructions
	Memory (b144)	Short-term (b1440)	Remember instruction content
		Long-term (b1441)	Remember task in broader context
		Retrieval (b1442)	Recall work context from previous steps
	Higher-level cognitive (b164)	Abstraction (b1640)	Understand conveyed task intention
		Judgment (b1645)	Evaluate different possible task interpretations
	Language (b167)	Reception (b1670)	Understand textual and acoustic instructions
Preparation	Memory (b144)	Short-term (b1440)	Remember locations of task-relevant objects
		Retrieval (b1442)	Recall instruction
	Perceptual (b156)	Visual (b1561)	Scan surrounding for task-relevant objects
		Visuospatial (b1565)	Localize task-relevant objects
	Higher-level cognitive (b164)	Organization (b1641)	Decompose task into action sequence
		Time management (b1642)	Estimate and allocate required time
		Cognitive flexibility (b1643)	Adapt plan to environmental constraints
		Judgment (b1645)	Evaluate feasibility of devised action plan
		Problem solving (b1646)	Come up with solution for instructed task
Action	Attention (b140)	Sustaining (b1400)	Concentrate on performed action
		Dividing (b1402)	Monitor action, virtual, and physical environment
	Memory (b144)	Retrieval (b1442)	Recall instructed task
	Psychomotor (b147)	Control (b1470)	Manage execution speed, motor response
		Quality (b1471)	Manage execution precision and quality
	Perceptual (b156)	Auditory (b1560)	Perceive acoustic action cues
		Visual (b1561)	Perceive visual action cues
		Tactile (b1564)	Perceive tactile action cues
		Visuospatial (b1565)	Perceive action-induced, changing spatial relations
	Higher-level cognitive (b164)	Cognitive flexibility (b1643)	Adapt execution to unexpected changes
		Problem solving (b1646)	Solve unexpected issues during execution
	Sequencing of complex movement (b176)		Coordinate involved movements
Validation	Attention (b140)	Sustaining (b1400)	Validate single area of interest
		Shifting (b1401)	Validate multiple areas of interest
		Dividing (b1402)	Jointly validate multiple areas of interest
	Memory (b144)	Retrieval (b1442)	Recall target outcome
	Perceptual (b156)	Visual (b1561)	Perceive visual changes
		Tactile (b1564)	Perceive tactile changes
		Visuospatial (b1565)	Perceive changed spatial relations
	Higher-level cognitive (b164)	Cognitive flexibility (b1643)	Adjust strategy in case of deviation
		Judgment (b1645)	Compare result with target outcome
		Problem-solving (b1646)	Develop strategy to address deviations
Task Engagement	Attention (b140)	Sustaining (b1400)	Keep up concentration without getting distracted
	Temperament and personality (b126)	Openness (b1264)	Curious towards task at hand
		Confidence (b1266)	Confidence or tentativeness while performing the task
	Energy and drive (b130)	Energy level (b1300)	Mental energy and effort spent to perform the task
		Motivation (b1301)	Inherent motivation to perform the task
		Impulse control (b1304)	Resisting urges irrelevant to the task
Stress Management	Temperament and personality (b126)	Extraversion (b1260)	Willingness to ask for help
		Psychic stability (b1263)	Remaining calm and patient
		Optimism (b1265)	Staying positive
		Confidence (b1266)	Remaining confident
	Emotional (b152)	Regulation (b1521)	Manage frustration
	Higher-level cognitive (b164)	Cognitive flexibility (b1643)	Adapt thinking to undesired outcome
System Interaction	Temperament and personality (b126)	Openness (b1264)	Curious to explore virtual environment
		Confidence (b1266)	Confidence or tentativeness when interacting with system
	Attention (b140)	Sustaining (b1400)	Focus on control elements
	Memory (b144)	Short-term (b1440)	Remember system functionalities and their location
	Psychomotor (b147)	Control (b1470)	Translate intention into correct system interaction
		Quality (b1471)	Precise system interaction (pinch, drag, gaze)
	Perceptual (b156)	Auditory (b1560)	Perceive acoustic interaction feedback
		Visual (b1561)	Perceive visual change on interaction
		Visuospatial (b1565)	Localize and perceive interaction elements
	Higher-level cognitive (b164)	Abstraction (b1640)	Understand system functions and their effect
	Language (b167)	Reception (b1670)	Understand textual labels or spoken system feedback
		Expression (b1671)	Produce voice commands
	Sequencing of complex movement (b176)		Perform multi-step UI navigation and manipulation

