HI-TAM, a hybrid intelligence framework for training and adoption of generative design assistants

Yaoli Mao[1]*, Janet Rafner[2,3], Yi Wang[4] and Jacob Sherson[2]

[1]Autodesk Inc., San Francisco, CA, United States, [2]Center for Hybrid Intelligence, Department of Management, School of Business and Social Science, Aarhus University, Aarhus, Denmark, [3]Aarhus Institute for Advanced Studies, Aarhus University, Aarhus, Denmark, [4]Autodesk Research, San Francisco, CA, United States

The Hybrid Intelligence Technology Acceptance Model (HI-TAM) presented in this paper offers a novel framework for training and adopting generative design (GD) assistants, facilitating co-creation between human experts and AI systems. Despite the promising outcomes of GD, such as augmented human cognition and highly creative design products, challenges remain in the perception, adoption, and sustained collaboration with AI, especially in creative design industries where personalized and specialized assistance is crucial for individual style and expression. In this two-study paper, we present a holistic hybrid intelligence (HI) approach for individual experts to train and personalize their GD assistants on-the-fly. Culminating in the HI-TAM, our contribution to human-AI interaction is 4-fold including (i) domain-specific suitability of the HI approach for real-world application design, (ii) a programmable common language that facilitates the clear communication of expert design goals to the generative algorithm, (iii) a human-centered continual training loop that seamlessly integrates AI training into the expert's workflow, (iv) a hybrid intelligence narrative that encourages the psychological willingness to invest time and effort in training a virtual assistant. This approach facilitates individuals' direct communication of design objectives to AI and fosters a psychologically safe environment for adopting, training, and improving AI systems without the fear of job-replacement. To demonstrate the suitability of HI-TAM, in Study 1 we surveyed 41 architectural professionals to identify the most preferred workflow scenario for an HI approach. In Study 2, we used mixed methods to empirically evaluate this approach with 8 architectural professionals, who individually co-created floor plan layouts of office buildings with a GD assistant through the lens of HI-TAM. Our results suggest that the HI-TAM enables professionals, even non-technical ones, to adopt and trust AI-enhanced co-creative tools.

KEYWORDS

generative design, co-creativity, hybrid intelligence, human-centered AI, architecture

1 Introduction

Generative Design (GD) has rapidly emerged as a powerful Artificial Intelligence (AI) enhanced design paradigm enabling human experts (e.g., architects) to augment their creativity and accelerate the design processes by suggesting new ideas and improving design quality in co-creation (Kazi et al., 2017; Keshavarzi et al., 2021; Demirel et al., 2023).

However, the adoption and integration of AI-support technologies into creative industries face significant challenges including the potential

for job displacement, deskilling, concerns over the transparency and accountability of AI systems, and the need for highly personalized assistance (Rafner et al., 2022b; Inie et al., 2023; Rafner et al., 2023). Additionally, up to 90% of AI-related projects fail to some extent in the real-world implementation stage (Boston Consulting Group, 2020) and in general only 10% of organizations are achieving significant financial benefits with AI (Boston Consulting Group, 2020). Given the specialized skills involved in GD, introducing AI into professional workflows in the creative design field is likely to create tension.

These challenges can be divided into two areas to conquer: (1) continually evolving human-centered interaction with AI and (2) holistic development and deployment frameworks taking into account domain-specific and organizational aspects of the introduction of the technology in professional work settings. The former can be addressed by incorporating principles from human-centered AI (HCAI) such as an emphasis on user control (Shneiderman and Maes, 1997; Shneiderman, 2020), mutual learning from the field of Hybrid intelligence (HI) (Dellermann et al., 2019), and active learning and feedback loops from the field of Interactive Machine Learning (IML) (Amershi et al., 2014). The latter is addressed particularly well by the HI framework which presents an integrated way of deploying human-centered AI solutions with appropriate information system management methodologies to optimize business, societal and human values for specific expert knowledge domains (Rafner et al., 2022a; Sherson et al., 2023).

