ImmersiProtor: A Collaborative Mixed-Prototype Tool Integrating Spatial Augmented Reality and Component-layered Generation

Hongbo Zhang
Zhejiang University
Hangzhou, China
hongbozhang@zju.edu.cn

Yifei Wu
Zhejiang University
Hangzhou, China
yifei_wu@zju.edu.cn

Chaoyi Lin
Zhejiang University
Hangzhou, China
chaoyilin@zju.edu.cn

Weitao You
Zhejiang University
Hangzhou, China
weitao_you@zju.edu.cn

Lingyun Sun
Zhejiang University
Hangzhou, China
sunly@zju.edu.cn

Pei Chen*
Zhejiang University
Hangzhou, China
chenpei@zju.edu.cn

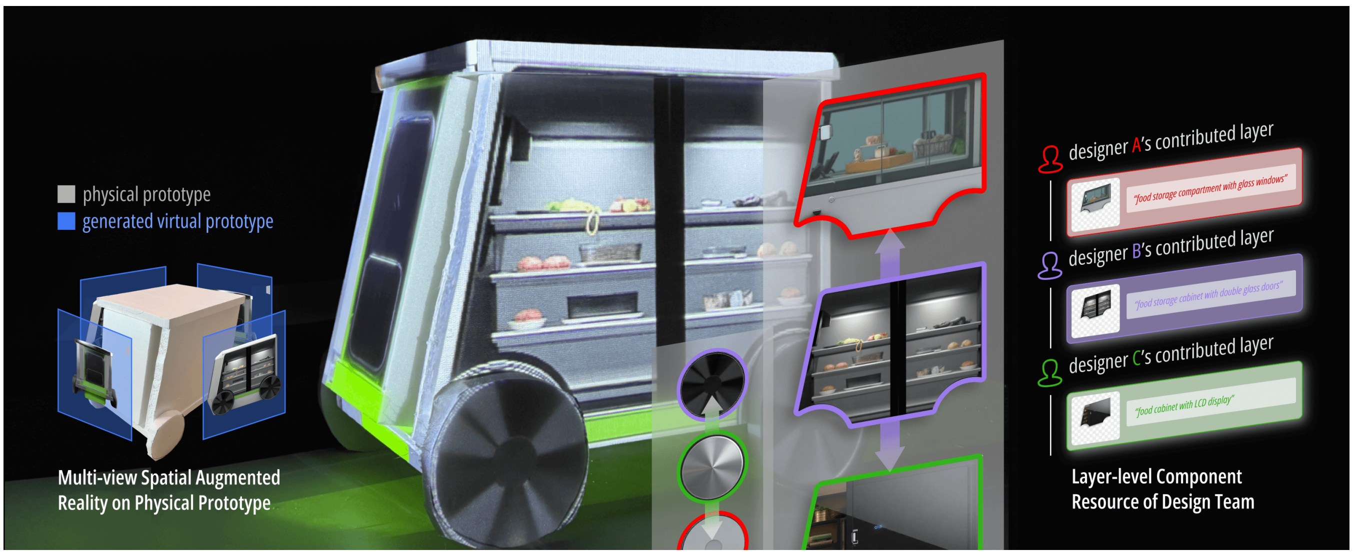

Figure 1: An autonomous delivery vehicle, created by participants using ImmersiProtor. ImmersiProtor introduces a component-layered generation and collaboration mode, offering both personal and shared team component resources. Furthermore, ImmersiProtor supports automatically generating multi-view and high-fidelity schemes, which are stuck onto the corresponding surfaces of the physical prototypes using SAR technology.

Abstract

Conceptual design is a critical stage in product development, which is a co-design process involving multidisciplinary collaboration based on prototypes. In this paper, we aim to propose a novel prototype paradigm that combines the distinct strengths of generative artificial intelligence (GAI) and spatial augmented reality (SAR), leveraging the expressive potential of SAR and the creative potential of GAI for co-design. To achieve this, we initially conducted a formative study with designers to explore how these technologies could be effectively combined to facilitate co-design. Based on our findings, we introduce ImmersiProtor, a prototype tool integrating multi-view SAR and component-layered GAI for co-design. On one hand, ImmersiProtor allows design team members to freely create and modify physical prototypes while automatically generating multi-view and high-fidelity renderings that are projected onto the surfaces of the physical prototype using SAR technology, enabling immersive communication and intuitive evaluation. On the other hand, ImmersiProtor introduces a component-layered generation and collaboration mode, offering both personal and shared team component resources. It ensures that individual team members can explore ideas independently without interference, while also supporting concept integration, evaluation, and iteration. We implemented ImmersiProtor, which involves a web-based application and an SAR design space. We conducted a user study to verify ImmersiProtor's usability in supporting prototype and collaboration. Our results highlighted ImmersiProtor's inherent strengths in

*The corresponding author.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790763

enhancing intuition, promoting collaboration, and strengthening GAI controllability. We also explored the effect of mixed interaction on design and critically discuss its best practices for the HCI community.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; • **Applied computing → Computer-aided design**; • **Computing methodologies → Artificial intelligence**.

Keywords

Prototype, Co-design, Spatial augmented reality, Generative artificial intelligence

ACM Reference Format:
Hongbo Zhang, Yifei Wu, Chaoyi Lin, Weitao You, Lingyun Sun, and Pei Chen. 2026. ImmersiProtor: A Collaborative Mixed-Prototype Tool Integrating Spatial Augmented Reality and Component-layered Generation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3772318.3790763

1 Introduction

Conceptual design is a critical stage in product development, inherently functioning as a multidisciplinary co-design process [19, 39]. In this process, prototypes serve as representations of artifacts, as well as the core language for externalizing concepts, communicating ideas, and iterating solutions [69]. Traditional prototyping faces a persistent dilemma: physical prototypes (e.g., foam or cardboard models) offer essential tactile feedback and spatial awareness but suffer from low fidelity and time-consuming [19, 51]. Conversely, virtual prototypes (e.g., renderings, virtual models) provide high visual fidelity but lack intuitive interactivity, often failing to support embodied face-to-face communication effectively [65]. To bridge this gap, the Human-Computer Interaction (HCI) community has proposed mixed prototypes based on Augmented Reality (AR) techniques, aiming to integrate the tangibility of physical materials with the flexibility of virtual information to support more efficient co-design [78, 80].

Among various mixed-prototyping technologies, Spatial Augmented Reality (SAR) is considered a highly promising interaction paradigm for co-design [80]. By projecting virtual content directly onto physical surfaces, SAR transcends traditional 2D sketches or screen-confined 3D models [10]. Its core advantages are twofold: First, it creates a shared, intuitive, and device-free "shared anchor", enabling natural eye contact and face-to-face discussion among team members [4]. Second, it supports embodied thinking, allowing designers to intuitively scrutinize designs by touching, moving, and physically sensing the scale and form of the prototype [58]. However, most existing SAR systems suffer from a critical limitation: they primarily function as passive displays, lacking intrinsic creative power [80]. Specifically, designers must still create content manually in time-consuming software, often constrained by predefined physical models [10, 50]. Crucially, predefined virtual content cannot self-adapt to the rapid changes in physical carriers that are frequent during conceptual exploration [4]. It conflicts with the rapid, fluid iterative needs of conceptual design [78], making it difficult to address complex co-design challenges.

Facing the creative limitations of SAR, the rise of Generative Artificial Intelligence (GAI) offers a new opportunity to dynamically reshape the creative power of prototyping [46]. GAI participation reconstructs the creative capabilities of prototyping from three key dimensions. First, in terms of efficiency, it can instantly generate massive, diverse, high-fidelity schemes [74], accelerating the creation and iteration in design process. Second, in terms of participation, it lowers the barrier to professional visual expression through natural language interaction, granting equal creative power to collaborators from different disciplinary backgrounds [72]. Third, in terms of roles, GAI shifts the designer's role from a laborious manual executor to a higher-level creative director [17]. This paradigm shift enables design teams to leverage AI's divergent capabilities to break through inherent mental fixations [7], vastly extending the boundaries of conceptual exploration in collaboration. In short, GAI provides the unprecedented creative drive for co-design.

Leveraging the high expressive potential of SAR and the high creative potential of GAI, we aim to explore a novel mixed-prototype interaction that integrates both capabilities. As this is not a simple additive process, we initially conducted a formative study to investigate how to combine GAI and SAR to serve actual co-design. It revealed two key interaction challenges when merging these technologies into a collaborative tool: (1) Inconsistent spatial experience in collaboration: Existing single-view SAR fails to provide a truly immersive experience across different viewing angles, forcing designers to cognitively imagine other perspectives; (2) Lack of controllability in collaboration: GAI's default holistic, one-off, end-to-end generation mode is incompatible with the need for fine-grained, controllable communication and iteration required by design teams, hindering deep collaborative refinement.

Based on insights from our formative study, we propose a novel interaction mode capable of satisfying both local iteration with fine-grained collaboration and immersive synergy. Accordingly, we designed, built, and evaluated ImmersiProtor, a prototype tool integrating component-layered generation and multi-view SAR for co-design. Specifically, the component-layered generation means that outputs from ImmersiProtor exist as independent component layers, supporting team members in iterating, replacing, modifying, and discussing specific parts independently, thus optimizing GAI's controllability in co-design. Multi-view SAR addresses the lack of immersion by achieving consistent multi-view scheme generation and presentation via four orthogonal projectors. These two core mechanisms disrupt traditional prototyping tools from the dimensions of co-design creation and expressiveness, delivering a fine-grained, highly immersive collaborative design tool.

The originality and contributions of this paper are demonstrated across three levels:

- At the Interaction Paradigm level: We propose an innovative co-design interaction combining component-layered generation and multi-view SAR, integrating their distinct expressive potential and creative potential.
- At the System Implementation level: We contribute a complete co-design system. We present technical pipelines supporting component-layered generation and highly consistent

multi-view generation, alongside a hardware design space for multi-view SAR, demonstrating the technical feasibility of the proposed interaction innovation.
- At the Empirical Support level: We conducted a user study with 36 designers, providing empirical evidence for proposed novel interaction. Beyond verifying usability. We critically discussed the novel mixed-prototype method through the perspective of AI controllability, design creativity, and team collaboration.

2 Related Work

2.1 Prototype in Conceptual Co-design

Prototypes are essential tools for designers to externalize ideas and understand design spaces [39]. Design prototypes are typically categorized into physical and digital types [29]. Physical prototypes facilitate design reasoning through rich multi-sensory stimulation, proving particularly valuable for co-design by reducing individual cognitive load and providing teams with an immediate, shared focal point [19]. Simultaneously, their tangible nature helps expose design flaws [27]. However, the high-fidelity production of physical prototypes is costly and difficult to version control [78], hindering rapid iteration and distributed collaboration. Conversely, digital prototypes are easy to iterate and distribute, but conveying concepts is ambiguous, nonintuitive and time-consuming, potentially slowing the rapid feedback essential for collaborative design [78]. For example, although recent work have enhanced VR prototyping by integrating GAI and multisensory interaction, purely virtual environments still lack the tangible feedback essential for embodied collaboration [37].

Given the limitations of the aforementioned approaches, the HCI community has begun focusing on mixed-prototype methods, aiming to fuse physical intuitiveness with digital flexibility to support conceptual co-design [80]. One major line of research focuses on enhancing both the intuition and fidelity in creation. This includes using AR to integrate physical and digital models [3], using everyday objects as physical proxies for animating 3D characters [36], or employing robotic arms to provide precise physical force feedback for simulating product interactions [32]. More advanced workflows, like Wizard of Props [83], integrate full-scale physical props with mixed reality to support prototyping interactions within large-scale environments. Another direction aims to accelerate the team's physical iteration loop. This includes robotic modeling assistants that rapidly fabricate AR sketches via 3D-printing pens [52], and new interaction techniques that use physical hand movements to reduce the difficulty of spatial sketching while digitally enhancing accuracy [31]. Collectively, these works demonstrate the potential of mixed-prototype to accelerate team iteration and enhance shared intuitive perception.

However, existing mixed-prototype workflows are often disjointed, requiring designers to frequently switch between physical and digital operations. Specifically, changes to physical forms cannot be instantly captured by the system to drive real-time updates and attachments of digital content. This creates a research gap for developing a more tightly-coupled workflow: one where iterations of physical forms are sensed in real time by the system, instantly triggering the creation of matching high-fidelity virtual content.

2.2 SAR Application in Prototype

While VR excels in immersion, AR interaction offers superior perceived informativeness, making it particularly suitable for design scenarios requiring detailed evaluation [70]. As a significant form of AR, SAR merges virtual content with the real world by projecting digital information directly onto the surface of physical objects via projectors [1]. SAR offers unique advantages in supporting face-to-face embodied collaboration. On the one hand, by attaching virtual content to physical prototypes, SAR provides a visual focal point that requires no wearable devices, facilitating natural eye contact and face-to-face discussions to form a shared anchor. For instance, Calixte and Leclercq [8] developed an Interactive Projection Mapping system enabling designers to make real-time graphical annotations on architectural models, enhancing teams' collective understanding of complex forms. Platforms like C-Space [66] also leverage SAR to support collaborative spatial design. On the other hand, SAR supports embodied thinking. Designers can touch, move, and perceive prototypes at real physical scale, integrating abstract digital information with bodily sensation. This is applied in product packaging design [55], material and texture visualization [56], and collaborative design of large-scale augmented environments via "SAR Miniatures" [44]. Compared to wearable or handheld AR displays, it offers advantages in comfort and ergonomics for prolonged use and promotes embodied thinking, which has been shown to reduce workload [25].