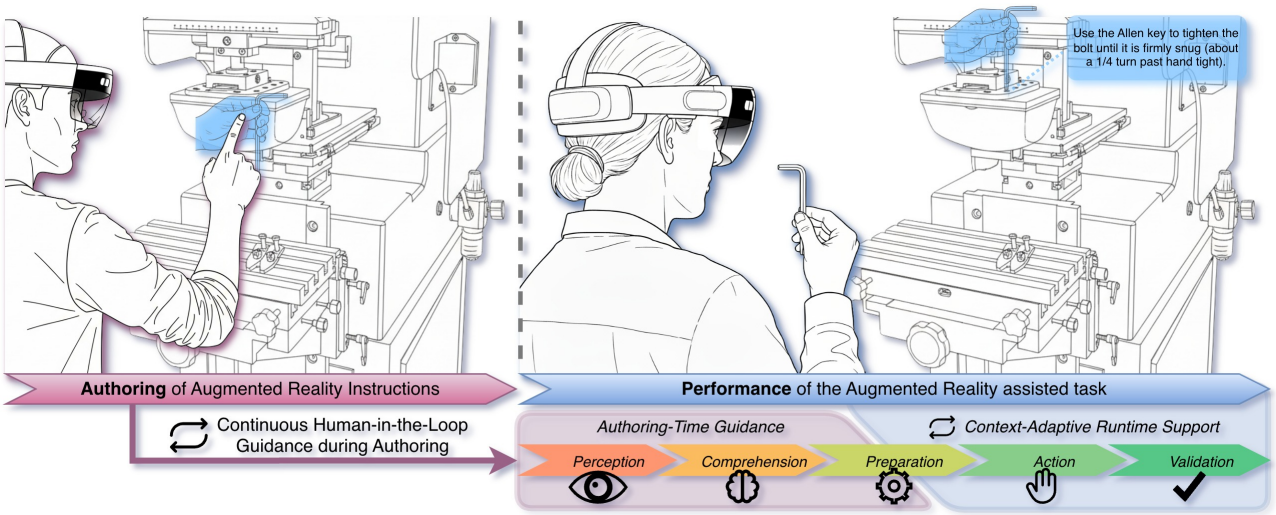

Fig. 1. Our dual-stage model for ensuring cognitive accessibility in AR task guidance. **Authoring-Time Guidance** involves AI-guidance *for authors*, helping them create clear content to support users' initial cognitive stages (Perception, Comprehension, Preparation). **Context-Adaptive Runtime Support** then delivers automated adaptations *driven by real-time task context*, supporting the dynamic phases of execution (Preparation, Action, Validation).

(b1641), the time for each step has to be estimated and allocated (b1642), and a feasibility evaluation of the formed plan has to be performed (b1645). Cognitive flexibility allows the user to adjust the plan in response to environmental constraints (b1643). For instance, the user may choose to engage the Allen key using either the long or short side. Considering our exemplary task, preparation may include localizing the Allen key, planning the grip and torque direction, and making sure there is enough space for engagement.

D. Action

Having formed a plan for execution and recalling the instructed task (b1442), the user enters the *action* stage where the task is performed. This phase operationalizes Norman's execution stage [30] and concludes Endsley's awareness loop by producing a new situation to be perceived and interpreted [29]. In our exemplary step, this involves grabbing the Allen key, engaging it on the bolt, fastening it finger-tight, and putting it back. To properly perform the task, the user has to coordinate and control overall movement speed and precision (b147, b176). Perceiving sensory feedback allows users to detect deviations due to visuospatial (b1561, b1565), auditory (b1560), or tactile cues (b1564). This, in turn, requires them to react dynamically (b1643) and come up with suitable corrective adjustments (b1646). For example, by sensing the physical resistance, the user can determine if the print plate is or is not sufficiently fastened according to the instruction. While the user's main focus will lie on the task being executed (b1400), they are required to simultaneously monitor the physical and virtual workspace for changes or external disturbances (b1402) such as people walking by.