In this paper, we present two studies investigating *how the HI approach could be designed and helps human experts build a partnership with a GD personal assistant in design co-creation*. We define partnership as a human expert's willingness to contribute to the co-creative tool during and after co-creation. Culminating in the presentation of the Hybrid Intelligence Technology Acceptance Model (HI-TAM), our contribution to human-AI interaction is 4-fold. First, we demonstrate the suitability of the HI approach in a specific knowledge domain by considering professionals' expectations and preferences for GD assistant across a broad range of workflow scenarios. This forms the basis for applying the HI approach to real-world HI application design as a starting point. Second, we have developed a novel grammar-based method for constructing a common language between human and algorithm allowing for the explicitation of individual experts' "design goals" and a method for feeding these in real-time into the generative algorithm. Third, we apply the human-centered AI interaction design principles to seamlessly integrate AI training into the expert's task workflow. These two algorithmic and human-computer interaction advances enable individuals to directly communicate design preferences and goals to AI and gradually grow an accumulated and personalized design knowledge library. Finally, we address the willingness to spend time and effort training such a virtual assistant by embedding the process in an HI narrative designed to create a psychologically safe space for co-creation without the fear of job-replacement that is so often an underlying perception of rapidly advancing AI.

2 Related work

2.1 Generative design tools—the technology

Generative Design (GD) (Shea et al., 2005) is a process where designers leverage AI to explore a broad array of options, achieving high-quality results that balance multiple objectives.

Various technologies can be used for implementing GD including simulation, optimization (e.g., genetic algorithm), deep learning models and a combination of those (e.g., Shea et al., 2005; Oh et al., 2019). GD has been applied in many domains, especially architecture design (Caetano et al., 2020; Nagy et al., 2018) and product design (Alcaide-Marzal et al., 2020). Tools have been developed to support GD processes, including the GD toolset for Autodesk Revit,[1] the Refinery toolkit for Dynamo,[2] and Grasshopper for Rhinoceros 3D.[3]

A GD process allows for computational expression of design goals through a parametric model and automatic generation of numerous design options, in contrast to traditional design processes where designers must internalize all design goals and constraints to create a single solution. The GD process is human-AI collaborative in nature as the algorithm can report back to the user promising design options for further analysis and refinement; the user can also revise their input parameters. Research along this line has so far been mostly focused on the representation of the design space, generation and evaluation of solutions, search algorithms and visualization of design options, with little discussion on the human-AI collaboration and user personalization.

In Study 1, we discuss GD broadly in different workflow scenarios of the architectural building lifecycle and gather information on individuals' expectations and preferences for using the technology.

In Study 2, we implemented a prototype system demonstrating a preferred GD workflow from Study 1 in a simplified design problem, with the intention to study human-AI co-creation behavior in a controllable research setting and discover useful insights that can be translated to improve workflows in practical GD software.

2.2 Virtual assistants—the application

Broadly, the goal of a virtual personal assistant is to provide support to users in a personalized and context-aware manner, thereby enhancing their productivity, satisfaction, and overall wellbeing (Kepuska and Bohouta, 2018; Dubiel et al., 2018). As an early example, the Microsoft Paperclip, also known as Clippy, was a virtual assistant introduced in the late 1990s to assist with tasks such as creating and formatting documents in Microsoft Office. It could be accessed by clicking on a small paperclip icon in the Office application window. However, its implementation was widely criticized for being intrusive, annoying, and unprofessional due to

---

1   http://www.autodesk.com/products/revit/overview

2   dynamobim.org/refinery-toolkit/

3   http://www.rhino3d.com/6/new/grasshopper/

unsolicited messages and its cartoonish appearance and behavior (Maedche et al., 2016). With the advancement of automated test and voice recognition and processing (Yuan et al., 2018), virtual assistants and chatbots have proliferated recently in both the commercial and private spheres providing helpful input and innovative interactions but always within quite restricted domains.

In contrast, ChatGPT seems to provide human-like conversation and assistance as a virtual assistant. However, in terms of output credibility there are still significant pitfalls (Borji, 2023) and in terms of user personalization, in its own words "*As an AI language model, I do not have the ability to adjust to individual preferences in the way that humans do... I can be trained on large datasets of text to learn how people typically communicate, which can inform my responses to some extent.*" This lack of information about individual users' preferences, needs, or domains and contexts can lead to generic or irrelevant responses that do not address the user's specific concerns or objectives (Sherson and Vinchon, 2024). Furthermore, because ChatGPT does not have a memory of previous interactions with a particular user beyond the current session ("*I do not store or transmit any personally identifiable information unless specifically instructed to do so by the user*"), it may not be able to provide a consistent and coherent conversation or maintain a sense of continuity in interaction over time.