Although SAR brings unprecedented expressiveness to prototyping, existing SAR co-design tools still face two key challenges at the creation level. First, content creation and editing remain heavily reliant on inefficient manual labor or predefined templates. Most SAR tools are primarily used for visualization, with content either sourced from predefined pattern libraries or other templates [45, 55], or dependent on manual creation [41]. This proves inefficient during early conceptual design phases requiring rapid, diverse exploration, limiting design flexibility [18]. Second, this rigid virtual content struggles to adapt to rapid changes in physical carriers during conceptual design. Existing SAR projections are typically bound to static physical forms. Once designers adjust physical prototypes, virtual content struggles to adaptively follow these changes, making on-site iteration difficult [45].

Consequently, while SAR is physically well-suited for tightly coupled collaborative workflows, its content creation capabilities lag significantly. A natural research direction is to integrate powerful dynamic creative engines into SAR to overcome manual creation bottlenecks and achieve dynamic content adaptation to physical prototype changes.

2.3 GAI Application in Prototype

The GAI advancement has opened new opportunities for creative capabilities in prototyping, with HCI researchers actively exploring how it can reshape the creation pattern [46]. GAI's contributions are threefold. First, GAI's involvement enhances the efficiency of creation and iteration. It can instantly produce diverse, high-fidelity solutions [74] and supports tight iterative workflows [40, 77, 79], which enhances iteration quality—not just speed—while maintaining creative coherence. Second, GAI also expands design participation by lowering barriers for non-professionals to express ideas [72,

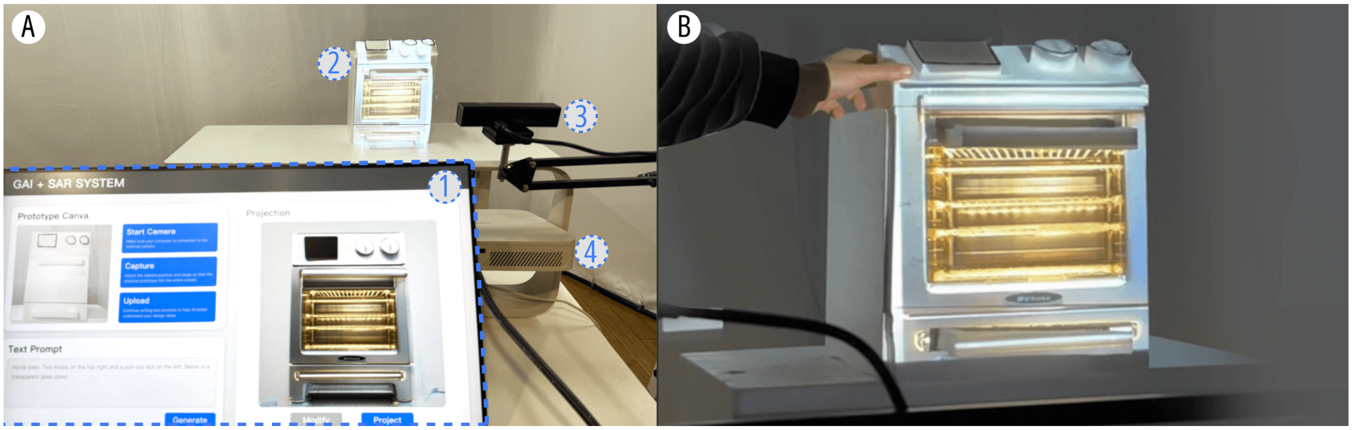

Figure 2: A: The constructed experimental environment for the formative study. 1: the developed generative systems. 2: the physical prototype. 3: the camera used to capture the physical prototype. 4: the projector used to project the generated scheme to the surface of the physical prototype. B: The scene of formative study. A participant is interacting with the physical component to conceive the design.

79]. To address the ambiguity of plain text prompts when conveying complex design intent, HCI research is exploring enhanced instructions beyond text [14, 60], such as materializing reference images into interactive "design tokens" [64]. Third, GAI is reshaping the creative process by helping teams break through mental patterns [67]. For this purpose, the HCI community is designing new tools that shift the focus from "Generation" to "Steering" and "Evaluation". For example, designers are leveraging GAI's reasoning capabilities to support concept evaluation [67], or as seen in FusAIn [53], where pen-based interaction constructs visual prompts, enabling designers to become active creative partners.

Despite this potential, GAI's integration also faces critical challenges in co-design. First, this one-shot generation logic is opaque, and its outputs are random and unpredictable, making it difficult for a team to build shared trust and consensus around an inscrutable solution [23, 47]. Second, GAI's end-to-end generation model is fundamentally incompatible with the incremental refinement innate to the design process. This conflict is magnified in collaborative workflows, because co-design is inherently a divide-and-conquer and merge process. This reveals a core conflict and a critical research gap in co-design: There is an unaddressed need for an interaction paradigm that fundamentally reshapes GAI's generation process—transforming it from a one-shot and black-box into a controllable, step-by-step workflow that supports the fine-grained and parallel iteration.

2.4 GAI Integration in AR Interaction

Recent HCI research is actively exploring the integration of GAI into AR devices to support smarter immersive creation. On the one hand, these efforts initially focus on immersive content generation and editing. Research encompasses rapid 3D generation in AR using multi-modal inputs [75], precise editing of emerging 3D representations [61], and contextualized fusion creation [30, 34]. Additionally, GAI is being applied to enhance collaborative content interaction, such as Thing2Reality [26], which enables users to instantly generate 3D objects from 2D content to enrich AR collaboration. On the other hand, beyond content generation, HCI is exploring GAI-driven AR interaction mechanisms. For instance, GazeNoter [71] combines gaze tracking with AI suggestions to assist note-taking, reducing intent bias. Guided Reality [86] and CARING-AI [63] leverage multi-modal inputs to generate context-aware AR task guidance, enhancing the fluidity and controllability of physical task interactions. Recent work has also begun to explore embodied design space that integrate AR and GAI for conceptual design, demonstrating how immersive environments can support ideation, evaluation, and human–AI co-creation in early-stage design [81].

While these studies highlight the growing trend of merging AR interactivity with GAI capabilities, a significant research gap remains in creation. Most existing approaches rely on head-mounted or handheld displays, often isolating users from the immersive physical environment. There is a lack of exploration into combining SAR with GAI. Addressing this gap, we aim to propose a novel interaction paradigm integrating SAR and GAI, leveraging the high expressive potential of SAR and the high creative potential of GAI.

3 Formative Study

3.1 Participants

Nine designers (five male and four female, average age 26) were recruited to participate in a formative study. This study aimed to identify practical co-design challenges and refine design requirements for a prototype tool integrating SAR interaction and GAI capabilities. All participants had relevant design experience with both GAI and SAR interaction. They were divided into three groups of three to collaboratively complete a design task.

3.2 Experimental Environment and System

For formative study, we constructed a dedicated experimental space (Figure 2A) and developed a supporting system (Figure 2B). The

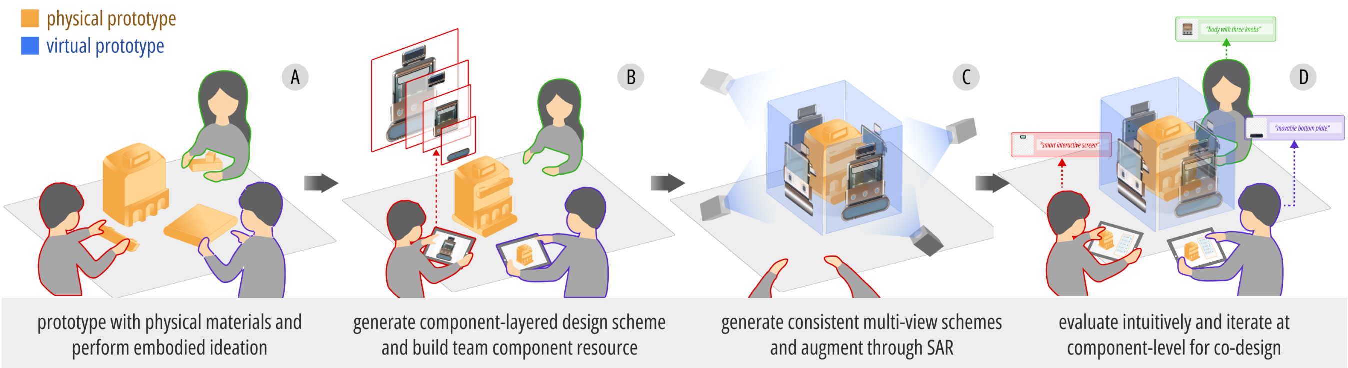

Figure 3: The flow chart using ImmersiProtor, which shows how designers use ImmersiProtor to interact with physical prototypes, GAI, and SAR. Orange represents the physical prototype, while blue represents the virtual form.

space, equipped with a computer, fixed camera, and projector, enabled designers to construct physical models while simultaneously projecting high-fidelity renderings onto their surfaces. The system featured an image generation model, enabling an iterative workflow: designers could input requirements, capture prototype photos, and generate initial high-fidelity proposals. They could then repeatedly refine these designs by either physically adjusting the prototype or modifying their prompts to regenerate new outputs, which could be projected back onto the prototype's surface at any time.

3.3 Task and Procedure

The formative study task was to design a *smart home oven*, focusing on innovations in its appearance, structure, and functionality. Participants were instructed to use both physical prototypes and GAI. The task required active collaboration, where designers shared concepts and engaged in group discussions to iteratively refine the design toward a consensus.

The one-hour experiment comprised four phases: introduction (5 min), independent ideation (10 min), collaborative prototyping (30 min), and a concluding semi-structured interview (15 min). The interview theme involves "design experience", "GAI interaction", "SAR interaction", and "co-design" to gather deeper insights into their experiences and perspectives during the co-design process.

3.4 Findings

Building upon the feedback gathered during the interviews, designers highlighted several key strengths of integrating SAR with GAI in co-design. These contributions included the ability to rapidly enhance prototype fidelity, accelerate the iteration and exploration of aesthetic styles, support tangible structural testing, provide a sense of presence and immersion in the design process, and make design discussions more focused and intuitive. These benefits underscored the potential of combining physical and digital tools to enrich the early stages of design.

However, interviews also identified several challenges (C) and areas for optimization in both design experience and usability. A significant issue was GAI's difficulty in accurately generating details based on the physical model's component positions (C1). For instance, P1 noted, "*GAI ignored my button design*" and P7 pointed out "*The content generated by GAI didn't match the physical model, which affected the projection quality*". Another limitation was the lack of precise control over iterative refinements (C2); as P3 explained, "*I struggled to make localized changes. Sometimes, I just wanted to update one part, but GAI altered everything*".

Further challenges included participant concerns about single-view SAR (C3), with comments like "*I had to imagine what the side view would look like*" and " *The projection on a single surface didn't fully leverage the advantages of the physical prototype*". Participants also found it unnecessary to manually design every surface, deeming it too time-consuming (C4). They instead emphasized leveraging GAI to extrapolate multi-view designs.

Collaboration challenges also emerged. The single-terminal setup hindered seamless teamwork (C5), as P5 noted "*When my partner was operating the system, I could only wait or watch*" and P9 observed their ideas were "*potentially influenced by the person in control*". Moreover, participants found it difficult to merge multiple design ideas into a cohesive version for evaluation (C6). P6 remarked "*Although we all contributed to the design, during discussions, we could only imagine how to combine our ideas, instead of building on each other's versions directly*".

3.5 Design Consideration and Goals

Drawing from the formative study's interviews and identified challenges, we defined the primary design goals for an ideal collaborative mixed-prototype tool integrating SAR and GAI.

G1: Support Fine-Grained Intention Conveyance, Generation, and Iteration (for C1&C2). This goal focuses on enabling precise communication of design intentions and flexible manipulation of design components. The system must accurately convey detailed designer inputs to GAI and support generating design solutions at the component level.

G2: Support Multi-view Immersive Presentation (for C3&C4). This goal aims to enhance immersion through multi-view presentation. The system must ensure that multi-view renderings are consistent and update as the physical prototype is modified. This ensures alignment between physical and virtual design elements and maintains consistency across all views.

Figure 4: ImmersiProtor's user interface involving five main working areas. A: the Design Requirement Input Area. B: Component Segmentation Area. C: Design Scheme Presentation Area. D: Component Resource Area. E: Multi-view SAR Area.

G3: Support Collaborative Iteration and Component Resource Sharing in Design Teams (for C5&C6). This goal emphasizes collaborative workflows, allowing team members to interact with the system independently and develop solutions iteratively. The system can support sharing and reusing component resources within the team to facilitate efficient evaluation and iteration.