E. Validation

Concluding the mental process behind an AR-guided operation, the user enters the *validation* stage, in which the achieved result is reviewed. This concludes Norman's seven stages of action by evaluating the outcome against the previously formed goal [30] and restarts Endsley's situation awareness loop by returning to environmental perception and comprehension of a new situation [29]. Similar to comprehension of a task, during validation of the outcome, the user not only has to focus on a single point of interest (b1400) but also has to shift their attention between multiple points (b1401) and evaluate them holistically (b1402). For example, the user replays the video instruction and compares the current physical state with the running video, requiring constant switching of context. Perceptual functions allow the user to detect changes produced through action (b156), while retrieval from memory (b1442) provides the mental reference of the desired goal state to compare to. Ultimately, the user decides whether the target outcome has been achieved (b1645) or whether corrective actions are needed. In that case, cognitive flexibility (b1643) enables the user to adapt their action strategy or come up with an alternative plan, solving the newly formed problem (b1646).

F. General

Beyond previously outlined key stages users go through when performing an AR-guided task, we found mental behaviors relevant throughout the whole workflow independent of the stage. We organized these into the following three groups:

1) Task Engagement: Being cognitively able to engage in a task does not yet mean a person is willing to do so. Inherent mental energy (b1300) and motivational drive (b1301) are necessary as well. Having control over sudden impulses (b1304) and being able to focus on the task without being

distracted (b1400) allows task engagement over a longer period of time. Overall disposition to be experience-seeking (b1264) and confident in performing actions (b1266) also has an effect on the workflow. For instance, while tentativeness may lead to inefficient execution and redundant requests for support, overly confident behavior can result in more errors being made.

2) Stress Management: Stress in an AR-guided workplace can form due to problems in either the virtual or physical environment where the outcome does not match expectations, e.g., if the user makes a mistake or triggers an User Interface (UI) event inadvertently. While stress itself is not an emotion or mental process, the response to it is, and as such determines how well stress is dealt with. A user's disposition to remain calm (b1263), optimistic (b1265), and confident (b1266) impacts how they cope with stressful situations. The ability to regulate emotions (b1521) and flexibly adapt one's thinking (b1643) to an unexpected situation can contribute to successful stress management. Ultimately, a user's willingness to ask for help (b1260) also decides if somebody else provides support.

3) System Interaction: Interaction with an AR system can itself become an accessibility barrier. Temperament and personality (b126) can shape a user's willingness (b1264) and confidence (b1266) to explore the interface, particularly when they are new to AR. Sustained attention (b1400) is relevant, e.g., in gaze-based scenarios where the user is required to dwell on an element for a certain amount of time, as was the case for navigating between steps. Understanding what control elements do (b1640) and remembering where they are and how they work (b1440) facilitates efficient interaction. Complex, multi-step interactions involving gaze, pinch, and drag, for example when spatially repositioning instructional panels, require control over movements (b1470), their quality (b1471), and sequence (b176). For speech interfaces, both the understanding of system text elements (b1670) and correct verbalization when issuing voice commands (b1671) are needed. Finally, perceptual functions (b156) aid the correct spatial perception of control elements and of multimodal feedback after UI events. An example would be the acoustic sound played after advancing to the next step.

IV. Discussion

The Authoring Tool Accessibility Guidelines (ATAG) by the W3C describe two key-aspects of accessibility: making the authoring interface accessible and supporting the production of accessible web content [31]. Combined with the Web Content Accessibility Guidelines (WCAG) [32], these perspectives map to three levels of addressing cognitive accessibility in the context of AR-guided work scenarios:

1) Making the authoring process itself accessible
2) Supporting the authoring of accessible instructions
3) Supporting accessibility during use of instructions

The first level targets the accessibility of the tool for the author. In contrast, the latter two levels aim to support the user of the authored content, aligning directly with the two phases of our model depicted in Fig. 1: supporting the creation of accessible instructions (authoring-time guidance) and supporting the accessible delivery of those (runtime support).