The importance of user feedback and preferences in designing virtual assistants, as well as the need for more advanced techniques such as personalized recommendation systems and user modeling, is highlighted by the limitations of ChatGPT. ChatGPT's inability to communicate residual uncertainty in its responses creates algorithmic overconfidence and a lack of transparency for the user. The study defines a GD virtual assistant as an AI system trained to assist human experts by generating design solutions based on their inputs and preferences. In both Study 1 and Study 2, the HI narrative is introduced as part of our HI approach to clarify that the AI in question is fallible in the beginning and can only improve with the user's continual training and feedback.
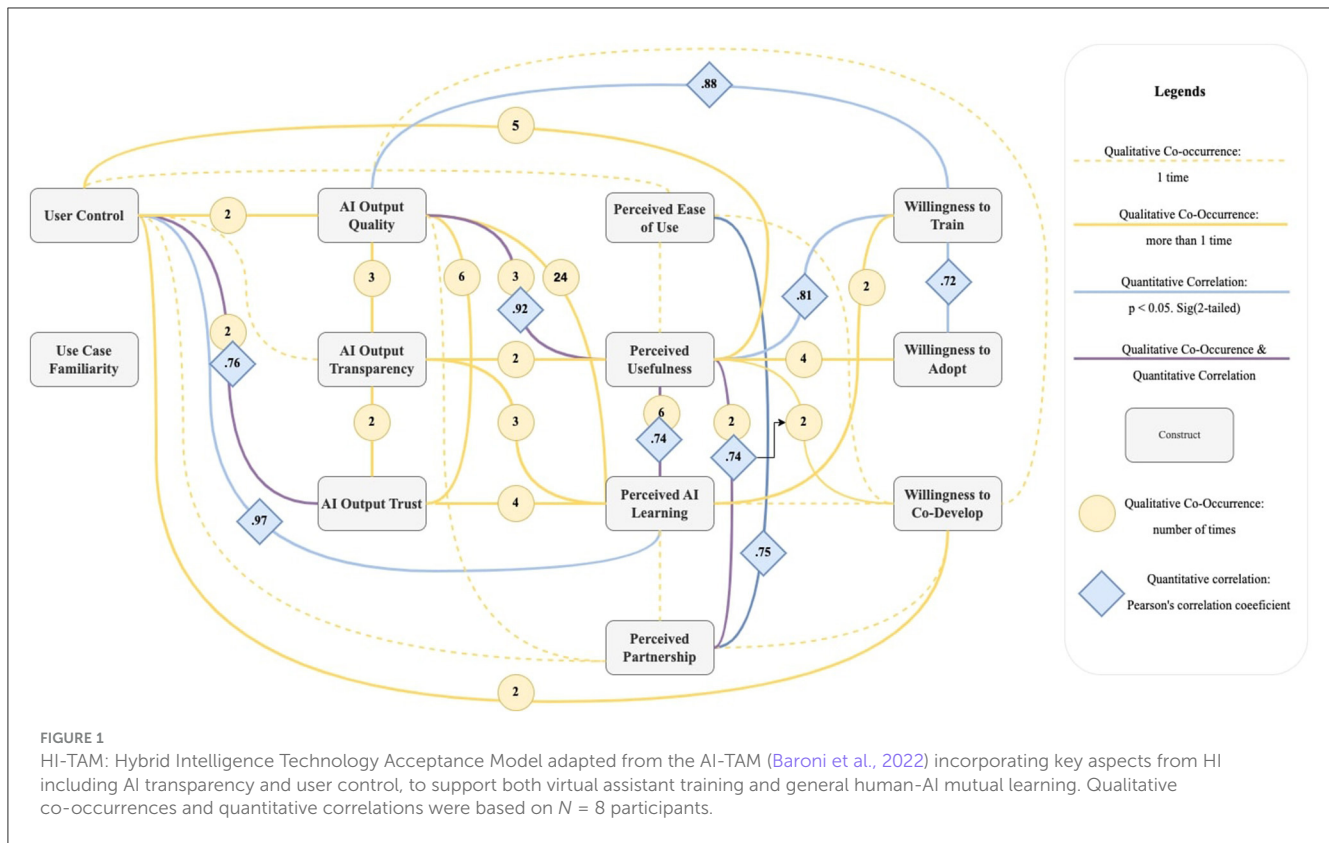
2.3 Hybrid intelligence