4 ImmersiProtor

Following the design goals we identified in the formative study, we developed ImmersiProtor, which is a prototype tool integrating multi-view SAR and component-layered generation for co-design. ImmersiProtor consists of an immersive SAR design space and a collaborative design system.

4.1 Usage Scenario: Designing a Movable Carrier Robot

To illustrate how ImmersiProtor bridges physical and virtual design, we present an illustrative scenario (Figure 3) of a team co-designing a *movable carrier robot*.

The process begins with the team physically constructing the robot's basic form using foam blocks (Figure 3A). A designer then uses a tablet to capture this physical model. In the ImmersiProtor web interface (Figure 4), the designer segments the front surface as a `smart interactive screen` (Figure 4B) and inputs corresponding design requirements (Figure 4A). The GAI module generates several high-fidelity *screen* components, which are displayed on the canvas (Figure 4C) and simultaneously added to the team's shared *Component Resource Area* (Figure 4D).

Simultaneously, other designers working on their own tablets see these new component resources appear in the shared resource pool. They can drag the *screen* component from team members onto their own canvas, combining it with a *movable base* component they created (Figure 3B).

The team then decides to evaluate this combined design in the immersive space (Figure 3C). ImmersiProtor's SAR module automatically generates consistent multi-view renderings for this new combination (Figure 4E). The entire team gathers around the physical foam prototype as projectors cast the complete, multi-view virtual design directly onto the model's surfaces. Seeing the design in-situ, the team intuitively discusses its proportions, providing direct, "what-you-see-is-what-you-get" feedback (Figure 3D). For instance, they notice the screen is too large, designers can iterate on the *screen* component, and the system immediately projects the updated design onto the physical model for review. This rapid, embodied iteration process continues as the team refines their concept.

4.2 Fine-Grained Interaction and Generation

This first core function addresses the goal of supporting fine-grained designer control (**G1**). Its primary contribution is to provide a *global & local* pipeline that translates a designer's high-level intentions and specific physical-model annotations into high-fidelity, component-layered digital designs. This function serves as the generative engine for the entire system.

4.2.1 Fine-Grained Intention Conveyance. To achieve fine-grained communication of design intent, ImmersiProtor adopts a *global & local* approach to design requirement input. For global requirements, it captures an overall image of the physical prototype (Figure 4 A) and allows designers to input global prompts (e.g., "`This is an Intelligent delivery robot, used in community smart logistics scenarios, supporting the delivery of goods to residents' home.`") to define the design object and its overall requirements. Designers can also select style reference images from six provided options or custom uploads to guide the overall aesthetic.

For local requirements, ImmersiProtor uses the Segment Anything Model (SAM) [57] to extract component segmented masks from the captured prototype image (Figure 4 B). Designers can select specific masks (Figure 5) and annotate their functional semantics (e.g., "`smart interactive screen`" or "`movable bottom plate`"). These masks and annotations are processed by a large language model (LLM) to generate localized prompts that precisely bind the masks' proportions, positions, and semantics, providing detailed generation guidance.

By combining user-provided global prompts with the LLM-refined local prompts, ImmersiProtor creates a detailed design requirement description (e.g., "`This is an [Intelligent delivery robot, used in community smart logistics scenarios, supporting the delivery of goods to residents' homes], with a [smart interactive screen] on the top and a [movable bottom plate] at the bottom`"). This *global & local* approach ensures consistency in overall design goals while enabling precise handling of local features, providing rich and structured input for subsequent design processes.

4.2.2 Component-Layered Generation. Design scheme generation in ImmersiProtor is guided by two core considerations: (1) ensuring the generated scheme's local details align precisely with the physical prototype's edges for subsequent SAR projection, and (2) supporting designer-specified, component-layered generation to enable iterative and combinable design workflows. This pipeline is illustrated in Figure 6.

To achieve the first goal, ImmersiProtor extracts edge information from the physical prototype. These edges, along with style references and the LLM-refined comprehensive prompts, are fed into the image-to-image generation model Flux [33] to generate the design scheme. To achieve the second goal, ImmersiProtor segments this generated scheme based on the designer-specified component masks. This produces independent component layers, which are then visually displayed on the canvas in the *Design Scheme Presentation Area* (Figure 4 C). Additionally, all generated component layers are consolidated in the *Component Resource Area* (Figure 4 D). Here, designers can adjust layer order and visibility to flexibly combine and visualize different design schemes.

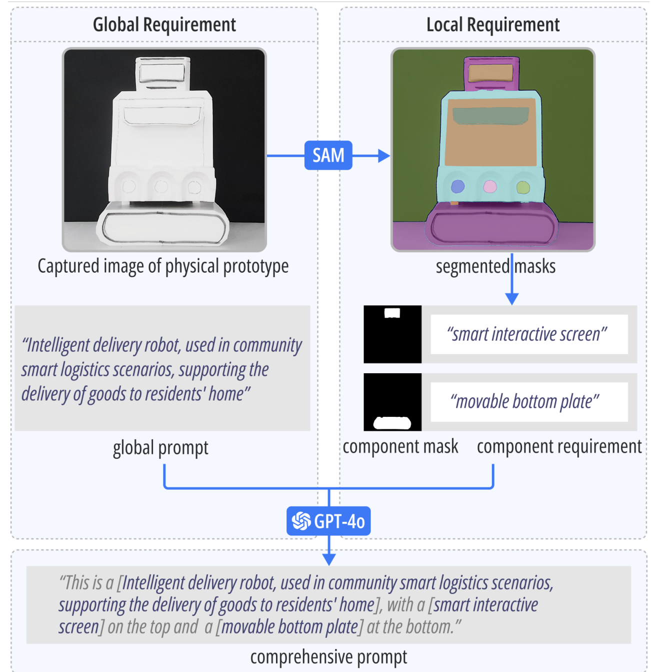

Figure 5: The fine-grained intention conveyance mechanism in ImmersiProtor, involving the global & local requirements.

4.3 Multi-view Immersive Presentation

With the digital components generated by Core Function 1, the second core function addresses their immersive presentation (**G2**). Its main contribution is a multi-view SAR system that automatically generates consistent multi-view renderings and projects them accurately onto the corresponding surfaces of the physical prototype. This function bridges the gap between the 2D virtual images and the 3D physical artifact.

4.3.1 Multi-view SAR Design Space. To achieve immersive effects in mixed-prototyping, we constructed a multi-view SAR design space for ImmersiProtor. This space includes two main workspaces: the *Prototype Construction Workspace* and the *Prototype SAR Workspace*. The *Prototype Construction Workspace* provides designers with materials and tools for building physical prototypes, enabling rapid construction and flexible adjustments. The *Prototype SAR Workspace* (Figure 7), equipped with projectors, cameras, and related hardware, supports design teams in generating proposals and conducting collaborative evaluations.

In this SAR design space, the system operates through the coordination of fixed spatial devices, portable terminals, and a remote server. The remote server hosts the generation models, processes image inputs, and outputs high-fidelity design proposals in real-time. Four fixed cameras, positioned at the prototype's front, back, left, and right, capture multi-view images to ensure generated schemes align precisely from all angles. Corresponding projectors are deployed at these four positions to project the matching high-fidelity schemes, enabling an immersive presentation experience. A local

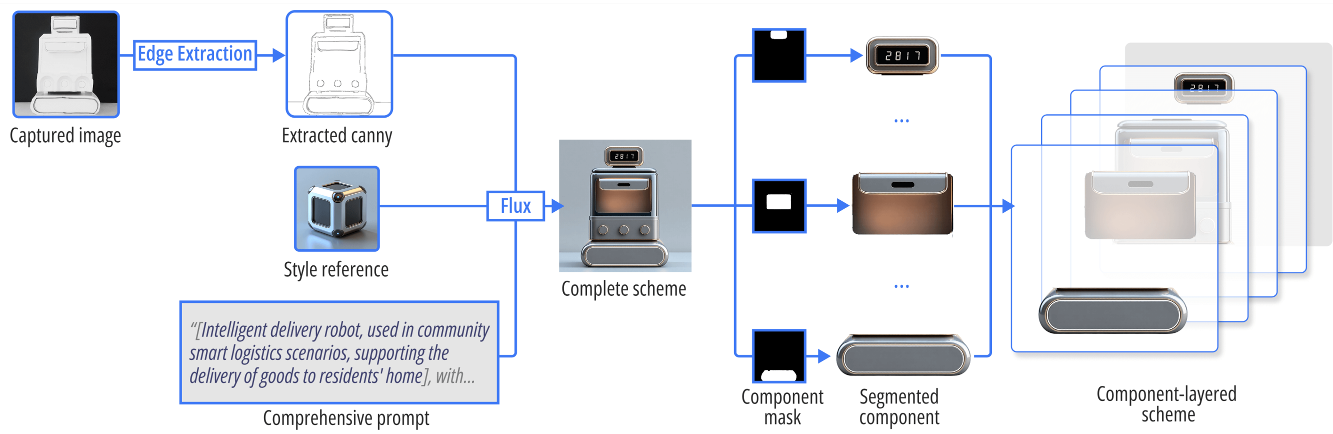

Figure 6: The component-layered generation pipeline in ImmersiProtor.

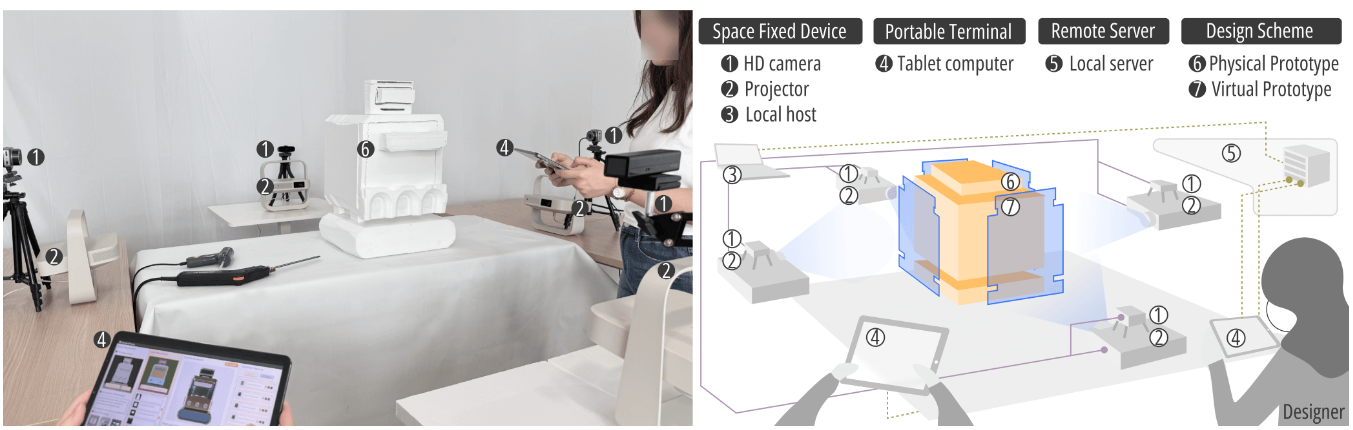

Figure 7: The constructed multi-view immersive SAR design space of ImmersiProtor, which is equipped with space fixed devices, portable terminals, and the remote server.

computer acts as the central host, managing data transmission between cameras, projectors, and the remote server. Additionally, mobile tablets serve as the interactive terminals, allowing designers to browse generated schemes in real-time and interact with the ImmersiProtor interface.

4.3.2 Consistent Multi-view Design Generation. A key prerequisite for multi-view SAR is obtaining high-fidelity designs from multiple perspectives. However, manually creating these designs is repetitive and labor-intensive. To address this, ImmersiProtor automates the generation of three additional orthogonal perspective designs, which are highly consistent with the current design on the canvas. These multi-view designs are presented in the *Multi-view SAR Area* (Figure 4 E) and projected onto the prototype. This approach not only ensures cross-view consistency but also enhances design efficiency. This multi-view generation process involves three steps (Figure 8).

1) *Orthogonal standard view synthesis*: ImmersiProtor employs the plug-and-play NVS-Adapter model [28] to synthesize coarse multi-view images from four standard views, using the current canvas image as input. By incorporating view-consistency cross-attention layers and global semantic conditional layers into a pre-trained text-to-image diffusion model, NVS-Adapter achieves rapid novel view synthesis. The prompt for this process combines the global prompt with all local component prompts. After this step, the resulting images have similar design styles and correct spatial relationships, but do not perfectly fit the physical model's edges.

2) *Canny-based image adaptation*: ImmersiProtor captures images of the physical model from four standard perspectives using four orthogonal cameras and converts them into canny images. It then uses ControlNet xanny [82] to align the generated image's canny shape with the physical model. Concurrently, IP-Adapter [76] encodes features from the NVS results (e.g., style, component structure), which are then injected into the image-to-image generation pipeline via decoupled cross-attention layers. In this process, the prompt includes the global prompt and a directional prompt (e.g., "`left view`", "`back view`"). Local prompts are removed to prevent generating components not visible from other views.