Although our HTA identified barriers in all five stages of the user's workflow, we center this discussion specifically on the early cognitive stages: *perception*, *comprehension*, and *preparation*. We believe, these stages involve mental functions that are most effectively addressed via authoring-time guidance. For instance, it appears that in our exemplary task, stages such as *action* or *system interaction* heavily depend on momentary context and are therefore less amenable to authoring-time measures. In contrast, the representation of instructions is effectively under the author's control. Thus, perception, understanding, and planning of an AR-guided task step can be aided substantially through proper authoring. Crucially, recent advances in multimodal AI could potentially enable the support of exactly those mental processes.

We argue addressing these initial barriers at the source is a frequently overlooked but necessary prerequisite for safety, serving as the reliable foundation upon which future runtime support must be built. Taking a practical perspective in view of recent and anticipated advances in AI, we believe authoring-time guidance remains a critical checkpoint for cognitive accessibility in the near future. Sole reliance on runtime adaptation cannot guarantee the factual and representational correctness of AR instructions, and mistakes quickly erode trust and acceptance, especially for vulnerable user groups [33]. Even if AI-based adaptations are technically correct, retaining a human-in-the-loop creates an auditable trail for safety and regulatory compliance. Furthermore, moving parts of adaptation to authoring-time reduces dependence on and use of computational resources at runtime, which AR HMDs typically lack and would need to acquire externally.

In the following, we outline authoring-time guidance features and discuss which are readily implementable today, e.g., with rule-based machine learning approaches, and which require more advanced but emerging multimodal AI capabilities. We organize these leverage points by key functional domains found in our analysis to show how they target specific cognitive barriers. By that, we underline our vision of a functional rather than diagnostic approach to cognitive accessibility in AR system design. As Johansson [11] correctly points out, a diagnosis alone is not informative in terms of how digital solutions can be designed accessibly. Instead, targeting the underlying functional barriers provides a more concrete implementation target, enabling the creation of adaptable features that effectively support diverse user groups with overlapping cognitive needs.

A. Managing perceptual and attentional demands

Perceptual (b156) and attentional (b140) barriers can arise from visual clutter, occlusion, or poor findability and quality of instructions. Straightforward approaches addressing this, which are already employable today, include quality assessment of acoustic or video instructions, e.g., for visual artifacts such as blur, noise, or flicker [34]. Beyond these unimodal

checks, solid cross-modal reasoning has the potential to enable more context-aware support. For instance, by combining task knowledge gained from a textual instruction with saliency prediction [35], the identification of attentional hotspots, the system can verify if the correct bolt and Allen key are visually prominent during the fastening action and whether users are likely to attend to these critical elements. Going even further, processing spatial information as another input modality, e.g., through dense representations [36] or point clouds [37], opens even more possibilities, such as recommending context-adapted anchor locations for instructional holograms instead of blind defaults, improving overall perceptibility of instructions.

B. Reducing memory load

AR-guided tasks rely heavily on the user's memory (b144), e.g., to remember instructions themselves, their broader task context, and locations, but also general system functioning. While runtime support for the memory is crucial, first barriers can be addressed at authoring-time already. For example, current LLMs can check if a textual instruction in fact describes multiple steps and decompose it into substeps. Integrating this textual context, the same can be done for an associated video instruction [38]. By that, the task not only becomes less complex, but the required memory load per task is also reduced. Apart from that, sometimes it is also helpful to recall information from the farther past, for example, whether the machine was correctly switched to setup mode, ensuring that reaching into it to fasten the print plate is safe. To alleviate this, relevant previous steps could be highlighted to authors at authoring-time for consideration or suggested as optional links to be included in the instruction itself. Additionally, the system could aid memory during validation by automatically extracting meaningful before-and-after snapshots from a video demonstration for comparison of the outcome to the target.

