Crossing Realities Has a Cost: Evaluating AR Second Screens for AI-Assisted Desktop Authoring

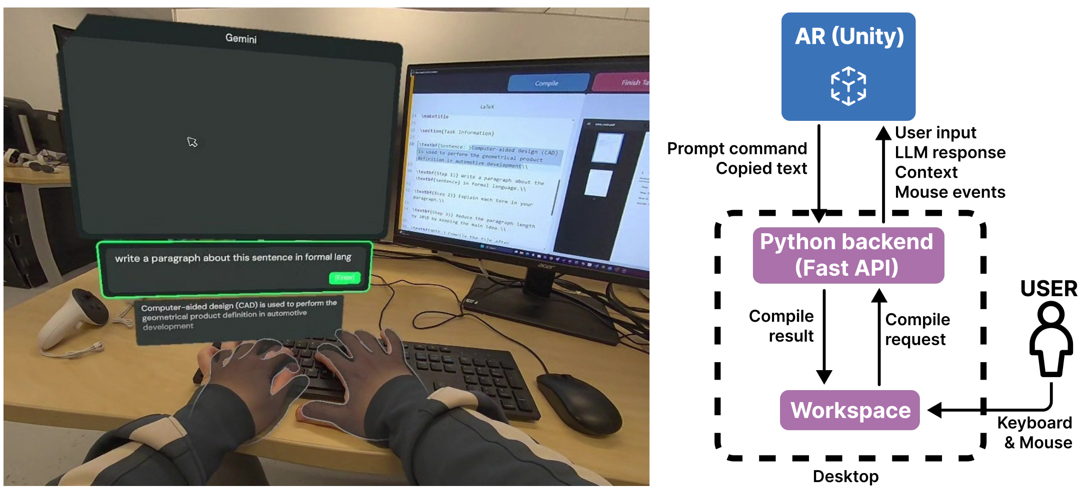
Mucahit Gemici
Department of Computer Science & Software Engineering
Concordia University
Montreal, Quebec, Canada
mucahit.gemici@mail.concordia.ca

Mohammad Raihanul Bashar
Department of Computer Science & Software Engineering
Concordia University
Montreal, Quebec, Canada
mohammadraihanul.bashar@mail.concordia.ca

Wolfgang Stuerzlinger
School of Interactive Arts + Technology (SIAT)
Simon Fraser University
Vancouver, British Columbia, Canada
w.s@sfu.ca

Vrushank Phadnis
Google
Mountain View, California, USA
vrushank@google.com

Anil Ufuk Batmaz
Department of Computer Science & Software Engineering
Concordia University
Montreal, Quebec, Canada
ufuk.batmaz@concordia.ca



(a) (b)

Figure 1: (a) User queries an LLM in AR with a physical keyboard. The context selected from the Workspace is shown under the input field. (b) A high-level system diagram of the system.

Abstract

AI-assisted authoring requires users to frequently exchange user-generated and AI-generated content between a primary workspace and a Large Language Model (LLM) assistant. Such repetitive actions can fragment attention and compete for limited display space. While Augmented Reality (AR) has been proposed as a way to extend desktop workspaces with virtual displays, just adding a spatial display area does not address the repeated attention shifts, input-focus changes, and content-transfer steps required in LLM-assisted workflows. This creates a gap in understanding whether AR second screens support such workflows or introduce cross-reality transition costs that limit their effectiveness. To investigate this gap, we designed an AR-Second-Screen system around a unified keyboard-and-mouse interaction model and evaluated it against Single-Screen and Dual-Screen desktop setups. Results show that Dual-Screen remained more efficient than AR-Second-Screen, with significantly lower task completion time and switching cost. AR-Second-Screen also produced higher task and LLM idle ratios, indicating re-engagement costs when users crossed between physical and virtual workspaces. At the same time, participants rated AR positively and preferred it significantly over Single-Screen, with no significant difference in preference compared to Dual-Screen. These findings show that spatially extending AI-assisted

SUI '26, Bari, Italy

ACM ISBN 979-8-4007-2812-9/26/10
https://doi.org/10.1145/3822518.3830046

workflows into AR can be appealing, yet effective cross-reality interfaces should minimize transition costs rather than simply add a virtual display space.

CCS Concepts

• **Human-centered computing → Interaction devices**; **Displays and imagers**.

Keywords

Augmented Reality, Cross-Reality, Human-AI Interaction, Input User Interfaces, Multi-Display Interaction, Authoring Workflows

ACM Reference Format:
Mucahit Gemici, Mohammad Raihanul Bashar, Wolfgang Stuerzlinger, Vrushank Phadnis, and Anil Ufuk Batmaz. 2026. Crossing Realities Has a Cost: Evaluating AR Second Screens for AI-Assisted Desktop Authoring. In *ACM Symposium on Spatial User Interaction (SUI '26), October 10–11, 2026, Bari, Italy.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3822518.3830046

1 Introduction

Large Language Model (LLM) workflows involve multi-turn conversations in which users must repeatedly ground the interaction in the relevant context, copy and paste portions of text into the LLM interface, inspect the response, and reintegrate useful content into their work. These iterations can fragment attention and compete for display space. Physical second monitors reduce this fragmentation by keeping the AI assistant persistently visible alongside the primary workspace [19]. However, physical displays are not always available, especially in mobile, shared-office, or space-constrained settings. As Augmented Reality (AR) systems and smart-glasses form factors continue to mature, virtual displays offer a spatial alternative [4], allowing auxiliary interfaces to be placed beside the physical workspace without occupying the primary screen or requiring an additional physical monitor. Yet placing an AI assistant in an AR interface raises a critical question: Does distributing an AI-assisted workflow between a desktop workspace and an AR panel introduce cross-reality transition costs that limit how effectively users can work across them?

Unlike adding a physical monitor, extending a desktop workflow into AR distributes interaction across two fundamentally different display surfaces [66]. Each time a user moves from the desktop workspace to the AR panel, e.g., to submit a prompt, they must *re-engage* with a different input context, recalibrate visual attention, and reorient their interaction focus before productive work can resume. These transitions may be especially consequential in LLM-assisted authoring, where information flows in both directions: from the document into the assistant as contextual input and from the assistant back into the document as reusable content. The central question is therefore not only whether AR can extend a workspace, but whether it can support repeated content exchange without re-engagement costs that undermine its value.

Previous work shows that virtual displays can extend interaction space beyond physical screens and support information access in mixed reality environments [44, 53, 55, 63]. Recent XR–LLM systems further demonstrate how AI assistants can support learning, accessibility, content generation, and situated interaction within immersive environments [11, 68, 69, 74]. However, these two bodies of work address different problems. AR workspace extension research focuses on display space and window management, treating content as relatively static, i.e., users access information but do not repeatedly exchange it with an active AI process. XR–LLM research, on the contrary, focuses on self-contained immersive experiences where users operate primarily within the AR or VR environment, without grounding in conventional desktop workflows. The hybrid scenario we investigate here involves a user anchored to a physical desktop who repeatedly exchanges context and generated content with an LLM panel placed in AR. This scenario falls between these two bodies of work and has received limited direct attention. Systems such as EmBARDiment [11] and InputJump [77] connect AR environments with desktop inputs, but they do not examine the repeated bidirectional transfer of context exchange that characterizes LLM-assisted authoring or quantify the transition overhead it creates. Understanding this overhead is necessary for designing AR systems that extend, rather than complicate, AI-assisted knowledge work.

To investigate this, we use LaTeX document authoring as a controlled experimental workflow. LaTeX was selected because its edit-compile-inspect cycle naturally involves repeated movement between authoring, verification, and external assistance, making it a useful controlled case for examining cross-reality transition costs. Rather than treating the AR interface as a direct replacement for a physical second monitor, we examine how placing an interactive AI assistant in AR changes the dynamics of the authoring workflow. Specifically, we ask the following research questions:

RQ1: How does placing an interactive LLM assistant in AR affect task performance and user experience relative to Single-Screen and Dual-Screen desktop configurations?

RQ2: What cross-reality transition costs emerge when users repeatedly exchange context and content between a physical desktop workspace and an AR-extended AI panel?

RQ3: What design considerations or guidelines follow from our work for future AR-extended, AI-assisted workspaces?

To answer these RQs, we conducted a formative study followed by a controlled within-subjects study with 18 participants. The formative study revealed interaction friction in the initial AR design that used head direction to focus input events, motivating a redesign with unified keyboard-and-mouse input and simplified bidirectional content transfer. We then compared the refined AR-Second-Screen system with Single-Screen and Dual-Screen desktop configurations.