3) *Image enhancement*: The images resulting from adaptation often lack fine appearance details. We improve the overall image

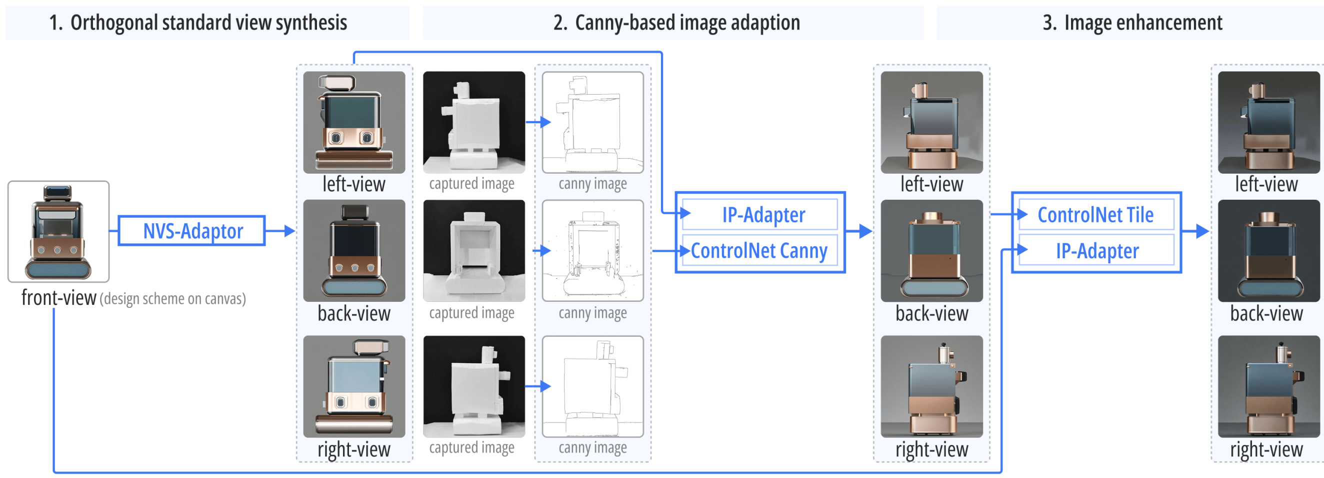

Figure 8: The consistent multi-view generation pipeline in ImmersiProtor.

quality by applying a Gaussian blur and using the ControlNet Tile Pipeline [82]. Finally, we apply an IP-Adapter [76] control, guided by the front-view image, to enhance multi-view style consistency that may have been compromised in the previous steps.

4.4 Embodied and Flexible Collaboration

Core Function 3 integrates the component-layered generation and the multi-view SAR to enable flexible and embodied team collaboration (**G3**). This function provides the critical infrastructure for co-design, consisting of (1) a web-based team resource pool for sharing and combining component resources, and (2) an embodied SAR interaction that allows the team to collaboratively iterate on components in real-time. Its primary contribution is balancing independent creation with shared, embodied evaluation.

4.4.1 Concurrently Visible Team Component Resource. ImmersiProtor, developed as a web-based platform, supports simultaneous access across multiple mobile devices to facilitate co-design. Its collaborative mechanism aims to balance individual creativity and team collaboration through independent creation and component resource sharing. ImmersiProtor's user interface divides the *Component Resource Area* (Figure 4 D) into personal and team component resources. In ImmersiProtor, we address this by providing equal access to both personal and team component resources, ensuring that all team members have equivalent creative opportunities regardless of their design expertise or background [84, 85]. Designers can independently generate, modify, and iterate components on their own devices without disrupting others' workflows. Meanwhile, the team resource section centrally displays all component resources created by team members for real-time browsing and selection.

To enrich team component resources (Figure 9), ImmersiProtor offers three flexible creation modes:

1) *Style Transfer*: Designers can apply one or more style reference images to generate components with diverse stylistic variations while maintaining consistent proportions and structures. This enables rapid exploration of components under different styles, introducing creativity and visual variety to team resources.

2) *Hand-Drawn Sketches*: To address the inefficiency and high trial-and-error cost of physical prototype editing, designers can sketch directly onto captured prototype images. These sketches, overlaid on the prototype, refine and enhance design details. The resulting images with sketches then serve as input for further design generation, injecting personalized and detailed elements into team resources.

3) *Layer Editing*: ImmersiProtor provides independent layer editing via an inpainting generative model [82], allowing designers to modify specific layers without affecting other layers. For example, a designer can adjust a robot's button design without altering other components. This precise editing improves efficiency and provides greater freedom for flexible component iteration.

4.4.2 Embodied Iteration and Collaboration. ImmersiProtor leverages its component-layered generation, concurrent team resources, and immersive SAR capabilities to enable embodied evaluation and iteration in co-design. Specifically, as shown in Figure 10, once the component resource library is constructed, the design team can collaboratively combine component resources from the team resource space to generate diverse design proposals. These schemes are then intuitively presented within the immersive SAR design space. Here, synchronized projectors from four orthogonal perspectives display the proposals onto the prototype, enabling team members to evaluate and adjust them in an embodied environment. It allows designers to validate proposal feasibility from multiple views, assess alignment with the physical prototype, and perform flexible, precise iterations.

4.5 Implementation

We implemented ImmersiProtor as a web application with a Python Flask back-end and a React Flow front-end. The back-end server is hosted on a local machine equipped with a GTX 3090 GPU.

The system integrates several generative models to support its core functions. We use GPT-4o [49] for prompt combination. For image-to-image generation, we employ Flux [33], a latent diffusion model. Style control is managed by IP-Adapter [76], which embeds

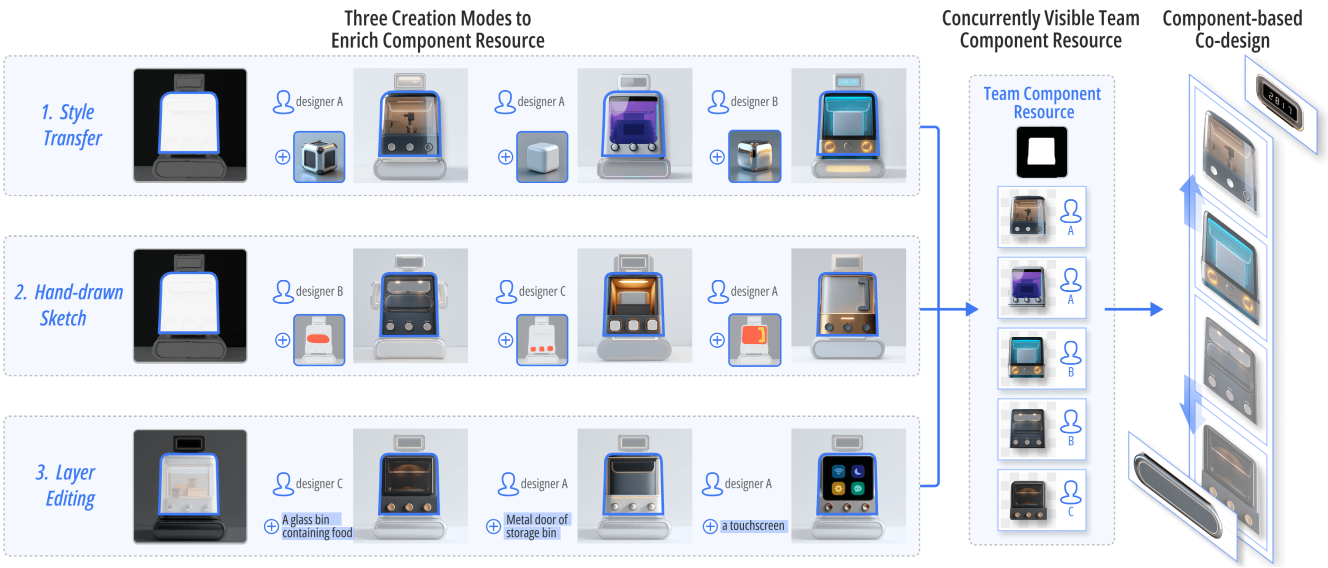

Figure 9: Three component generation and edition modes in ImmersiProtor to enrich team component resources, which contributes to the component-based co-design process.

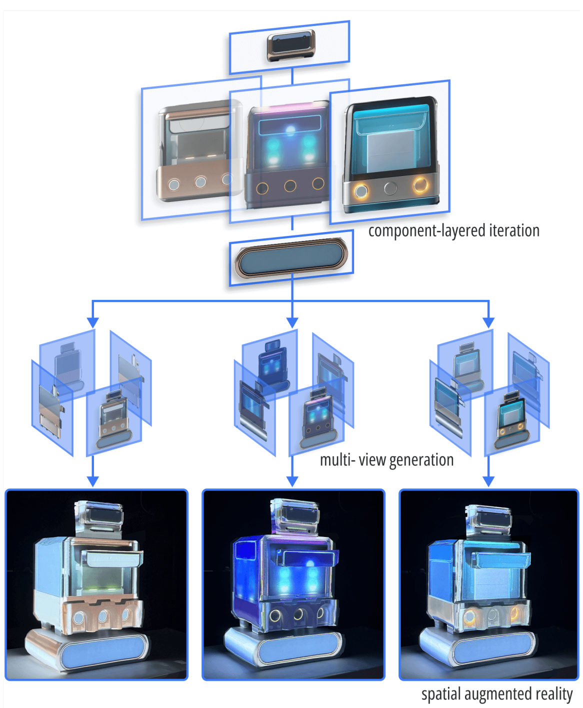

Figure 10: Embodied component-layered iteration with the support of ImmersiProtor.

target style features to ensure design consistency. Canny edge extraction and component iteration are handled by the ControlNet Canny and ControlNet Inpaint models, respectively [82]. Finally, we use SAM2 [57] for component segmentation and NVS-Adapter [28] for multi-view generation.

Regarding performance, core operations in the ImmersiProtor pipeline are timed as follows: image-to-image translation takes approximately 20s, and component segmentation takes 10s. The multi-view generation process includes 30s for orthogonal standard view synthesis, 10s per view for canny-based image adaptation, and 20s per view for image enhancement.

5 Evaluation Study

We address the following research questions in the user evaluation study.

RQ1: Is ImmersiProtor usable? Can ImmersiProtor support conceptual design?

RQ2: What is the influence of ImmersiProtor on creation? How do designers use ImmersiProtor for design ideation?

RQ3: What effect does the interactive mode of GAI and SAR integration have on design communication and collaboration?

5.1 Participants

To reflect the interdisciplinary nature of co-design and enrich stakeholder perspectives, we recruited 36 design professionals from diverse fields, including industrial, interaction, and engineering design. The cohort consisted of 16 males and 20 females, with a mean age of 24. Participants' qualifications included at least three years of design experience, encompassing both practical training in school and professional work. These participants were distinct from those involved in our formative research. Recruited via online advertisements and forums, all participants were pre-screened for specific expertise. All participants had prior experience with GAI tools. As SAR is a less common interaction paradigm, we did not require prior SAR experience. However, we did require all participants to

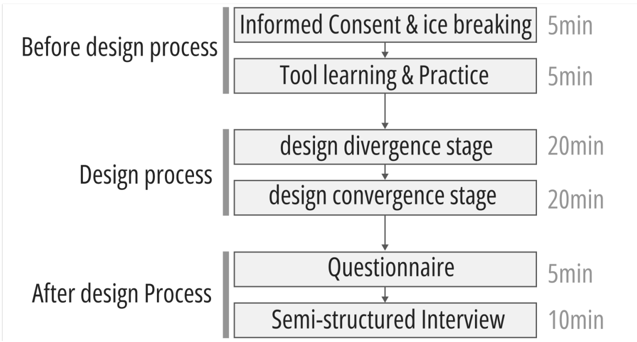

Figure 11: The procedure of the user study.

have experience in both physical prototyping and AR interaction. This screening allowed them to focus on concept exploration rather than tool learning.

For a fair comparison, we divided the 36 designers into two between-subjects conditions: an experimental group (using ImmersiProtor) and a control group (using a baseline tool). Participants were organized into teams of three to collaboratively complete a design task. We balanced these teams across both conditions by gender, age, design experience, and design domain to mitigate the influence of individual differences.

The study was conducted with no other ethical or privacy impacts, and all participants signed an institutionally-approved consent form. All participants were compensated for their participation.

5.2 Task and Procedure

The procedure of the user study is shown in Figure 11. Each participant underwent an individual orientation session, beginning with a 5-minute informed consent process, followed by a 5-minute ice-breaking stage with design collaborators, and concluding with a 5-minute demonstration. This initial phase was designed to familiarize participants with ImmersiProtor and the Think-Aloud method. Subsequently, participants engaged in a prototyping and co-design session lasting approximately 40 minutes. During this phase, they were tasked with completing a prototype design and presenting a design concept based on the assigned design task. The creative process was structured into two stages: a 20-minute divergence stage and a 20-minute convergence stage. In the divergence stage, participants constructed physical models using tangible materials and interacted with GAI to derive design inspiration and concepts. In the convergence stage, participants evaluated all proposed concepts, made design decisions, selected an optimal design concept, and documented it through images and verbal descriptions.