Hybrid Intelligence (HI) is often defined in broad terms (Akata et al., 2020; Prakash and Mathewson, 2020), though a clearer formulation was proposed by Dellermann et al. (2019), focusing on three key criteria: collectiveness, solution superiority, and mutual human-AI learning. Despite this, practical implementation guidelines and distinction from other forms of human-AI interaction remains challenging. Recent efforts have shifted HI from a conceptual term toward a holistic, interdisciplinary framework that supports both development and deployment, integrating the organizational strategies like deskilling and upskilling resulting from AI system implementation (Rafner et al., 2022b). Additionally, HI blends interface design considerations with established principles of information systems and human resource management, such as business process re-engineering and psychological safety and employee co-creation, forming twelve distinct design and deployment dimensions (Sherson et al., 2023).

As HI extends into the creation of generative AI virtual assistants, a change management framework to HI has emerged (Sherson, 2024) emphasizing systematic end-user empowerment through a series of expert rated microinnovations. Finally, preliminary design guidelines (Sherson et al., 2024) separate HI at the interface, the organizational and the societal/impact levels. These guidelines emphasize incorporating human-centered goals into both the narrative and visual aspects of AI applications, fostering psychological safety and encouraging user engagement in the co-creation process. This approach supports continuous adaptation and innovation post-deployment (Sherson et al., 2024).

2.4 HI-TAM—the analysis method

Researchers in the field of information systems management have developed models to understand factors influencing technology acceptance, including the widely used Technology Acceptance Model (TAM) which links user acceptance of technology to perceived usefulness and ease of use (Davis, 1989). The TAM has been applied to various contexts such as mobile apps and e-commerce systems. An extension of TAM, the AI-TAM, has been proposed to evaluate user acceptance and collaborative intention in human-in-the-loop AI applications (Baroni et al., 2022). Human-in-the-loop AI involves integrating human input and feedback into AI systems to improve their accuracy and efficiency (Zanzotto, 2019). The AI-TAM includes constructs related to human-AI interaction such as functioning, quality, trust, familiarity, and collaborative intention. Despite adding the collaborative variable, the scenario investigated in the AI-TAM development work involved single-shot interactions with a fully developed product, which is far from the continuous mutual learning relationship with an HI system.

Although originally perceived as a useful framework for understanding the early stages of product design, the existing TAM approach has been heavily criticized concerning product design (Salovaara and Tamminen, 2009). Moreover, subsequent literature such as the New Product Development TAM (NPD-TAM) (Diamond et al., 2018) as well as case studies of developing smart payment card and adopting virtual assistants (Song et al., 2019; Gunadi et al., 2019) have added no new variables to provide insight into the dynamic development and training process. In order to capture the dynamic mutual learning process of HI systems, we here introduce the HI-TAM (see Figure 1 for an overview and Table 1 for detailed variable definitions). Taking the AI-TAM as a point of departure, we added the process variables of *user control*, *AI output transparency*, *perceived partnership* and replaced the output variables *collaborative intention* and *behavioral intention* with *willingness to train*, *willingness to co-develop* and *willingness to adopt*. These variables were inspired by fundamental principles of HI (mutual learning), HCAI (pursuit of high levels of automation and control simultaneously), and IML (continuous interactivity) as well as the Co-Creative Framework for Interaction Design (COFI) which emphasizes the importance of establishing a partnership between end users and AI support tool (Rezwana and Maher, 2022).

FIGURE 1
HI-TAM: Hybrid Intelligence Technology Acceptance Model adapted from the AI-TAM (Baroni et al., 2022) incorporating key aspects from HI including AI transparency and user control, to support both virtual assistant training and general human-AI mutual learning. Qualitative co-occurrences and quantitative correlations were based on $N = 8$ participants.