Our results show that AR-extended LLM workspaces are subjectively appealing but do not yet match the efficiency of physical dual-screen setups. Participants completed tasks faster with Dual-Screen than with AR-Second-Screen, while AR-Second-Screen received positive ratings, outperformed Dual-Screen on Stimulation and Novelty scales, and was preferred significantly more than Single-Screen. However, AR-Second-Screen also produced higher switching costs and idle ratios, suggesting re-engagement overhead when users crossed between the workspace and the AR panel. In summary, we make the following contributions:

- Revealed a UX–performance tradeoff in AR-extended LLM workspaces: AR-Second-Screen produced subjective ratings and preference scores comparable to a physical Dual-Screen setup, while remaining slower, showing that display extension alone is insufficient for productivity parity.
- Identified cross-reality switching cost as a measurable bottleneck in AR-extended LLM authoring: transitions between the physical desktop and AR panel produced higher re-engagement latency and idle time than transitions between two physical monitors, even under a unified keyboard-and-mouse input model.
- Derived design implications for spatial AI interfaces, suggesting that reducing cross-reality re-engagement overhead, rather than maximizing virtual display space, is a primary challenge for AR systems that support high-frequency human-AI collaboration workflows.

2 Related Work

2.1 AR-Extended and Hybrid Display Workspaces

Multiple or large-sized physical displays can improve performance in cognitively demanding tasks by reducing navigation and allowing users to view more information simultaneously [5, 19]. However, added screen space can also create navigational challenges, including difficulty locating cursors, managing windows, and shifting attention across displays [2, 56, 60]. Prior work has shown that physical navigation through head or body movement can mitigate these issues by enabling access to additional screen space through embodied motion rather than explicit window switching [6, 56].

Extending a workspace through AR has been explored previously. Early work on hybrid user interfaces explored augmenting small physical displays with virtual interface elements [24]. More recent work has investigated virtually extended displays around mobile devices and smartphones [35, 54]. Other systems distribute interaction across multiple displays on and around the body [28] or augment mobile interfaces with spatial AR widgets and interactions [13, 71]. These approaches demonstrate how AR can expand interaction space beyond traditional displays and support multitasking workflows that benefit from additional visual context.

Several studies have explored hybrid workspaces that combine physical and virtual displays [34], including AR virtual monitors [56], cross-device interaction systems [59, 78], see-through AR tabletop environments [57], and hybrid reading and analytics interfaces [7, 80]. However, many of these systems treat AR displays as passive extensions of existing monitors. In contrast, our work studies an AR screen as an interactive, task-specific interface for LLM assistance in a desktop authoring workflow. We situate this scenario within a cross-reality framework, where users transition between physical and virtual workspaces during task execution [17, 27, 81].

2.2 AI Assistants in XR Environments

LLMs are increasingly integrated into XR systems because they support natural language interaction, contextual reasoning, and adaptive assistance [10, 14, 47]. Prior work shows that incorporating LLMs into AR systems can reduce task duration and cognitive load while increasing perceived trust in interactive workflows [73]. Consequently, LLM-powered XR systems are increasingly explored across domains such as language learning [31, 41], education [25, 46], accessibility [15, 26, 45, 52], healthcare training [37, 62, 75], and XR content generation [20, 39, 64, 69].

Beyond domain applications, recent work has investigated how LLMs can support interaction in XR through contextual reasoning, personalized feedback, and simulated agents that adapt to user intent [12, 21, 22, 58, 73]. However, researchers have also noted that effectively integrating LLMs into spatial computing environments requires novel interaction paradigms beyond conventional text-based prompting [20, 48]. Several systems, therefore, explore embodied or multimodal approaches, including attention-aware interfaces such as EyeClick [29], embodied conversational agents such as SocialMind [74], and persona-based XR assistants [65]. Only a limited number of studies connect AR, LLMs, and conventional desktop or laptop environments. InputJump [77], for example, uses LLMs to support cross-device input mapping, but does not study an AI assistant that users repeatedly consult while authoring content.

Together, these studies show growing interest in AR workspace extension and LLM-supported XR interaction, but they leave open a specific interaction problem: how users manage repeated bidirectional exchanges between a conventional desktop workspace and an AI assistant placed in AR. This matters because LLM-assisted authoring is not only a display-space problem; it also requires users to transfer context, inspect generated responses, and reintegrate selected content into the primary workspace. By comparing AR-Second-Screen against both Single-Screen and Dual-Screen baselines, our study examines whether AR provides benefits beyond additional display space and quantifies the transition costs introduced by cross-reality exchange.

3 Methodology

The study integrates AR interaction, LLM communication, and desktop authoring through controlled LaTeX writing tasks. We conducted a formative study with an initial prototype to identify usability gaps, such as input-modality confusion and extra steps for transferring content across the AR and desktop interfaces. Based on these findings, we refined the system to reduce cross-reality interaction overheads and then conducted a controlled user study to evaluate the redesigned workflow.

3.1 Tasks

The experimental tasks used LaTeX authoring because it requires repeated alternation between editing, compiling, inspecting output, and seeking external assistance. To ensure participants actively interacted with the LLM during the experiment, each task consisted of three sequential subtasks that required LLM assistance to complete. Learning effects were mitigated by creating one training task and three primary task types: Task-TR, Task-1, Task-2, and Task-3.

Task-TR introduced the system through list-formatting exercises. Participants could repeat the training until they were comfortable with the task and interface. **Task-1** involved text editing and summarization, where participants expanded a sentence, explained key terms, and then condensed the paragraph. **Task-2** focused on figure management in LaTeX, requiring participants to insert, resize, and combine figures. **Task-3** addressed structured formatting by reorganizing variable–value statements, generating a table, and applying formatting changes. Complete task details are provided in the supplemental materials.

3.2 Participants

The study was approved by the Concordia University research ethics board, with the certification number 30023195. All participants provided informed consent before participation and were informed that they could withdraw from the study at any time. We conducted an a priori power analysis using G*Power [23] with $\alpha = 0.05$ and power = 0.90. Assuming a large effect size ($\eta^2 = 0.14$) [76], the analysis indicated a minimum sample of 15 participants for repeated-measures (RM) ANOVA. We recruited 18 participants for the formative study and 18 different participants for the main user study. Using the same sample size across studies kept the two phases comparable while avoiding participant overlap. The order of Screen-Condition was Latin-square counterbalanced to reduce order effects.

3.3 Apparatus

Both studies were conducted on a desktop PC with an Intel® Core™ i7-12700F processor, 16GB RAM, and an NVIDIA RTX 3070 GPU. The system was connected to two 27-inch Acer V277U monitors with a standard keyboard (Dell KB216) and mouse (Dell MS116). For the AR condition, we used a Meta Quest 3 video see-through headset. The AR interface was implemented in Unity (version 6000.0.39f1) using the Meta XR All-In-One SDK (v83.0.0).

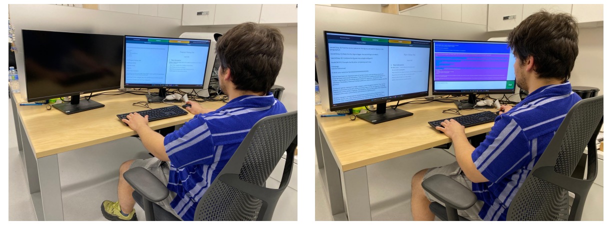
3.4 Experimental Design

We used a within-subjects design with one factor: (Screen-Condition: AR-Second-Screen, Single-Screen, Dual-Screen). In each Screen-Condition, the participants completed three task types (Task-1, Task-2, Task-3), resulting in nine task instances per participant.

3.5 Study 1: Formative Study and Evaluation

3.5.1 Prototype. To study AI-assisted knowledge work in a complex setting, we developed a custom LaTeX authoring environment referred to as the *Workspace*. The Workspace allows users to write LaTeX code, compile the document, and review the generated PDF output. We developed a custom editor rather than using existing tools such as Overleaf or TeXShop to enable detailed logging of user interactions during the experiment and to remove extraneous features that could increase task variability. LaTeX code entered by the user is compiled using the "pdflatex"[1] Python library to generate a PDF preview displayed alongside the code editor.

To provide AI assistance, the Workspace communicates with an LLM through the Google Gemini API [36], using Gemini 3 Flash. The system maintains conversational context so that prompts and responses within the same session remain available to the model. To prevent excessive cognitive load during the study, LLM responses were constrained to short outputs, following prior guidelines [11, 69]. An exception was introduced for LaTeX code generation to ensure that generated code retained proper formatting and line breaks. The exact system prompt sent to the LLM was: "You are an assistant to help the user complete his/her assignment. LaTeX code is being used to create a PDF file. Respond with 6 sentences max, keep it under 400 characters maximum. However, for LaTeX code, do not limit characters and give the LaTeX code properly."



(a) (b)

Figure 2: Experiment setup for (a) Single-Screen and (b) Dual-Screen conditions in formative user study. In Dual-Screen, the Workspace was opened on the left monitor, while the LLM interface was opened on the right monitor.

The AR-Second-Screen was implemented as an AR panel displayed next to the physical monitor and could be manually repositioned through pinching gestures. The panel contained a chat history view, a text input field, a loading indicator, and a transfer button (Figure 3). Users interacted with the panel using head-directed raycasting. When the user looked at the panel, it became active and accepted input from the physical Workspace keyboard. Pressing the Enter key submitted the prompt to the LLM, after which the response was displayed in the panel's chat history.