The designated task involved creating an *intelligent delivery vehicle*, such a food delivery robot, a service express car, and so on. A potential theme was provided to ensure participants to explore the design context and complete the task under conditions similar to real-world design scenarios [38]. Following the prototype and design process, participants were asked to complete several questionnaires and participate in a 10-min semi-structured interview.

Table 1: The questionnaire for design collaboration.

Theme	Questions
Collaboration and Awareness [20]	**Q1:** *I can effectively direct my collaborators' attention to a specific location within the shared workspace.* **Q2:** *The tool facilitates a shared visual focus, which makes our discussions more concentrated and efficient.*
Coordination and Management [42]	**Q3:** *Integrating my contributions with those of my collaborators is effortless.* **Q4:** *The tool helps us avoid conflicts and the accidental overwriting of each other's work.*
Shared Understanding and Consensus [15]	**Q5:** *I feel that my collaborators and I have a clear consensus on our shared objectives when using this tool.* **Q6:** *This tool provides effective ways to communicate ideas and clarify intentions to each other.*
Authorship and Attribution [16]	**Q7:** *The tool clearly attributes each contribution or modification to the respective author.* **Q8:** *I feel a sense of authorship over my portion of the final collaborative work.*

5.3 Baseline

To represent a common, non-integrated hybrid workflow, we deliberately selected Vizcom as the baseline. As a widely used and representative screen-based AI design tool, Vizcom excels at rapidly generating high-fidelity renderings from user inputs, including digital sketches and—crucially for our study—uploaded images via its image-to-image functionality. This makes it an ideal proxy for how designers might currently attempt to bridge the physical-to-digital gap using standard, screen-based AI tools.

The baseline workflow required participants to first create physical models, then captured images of these prototypes with a tablet as input for Vizcom. We posit this approach is both methodologically necessary for a fair comparison and increasingly representative of emerging design practices. Previous work has already demonstrated (e.g., Neural Canvas [62], Tangible Diffusion [21]), using low-fidelity physical models as tangible "scaffolds" for GAI is a valid and growing trend, which our formative study also employed. Therefore, this setup intends to model a plausible workflow. While we acknowledge designers use Vizcom in many different ways, this specific approach aligns with the emerging trend of physical-to-digital generation.

More importantly, This baseline establishes the fair, controlled comparison our study requires. Our objective was not to compare fine-grained interface features or to pit a mixed-prototype tool against a purely virtual one. Instead, our goal was to assess the impact of two different types of mixed-prototype workflows. To ensure this, participants in both groups were provided with the same tangible materials for physical prototyping as those in the experimental group. This setup ensured that both groups began with the same physical ideation process, enabling us to isolate the key variable: the method of integration between the physical prototype and the digital AI generation. The baseline represents a disconnected workflow (where the team evaluated physical models on the table and generated virtual schemes on separate screens), while ImmersiProtor represents our proposed integrated workflow (evaluating the virtual scheme directly on the physical model via SAR). This enabled a meaningful comparison of high-level collaborative and creative outcomes.

Figure 12: ImmersiProtor's design outcome in the user study.

5.4 Measurement

We combined quantitative and qualitative analysis to answer three RQs. Specifically, the System Usability Scale (SUS) [6] was applied to evaluate the usability of design tools for RQ1. The NASA Task Load Index (TLX) [22] was also adopted to assess the workload of the design process for RQ1. It is an overall workload score based on weighted average ratings of *mental demand*, *physical demand*, *temporal demand*, *effort*, *performance*, and *frustration level*.

To answer RQ2, the Creativity Support Index (CSI) [13] was utilized to evaluate the creativity support. It measures six dimensions of creativity support: *collaboration*, *enjoyment*, *exploration*, *expressiveness*, *immersion*, and *results worth effort*. In addition, we adopted the think-aloud protocol and behavior analysis to explore the mixed creation and ideation mode under the ImmersiProtor's support. Three researchers reviewed participants' design process and design outcomes. We focused on 1) how designers used physical and virtual prototypes to conceive ideas, and 2) how they switched between physical and virtual modalities.

To answer RQ3, we developed a novel questionnaire (Table 1) to evaluate the communication and collaboration dynamics within design teams during the design process. Based on a synthesis of existing collaboration instruments [15, 16, 20, 42], our questionnaire is specifically designed to assess four key dimensions: *Collaboration and Awareness*, *Coordination and Management*, *Shared Understanding and Consensus*, and *Authorship and Attribution*. Each of these dimensions is evaluated by two dedicated questions, with responses recorded on a 10-point rating scale. In addition, all participants were invited to a semi-structured interview. The interview focused on four key issues, including 1) *overall design collaboration*, 2) *GAI-supported communication*, and 3) *SAR-supported communication (experimental group only)*. Three researchers used thematic analysis [68] to extract prevalent codes. They independently coded all raw data and shared their codes, resolving disagreements and merging similar codes until they reached a consensus.

Statistical analysis was performed on some questionnaire data. We first tested for normality and homogeneity of variance. A t-test was used if both conditions were satisfied to analyze the differences between ImmersiProtor and the baseline; otherwise, a Mann-Whitney U test was applied, with a significance level of $p < 0.05$.

6 Results and Findings

6.1 ImmersiProtor's Usability (for RQ1)

Design Outcome. The final design outcomes of the six teams in the experimental group are presented in Figure 12. Designers subjectively indicated that ImmersiProtor successfully assisted them in completing the prototype and evaluation task.

SUS Results. The SUS questionnaire results for ImmersiProtor are presented in Figure 13. The average SUS score for ImmersiProtor was 84.44 ($SD = 8.92$), which is classified as "Excellent" according to Bangor's standard [2]. Furthermore, the usability score and learnability score were 81.94 ($SD = 16.78$) and 85.07 ($SD = 8.56$), respectively. Specifically, most participants in the experimental group reported that the ImmersiProtor system demonstrates high usability and strong functional consistency, effectively supporting them in efficiently completing conceptual design tasks.

NASA TLX Results. We reported the NASA-TLX results to compare the workload of ImmersiProtor against the baseline design method in Figure 14. Specifically, ImmersiProtor significantly outperformed the baseline in reducing cognitive load. It showed a statistically significant advantage in mental demand ($U = 10.5, p < 0.01$) and performance ($U = 60.0, p < 0.01$). This suggests that its intuitive interaction and expressive capabilities created a more user-friendly design environment, lowering frustration and enhancing perceived productivity [22].

However, ImmersiProtor produced significantly higher physical demand ($U = 266.0, p < 0.01$). This increased physical effort is an inherent consequence of its reliance on physical prototyping, where manual construction and modification of tangible models are central to the design process, particularly for high-precision SAR interactions [80].

Dissatisfied Cases Produced by ImmersiProtor. We also report cases where designers were dissatisfied in the experimental group to critically report ImmersiProtor's usability. These cases were extracted from the user study, design processes, and post-design interviews, where designers mentioned dissatisfaction with either the design outcomes or process. All information was collected through video recordings of the experimental sessions and backend server logs.

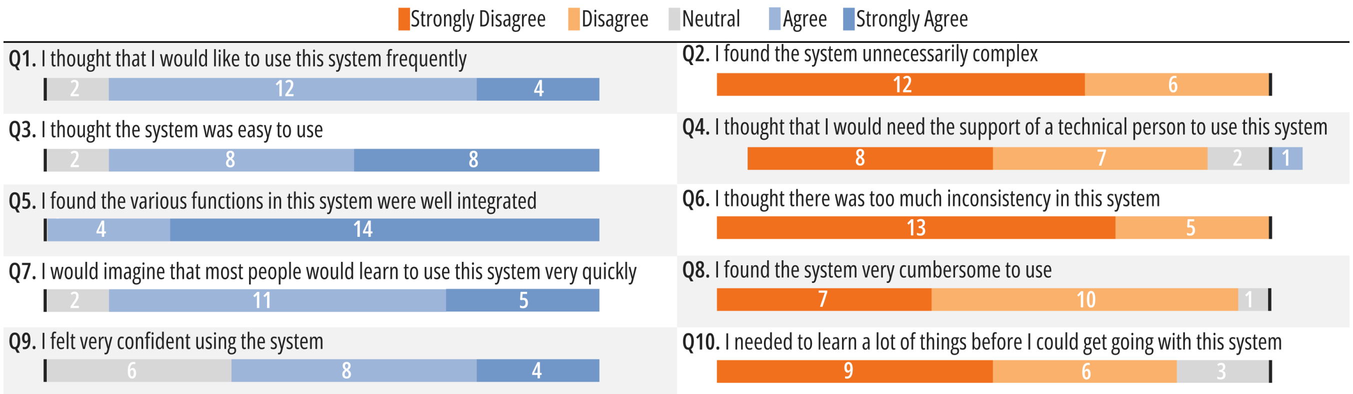

Figure 13: The SUS results of ImmersiProtor.

We categorized these issues into three major types: *Problems caused by GAI interaction*: These issues arose during the interaction between designers and the GAI system. For example, one designer provided design requirements that included a significant number of abstract functional needs (e.g., "`An advanced intelligent food delivery robot. The exterior design follows the principle of simple and elegant aesthetics, with exquisite details, harmonious overall proportions.`") (Figure 15 A). These prompts contained limited visual elements or detailed appearance requirements, making it challenging for the image generation model deployed in ImmersiProtor to produce design outcomes with concrete semantic information. As a result, the generated design proposal failed to meet the designer's expectations.

Problems caused by SAR interaction: These issues resulted from the interaction between designers and the SAR. For instance, difficulties were reported when designers created physical prototypes with hollow structures, which made it challenging to project visual content onto them (Figure 15 B). Similarly, some designers used black markers to add detailed graffiti to the physical prototype, which led to suboptimal SAR projection performance (Figure 15 C).

Inherent limitations of ImmersiProtor's interaction mechanisms: Beyond the issues caused by designers' interaction patterns, we also identified inherent constraints within ImmersiProtor's interaction mechanisms. These included challenges in understanding complex structures and handling intricate appearances. Specifically, when physical prototypes involve numerous structural elements (e.g., hinge mechanisms), the GAI often misinterpreted or omitted structural information during the refinement process (Figure 15 D). Additionally, when the physical prototype's appearance is highly complex (Figure 15 E) or includes curved surfaces (Figure 15 F), the SAR interaction mechanism struggled to adapt and provide effective support. By analyzing these issues, we aim to identify potential areas for improvement and best practices for ImmersiProtor.

6.2 Ideation Under ImmersiProtor's Support (for RQ2)

CSI Results. To assess its effectiveness in fostering creativity, we evaluated ImmersiProtor using the Creativity Support Index (CSI) and compared it against a baseline method (Figure 16). It showed that ImmersiProtor provided significantly greater creativity support overall, achieving a higher total CSI score than the baseline ($t(df) = 5.01, p < 0.01$). On average, ImmersiProtor's score was 87.93 ($SD = 2.33$).

A breakdown of the CSI sub-dimensions reveals the sources of this advantage. ImmersiProtor demonstrated significantly stronger support in fostering expressiveness ($U = 274.0, p < 0.01$), collaboration ($t(df) = 2.09, p = 0.04$), and user immersion ($U = 227.5, p = 0.04$). These findings highlight ImmersiProtor's capacity to empower designers with more potent creative tools and a more engaging collaborative environment.

However, despite these strengths, participants rated ImmersiProtor significantly lower on the satisfaction dimension ($t(df) = -2.05, p = 0.05$). This suggests a trade-off between powerful creative support and immediate user comfort. The lower satisfaction may be attributed to the cognitive friction from frequent transitions between physical and virtual spaces [79], as well as a steeper learning curve associated with its unconventional interaction modes, which could temper the overall enjoyment.

Creation Modes. Based on the analysis of user behavior and Think-aloud protocols during the experiment, we identified the patterns of creation and ideation in conceptual design supported by ImmersiProtor (Figure 17). These patterns were categorized based on the distinct operational domains in physical and virtual environments, as designers interacted with different objects in each context.