3 Research question and study design

The primary research question guiding this work is: *How can the HI approach be designed to help human experts build a partnership with a GD personal assistant in design co-creation?* This question motivates both Study 1 and Study 2.

To evaluate the proposed HI approach and the underlying HI-TAM framework, we conducted two studies using a descriptive and correlational design with architectural professionals. Study 1 consisted of individual surveys and brainstorming to understand architectural professionals' expectations and preferences for personalizing their GD assistants. The focus was on how GD assistants and human experts divide and integrate their tasks in various design-build scenarios, reflecting typical activities throughout the architectural building lifecycle. Study 2 was a case study (Creswell and Creswell, 2017) that explored an early conceptual design scenario, where participants designed a floor plan for a typical office building. They interacted with a GD assistant prototype through an HI approach tailored to the collaboration preferences identified in Study 1. Both qualitative and quantitative data were collected to examine the HI-TAM framework, with exploratory data analyses supporting the links shown in Figure 1.

Given the multidisciplinary nature of human-AI co-creativity, spanning cognitive science, psychology, and HCI, this research adopts a mixed-methods approach (Van Turnhout et al., 2014). The quantitative analysis in Study 2 involved seven participants using a generative AI design tool. Although the small sample limits statistical significance, it revealed valuable initial usage trends.

The qualitative analysis is more extensive, drawing from both studies. Study 1 explored various use cases for generative AI in architecture, while Study 2 analyzed participants' think-aloud interviews during the task, along with pre- and post-task responses. This qualitative data provides insights into their cognitive processes and the impact of AI on their design methods. By synthesizing qualitative observations with quantitative findings, this research presents a comprehensive view of how generative AI is integrated into architectural practice, highlighting both broad usage patterns and individual creative interactions.

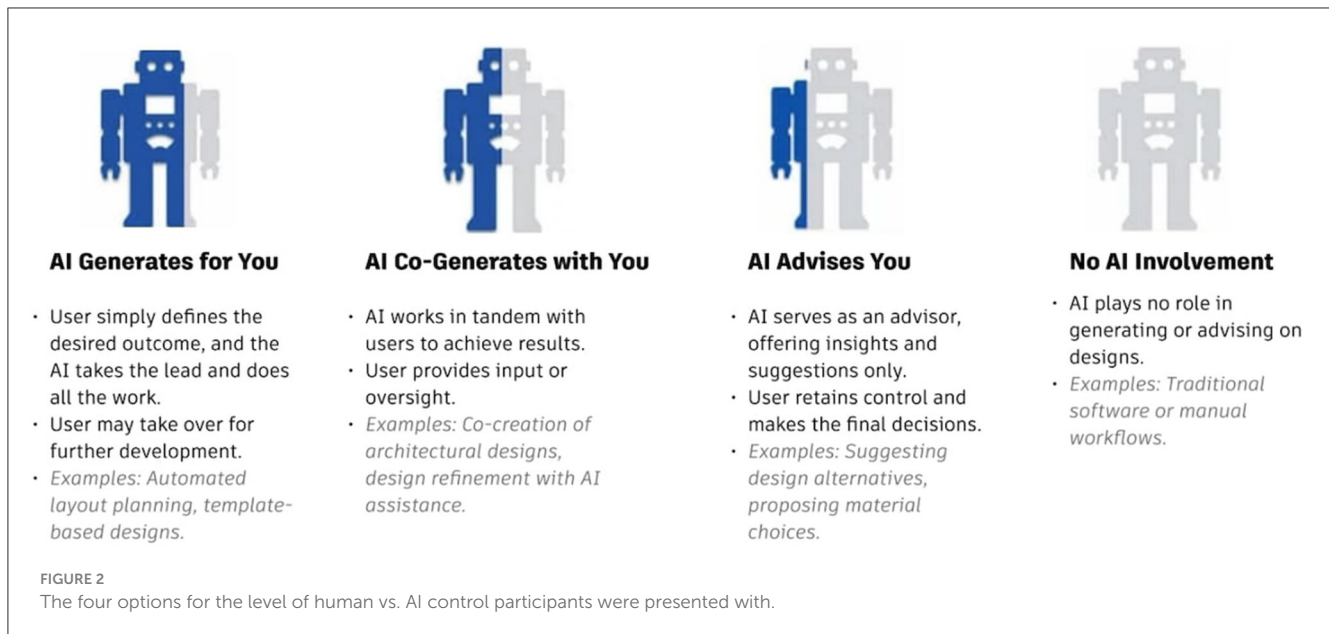# 4 Study 1: professionals' expectations and preferences for personalizing GD assistants