Users could select LaTeX text in the Workspace before activating the AR panel, which was *automatically* appended to the prompt as context for the LLM when the panel was activated. Users could scroll through previous messages via hand interaction, and a transfer button (see Figure 3) sent the latest LLM output back to the Workspace in a pop-up window for reuse or editing. Together, these components formed the AR-Second-Screen condition, where the Workspace, LLM backend, and AR interface act as a cross-reality system supporting interactive LaTeX authoring.

For the Single-Screen and Dual-Screen conditions, we designed a conventional graphical LLM chat interface that handled prompt entry and response display while communicating with the same backend infrastructure. Unlike in the AR condition, participants were required to copy and paste text content between the LLM and Workspace as in conventional LLM interfaces. In Single-Screen and Dual-Screen conditions, the chat interface was displayed on one or two physical monitors, whereas the AR condition extended the workflow into a cross-reality environment through the AR panel. The formative study experiment setup for Single-Screen and Dual-Screen is shown in Figure 2.

3.5.2 Method. We recruited 18 participants (9 female, 9 male) aged between 19 and 31 (M = 25, SD = 3.14) for the formative study. Participants had varied prior experience with AR, ranging from minimal exposure to frequent use. Approximately half of the participants reported little or no prior LaTeX experience, while the

[1]pdflatex: https://pypi.org/project/pdflatex

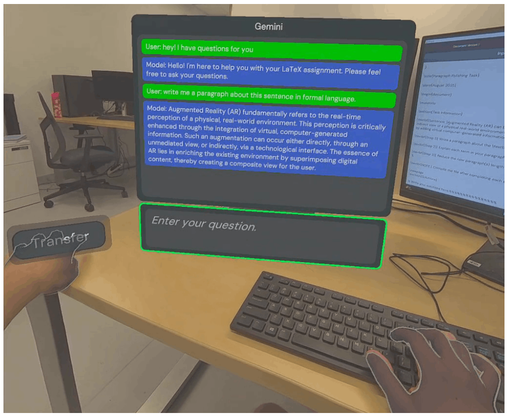


Figure 3: The AR panel design in the formative study. Clicking the “Transfer” button via direct hand interaction copied the most recent LLM response into the Workspace in a pop-up window.

remainder reported moderate to frequent use. In contrast, most participants were familiar with LLM-based tools.

Participants first completed a consent form and a demographic questionnaire covering prior experience with AR, LaTeX, and LLM-based tools. They then completed a training session to familiarize themselves with the AR headset, input devices, Workspace, and LLM interface. In the AR-Second-Screen condition, participants positioned the AR panel beside the physical monitor. In the Single-Screen condition, both the Workspace and LLM chat interface were shown on a single monitor. In the Dual-Screen condition, the Workspace was displayed on the left monitor, and the LLM chat interface on the right. After completing tasks for each Screen-Condition, participants completed the System Usability Scale (SUS) [42], NASA-TLX [32], and the User Experience Questionnaire (UEQ-S) [33]. The formative study lasted approximately 60 minutes.

3.5.3 Findings. The complete results of the formative study are given in the supplemental material. From the formative study, we obtained the following findings that informed the next iteration of the system and the main study.

Cross-Reality Transition Overhead: We found that participants were significantly slower with AR-Second-Screen compared to Single-Screen ($p < 0.05$) and Dual-Screen ($p < 0.05$). Additionally, the number of cursor changes between LLM and Workspace was also significantly higher with AR-Second-Screen compared to Single-Screen ($p < 0.001$) and Dual-Screen ($p < 0.001$). These results indicated that the initial prototype introduced transition overhead through frequent switches between the Workspace and LLM interface. We therefore replaced the head rotation-based transition mechanism with a mouse position-based mechanism to reduce unintentional switches.

Input Modality Mismatch: The initial prototype combined head-directed activation, hand interaction in AR, and keyboard-and-mouse interaction in the desktop Workspace. Participants needed to use their hands for the Transfer button and to scroll the AR panel, a mix of body movement and mouse for giving context, and only keyboard and mouse for Workspace interaction. Qualitative feedback showed confusion about interaction with the system. Some participants tried to click the Transfer button with the mouse instead of their hand. We also observed unintentional keyboard forwarding switches. Therefore, we redesigned the interaction for the user study, focusing on the keyboard and the mouse as the main input modalities for system interaction.

Explicit Content Transfer: We also identified issues in the content transfer in both directions between the AR interface and the desktop Workspace. To use an LLM response generated in AR, users had to click the Transfer button, which sent the response to the Workspace. There, they could edit or copy the response from a pop-up before using it in their document. Conversely, users had to select a portion of text on the desktop, then turn their head toward the AR panel to copy it as context for the LLM. These processes introduced extra steps for transferring content, which may have contributed to longer task completion times. Therefore, we redesigned the content transfer workflow so that participants could directly select content from the AR chat history for later use in the workspace. We also simplified context selection by replacing head rotation with button clicks.

The initial prototype revealed the costs of cross-reality interaction, but participants still rated AR-supported LLM assistance positively. Therefore, we used the findings from the formative study to redesign the system before evaluating it in the main study.

3.6 Study 2: Redesign and Controlled Evaluation

The formative study revealed several limitations that motivated the system redesign. Although participants appreciated the concept of AR-extended authoring, the initial design introduced cross-reality interaction friction that reduced efficiency compared to the Dual-Screen. In particular, users had to complete additional steps to transfer LLM output, switch between mixed input modalities, and repeatedly re-engage with the Workspace and the AR panel. Based on these findings, we redesigned the system around a unified keyboard and mouse input model. This removed the AR hand-based transfer interaction to make interaction more seamless, reduce input modality confusion, and streamline content transfer.

3.6.1 System Design Revisions. In the formative study, the system exhibited GUI limitations, including restricted text input fields, limited multi-window support, and a lack of PDF rendering. To address these issues, we transitioned to a web-based interface while keeping the overall LaTeX editor/compiler layout similar. The compiled output is now rendered as a live PDF within the Workspace, similar to commercial LaTeX editors. This transition also simplified the architecture (see Figure 1(b)) by reducing the system to two main applications: the Workspace and the AR interface.

Considering the lower usability and higher workload associated with the head-rotation-based mode switch [9] in the formative study, as well as the input modality confusion observed among participants, we redesigned the interaction model. Instead of relying on separate input modalities for the two realities, users interacted with the AR LLM panel using the same keyboard and mouse as the desktop workspace [70]. To support this, we introduced a keyboard–mouse forwarding mechanism. An orange bar placed between the Workspace and AR panel allowed users to switch the focus of the input by moving the cursor into this region, enabling typing, scrolling, and clicking directly within the AR interface. This technique was designed to provide a consistent way to switch input between Workspace and AR without changing the physical input device. The orange ‘’switching” bar in AR is shown in Figure 4.

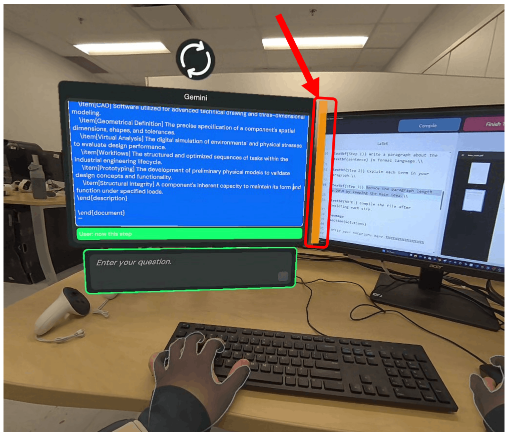
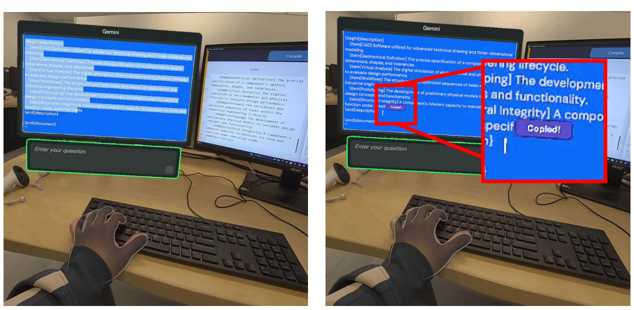


Figure 4: In the refined system design, switching between Workspace and AR was handled by simply moving the mouse cursor to predefined orange areas, keeping the input modality consistent through the experiment.

With this setup, users could write prompts and submit them to the LLM using the keyboard (Figure 1(a)). LLM responses could be selected in AR using familiar desktop interactions, i.e., click–drag–release, as shown in Figure 5(a). When the user released the cursor, the selected text was automatically copied to the clipboard (Figure 5(b)) and could be pasted into the Workspace. This mechanism reduced the steps required to use LLM output by allowing users to select only the relevant portion of a response.