In the physical domain, designers primarily engaged in two modes of interaction with physical prototypes: *Physical Editing and Iteration*, and *Physical Combination and Adjustment*. In the first mode (Figure 17 A), designers iteratively modified physical prototypes through collaborative discussion, making changes such as adding new physical components, cutting materials, or altering proportions. For instance, in one design scenario, participants extended the length of a delivery vehicle and adjusted the style of its storage box after deliberation. In the second mode (Figure 17 B), designers focused on adjusting the spatial relationships of existing physical components and assembling them without altering their fundamental properties. For example, the design team created

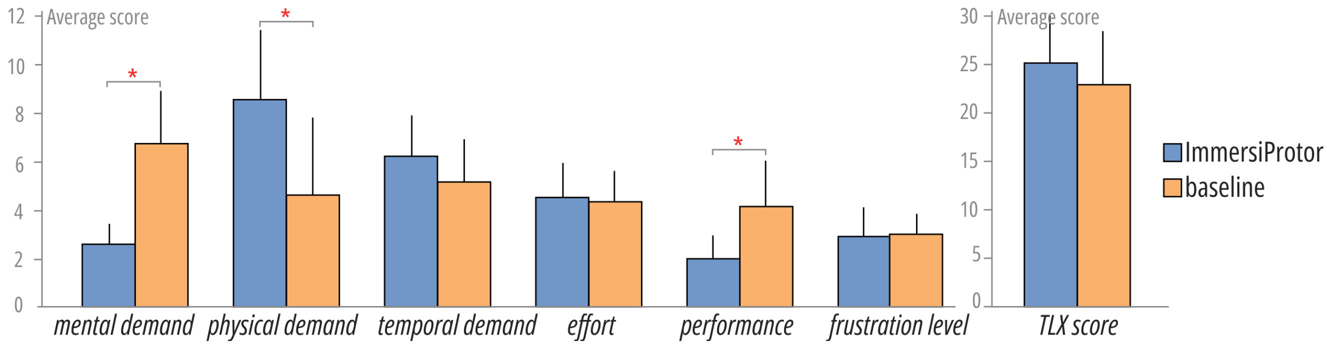

Figure 14: The NASA TLX results of ImmersiProtor and baseline. The lower value is better. * indicated significant difference.

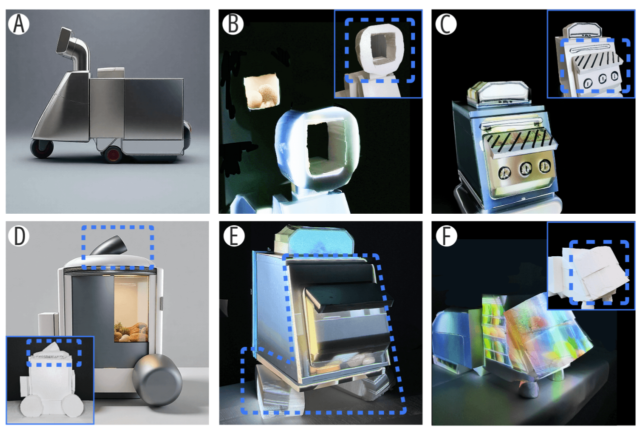

Figure 15: The process scheme that participants are not satisfied with in the user study. All image data was obtained from ImmersiProtor's system log and video recording.

an adjustable "front shield" for a robot by reconfiguring it to test various situations through tangible adjustment.

In the virtual domain, designers exhibited two primary creation modes centered on the manipulation of components: *Component Iteration and Optimization* and *Component Combination*. In the first mode (Figure 17 C), the design team demonstrated satisfaction with the overall design but concentrated on refining specific components while maintaining its overall style. A typical instance involved the iterative modification of the texture for a storage compartment door to achieve a more desirable aesthetic. In the second mode (Figure 17 D), designers explored new possibilities by freely combining components from the resource library. This process involved mixing and matching product parts, comparing configurations, and seeking optimal directions for design improvement through iterative experimentation.

It is worth noting that these four modes of interaction are not independent or isolated. Instead, they frequently overlap and alternate within a complete design task, dynamically transitioning between the physical and virtual domains. In addition, when addressing how design teams transfer between virtual and physical transformation modes, we were unable to identify meaningful common transformation patterns due to the limited experimental groups. Instead, we observed that design teams transitioned between physical and virtual modes at arbitrary points in the design process. This aligns with findings from prior research on hybrid prototyping, which suggests that transformation patterns vary significantly depending on individual design preferences and workflows [78, 79]. For instance, some designers showed a tendency to transition to virtual prototypes in the later stages of the design process, leveraging GAI for fine-tuning and detailed adjustments. Conversely, other designers preferred an early transition to virtual prototypes, using GAI as a tool for exploratory, randomness-driven ideation.

6.3 Collaboration Under ImmersiProtor's Support (for RQ3)

For the collaboration questionnaire, we initially examined the correlations between the two items under each theme to evaluate their internal consistency. The results showed that *Collaboration and Awareness* (Q1, Q2) yielded a moderate positive correlation ($r = 0.53, p < 0.01$) with a Spearman-Brown coefficient of 0.69, approaching the acceptable threshold. *Coordination and Management* (Q3, Q4) demonstrated a similar pattern ($r = 0.55, p < 0.001$); with a Spearman-Brown coefficient of 0.71, indicating good reliability. *Shared Understanding and Consensus* (Q5, Q6) exhibited a strong correlation ($r = 0.74, p < 0.001$) with a coefficient of 0.85, reflecting high consistency. Finally, *Authorship and Attribution* (Q7, Q8) showed an exceptionally strong correlation ($r = 0.93, p < 0.001$) with a coefficient of 0.97, suggesting excellent reliability. Taken together, these results indicate that the paired items under each theme consistently measure the same underlying construct. Therefore, it is statistically justified to compute each theme score as the sum of its two items (Figure 18).

For *Collaboration and Awareness*, ImmersiProtor demonstrated significantly higher values than baseline($U = 228.5, p = 0.034$). Similarly, at *Shared Understanding and Consensus*, ImmersiProtor exceeded the baseline ($U = 239.0, p = 0.014$). ImmersiProtor was also greater than the baseline at *Shared Understanding and Consensus* ($t(34) = 2.55, p = 0.016$) and *Authorship and Attribution* ($U = 252.0, p = 0.004$). The results indicated that ImmersiProtor showed potential in improving communication and collaboration.

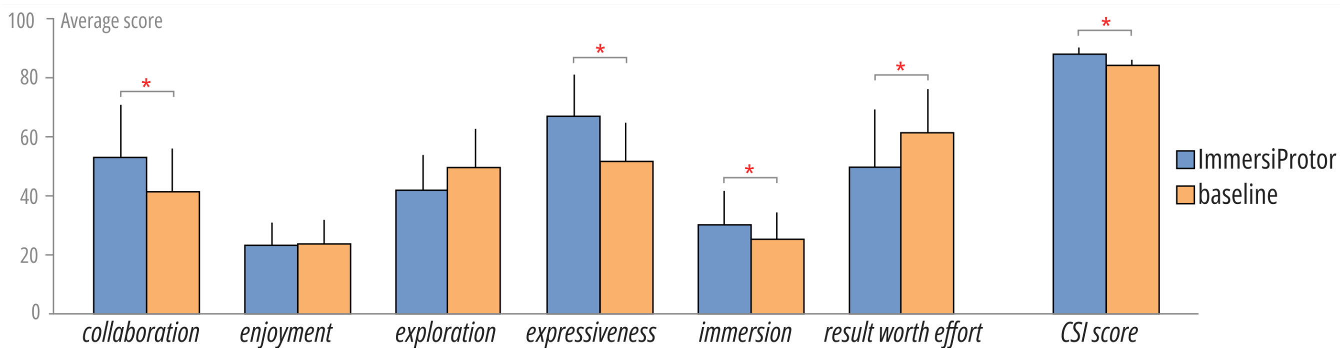

Figure 16: The CSI results of ImmersiProtor. The higher value is better. * indicated significant difference.

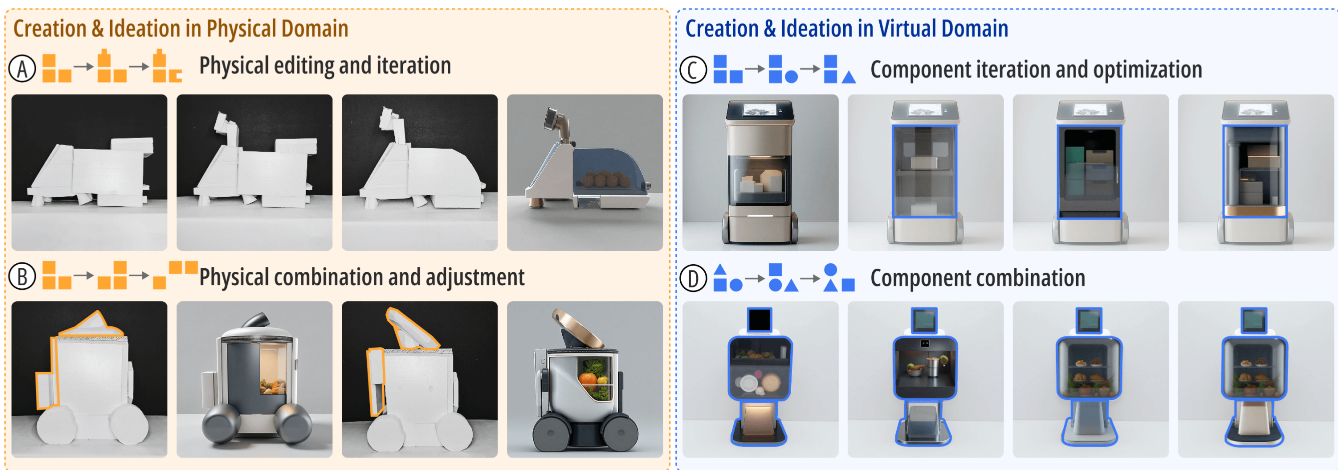

Figure 17: Four extracted creation modes during using ImmersiProtor through behavior analysis and Think-aloud method. All image data was obtained from ImmersiProtor's system log, involving the design scheme on the canvas and captured image of the physical prototype.

In addition to the quantitative analysis, we conducted a qualitative analysis of the transcribed interview data from the experimental group ($N = 18$). Three researchers coded the data using protocol analysis. After resolving discrepancies through discussion, we identified six prominent codes that characterized the participants' collaborative experience with ImmersiProtor. These themes, along with their definitions and supporting participant quotes, are summarized in Table 2.

7 Discussion

7.1 The Comprehensive Effectiveness of the ImmersiProtor Interaction

We comprehensively evaluate the effectiveness of its core interaction paradigm. Our empirical study provides an affirmative answer from multiple dimensions. First, on usability and cognitive load, ImmersiProtor demonstrated high usability. More importantly, NASA-TLX results revealed that, compared to the baseline, it significantly reduced users' mental demand. This indicates that its intuitive "what-you-see-is-what-you-get" interaction paradigm is efficient and cognitively friendly. Second, on creativity support, the CSI results indicated that ImmersiProtor provided superior creativity support. Specifically, it showcased significant advantages in the expressiveness and collaboration dimensions. This confirms the value of the component-layered generation and multi-view SAR space in empowering designers' creative expression and team interaction. Third, on collaborative experience, ImmersiProtor paradigm successfully addressed the two core challenges identified in the introduction. It enhanced GAI Controllability. The component-layered paradigm significantly enhanced designers' authorship and attribution, echoing recent findings in 2D sketching where iterative scaffolding improved user agency [40]. This demonstrates that our paradigm successfully ensured designers' creative dominance, aligning with recent calls for human-centered AI tools that prioritize user control over automated generation [73]. It also optimized AR collaborative experience. The multi-view SAR paradigm significantly elevated "Collaboration & Awareness" and "Shared Understanding & Consensus". This confirms its core value as a shared anchor, successfully unifying the team's discussion onto a physical entity.

Table 2: Key codes extracted from the transcribed interview data through protocol analysis.

Theme (Code)	Description	Representative Quotes
`design focus`	Provides a concrete focal point for team discussion and collaboration.	P15: "*The component-layered creation mode of ImmersiProtor served as an 'anchor' for our collaboration... avoiding scattered thinking.*"
`intuitive communication`	Supports clear and direct communication, avoiding vague verbal descriptions.	P7: "*ImmersiProtor allows me to precisely convey my intentions using the physical model... avoiding vague verbal descriptions like 'make it a bit longer' or 'a little higher'.*"
`design presence`	The immersive environment enabled by SAR brings the team to a shared, high-fidelity understanding of the design.	P11: "*With SAR, I found the effect very realistic and impressive, like looking at a real product... avoiding evaluation gaps caused by interpretive differences.*"
`low-threshold expression`	Allows designers to directly convey ideas using physical models without relying on technical modeling skills.	P13: "*Expressing design intentions through physical materials is an approach friendly to team members without technical backgrounds... fostering closer team communication.*"
`efficient feedback`	The layer-level generation by GAI accelerates the design iteration process.	P17: "*ImmersiProtor enables everyone to create independently and then combine individual outputs...*"
`creative combination`	Enables team members to independently create component resources and then combine them, which inspires creativity and maintains the overall design direction.	P17: "*The component resources contributed by collaborators not only inspire me but also ensure that individual contributions at the component level do not disrupt the overall design direction.*"

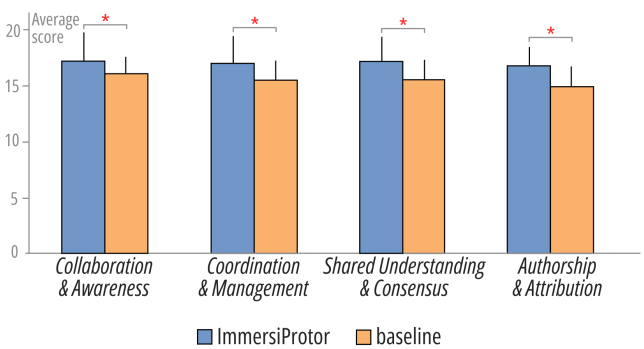

Figure 18: The results of the collaboration questionnaire, the higher value is better. * indicated significant difference.