C. Supporting cognition

Higher-level cognitive (b164) and language (b167) functions are critical to comprehension and preparation of an AR-guided task (see Table I). We can support these through semantically clear, easy-to-understand, and coherent instructions. Pillars of cognitively accessible language are well established. Important factors include: keeping sentences short, addressing the reader, using active voice, and avoiding abbreviations or negations [39]. Consequently, LLMs can already audit authored text, apply policy-based adaptations, or generate variants different in detail. Importantly, multimodal reasoning also enables the authoring tool to verify cross-modal instructional coherence (e.g., whether a named tool in the textual instruction, such as the Allen key, appears in the video instruction or vice versa), which was shown to positively impact human learning [40]. Due to the extensive pre-training of modern multimodal models on textual and 2D visual data, correlating these two modalities is already a feasible step. However, accurately interpreting and assessing coherence against 3D instructions poses a remaining challenge.

D. Addressing psychological factors

Psychological factors, such as lack of confidence (b1266) or motivation (b1301), can pose barriers throughout a task, e.g., for general engagement and stress management. A system should therefore aim to minimize stress-inducing situations and actively motivate the user so that mistakes and unwanted scenarios are avoided and the user remains inherently open to the AR-guided experience (b1264). Lightweight support without deep task understanding can already be supplied through periodic suggestions of confirmation or "take a break" prompts for authors to include. However, with contextual understanding, the system can go further by automatically identifying potentially stressful, dangerous, or critical steps from the authored assets to then propose task-specific safety tips and motivational prompts. Research has shown that contextual specificity is highly effective in terms of motivating people [41].

V. Limitations & Future Work

These findings are a preliminary first step to generate research questions and analyze the design space rather than a representative or definitive framework. Our analysis focuses on a single exemplary step of one specific AR-guided task. While derived systematically, the mapping relies on the judgment of a single reviewer. In future work, we plan to verify these theoretical barriers by reviewing them with additional experts and by broadening the scope of the analysis to include diverse task types. Beyond this theoretical perspective, we acknowledge significant practical challenges regarding technical feasibility on mobile hardware, reliability of generative models in safety-critical contexts, data privacy, and latency. Despite these limitations, we consider these findings a valuable contribution to initiate the necessary discourse within the research community.

VI. Conclusion

In this paper, we present a functional analysis of cognitive demands in an exemplary AR-guided task using the ICF and derived a dual-stage model encompassing authoring-time guidance and runtime support. Highlighting the importance of authoring-time guidance, we outline a set of leverage points (e.g., cross-modal coherence audit, instructional simplification, contextual motivation and safety prompts) that can help reduce barriers to perception, attention, memory, cognition, and psychological engagement, key functional components of the analyzed AR-guided task. At the same time, many proposals depend on robust multimodal scene and task understanding, demanding further technical validation, user studies, and careful treatment of reliability, privacy, and regulatory concerns. With our work, we hope to motivate a research roadmap for embedding multimodal AI into AR authoring tools and conducting empirical evaluation to turn the identified leverage points into trustworthy accessibility practices.

References

[1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transactions on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[2] M. Funk, A. Bächler, L. Bächler, T. Kosch, T. Heidenreich, and A. Schmidt, "Working with augmented reality? A long-term analysis of in-situ instructions at the assembly workplace," in *Proceedings of the 10th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2017, pp. 222–229.

[3] J. Blattgerste, P. Renner, and T. Pfeiffer, "Augmented reality action assistance and learning for cognitively impaired people: A systematic literature review," in *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, ser. PETRA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 270–279. [Online]. Available: https://doi.org/10.1145/3316782.3316789

[4] M. Nebeling and M. Speicher, "The trouble with augmented reality/virtual reality authoring tools," in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2018.

[5] J. Blattgerste, "The design space of augmented reality authoring tools and its exploration for the procedural training context," Ph.D. dissertation, Ph. D thesis, Bielefeld University, 2023.

[6] A. Hampshire, H. Seichter, R. Grasset, and M. Billinghurst, "Augmented reality authoring: Generic context from programmer to designer," in *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, 2006.