(a) (b)

Figure 5: (a) User selects text in the chat history by holding and moving the mouse cursor in AR. (b) When the user releases the cursor, the system automatically copies the selected text into the clipboard.

To provide context to the LLM, users could highlight LaTeX code in the Workspace and click an “Append” button to add the selected text to the AR chat interface (Figure 6). We also updated the navigation in the AR panel by replacing gesture-based scrolling with

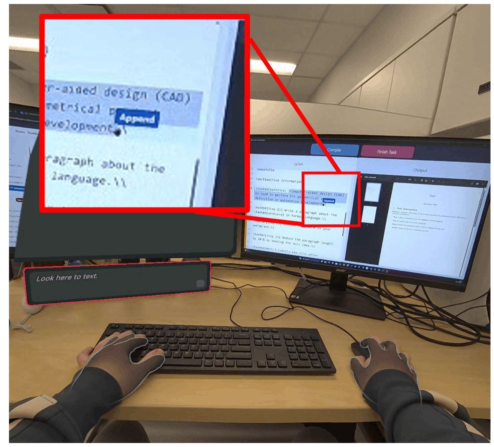


Figure 6: User gives context to the LLM in the AR-Second-Screen condition. After selecting a text in the Workspace, the user clicks the “Append” button to use the selected text as context.

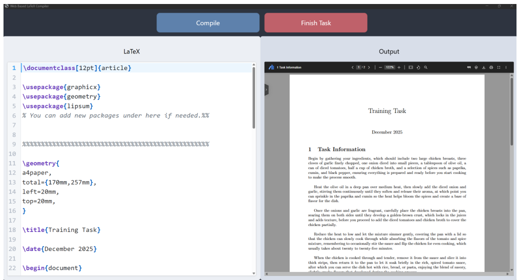
mouse-wheel scrolling, enabling easier review of longer responses with a well-known action and affording consistent behavior between AR and Workspace. Together, these mechanisms formed the revised AR-Second-Screen condition, designed to support smoother transitions between the Workspace and the AR LLM panel. Additionally, the updated web-based GUI enabled the same text-editing capabilities in the Single-Screen and Dual-Screen conditions using standard editing tools. The final version of the Workspace and LLM pages is shown in Figure 7(a-b). The source code of the system is publicly available under the MIT License [2].

Task & Experimental Design: We followed the same task structure and within-subjects experiment design as in the formative user study.

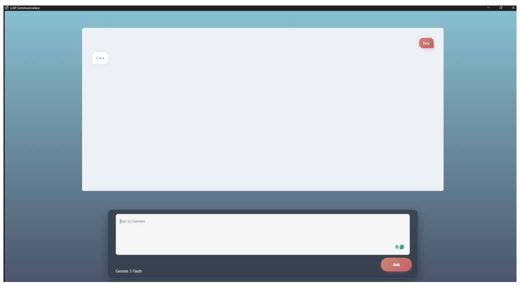
3.6.2 Participants. We recruited 18 different participants (8 female, 10 male) aged 24–30 (M = 26.17, SD = 1.72), matching the sample size used in the formative study. Eight participants had corrected-to-normal vision, and the rest had normal vision. Eight had used AR ten or more times, two had used it four to eight times, and eight had used it fewer than four times. Regarding LaTeX, nine participants used LaTeX weekly or more often, five monthly, three a few times per year, and one had heard of it but never used it. Seventeen participants used LLM tools weekly or more often, and one monthly.

3.6.3 Procedure. The procedure followed in the second study was consistent with the formative study structure, with the full UEQ administered after each condition. In the AR-Second-Screen condition, the AR panel was positioned in front of the left monitor, adjacent to the monitor displaying the Workspace. This placement was chosen to approximate a familiar dual-monitor layout, with the AR panel acting as a virtual second screen. After all conditions, they also completed a post-experiment feedback questionnaire, including preference ranking and overall feedback on receiving LLM assistance in AR. The experiment lasted approximately 60 minutes.

[2]https://github.com/mucahitGemici/Crossing-Realities-Has-a-Cost



(a)



(b)

Figure 7: The desktop GUI for the (a) Workspace and (b) LLM chat.

3.6.4 Evaluation Metrics. We gathered performance, transition, workload, usability, and user experience measures. Task time was used as the primary performance metric and defined as the total time in seconds required to complete each authoring task. Because our main research question concerns the cost of moving between the desktop workspace and the AR LLM panel, we also computed three transition-level measures from the interaction logs: Switching Cost, Task Idle Ratio, and LLM Idle Ratio.

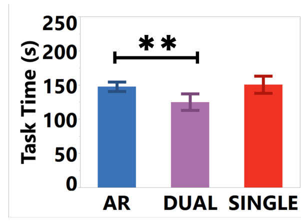
We defined a switch as a change in active interaction context between the Workspace and the LLM interface. In the AR-Second-Screen condition, this corresponded to moving the interaction between the physical desktop workspace and the AR panel. In the Single-Screen and Dual-Screen conditions, it corresponded to moving interaction between the LaTeX Workspace and the desktop LLM interface. Switching Cost was computed as the latency from a detected switch to the first subsequent user action in the new context, such as typing, clicking, scrolling, or text selection. This metric captures the re-engagement time required before users resumed productive interaction after moving between the interfaces.

We also computed idle-time measures to capture periods where users were not actively interacting with the system for more than 2 seconds. We used this threshold because pauses longer than 2 seconds have been interpreted as indicators of higher-order cognitive processes, such as planning or revising, in writing-process research [43]. The threshold was implemented in the logging system before data collection. Task Idle Ratio was defined as the proportion of total task time during which no user input was detected. LLM Idle Ratio was defined as the proportion of time spent in the LLM interface during which no active input was detected. These metrics were included to capture transition-related pauses that may not be visible from task completion time alone, especially when users had to reorient attention, redirect input, or decide how to continue after moving between interaction contexts.

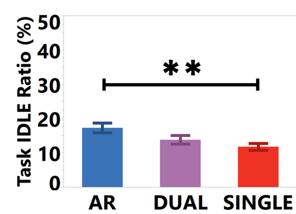
After each condition, participants completed the NASA-TLX [32], SUS [42], and UEQ [40] to assess perceived workload, usability, and user experience, respectively. In this, we used the full UEQ with six scales (instead of the UEQ-S as in the formative study). Attractiveness reflects overall valence, Perspicuity, Efficiency, and Dependability are pragmatic (goal-directed), while Stimulation and Novelty are hedonic (not goal-directed) quality aspects [40].

3.6.5 Results. The data were analyzed using SPSS 29. Normality was assumed for Skewness (S) and Kurtosis (K) within ±1 [30, 49]; otherwise, data were log-transformed. If the data still failed normality assumptions, we used the Aligned Rank Transform (ART) method [72]. One-Way Repeated Measures ANOVA with Bonferroni-corrected post hoc tests was used to analyze the effect of Screen-Condition. Questionnaire data were analyzed using the Friedman test with Wilcoxon Signed-Rank post hoc comparisons. Statistical significance is annotated with (*) (p<0.05), (**) (p<0.01), and (***) (p<0.001). Task Idle Ratio ($S = 0.68, K = 0.22$) and LLM Idle Ratio ($S = 0.6, K = -0.25$) were normally distributed, while Task Time and Switching Cost were normally distributed after log transformation, with ($S = 0.51, K = 0.34$) and ($S = -0.64, K = -0.26$) after the transformation.

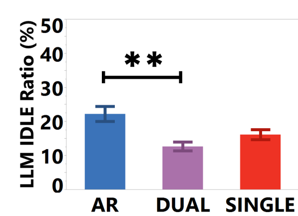
Task Time: Screen-Condition had a significant effect on Task Time ($F(1.493, 25.378) = 5.934$, $p < 0.05$, $\eta^2 = 0.259$). Post-hoc analysis showed that participants were faster with Dual-Screen than AR-Second-Screen ($\Delta M = 0.266$, $SE = 0.062$, $p < 0.01$), as shown in Figure 8(a). We also logged LLM response delay; mean delays were comparable across conditions (AR-Second-Screen: 7.6s, Dual-Screen: 8.0s, Single-Screen: 6.9s).



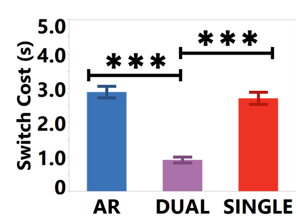
(a)



(b)



(c)



(d)

Figure 8: Graphs illustrating the data for (a) Task Time, (b) Task Idle Ratio, (c) LLM Idle Ratio, and (d) Switching Cost metrics.

Task Idle Ratio: A significant effect of Screen-Condition was found for Task Idle Ratio ($F(2, 34) = 6.857$, $p < 0.01$, $\eta^2 = 0.287$). The ratio was higher in AR-Second-Screen than Single-Screen ($M = 5.976$, $SE = 1.607$, $p < 0.01$), as shown in Figure 8(b).