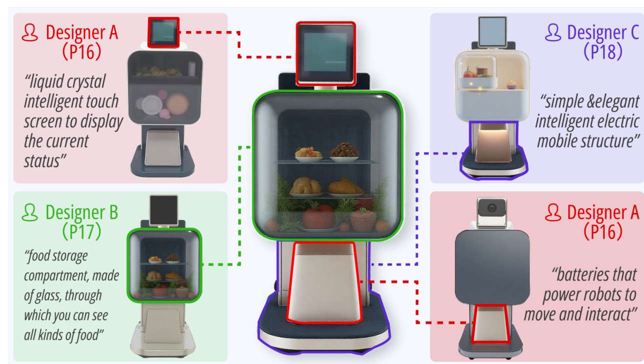

Figure 19: Example outcome (Group 6) from the user study, which shows the component-based collaboration under ImmersiProtor's support.

In summary, the effectiveness of the ImmersiProtor paradigm is holistic, not additive. It creates a virtuous cycle: The high usability and low cognitive load unleash designers' creative expression, and this fine-grained, controllable creativity, in turn directly fosters a more efficient and unified team consensus, similar to the spatial steering mechanisms [43] in previous work.

7.2 Implications for Human-AI Co-creation: Balancing "Fine-grained Control" and "Holistic Cognition"

One of this study's contributions is the validation of the component-layered paradigm's exceptional effectiveness as a GAI collaborative control mechanism. Traditional GAI holistic generation often functions as an opaque "black box", making it difficult to explain, control, and iterate upon, often resulting in outputs that stray from the designer's specific intent or lack explainability [48]. In contrast, the component-layered paradigm offered by ImmersiProtor demonstrates significant advantages in empowering both the individual and the team.

On the one hand, the fine-grained interaction enhanced control and authorship for the individual designer. This paradigm first resolves the core anxiety of losing designer dominance in GAI collaboration. By anchoring GAI's generative capabilities to specific components, designers gained unprecedented precise control. This is not merely an operational convenience but also fosters a psychological sense of authorship, further demonstrating that meticulous control and critical integration are essential for maintaining a sense of agency [24]. On the other hand, the fine-grained interaction supported a parallel collaborative mechanism for the design team. The component-layered paradigm provides a powerful mechanism to support team collaboration. As previous studies have clarified that team involvement can increase the originality of the final ideas even when the initial individually generated ideas are not very original [5, 54], it inherently supports an efficient workflow of asynchronous-creation and synchronous-integration. Designers can in parallel create and iterate on personal components on their respective terminals without interference. Subsequently, the team can instantly call upon these outcomes from the team component resource library for combination, evaluation, and discussion (Figure 19), similar to how structured idea representation enables traceable trajectories of creative evolution [59].

However, our study reveals an unanticipated, deeper trade-off: powerful "micro-execution" capability may narrow down designers' "macro-strategic thinking". We observed that when teams began optimizing an isolated component, they sometimes became overly-focused on its refinement, while temporarily overlooking higher-level, more fundamental design decisions (e.g., Is this component segmentation appropriate?). This reveals a new design challenge not just for embodied design tools, but for the broader HCI community: balancing the "Fine-grained Control" and "Holistic Cognition" in the creation process. While previous work suggests that compositional structures can help creators stay oriented [9], our findings indicate that distinct interaction mechanisms are needed to prevent "micro-execution" from overshadowing "macro-strategic thinking". The next agenda for GAI tools—from screen-based applications like Photoshop to spatial systems like ours—should not merely be to provide control, but to design interaction mechanisms that help designers maintain a cognitive balance between "micro-component execution" and "macro-conceptual integrity".

7.3 Implications for Embodied Collaboration: Navigating "Intuitive Collaboration" and "Fidelity Fixation"

The core value of ImmersiProtor's multi-view SAR paradigm lies in the efficient embodied collaboration experience it provides. As our study confirmed, physical models provide a spatial awareness that 2D screens cannot. Designers can intuitively evaluate proportions, ergonomics, and sightlines [19]. Furthermore, the tangible nature of these models facilitates structural ideation and modular thinking, allowing teams to test component relationships through hands-on assembly [29]. Besides, our collaboration results strongly supports that the "Shared Understanding & Consensus" were significantly higher than the baseline. This confirms its value as a communication anchor, allowing team members to simultaneously and synchronously engage in embodied collaboration, thus avoiding vague verbal descriptions and improving collaboration efficiency [79].

ImmersiProtor's multi-view SAR paradigm builds upon this fundamental value and amplifies it into a fluid team experience. When discussing the interaction trade-off regarding the medium for embodied interaction, other AR forms (e.g., Head-mounted Display(HMD) and Handheld Display(HHD)) can also integrate GAI with physical prototypes, but their emphases differ. HMD offers powerful individual spatial editing, but at the cost of a device barrier and high physical load, which might hinder team collaboration [80]. HHD conforms to existing habits, but its interaction is indirect, forcing designers to imagine 3D spatial relationships [80]. In contrast, ImmersiProtor's adoption of the SAR paradigm represents a deliberate trade-off. It sacrifices complex individual virtual spatial operation to prioritize intuitive team collaboration.

However, this high-fidelity embodied presence also comes with its own cost. Participants astutely noted that this "what-you-see-is-what-you-get" realism is ideal for design evaluation, but if introduced too early in the early concept exploration stage, it risks causing design fixation [12]. It prematurely eliminates the beneficial ambiguity inherent in low-fidelity prototypes [11]. The implication for all SAR embodied design tools is that the goal should not be the blind pursuit of highest fidelity, but rather variable and adaptive fidelity. Recent work has demonstrated the efficacy of workflows that bridge these levels, using low-fidelity 3D inputs to structurally guide high-fidelity AI generations, thereby allowing designers to maintain control throughout the progressive development of the concept [35]. Future prototype systems should be able to navigate the "Fidelity-Ambiguity" spectrum. They should sense the team's design stage: offering low-fidelity physical prototype during exploration to stimulate creativity, and only switching to high-fidelity projections for evaluation to aid decision-making. This opens up a specific and crucial new direction for embodied design tools.

7.4 Limitation

We acknowledge several potential limitations in this study. First, while our research introduces a novel paradigm integrating multi-view SAR and component-layered generation, we did not evaluate its applicability across diverse design contexts or tasks. Future work should aim to disentangle these variables through more controlled experiments. Second, more comprehensive quantitative research is needed to fully understand the factors that enhance the prototyping process. This would involve comparative studies against existing benchmarks, rigorous statistical analysis of key performance indicators, and experiments under strictly controlled conditions (e.g., standardized tasks and time constraints). Third, while our study recruited from diverse design fields, it lacked non-design stakeholders (e.g., product managers, clients). Future work should involve these more diverse collaborators to fully explore ImmersiProtor's potential for facilitating broader interdisciplinary co-design. Finally, our study was constrained by a short duration and a limited participant sample. A more realistic evaluation would require testing ImmersiProtor in real-world design workflows with larger teams over extended periods. We also plan an in-the-wild release for broader adoption to explore its effectiveness in diverse, uncontrolled environments.

8 Conclusion

We propose ImmersiProtor, a prototype tool integrating multi-view SAR and component-layered GAI for co-design, in this paper. ImmersiProtor allows design team members to freely create and modify physical prototypes while automatically generating multi-view and high-fidelity renderings that align with the physical iteration. These renderings are stuck onto the corresponding surfaces of the physical prototypes using SAR technology. Besides, ImmersiProtor introduces a component-layered generation and collaboration mode, offering both personal and shared team component spaces. This mode ensures that individual team members can explore ideas independently without interference, while also supporting concept integration, evaluation, and iteration. We contribute a complete co-design system, presenting hardware design space and technical pipelines. We verified ImmersiProtor's usability and inherent strength through a user study, including enhancing intuition, promoting embodied collaboration, and strengthening GAI controllability. We also critically discussed the applications and implications of the innovated mixed interaction for the HCI community.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No.62502436) and the Zhejiang Provincial Natural Science Foundation of China under Grant No.LMS26F020004.

References

[1] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent advances in augmented reality. *IEEE computer graphics and applications* 21, 6 (2001), 34–47.

[2] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[3] Loris Barbieri, Agostino Angilica, Fabio Bruno, and Maurizio Muzzupappa. 2013. Mixed prototyping with configurable physical archetype for usability evaluation of product interfaces. *Computers in Industry* 64, 3 (2013), 310–323.

[4] Fatma Ben Guefrech, Jean-François Boujut, Elies Dekoninck, and Gaetano Cascini. 2023. Studying interaction density in co-design sessions involving spatial augmented reality. *Research in Engineering Design* 34, 2 (2023), 201–220.

[5] Bryan M Blair and Katja Hölttä-Otto. 2012. Comparing the contribution of the group to the initial idea in progressive idea generation. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 45066. American Society of Mechanical Engineers, 425–436.

[6] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[7] Alex Brown, Molly H Goldstein, John Clay, H Onan Demirel, Xingang Li, and Zhenghui Sha. 2024. A study on generative design reasoning and students' divergent and convergent thinking. *Journal of Mechanical Design* 146, 3 (2024), 031405.

[8] Xaviéra Calixte and Pierre Leclercq. 2017. The interactive projection mapping as a spatial augmented reality to help collaborative design: Case study in architectural design. In *International Conference on Cooperative Design, Visualization and Engineering*. Springer, 143–152.

[9] Yining Cao, Yiyi Huang, Anh Truong, Hijung Valentina Shin, and Haijun Xia. 2025. Compositional Structures as Substrates for Human-AI Co-creation Environment: A Design Approach and A Case Study. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–25.

[10] Gaetano Cascini, Jamie O'Hare, Elies Dekoninck, Niccolo Becattini, Jean-François Boujut, Fatma Ben Guefrache, Iacopo Carli, Giandomenico Caruso, Lorenzo Giunta, and Federico Morosi. 2020. Exploring the use of AR technology for co-creative product and packaging design. *Computers in Industry* 123 (2020), 103308.

[11] Pei Chen, Zhuoshu Li, Yexinrui Wu, Hongbo Zhang, Jiaxuan Zhou, and Lingyun Sun. 2024. Modality and fidelity: Understanding how creative stimulus combinations impact design outcomes and process across different conceptual design phases. *Thinking Skills and Creativity* 54 (2024), 101657.

[12] Peiyao Cheng, Ruth Mugge, and Jan PL Schoormans. 2014. A new strategy to reduce design fixation: Presenting partial photographs to designers. *Design Studies* 35, 4 (2014), 374–391.

[13] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.

[14] DaEun Choi, Kihoon Son, HyunJoon Jung, and Juho Kim. 2025. Expandora: Broadening Design Exploration with Text-to-Image Model. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–10.

[15] Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. (1991).

[16] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7, 1 (2000), 59–83.

[17] Cong Fang, Yujie Zhu, Le Fang, Yonghao Long, Huan Lin, Yangfan Cong, and Stephen Jia Wang. 2025. Generative AI-enhanced human-AI collaborative conceptual design: A systematic literature review. *Design Studies* 97 (2025), 101300.

[18] Danilo Gasques, Janet G Johnson, Tommy Sharkey, and Nadir Weibel. 2019. What you sketch is what you get: Quick and easy augmented reality prototyping with pintar. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.

[19] Matheus Galvão Gomes, André Ogliari, Rodrigo Bastos Fernandes, and Karuliny Oliveira Marques. 2022. Evaluation of physical models as creative stimuli in conceptual design of products. *Design Studies* 81 (2022), 101119.

[20] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)* 11, 3 (2002), 411–446.

[21] Kuntong Han, Keyang Tang, and Meng Wang. 2024. Tangible diffusion: Exploring artwork generation via tangible elements and AI generative models in arts and design education. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–13.

[22] 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.

[23] Vikas Hassija, Vinay Chamola, Atmesh Mahapatra, Abhinandan Singal, Divyansh Goel, Kaizhu Huang, Simone Scardapane, Indro Spinelli, Mufti Mahmud, and Amir Hussain. 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation* 16, 1 (2024), 45–74.

[24] Jessica He, Stephanie Houde, and Justin D Weisz. 2025. Which contributions deserve credit? perceptions of attribution in human-ai co-creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[25] Florian Heinrich, Danny Schott, Lovis Schwenderling, and Christian Hansen. 2025. Do You See What I See? Evaluating Relative Depth Judgments Between Real and Virtual Projections. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[26] Erzhen Hu, Mingyi Li, Jungtaek Hong, Xun Qian, Alex Olwal, David Kim, Seongkook Heo, and Ruofei Du. 2025. Thing2Reality: Enabling Spontaneous Creation of 3D Objects from 2D Content using Generative AI in XR Meetings. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[27] Matilde B Jensen, Christer W Elverum, and Martin Steinert. 2017. Eliciting unknown unknowns with prototypes: Introducing prototrials and prototrial-driven cultures. *Design Studies* 49 (2017), 1–31.