[7] J. Dudley, L. Yin, V. Garaj, and P. O. Kristensson, "Inclusive immersion: A review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse," *Virtual Reality*, Sep. 2023. [Online]. Available: https://doi.org/10.1007/s10055-023-00850-8

[8] D. Killough, T. F. Ji, K. Zhang, Y. Hu, Y. Huang, R. Du, and Y. Zhao, "XR for all: Understanding developer perspectives on accessibility integration in extended reality," *Preprint arXiv:2412.16321*, 2024.

[9] WebAIM, "The WebAIM Million - The 2025 report on the accessibility of the top 1,000,000 home pages," https://webaim.org/projects/million/, 2024, accessed: 2025-09-19.

[10] M. Cooper and L. Seeman-Horwitz, "Cognitive accessibility user research," W3C, W3C Working Draft, Jan. 2015, https://www.w3.org/TR/2015/WD-coga-user-research-20150115/.

[11] S. Johansson, "Design for participation and inclusion will follow: Disabled people and the digital society," Ph.D. dissertation, KTH Royal Institute of Technology, 2019.

[12] WHO, *International classification of functioning, disability and health : ICF*. World Health Organization, 2001. [Online]. Available: https://iris.who.int/handle/10665/42407

[13] J. Lee, F. Aleotti, D. Mazala, G. Garcia-Hernando, S. Vicente, O. J. Johnston, I. Kraus-Liang, J. Powierza, D. Shin, J. E. Froehlich *et al.*, "Imaginatear: AI-assisted in-situ authoring in augmented reality," in *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, 2025, pp. 1–21.

[14] M. Mott, E. Cutrell, M. G. Franco, C. Holz, E. Ofek, R. Stoakley, and M. R. Morris, "Accessible by design: An opportunity for virtual reality," in *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2019, pp. 451–454.

[15] C. Creed, M. Al-Kalbani, A. Theil, S. Sarcar, and I. Williams, "Inclusive augmented and virtual reality: A research agenda," *International Journal of Human–Computer Interaction*, vol. 40, no. 20, pp. 6200–6219, 2024.

[16] K. Gerling, A.-L. Meiners, L. Schumm, J. Rixen, M. Wolf, Z. Yildiz, D. Alexandrovsky, and M. Opp, "An equitable experience? How HCI research conceptualizes accessibility of virtual reality in the context of disability," *ACM Transactions on Accessible Computing*, 2024.

[17] F. Heilemann, G. Zimmermann, and P. Münster, "Accessibility guidelines for VR games-A comparison and synthesis of a comprehensive set," *Frontiers in Virtual Reality*, vol. 2, p. 697504, 2021.

[18] D. Fox and I. G. Thornton, "White paper-The ieee global initiative on ethics of Extended Reality (XR) report–Extended Reality (XR) ethics and diversity, inclusion, and accessibility," *The IEEE Global Initiative on Ethics of Extended Reality (XR) Report–Extended Reality (XR) Ethics and Diversity, Inclusion, and Accessibility*, pp. 1–25, 2022.

[19] G. Makamara and M. Adolph, "A Survey of Extended Reality (XR) standards," *2022 ITU Kaleidoscope-Extended Reality–How to Boost Quality of Experience and Interoperability*, pp. 1–11, 2022.

[20] S. Hollier, M. Cooper, J. Sajka, J. O'Connor, and J. White, "XR accessibility user requirements," W3C, W3C Note, Aug. 2021.

[21] B. Bhavana, K. Ankolekar, and B. Usha, "Artificial intelligence for accessibility: A comprehensive systematic review and impact framework for assistive technologies," *International Journal of Advanced Research in Computer and Communication Engineering*, 2025.

[22] R. Deetjen-Ruiz, M. P. Daniel, J. Telus, and L. Deetjen, "Advancing cognitive accessibility: The role of artificial intelligence in enhancing inclusivity," *PriMera Scientific Engineering*, vol. 4, no. 2, p. 58, 2024.

[23] F. Ledoyen, G. Dias, J. Pantin, A. Lechervy, F. Maurel, and Y. Chahir, "Facilitating cognitive accessibility with llms: A multi-task approach to easy-to-read text generation," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 2025.