LLM Idle Ratio: Screen-Condition significantly affected LLM Idle Ratio ($F(1.523, 25.884) = 8.793$, $p < 0.01$, $\eta^2 = 0.341$), which was higher in AR-Second-Screen than Dual-Screen ($M = 10.087$, $SE = 2.779$, $p < 0.01$), as shown in Figure 8(c).

Switching Cost: A significant main effect of Screen-Condition was observed for Switching Cost ($F(2, 34) = 133.366$, $p < 0.001$, $\eta^2 = 0.887$). Switching Cost was lower in Dual-Screen than in AR-Second-Screen ($M = 1.276$, $SE = 0.087$, $p < 0.001$) and Single-Screen ($M = 1.128$, $SE = 0.104$, $p < 0.001$), as shown in Figure 8(d).

NASA TLX: Significant effects of Screen-Condition were found for Mental Demand ($\chi^2(2, 18) = 9.926$, $p < 0.01$), Physical Demand ($\chi^2(2, 18) = 6.698$, $p < 0.05$), Performance ($\chi^2(2, 18) = 7.277$, $p < 0.05$), Effort ($\chi^2(2, 18) = 19.231$, $p < 0.001$), Frustration ($\chi^2(2, 18) = 14.358$, $p < 0.001$), and Overall Score ($\chi^2(2, 18) = 18.299$, $p < 0.001$). Wilcoxon post-hoc comparisons are shown in Figure 9(a).

SUS: A significant effect was found for the overall SUS score ($\chi^2(2, 18) = 16.676$, $p < 0.001$). Post-hoc analysis showed that Dual-Screen achieved higher SUS scores than Single-Screen ($Z = -3.628$, $p < 0.001$) and AR-Second-Screen ($Z = -2.641$, $p < 0.01$). Mean SUS scores were 87.92 (*A*, *Excellent*) for Dual-Screen, 75.56 (*B*, *Good*) for AR-Second-Screen, and 72 (*B*, *Good*) for Single-Screen.

UEQ: Significant differences were observed across all UEQ scales: Attractiveness ($\chi^2(2, 18) = 20.029$, $p < 0.001$), Perspicuity ($\chi^2(2, 18) = 10.478$, $p < 0.01$), Efficiency ($\chi^2(2, 18) = 13.848$, $p < 0.001$), Dependability ($\chi^2(2, 18) = 10.125$, $p < 0.01$), Stimulation ($\chi^2(2, 18) = 23.771$, $p < 0.001$), and Novelty ($\chi^2(2, 18) = 30.789$, $p < 0.001$). Benchmark results [61] are shown in Table 1.

Table 1: The evaluation of the Screen-Condition UEQ scales with the benchmark.

	AR	**DUAL**	**Single-Screen**
Attractiveness	*Good*	*Above average*	*Bad*
Perspicuity	*Good*	*Excellent*	*Above average*
Efficiency	*Good*	*Good*	*Bad*
Dependability	*Above average*	*Good*	*Below average*
Stimulation	*Excellent*	*Below average*	*Bad*
Novelty	*Excellent*	*Bad*	*Bad*

3.6.6 Post-Experiment Feedback Results. Participants rated how much they liked receiving LLM assistance in AR on a 1–5 scale (1 = Not at all, 5 = Very much), resulting in an average rating of 4, indicating generally positive perceptions. Participants also ranked the Screen-Condition by preference. Dual-Screen was most preferred, with 50% (9/18) ranking it first and 44% (8/18) second. AR-Second-Screen followed closely (8 first-place and 6 second-place votes), while Single-Screen was least preferred, with 72% (13/18) ranking it last. A Friedman test revealed a significant difference in rankings ($\chi^2(2, 18) = 14.333$, $p < 0.001$). Post-hoc Wilcoxon tests showed that both AR-Second-Screen ($Z = -2.851$, $p < 0.01$) and Dual-Screen ($Z = -3.176$, $p = 0.01$) were preferred significantly more than Single-Screen, with no significant difference between AR-Second-Screen and Dual-Screen.

4 Discussion

We investigated whether an AR second screen can support AI-assisted desktop authoring. Across a formative and controlled user study, AR-Second-Screen was subjectively appealing but slower than Dual-Screen. Transition-level measures suggest that this performance gap reflected both interface design and the cost of shifting attention, input focus, and content across the physical-virtual boundary. The key challenge for AR-extended AI workspaces is therefore not just adding display space but enabling efficient transitions across realities.

4.1 Task Performance and User Experience in AR-Extended AI Workflows (RQ1)

AR-Second-Screen did not match the efficiency of Dual-Screen. Participants completed tasks faster with Dual-Screen, and this advantage persisted even after redesigning the AR system around a unified input model. Although we did not analyze Task as a formal factor, AR-Second-Screen was descriptively slower than Dual-Screen across all three tasks: 38.8%, 33.5%, and 43.8% in Task-1, Task-2, and Task-3, respectively. This pattern suggests that the performance gap was not confined to a single task type. In contrast, AR-Second-Screen performed better subjectively: participants preferred it significantly more than Single-Screen, with no significant preference difference from Dual-Screen, and AR received stronger ratings on the Stimulation and Novelty scales.

These results reveal a UX–performance tradeoff. Although AR-Second-Screen and Dual-Screen share a similar high-level workspace structure, they are not equivalent display surfaces. A physical second monitor extends the workspace while preserving visual continuity, input context, and interaction state [18, 60]. By contrast, an AR panel adds a different perceptual layer and an additional boundary between interaction contexts. In LLM-assisted authoring, where users switch frequently, this boundary-related cost appears to be a key factor separating AR-Second-Screen from Dual-Screen, beyond the amount of display space available.

This distinction matters for evaluating and deploying AR productivity systems. AR displays may still be valuable when physical displays are unavailable, when portability matters, or when auxiliary information should remain visible only to the headset wearer. However, they should not be treated as exact replacements for physical monitors in complex authoring workflows without accounting for cross-reality interaction costs [56].

4.2 Cross-Reality Transition Cost as a Structural Bottleneck (RQ2)

To better understand why AR-Second-Screen was slower than Dual-Screen, we examined transition-level measures. Switching cost was significantly higher in AR-Second-Screen than in Dual-Screen, and Task and LLM idle ratios were also larger, indicating that users needed more time to resume productive activity after moving between the physical workspace and the AR panel. These

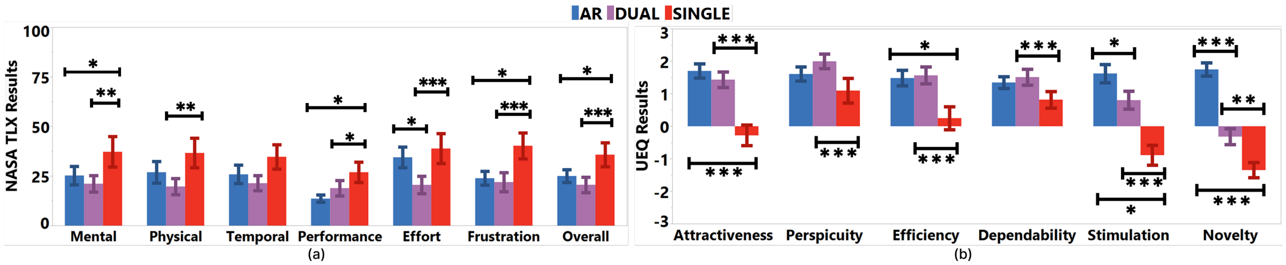


Figure 9: (a) NASA TLX results with statistically significant pairwise comparisons. We found significant differences for Mental Demand, Physical Demand, Performance, Effort, Frustration, and Overall scores. (b) UEQ Results with pairwise comparisons.

measures show that the performance gap was concentrated at moments of cross-reality transition rather than distributed evenly across the task. This finding aligns with recent cross-reality work identifying transitions across the reality–virtuality continuum and the evaluation of such switches as central open challenges [3].

This pattern is also consistent with cognitive switching and interruption research on re-engagement cost: the time required to reorient attention, reactivate task context, and resume productive behavior after an interruption or environment shift [1, 38, 50, 51]. Compared with dual-monitor switching, which preserves the same display medium and input context, switching in AR-Second-Screen additionally required users to recalibrate visual attention across physical and virtual display surfaces, redirect keyboard-and-mouse input through the forwarding mechanism, and resume the task context in a spatially distinct environment. Even with a unified keyboard-and-mouse model, these factors likely contributed to the higher switching costs observed in AR-Second-Screen. The spatial placement of the AR panel may also have shaped these costs. Our side-mounted placement approximated a familiar dual-monitor layout, but alternative placements could change the balance between visual travel, occlusion, and posture-related costs. For example, placing the panel closer to the Workspace might reduce visual travel but occlude the document, hands, or keyboard, while above-monitor placement may preserve workspace visibility but introduce vertical gaze shifts. Because the switching cost we observed reflects re-engagement across the cross-reality boundary itself, we would expect alternative placements to change the magnitude of this cost rather than remove it. Since we did not systematically vary panel placement, future work should examine how virtual-screen layout affects switching cost and re-engagement latency.