4.1 Materials and methods

Study 1 was conducted during a 60-min workshop at a major architecture, engineering, and construction industry conference, where professionals voluntarily participated in exploring the future of generative AI in their workflows. The workshop consisted of two integrated parts (see the detailed worksheets and questions in Supplementary Figure S1). In part one, architectural professionals were surveyed to choose their preferred level(s) of control for GD assistants throughout the building lifecycle from the following four options (Figure 2).

These four options were chosen specifically for their relevance to GD assistants. However, they were inspired by literature which differentiates relationships between human and machine intelligent

TABLE 1 Constructs with respective qualitative codes and quantitative measures.

Construct	Qualitative codes	Quantitative measures (data range)
Familiarity	Both familiarity toward AI, and familiarity toward AI and Generative Design tools.	I don't use it—Expert (1–6) • Proficiency with machine learning or artificial intelligence for automated design. • Proficiency with generative design tools such as Grasshopper or Dynamo for automated design.
User control	The level of control and autonomy the user feels they have over the tool. It also includes comments about input features that the user appreciates or feels are lacking. This code could be applied to any mention of the user's ability to direct or influence the tool's behavior or output.	Adapted from controllability (Shneiderman and Maes, 1997; Hartmann, 2009; Oh et al., 2018) (1–5) • I am able to let it do its work. • I do not feel out of control in working with it.
AI output quality	The system provides accurate and complete information. This construct refers to subjective and objective evaluations of the quality of concrete and identified outputs of the system.	Adapted from ability (Mayer and Davis, 1999) (1–5) • It is very capable of performing its job. • It has specialized capabilities that can increase our performance.
AI output transparency	The degree to which the output generated by an artificial intelligence system is understandable and interpretable to the human user. It pertains to the system's ability to provide clear and coherent feedback or explanations to the user on how it arrived at a particular output	Adapted from comprehensibility (Shneiderman and Maes, 1997; Jameson, 2009; Oh et al., 2018) (1–5) • I understand its inputs and outputs. • I'm familiar with it.
AI output trust	The user finds it predictable and trustworthy, and takes the assertions as valid and true.	Adapted from predictability (Shneiderman and Maes, 1997; Jameson, 2009; Oh et al., 2018) (1–5) • I am certain about how my interaction with it affects the generated designs. Adapted from integrity (Mayer and Davis, 1999) • Sound principles seem to guide its behaviors. Adapted from trust (Jian et al., 2000) • It is reliable. • I can trust it.
Perceived ease of use	Using the system does not require a lot of mental effort, the system is easy to use and understand.	Adapted from TAM - perceived ease of use (Mayer and Davis, 1999) (1–5) • Working with it is easy for me. • The interface is difficult to understand.
Perceived usefulness	Using the system helps the user's work, achieve tasks, and make better choices. The user finds the system useful. This construct refers to subjective evaluations of the system as a whole. This includes when the tool provides a positive output which is surprising, creative, or novel to the user.	Adapted from TAM—perceived usefulness (Mayer and Davis, 1999) (1–5) • I feel more involved in the generative design process working with it compared to other generative design tools. • I feel responsible for marking my likes and dislikes working with it. • It is clear to me why I need to work with it. • Working with it makes me feel in control of the design process. • I think it is useful. • I think it is useless to train it with my likes, dislikes and reasons.
Perceived partnership	The user's subjective perception of the degree to which an AI is perceived as a collaborative and cooperative partner in the interaction. In other words, it describes how much the user feels that the AI assistant is working with them, rather than just functioning as a tool that responds to their commands.	• Overall, how much do you feel that you are building a meaningful partnership i.e., task distribution and integration and tool customization? (1–5)
Perceived AI learning	The user's perception of an AI's ability to learn from their interactions and adapt to their needs and preferences. It refers to the degree to which the user believes the AI assistant is capable of improving its performance over time by learning from the user's behavior, feedback, and input. This code covers anything that is descriptive of the process over several inputs.	• To which extent did you feel that your interactions with the tool yielded improved designs? (1–5) • Please rate on a scale of 1–5, how much do you see your contributed reasons are reflected in the newly generated designs by the tool? [Subjective goal satisfaction (1–5)] • Objective goal satisfaction: see details in Section 5.1.1 (0–1)
Willingness to train	The user's willingness and openness to invest time and effort in training an AI tool to better understand their needs and preferences. It refers to the degree to which the user is willing to provide input and feedback to the AI tool, with the goal of improving its performance and adapting it to their specific needs.	• Would you invest your time in training to use this tool in your work practice? (1–5) • The number of design goals.
Willingness to adopt	The user's intention and openness to use and incorporate the AI tool into their workflow. This also includes alternative or future uses of the tool.	• Would you use such a tool in your work practice? (1–5)
Willingness to co-develop	The user's willingness and openness to actively provide feedback and input to help shape the development and improvement of the AI tool. This also includes suggestions for feature improvements.	• Please list as many suggestions as possible to improve the tool (the number of suggestions).