[24] L. Moreno, H. Petrie, P. Martínez, and R. Alarcon, "Designing user interfaces for content simplification aimed at people with cognitive impairments," *Universal Access in the Information Society*, vol. 23, no. 1, pp. 99–117, 2024.

[25] A. Pascual Almenara and S. Sayago Barrantes, "Towards designing a set of usability and accessibility heuristics focused on cognitive diversity: An exploratory case study with generative artificial intelligence," *Information*, vol. 15, no. 12, p. 769, 2024.

[26] K. Gerling, M. Rauschenberger, B. Tannert, and G. Weber, "The next decade in accessibility research," *I-Com*, vol. 23, no. 2, 2024.

[27] N. A. Stanton, "Hierarchical task analysis: Developments, applications, and extensions," *Applied Ergonomics*, vol. 37, no. 1, pp. 55–79, Jan. 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687005000980

[28] Microsoft, "Microsoft Dynamics 365 Guides," https://learn.microsoft.com/dynamics365/mixed-reality/guides, 2024, accessed: 2024-08-01.

[29] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems: Situation awareness," *Human Factors*, vol. 37, no. 1, 1995.

[30] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[31] J. Treviranus, J. Richards, and J. F. Spellman, "Authoring Tool Accessibility Guidelines (ATAG) 2.0," W3C, W3C Recommendation, Sep. 2015, https://www.w3.org/TR/2015/REC-ATAG20-20150924/.

[32] J. F. Spellman, F. Storr, A. Campbell, K. White, C. Adams, and R. B. Montgomery, "W3C Accessibility Guidelines (WCAG) 3.0," W3C, W3C Working Draft, Dec. 2024, https://www.w3.org/TR/2024/WD-wcag-3.0-20241212/.

[33] B. Thordardottir, A. Malmgren Fänge, C. Lethin, D. Rodriguez Gatta, and C. Chiatti, "Acceptance and use of innovative assistive technologies among people with cognitive impairment and their caregivers: a systematic review," *BioMed Research International*, vol. 2019, no. 1, p. 9196729, 2019, publisher: Wiley Online Library.

[34] Q. Zheng, Y. Fan, L. Huang, T. Zhu, J. Liu, Z. Hao, S. Xing, C.-J. Chen, X. Min, A. C. Bovik *et al.*, "Video quality assessment: A comprehensive survey," *Preprint arXiv:2412.04508*, 2024.

[35] W. Wang, J. Shen, J. Xie, M.-M. Cheng, H. Ling, and A. Borji, "Revisiting video saliency prediction in the deep learning era," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 220–237, 2019.

[36] R. Fu, J. Liu, X. Chen, Y. Nie, and W. Xiong, "Scene-llm: Extending language model for 3d visual understanding and reasoning," *Preprint arXiv:2403.11401*, 2024.

[37] X. Linghu, J. Huang, X. Niu, X. S. Ma, B. Jia, and S. Huang, "Multi-modal situated reasoning in 3d scenes," *Advances in Neural Information Processing Systems*, vol. 37, pp. 140 903–140 936, 2024.

[38] S. Nag, X. Zhu, Y.-Z. Song, and T. Xiang, "Zero-shot temporal action detection via vision-language prompting," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham: Springer Nature Switzerland, 2022, pp. 681–697.

[39] G. Freyhoff, G. Hess, L. Kerr, E. Menzel, B. Tronbacke, and K. Van Der Veken, *Make it Simple: European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability*. Brussels: ILSMH European Association, 1998.

[40] T. Seufert, "Supporting coherence formation in learning from multiple representations," *Learning and Instruction*, vol. 13, no. 2, Apr. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959475202000221

[41] K. Joyal-Desmarais, A. K. Scharmer, M. K. Madzelan, J. V. See, A. J. Rothman, and M. Snyder, "Appealing to motivation to change attitudes, intentions, and behavior: A systematic review and meta-analysis of 702 experimental tests of the effects of motivational message matching on persuasion." *Psychological Bulletin*, vol. 148, no. 7-8, p. 465, 2022.