Importantly, these costs were not eliminated by interaction redesign. The formative study identified mixed input modalities and explicit content-transfer steps as sources of friction; removing them narrowed the performance gap between AR-Second-Screen and Single-Screen. However, Dual-Screen remained faster, suggesting that some transition overhead may arise from the cross-reality boundary itself, not only from the initial interface implementation. Switching cost and idle ratio measures were therefore useful for isolating where time was lost, and may support future evaluations of cross-reality and hybrid display systems.

4.3 AR May Be Better Suited for Peripheral AI Awareness Than High-Frequency Content Transfer (RQ3)

Our results suggest that the type of AI-assisted workflow matters as much as the display configuration. In our task, users repeatedly selected context from the document, submitted it to the LLM, reviewed the response, extracted relevant content, and reinserted it into the workspace. This made the AR panel an active, high-frequency interaction target rather than a passive display. As a result, the cost of entering, operating within, and leaving the AR panel was a key performance indicator. When the AR interface replicates the same dense text manipulation and content-transfer operations already well supported by physical monitors and desktop tools, users may absorb the cross-reality transition cost without receiving enough benefit to offset it.

This suggests a design distinction with broader implications: AR may be less effective when it merely mirrors desktop interaction patterns, and more effective when it provides a qualitatively different kind of support. An AR assistant that offers persistent peripheral awareness, such as high-level document summaries, citation cues, structural overviews, conversational history, or ambient reminders, may reduce the need for users to shift input focus or transfer text across realities frequently. Instead, users could glance at the AR screen, extract useful information, and return to the primary workspace with fewer full cross-reality re-engagement cycles. Although we did not test such an awareness interface, this interpretation is consistent with our subjective findings: participants rated AR positively and showed high stimulation and novelty scores, even though it was slower. The appeal of AR may reflect users' intuition that spatially placed AI support is valuable, even when the current interaction design does not yet realize that value in a manner that equals dual display setups.

The design opportunity, therefore, is not simply to place a chat window in AR. This view is consistent with desktop–AR and cross-reality design work that treats AR and desktop environments as complementary interaction spaces rather than direct substitutes [16]. Instead, the goal is to identify which components of an AI-assisted workflow are appropriate for spatial, peripheral presentation, and which ones should remain tightly integrated with the desktop workspace.

4.4 Design Implications for Future Cross-Reality AI Workspaces (RQ3)

Building on these findings, we propose four design implications for future AR-extended AI productivity systems.

First, reduce explicit input-focus transfer across realities. Even the improved keyboard-and-mouse forwarding mechanism required users to move the cursor to a designated region to redirect input. This step was small but repeated at every transition. Future systems should explore implicit input methods, including gaze-aware activation, cursor-proximity prediction, or context-sensitive forwarding, all of which could potentially reduce the impact of the explicit mode-switching step in our cross-reality workflow.

Second, keep high-frequency content transfer within the primary workspace. LLM-assisted authoring requires repeated bidirectional movement of text between the document and the assistant; transition costs accumulate when this exchange crosses a reality boundary. Future systems could reduce this cost through in-place suggestions, inline generation, or context-injection shortcuts, reserving the AR panel for responses that benefit from spatial separation.

Third, design AR-AI interfaces for awareness and persistence, not the display area. Our results show that the additional spatial display area did not make AR more efficient than Dual-Screen. AR's subjective value may instead come from keeping information visible, spatially organized, and privately available without occupying the primary screen. Designers should identify which AI outputs, such as overviews, alternatives, references, and structural cues, are suited for persistent spatial display, rather than routing all LLM interaction through the AR panel.

Fourth, include transition-level measures in evaluations of cross-reality productivity systems. Task completion time, workload, and usability scales capture overall outcomes, but switching cost and idle ratios reveal where time is lost at cross-reality boundaries. Future evaluations of AR workspaces, hybrid displays, and spatial AI interfaces should therefore measure transition costs directly.

Together, these implications reframe the overall design goal for AR productivity support systems with LLMs from adding virtual screen space to reducing the cost of moving across realities. AR can make AI assistance more visible, spatially organized, and subjectively appealing, but these benefits can translate into efficient workflows only if systems minimize the attention, input, and content-transfer costs introduced at the cross-reality boundary.

5 Limitations and Future Work

In this study, we used sequential multi-step LaTeX authoring as our controlled instance of structured AI-assisted authoring. This choice allowed us to study a workflow with repeated transitions between writing, verification, LLM consultation, and content integration while maintaining experimental control; however, we acknowledge that our task is not as nonlinear as a real document-writing workflow. The findings therefore may generalize more directly to workflows with a similar interaction structure, where users frequently move task context and generated content between a primary workspace and an assistant. Similar transition costs may also appear in other structured workflows involving iterative editing, checking, and repair, such as programming, technical writing, and report preparation. At the same time, LaTeX authoring is more structured and syntax driven than many forms of writing, so less structured activities, such as long-form writing, open-ended ideation, other authoring systems, or use by non-technical users, may produce different transition patterns and should be examined in future studies. The study also used a constrained LLM response format: responses were limited to 6 sentences and fewer than 400 characters, with an exception for LaTeX code. While we followed previous work in this choice, we recommend future work to support larger or adaptable AR screen sizes for long LLM responses instead of limiting the response size, as this choice can result in low LLM performance [67]. Although Meta Quest 3 provided usable video see-through performance for our study, future work should investigate AR LLM use with smart glasses as this technology continues to improve [79]. Although we did not encounter readability issues because we used large fonts, video-see-through techniques may still introduce readability challenges. Future work should therefore consider AR glasses for text-based cross-reality applications, or optical-see-through devices when a lower field of view is acceptable. Our AR panel intentionally used desktop-style text selection and keyboard-and-mouse input. Future AR-AI interfaces that rely on richer spatial selection or object manipulation should consider task-environment factors such as target distance, target size, density, occlusion, depth, and keyboard-based selection contexts [8]. More broadly, our findings suggest that future systems should minimize the switching cost and re-engagement latency, especially when users must frequently move attention, input focus, or generated content between workspaces. This recommendation also applies to emerging IDE-integrated and agentic AI authoring workflows, where the assistant can directly modify files within the primary editor. In such settings, AR second screens may be more appropriate for peripheral awareness of agent plans, progress, generated alternatives, or test results, while high-frequency interactions, such as prompting, editing, and accept/reject actions, remain in the desktop workspace. Although our sample size of eighteen was based on a priori power analysis, we acknowledge that it is still relatively small and findings should be generalized carefully. Participants' prior experience, particularly with LaTeX, also varied; while each participant completed all conditions, this variation may limit generalizability across experience levels, which we did not analyze separately in this study.

6 Conclusion

In this paper, we investigated AR-Second-Screen as a workspace extension for LLM-assisted desktop authoring, focusing on workflows in which users repeatedly move task context and generated content between a primary workspace and an LLM assistant. Our results indicate that the value of AR-extended AI workspaces cannot be evaluated through display space alone. Although AR-Second-Screen was subjectively appealing, it did not match the efficiency of Dual-Screen when the workflow required frequent bidirectional content transfer. Transition-level measures showed that this performance gap was associated with re-engagement costs when users moved attention, input focus, and content across the physical-virtual boundary. These findings suggest that AR-Second-Screen is not a direct replacement for physical monitors in high-frequency authoring workflows. However, its subjective appeal may make it a promising alternative when physical displays are unavailable or impractical. Ultimately, our results motivate a shift in design priorities for AR-extended AI workspaces: away from merely expanding virtual display areas and toward minimizing the cost of transitioning attention, input focus, and content across realities.

References

[1] Erik M Altmann and J Gregory Trafton. 2002. Memory for goals: An activation-based model. *Cognitive science* 26, 1 (2002), 39–83.

[2] Christopher Andrews, Alex Endert, and Chris North. 2010. Space to Think: Large High-Resolution Displays for Sensemaking. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 55–64.

[3] Christoph Anthes, Mark Billinghurst, Uwe Grünefeld, Hans-Christian Jetter, Hai-Ning Liang, Frank Maurer, David Aigner, Craig Anslow, Guillaume Bataille, Abraham G Campbell, et al. 2026. Grand Challenges in Cross Reality. *IEEE Transactions on Visualization and Computer Graphics* (2026).

[4] Fabio Arena, Mario Collotta, Giovanni Pau, and Francesco Termine. 2022. An overview of augmented reality. *Computers* 11, 2 (2022), 28.

[5] Robert Ball and Chris North. 2005. Effects of tiled high-resolution display on basic visualization and navigation tasks. In *CHI'05 extended abstracts on Human factors in computing systems*. 1196–1199.

[6] Robert Ball, Chris North, and Doug A Bowman. 2007. Move to improve: promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 191–200.