[28] Yoonwoo Jeong, Jinwoo Lee, Chiheon Kim, Minsu Cho, and Doyup Lee. 2024. NVS-Adapter: Plug-and-Play Novel View Synthesis from a Single Image. In *European Conference on Computer Vision*. Springer, 449–466.

[29] Lee Kent, Chris Snider, James Gopsill, and Ben Hicks. 2021. Mixed reality in design prototyping: A systematic review. *Design Studies* 77 (2021), 101046.

[30] Seokyoung Kim, Dooyoung Kim, Taejun Son, and Woontack Woo. 2025. RealityCrafter: User-guided Editable 3D Scene Generation from a Single Image in Mixed Reality. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–3.

[31] Yongkwan Kim, Sang-Gyun An, Joon Hyub Lee, and Seok-Hyung Bae. 2018. Agile 3D sketching with air scaffolding. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[32] Nedim Kovacevic and Rainer Stark. 2024. User Experience Assessment on Virtual-Reality Robot-Assisted Smart Hybrid Prototype of a Vehicle Trunk Lid. *Procedia CIRP* 128 (2024), 900–904.

[33] Black Forest Labs. 2024. FLUX: Official inference repo for FLUX.1 models. https://github.com/black-forest-labs/flux.

[34] Jaewook Lee, Filippo Aleotti, Diego Mazala, Guillermo Garcia-Hernando, Sara Vicente, Oliver James Johnston, Isabel Kraus-Liang, Jakub Powierza, Donghoon Shin, Jon E Froehlich, et al. 2025. Imaginatear: Ai-assisted in-situ authoring in augmented reality. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–21.

[35] Seung-Jun Lee, Jeongche Yoon, Sang-Hyun Lee, Joon Hyub Lee, and Seok-Hyung Bae. 2025. 3D Sketching+ 2D Generative AI for Car Exterior Design. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[36] Boyu Li, Linping Yuan, Zhe Yan, Qianxi Liu, Yulin Shen, and Zeyu Wang. 2024. Anicraft: Crafting everyday objects as physical proxies for prototyping 3d character animation in mixed reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[37] Zhuoshu Li, Pei Chen, Hongbo Zhang, Yexinrui Wu, and Xuanhui Liu. 2025. Painting++: Human-Computer Collaborative Painting in VR with Multisensory Interaction. *Design and Artificial Intelligence* (2025), 100019.

[38] Jing Liao, Preben Hansen, and Chunlei Chai. 2020. A framework of artificial intelligence augmented design support. *Human–Computer Interaction* 35, 5-6 (2020), 511–544.

[39] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 2 (2008), 1–27.

[40] David Chuan-En Lin, Hyeonsu B Kang, Nikolas Martelaro, Aniket Kittur, Yan-Ying Chen, and Matthew K Hong. 2025. Inkspire: supporting design exploration with generative ai through analogical sketching. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[41] Mille Skovhus Lunding, Germán Leiva, Jens Emil Sloth Grønbæk, and Marianne Graves Petersen. 2022. ProjectAR: Rapid Prototyping of Projection Mapping with Mobile Augmented Reality. In *Adjunct Proceedings of the 2022 Nordic Human-Computer Interaction Conference*. 1–5.

[42] Thomas W Malone and Kevin Crowston. 1994. The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR)* 26, 1 (1994), 87–119.

[43] Nicolai Marquardt, Asta Roseway, Hugo Romat, Payod Panda, Michel Pahud, Gonzalo Ramos, Steven M Drucker, Andrew D Wilson, Ken Hinckley, and Nathalie Riche. 2025. ImaginationVellum: Generative-AI Ideation Canvas with Spatial Prompts, Generative Strokes, and Ideation History. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–19.

[44] Omar Martinez Gasca, Lucas Van Dorpe, Ivo Dewit, and Lukas Van Campenhout. 2025. SAR Miniatures: Physical Scale Models as Immersive Prototypes for Spatially Augmented Environments. *Designs* 9, 1 (2025), 10.
[45] Feederico Morosi, Iacopo Carli, Giandomenico Caruso, Gaetano Cascini, Elies Dekoninck, and Jean-François Boujut. 2021. Exploring tablet interfaces for product appearance authoring in spatial augmented reality. *International Journal of Human-Computer Studies* 156 (2021), 102719.
[46] Michael Muller, Lydia B Chilton, Mary Lou Maher, Charles Patrick Martin, Minsik Choi, Greg Walsh, and Anna Kantosalo. 2025. GenAICHI 2025: Generative AI and HCI at CHI 2025. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.
[47] Suchismita Naik, Austin L Toombs, Amanda Snellinger, Scott Saponas, and Amanda K Hall. 2025. Designing with Multi-Agent Generative AI: Insights from Industry Early Adopters. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 1961–1972.
[48] Syeda Masooma Naqvi, Ruichen He, and Harmanpreet Kaur. 2025. Catalyst for Creativity or a Hollow Trend?: A Cross-Level Perspective on The Role of Generative AI in Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.
[49] OpenAI. 2023. GPT-4V(ision) system card. https://chat.openai.com/.
[50] Jamie O'Hare, Elies Dekoninck, Mendy Mombeshora, Philippe Martens, Niccolò Becattini, and Jean-Francois Boujut. 2020. Defining requirements for an Augmented Reality system to overcome the challenges of creating and using design representations in co-design sessions. *CoDesign* 16, 2 (2020), 111–134.
[51] Eujin Pei, Ian Campbell, and Mark Evans. 2011. A taxonomic classification of visual design representations used by industrial designers and engineering designers. *The Design Journal* 14, 1 (2011), 64–91.
[52] Huaishu Peng, Jimmy Briggs, Cheng-Yao Wang, Kevin Guo, Joseph Kider, Stefanie Mueller, Patrick Baudisch, and François Guimbretière. 2018. RoMA: Interactive fabrication with augmented reality and a robotic 3D printer. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
[53] Xiaohan Peng, Janin Koch, and Wendy E Mackay. 2025. FusAIn: Composing Generative AI Visual Prompts Using Pen-based Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.
[54] Cecil Piya, Vinayak, Senthil Chandrasegaran, Niklas Elmqvist, and Karthik Ramani. 2017. Co-3Deator: A team-first collaborative 3D design ideation tool. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6581–6592.
[55] Maud Poulin, Cédric Masclet, and Jean-François Boujut. 2024. Investigating the effects of spatial augmented reality on user participation in co-design sessions: a case study. *Computers in Industry* 154 (2024), 104023.
[56] Parinya Punpongsanon, Daisuke Iwai, and Kosuke Sato. 2015. Softar: Visually manipulating haptic softness perception in spatial augmented reality. *IEEE transactions on visualization and computer graphics* 21, 11 (2015), 1279–1288.
[57] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. 2024. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024).
[58] Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J Mcguffin, and Tobias Isenberg. 2020. Collaborative work in augmented reality: A survey. *IEEE transactions on visualization and computer graphics* 28, 6 (2020), 2530–2549.
[59] Hanshu Shen, Lyukesheng Shen, Wenqi Wu, and Kejun Zhang. 2025. Ideationweb: Tracking the evolution of design ideas in human-ai co-creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
[60] Leixian Shen, Haotian Li, Yifang Wang, Xing Xie, and Huamin Qu. 2025. Prompting generative AI with interaction-augmented instructions. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.
[61] Yulin Shen, Boyu Li, Jiayang Huang, David Yip, and Zeyu Wang. 2025. GaussianShopVR: Facilitating Immersive 3D Authoring Using Gaussian Splatting in VR. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
[62] Yulin Shen, Yifei Shen, Jiawen Cheng, Chutian Jiang, Mingming Fan, and Zeyu Wang. 2024. Neural canvas: Supporting scenic design prototyping by integrating 3D sketching and generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
[63] Jingyu Shi, Rahul Jain, Seunggeun Chi, Hyungjun Doh, Hyung-gun Chi, Alexander J Quinn, and Karthik Ramani. 2025. CARING-AI: Towards Authoring Context-aware Augmented Reality INstruction through Generative Artificial Intelligence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23.
[64] Xinyu Shi, Yinghou Wang, Ryan Rossi, and Jian Zhao. 2025. Brickify: Enabling Expressive Design Intent Specification through Direct Manipulation on Design Tokens. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.
[65] Dorothé Smit, Bart Hengeveld, Martin Murer, and Manfred Tscheligi. 2022. Hybrid design tools for participatory, embodied sensemaking: an applied framework. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–10.
[66] Kihoon Son, Hwiwon Chun, Sojin Park, and Kyung Hoon Hyun. 2020. C-Space: an interactive prototyping platform for collaborative spatial design exploration. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
[67] Francesco Stella, Cosimo Della Santina, and Josie Hughes. 2023. How can LLMs transform the robotic design process? *Nature machine intelligence* 5, 6 (2023), 561–564.
[68] Jon Swain. 2018. A hybrid approach to thematic analysis in qualitative research: Using a practical example. *Sage research methods* (2018).
[69] John C Tang. 1991. Findings from observational studies of collaborative work. *International Journal of Man-machine studies* 34, 2 (1991), 143–160.
[70] Yilin Tang, Jiaxiang You, and Liuqing Chen. 2025. Exploring user perception in mobile shopping: A comparison between augmented reality and virtual reality. *Design and Artificial Intelligence* (2025), 100039.
[71] Hsin-Ruey Tsai, Shih-Kang Chiu, and Bryan Wang. 2025. GazeNoter: Co-Piloted AR Note-Taking via Gaze Selection of LLM Suggestions to Match Users' Intentions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.
[72] Yun Wan and Yoram M Kalman. 2025. Using Generative AI Personas Increases Collective Diversity in Human Ideation. *arXiv preprint arXiv:2504.13868* (2025).
[73] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa i Campanyà, Bing-Yu Chen, and Mike Y Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–28.
[74] Justin D Weisz, Michael Muller, Jessica He, and Stephanie Houde. 2023. Toward general design principles for generative AI applications. *arXiv preprint arXiv:2301.05578* (2023).
[75] Zhihao Yao, Xiwen Yao, Qirui Sun, Haowei Xiong, Kin Wang Lau, and Haipeng Mi. 2025. Gesture2Prop: Conditional VR Prop Generation from Hand Gestures and Speech. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–4.
[76] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. 2023. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721* (2023).
[77] hongbo Zhang, Pei Chen, Xuelong Xie, Zhaoqu Jiang, Yifei Wu, Zejian Li, and Lingyun Chen, Sun. 2025. FusionProtor: A Mixed-Prototype Tool for Component-level Physical-to-Virtual 3D Transition and Simulation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.
[78] Hongbo Zhang, Pei Chen, Xuelong Xie, Zhaoqu Jiang, Zihong Zhou, and Lingyun Sun. 2024. A hybrid prototype method combining physical models and generative artificial intelligence to support creativity in conceptual design. *ACM Transactions on Computer-Human Interaction* 31, 5 (2024), 1–34.
[79] Hongbo Zhang, Pei Chen, Xuelong Xie, Chaoyi Lin, Lianyan Liu, Zhuoshu Li, Weitao You, and Lingyun Sun. 2024. Protodreamer: A mixed-prototype tool combining physical model and generative AI to support conceptual design. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.
[80] Hongbo Zhang, Pei Chen, Jingwen Yang, Yifei Wu, Zhaoqu Jiang, Xuelong Xie, Weitao You, and Lingyun Sun. 2025. IEDS: Exploring an Intelli-Embodied Design Space Combining Designer, AR, and GAI to Support Industrial Conceptual Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–25.
[81] Hongbo Zhang, Yifei Wu, and Pei Chen. 2025. IEDS in Practice: A Comparative Study of an Intelli-Embodied Design Space Combining AR and GAI for Conceptual Design. *Design and Artificial Intelligence* (2025), 100049.
[82] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*. 3836–3847.
[83] Yuzhen Zhang, Ruixiang Han, Ran Zhou, Peter Gyory, Clement Zheng, Patrick C Shih, Ellen Yi-Luen Do, Malte F Jung, Wendy Ju, and Daniel Leithinger. 2024. Wizard of props: mixed reality prototyping with physical props to design responsive environments. In *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–15.
[84] Zheng Zhang, Ning Li, Qi Liu, Rui Li, Weibo Gao, Qingyang Mao, Zhenya Huang, Baosheng Yu, and Dacheng Tao. 2025. The other side of the coin: Exploring fairness in retrieval-augmented generation. *arXiv preprint arXiv:2504.12323* (2025).
[85] Zheng Zhang, Qi Liu, Hao Jiang, Fei Wang, Yan Zhuang, Le Wu, Weibo Gao, and Enhong Chen. [n. d.]. FairLISA: Fair User Modeling with Limited Sensitive Attributes Information. In *Thirty-seventh Conference on Neural Information Processing Systems*.
[86] Ada Yi Zhao, Aditya Gunturu, Ellen Yi-Luen Do, and Ryo Suzuki. 2025. Guided Reality: Generating Visually-Enriched AR Task Guidance with LLMs and Vision Models. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–15.