[7] Sunyoung Bang, Hyunjin Lee, Seo Young Oh, and Woontack Woo. 2025. AReading with Smartphones: Understanding the Trade-offs between Enhanced Legibility and Display Switching Costs in Hybrid AR Interfaces. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[8] Mohammad Raihanul Bashar and Anil Ufuk Batmaz. 2024. Virtual task environments factors explored in 3d selection studies. In *Proceedings of the 50th Graphics Interface Conference*. 1–16.

[9] Mohammad Raihanul Bashar, Aunnoy K Mutasim, Ken Pfeuffer, and Anil Ufuk Batmaz. 2026. Eyes on Many: Evaluating Gaze, Hand, and Voice for Multi-Object Selection in Extended Reality. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (Barcelona, Spain) *(CHI '26)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3772318.3790513

[10] Jair Augusto Bottega, Victor Augusto Kich, Junior Costa de Jesus, Raul Steinmetz, Alisson Henrique Kolling, Ricardo Bedin Grando, Rodrigo da Silva Guerra, and Daniel Fernando Tello Gamarra. 2023. Jubileo: An Immersive Simulation Framework for Social Robot Design. *Journal of Intelligent & Robotic Systems* 109, 4 (2023), 91.

[11] Riccardo Bovo, Steven Abreu, Karan Ahuja, Eric J Gonzalez, Li-Te Cheng, and Mar Gonzalez-Franco. 2025. Embardiment: An Embodied AI Agent for Productivity in XR. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 708–717.

[12] Efe Bozkir, Süleyman Özdel, Ka Hei Carrie Lau, Mengdi Wang, Hong Gao, and Enkelejda Kasneci. 2024. Embedding Large Language Models into Extended Reality: Opportunities and Challenges for Inclusion, Engagement, and Privacy. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*. 1–7.

[13] Eugenie Brasier, Emmanuel Pietriga, and Caroline Appert. 2021. AR-enhanced Widgets for Smartphone-centric Interaction. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–12.

[14] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2024. A Survey on Evaluation of Large Language Models. *ACM transactions on intelligent systems and technology* 15, 3 (2024), 1–45.

[15] Chi-Hui Chiang and Hsin-Yu Chiang. 2023. Enhancing Elderly Leisure Experience Through Innovative VTuber Interaction in VR With ChatGPT. In *International Conference on Artificial Intelligence and Virtual Reality*. Springer, 163–178.

[16] Robbe Cools, Matt Gottsacker, Adalberto Simeone, Gerd Bruder, Greg Welch, and Steven Feiner. 2022. Towards a desktop-AR prototyping framework: Prototyping cross-reality between desktops and augmented reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 175–182.

[17] Robbe Cools, Jihae Han, Augusto Esteves, and Adalberto L Simeone. 2025. From Display to Interaction: Design Patterns for Cross-Reality Systems. *IEEE Transactions on Visualization and Computer Graphics* (2025).

[18] Mary Czerwinski, George Robertson, Brian Meyers, Greg Smith, Daniel Robbins, and Desney Tan. 2006. Large display research overview. In *CHI'06 extended abstracts on Human factors in computing systems*. 69–74.

[19] Mary Czerwinski, Greg Smith, Tim Regan, Brian Meyers, George G Robertson, and Gary K Starkweather. 2003. Toward Characterizing the Productivity Benefits of Very Large Displays.. In *Interact*, Vol. 3. 9–16.

[20] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. 2024. LLMR: Real-time Prompting of Interactive Worlds Using Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.

[21] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented object intelligence with xr-objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[22] Xiaonuo Dongye, Dongdong Weng, Haiyan Jiang, and Pukun Chen. 2024. Learning personalized agent for real-time face-to-face interaction in vr. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 759–760.

[23] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[24] Steven Feiner and Ari Shamash. 1991. Hybrid user interfaces: Breeding virtually bigger interfaces for physically smaller computers. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*. 9–17.

[25] Marco Fiore, Michele Gattullo, and Marina Mongiello. 2024. First steps in constructing an ai-powered digital twin teacher: Harnessing large language models in a metaverse classroom. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 939–940.

[26] Brittany Garcia-Pi, Rohan Chaudhury, Miles Versaw, Jonathan Back, Dongjin Kwon, Caleb Kicklighter, Paul Taele, and Jinsil Hwaryoung Seo. 2023. Ally Chat: Developing a VR Conversational AI Agent Using Few-Shot Learning to Support Individuals with Intellectual Disabilities. In *IFIP Conference on Human-Computer Interaction*. Springer, 402–407.

[27] Mucahit Gemici, Kyle Johnsen, Mayra Donaji Barrera Machuca, and Anil Ufuk Batmaz. 2026. Modeling Across Realities (MAR): Extending Desktop 3D Modeling with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* (2026). To appear.

[28] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. 2015. Multifi: Multi fidelity interaction with displays on and around the body. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3933–3942.

[29] Jiajing Guo, Andrew Benton, Nan Tian, William Ma, Nicholas Feffer, Zhengyu Zhou, and Liu Ren. 2023. EyeClick: A Robust Two-Step Eye-Hand Interaction for Text Entry in Augmented Reality Glasses. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–4.

[30] Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, Ronald L Tatham, et al. 2019. Multivariate Data Analysis.

[31] Shirin Hajahmadi, Luca Clementi, María Dolores Jiménez López, and Gustavo Marfia. 2024. Arele-bot: Inclusive Learning of Spanish as a Foreign Language Through a Mobile app Integrating Augmented Reality and CHATGPT. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 335–340.

[32] Sandra G Hart. 2006. NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[33] Andreas Hinderks. 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* (2017).

[34] Sebastian Hubenschmid, Marc Satkowski, Johannes Zagermann, Julián Méndez, Niklas Elmqvist, Steven Feiner, Tiare Feuchtner, Jens Emil Grønbæk, Benjamin Lee, Dieter Schmalstieg, et al. 2026. Hybrid User Interfaces: Past, Present, and Future of Complementary Cross-Device Interaction in Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* (2026).

[35] Sebastian Hubenschmid, Johannes Zagermann, Daniel Leicht, Harald Reiterer, and Tiare Feuchtner. 2023. Around the smartphone: Investigating the effects of virtually-extended display size on spatial memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[36] Raisa Islam and Imtiaz Ahmed. 2024. Gemini-the most powerful LLM: Myth or Truth. In *2024 5th Information Communication Technologies Conference (ICTC)*. IEEE, 303–308.

[37] Nimit Kapadia, Shreekant Gokhale, Anthony Nepomuceno, Wanning Cheng, Samantha Bothwell, Maureen Mathews, John S Shallat, Celeste Schultz, and Avinash Gupta. 2024. Evaluation of Large Language Model Generated Dialogues for an AI based VR Nurse Training Simulator. In *International Conference on Human-Computer Interaction*. Springer, 200–212.

[38] Iring Koch and Andrea Kiesel. 2022. Task switching: Cognitive control in sequential multitasking. *Handbook of human multitasking* (2022), 85–143.

[39] Ryutaro Kurai, Takefumi Hiraki, Yuichi Hiroi, Yutaro Hirao, Monica Perusquia-Hernandez, Hideaki Uchiyama, and Kiyoshi Kiyokawa. 2024. Magicitem: Dynamic behavior design of virtual objects with large language models in a consumer metaverse platform. *arXiv preprint arXiv:2406.13242* (2024).

[40] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76.

[41] Hyungmin Lee, Chen-Chun Hsia, Aleksandr Tsoy, Sungmin Choi, Hanchao Hou, and Shiguang Ni. 2023. VisionARy: Exploratory Research on Contextual Language Learning using AR Glasses with ChatGPT. In *Proceedings of the 15th biannual conference of the Italian SIGCHI chapter*. 1–6.

[42] James R Lewis. 2018. The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590.

[43] Teresa Limpo and Rui A Alves. 2017. Written language bursts mediate the relationship between transcription skills and writing performance. *Written Communication* 34, 3 (2017), 306–332.

[44] Richen Liu, Min Gao, Lijun Wang, Xiaohan Wang, Yuzhe Xiang, Aolin Zhang, Jiazhi Xia, Yi Chen, and Siming Chen. 2022. Interactive extended reality techniques in information visualization. *IEEE Transactions on Human-Machine Systems* 52, 6 (2022), 1338–1351.

[45] Xingyu Bruce Liu, Jiahao Nick Li, David Kim, Xiang'Anthony' Chen, and Ruofei Du. 2024. Human I/O: Towards a Unified Approach to Detecting Situational Impairments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[46] Ziyi Liu, Zhengzhe Zhu, Lijun Zhu, Enze Jiang, Xiyun Hu, Kylie A Peppler, and Karthik Ramani. 2024. Classmeta: Designing Interactive Virtual Classmate to Promote VR Classroom Participation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[47] Jose Llanes-Jurado, Lucía Gómez-Zaragozá, Maria Eleonora Minissi, Mariano Alcañiz, and Javier Marín-Morales. 2024. Developing Conversational Virtual Humans for Social Emotion Elicitation based on Large Language Models. *Expert Systems with Applications* 246 (2024), 123261.

[48] Xianzheng Ma, Yash Bhalgat, Brandon Smart, Shuai Chen, Xinghui Li, Jian Ding, Jindong Gu, Dave Zhenyu Chen, Songyou Peng, Jia-Wang Bian, et al. 2024. When LMMs Step into the 3D World: A Survey and Meta-analysis of 3D Tasks via Multi-Modal Large Language Models. *arXiv preprint arXiv:2405.10255* (2024).

[49] Paul Mallery and Darren George. 2000. SPSS for Windows Step by Step.

[50] Christopher A Monk, J Gregory Trafton, and Deborah A Boehm-Davis. 2008. The effect of interruption duration and demand on resuming suspended goals. *Journal of experimental psychology: Applied* 14, 4 (2008), 299.

[51] Stephen Monsell. 2003. Task switching. *Trends in cognitive sciences* 7, 3 (2003), 134–140.

[52] Shogo Morita, Yan Zhang, Takuto Yamauchi, Sinan Chen, Jialong Li, and Kenji Tei. 2024. Towards Context-Aware Support for Color Vision Deficiency: An Approach Integrating LLM and AR. In *2024 IEEE 13th Global Conference on Consumer Electronics (GCCE)*. IEEE, 188–189.

[53] Arthur Nishimoto and Andrew E Johnson. 2019. Extending virtual reality display wall environments using augmented reality. In *Symposium on spatial user interaction*. 1–5.

[54] Erwan Normand and Michael J McGuffin. 2018. Enlarging a smartphone with ar to create a handheld vesad (virtually extended screen-aligned display). In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 123–133.

[55] Leonardo Pavanatto, Feiyu Lu, Chris North, and Doug A Bowman. 2024. Multiple Monitors or Single Canvas? Evaluating Window Management and Layout Strategies on Virtual Displays. *IEEE Transactions on Visualization and Computer Graphics* 31, 3 (2024), 1713–1730.

[56] Leonardo Pavanatto, Chris North, Doug A Bowman, Carmen Badea, and Richard Stoakley. 2021. Do we still need physical monitors? An Evaluation of the Usability of AR Virtual Monitors for Productivity Work. In *2021 IEEE virtual reality and 3D user interfaces (VR)*. IEEE, 759–767.

[57] Gary Perelman, Emmanuel Dubois, Alice Probst, and Marcos Serrano. 2022. Visual Transitions Around Tabletops in Mixed Reality: Study on a Visual Acquisition Task Between Vertical Virtual Displays and Horizontal Tabletops. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (2022), 660–679.

[58] Francesca Perillo, Marco Romano, and Giuliana Vitiello. 2024. Enhancing Historical Understanding in School Students: Designing a VR Application with AI-Animated Characters. In *International Conference on Human-Computer Interaction*. Springer, 84–93.

[59] Carolin Reichherzer, Jack Fraser, Damien Constantine Rompapas, and Mark Billinghurst. 2021. Secondsight: A framework for Cross-Device Augmented Reality Interfaces. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.

[60] George Robertson, Mary Czerwinski, Patrick Baudisch, Brian Meyers, Daniel Robbins, Greg Smith, and Desney Tan. 2005. The Large-Display User Experience. *IEEE computer graphics and applications* 25, 4 (2005), 44–51.

[61] Martin Schrepp, Andreas Hinderks, and Jorg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40–44.

[62] Ashley JS Scott, Fairleth McCuaig, Victoria Lim, William Watkins, Jerry Wang, and Graydon Strachan. 2024. Revolutionizing Nurse practitioner Training: Integrating Virtual Reality and Large Language Models for Enhanced Clinical Education. In *Innovation in Applied Nursing Informatics*. IOS Press, 671–672.

[63] Mohammad Rajabi Seraji, Parastoo Piray, Vahid Zahednejad, and Wolfgang Stuerzlinger. 2024. Analyzing user behaviour patterns in a cross-virtuality immersive analytics system. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (2024), 2613–2623.

[64] Xiaoyu Shi, Zhaoyang Huang, Fu-Yun Wang, Weikang Bian, Dasong Li, Yi Zhang, Manyuan Zhang, Ka Chun Cheung, Simon See, Hongwei Qin, et al. 2024. Motion-i2v: Consistent and controllable image-to-video generation with explicit motion modeling. In *ACM SIGGRAPH 2024 Conference Papers*. 1–11.

[65] Alon Shoa, Ramon Oliva, Mel Slater, and Doron Friedman. 2023. Sushi with einstein: Enhancing hybrid live events with llm-based virtual humans. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. 1–6.

[66] Toqeer Ali Syed, Muhammad Shoaib Siddiqui, Hurria Binte Abdullah, Salman Jan, Abdallah Namoun, Ali Alzahrani, Adnan Nadeem, and Ahmad B Alkhodre. 2022. In-depth review of augmented reality: Tracking technologies, development tools, AR displays, collaborative AR, and security concerns. *Sensors* 23, 1 (2022), 146.

[67] Zhi Rui Tam, Cheng-Kuang Wu, Yi-Lin Tsai, Chieh-Yen Lin, Hung-yi Lee, and Yun-Nung Chen. 2024. Let me speak freely? A Study on the Impact of Format Restrictions on Performance of Large Language Models. *arXiv preprint arXiv:2408.02442* (2024).

[68] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 73–80.

[69] Yiliu Tang, Jason Situ, Andrea Yaoyun Cui, Mengke Wu, and Yun Huang. 2025. LLM Integration in Extended Reality: A Comprehensive Review of Current Trends, Challenges, and Future Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–24.

[70] Esen K Tütüncü, Mar Gonzalez-Franco, Khushman Patel, and Eric J Gonzalez. 2026. World Mouse: Exploring Interactions with a Cross-Reality Cursor. *arXiv preprint arXiv:2603.10984* (2026).

[71] Jonathan Wieland, Hyunsung Cho, Sebastian Hubenschmid, Akihiro Kiuchi, Harald Reiterer, and David Lindlbauer. 2024. Push2ar: enhancing mobile list interactions using augmented reality. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 671–680.

[72] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[73] Fang Xu, Tri Nguyen, and Jing Du. 2024. Augmented Reality for Maintenance Tasks with ChatGPT for Automated Text-to-Action. *Journal of Construction Engineering and Management* 150, 4 (2024), 04024015.

[74] Bufang Yang, Yunqi Guo, Lilin Xu, Zhenyu Yan, Hongkai Chen, Guoliang Xing, and Xiaofan Jiang. 2025. Socialmind: LLM-based Proactive AR Social Assistive System with Human-like Perception for In-situ Live Interactions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 1 (2025), 1–30.

[75] Yiduo Yao, WZW Hasan, Wenlong Jiao, Xianling Dong, HR Ramli, Nor Mohd Haziq Norsahperi, and Dong Wen. 2024. ChatGPT and BCI-VR: A New Integrated Diagnostic and Therapeutic Perspective for the Accurate Diagnosis and Personalized Treatment of Mild Cognitive Impairment. *Frontiers in Human Neuroscience* 18 (2024), 1426055.

[76] Koji Yatani. 2016. Effect Sizes and Power Analysis in HCI. In *Modern Statistical Methods for HCI*. Springer, 87–110.

[77] Xin Zeng, Xiaoyu Wang, Tengxiang Zhang, Yukang Yan, and Yiqiang Chen. 2024. InputJump: Augmented Reality-Facilitated Cross-Device Input Fusion based on Spatial and Semantic Information. *Virtual Reality & Intelligent Hardware* 6, 6 (2024), 502–526.

[78] Xiaotian Zhang, Weiping He, Mark Billinghurst, Daisong Liu, Lingxiao Yang, Shuo Feng, and Yizhe Liu. 2024. Usability of Cross-Device Interaction Interfaces for Augmented Reality in Physical Tasks. *International Journal of Human–Computer Interaction* 40, 9 (2024), 2361–2379.

[79] Zheng Zhang, Mengjie Yu, Tianyi Wang, Kashyap Todi, Ajoy Savio Fernandes, Yue Liu, Haijun Xia, Tovi Grossman, and Tanya Jonker. 2026. Gazeify Then Voiceify: Physical Object Referencing Through Gaze and Voice Interaction with Displayless Smart Glasses. *arXiv preprint arXiv:2601.19281* (2026).

[80] Xiaoyan Zhou, Benjamin Lee, Francisco R Ortega, Anil Ufuk Batmaz, and Yalong Yang. 2024. Lights, Headset, Tablet, Action: Exploring the use of Hybrid User Interfaces for Immersive Situated Analytics. *Proceedings of the ACM on Human-Computer Interaction* 8, ISS (2024), 517–539.

[81] Cindy Ziker, Barbara Truman, and Heather Dodds. 2021. Cross Reality (XR): Challenges and Opportunities Across the Spectrum. *Innovative learning environments in STEM higher education: Opportunities, challenges, and looking forward* (2021), 55–77.