The four options for the level of human vs. AI control participants were presented with.

systems (Rafner et al., 2022b; Berditchevskaia and Baeck, 2020) such as human in the loop vs. human on the loop.

In part two, they were guided to brainstorm personalized GD assistant solutions for their specific use cases and preferences. These expectations, solutions and preferences are critical in designing for training and adoption of GD assistants.

A total of 41 individuals participated in the study. All of them completed part 1 and 26 completed part 2. The participants included architects, CAD/BIM managers, engineers, and managers/directors from relevant industry fields, with 1–33 years (mean = 15.7, median = 16), and varying levels of experience with generative AI solutions at work: 26% did not use it, 43% were just getting started, and 31% had some to intermediate to expert experience. Most were from North America (80%), with the remainder from South America, Europe, and Asia. All participants provided informed consent before the workshop.

4.2 Analysis

We present participants' preferred level of control for each architectural lifecycle phase, using the raw count for each of the four options (see bars in Figure 3). Furthermore, we distinguish *AI Co-Generates with You* from *AI Generates for You* and *AI Advises You* because the former represents a closely coupled, integrated, and iterative collaboration process between AI and humans, as advocated by the HI approach. In contrast, the latter two options represent loosely coupled and clearly divided work or responsibilities, resulting in task separation between humans and AI.

To quantify this, we assign a score of 2 to *AI Co-Generates with You*, a score of 1 to both *AI Generates for You* and *AI Advises You*, and a score of 0 to *No AI Involvement*. For each phase in the building lifecycle, we calculate the average score of all participants' choices as the human-AI integration score, along with

the standard deviation. Additionally, we computed a cumulative human-AI integration score across the building lifecycle for each participant and then performed Pearson's correlation analysis to investigate the relationships between this integration score and participants' years of industry experience, as well as their level of familiarity with generative AI at work.

4.3 Results

With respect to Part 1 of Study 1, relatively higher human-AI integration was observed in the early phases of the building lifecycle, such as conceptual design and schematic design, whereas integration was lower in later phases like design development & refinement, procurement, and construction (see the dotted line in Figure 3). Specifically, human-AI integration was significantly higher in site analysis & feasibility study ($p = 0.03 < 0.05^*$), conceptual design ($p = 0.002 < 0.05^{**}$), and schematic design ($p = 0.001 < 0.05^{***}$) compared to that in construction. These early stages involve more information gathering, problem defining, divergent thinking and tightly coupled considerations from different disciplines (e.g., aesthetic and technical), whereas later stages tend to involve more well-defined tasks that fall into more specialized expertise.

Additionally, we identified a significant correlation between the human-AI integration score and participants' familiarity with generative AI at work ($r = 0.39, p = 0.02 < 0.05^*$). However, no significant correlation was found between the human-AI integration score and years of industry experience ($p = 0.2 > 0.05$). This suggests that professionals with greater familiarity with generative AI tend to prefer more closely integrated collaboration with AI, rather than loosely separated sequential task division.

With respect to Part 2 of Study 1, we delved into participants' specific preferences regarding how they personalize the integration of GD assistant within their selected workflow scenarios. Methods preferred by participants for training and improving GD assistants
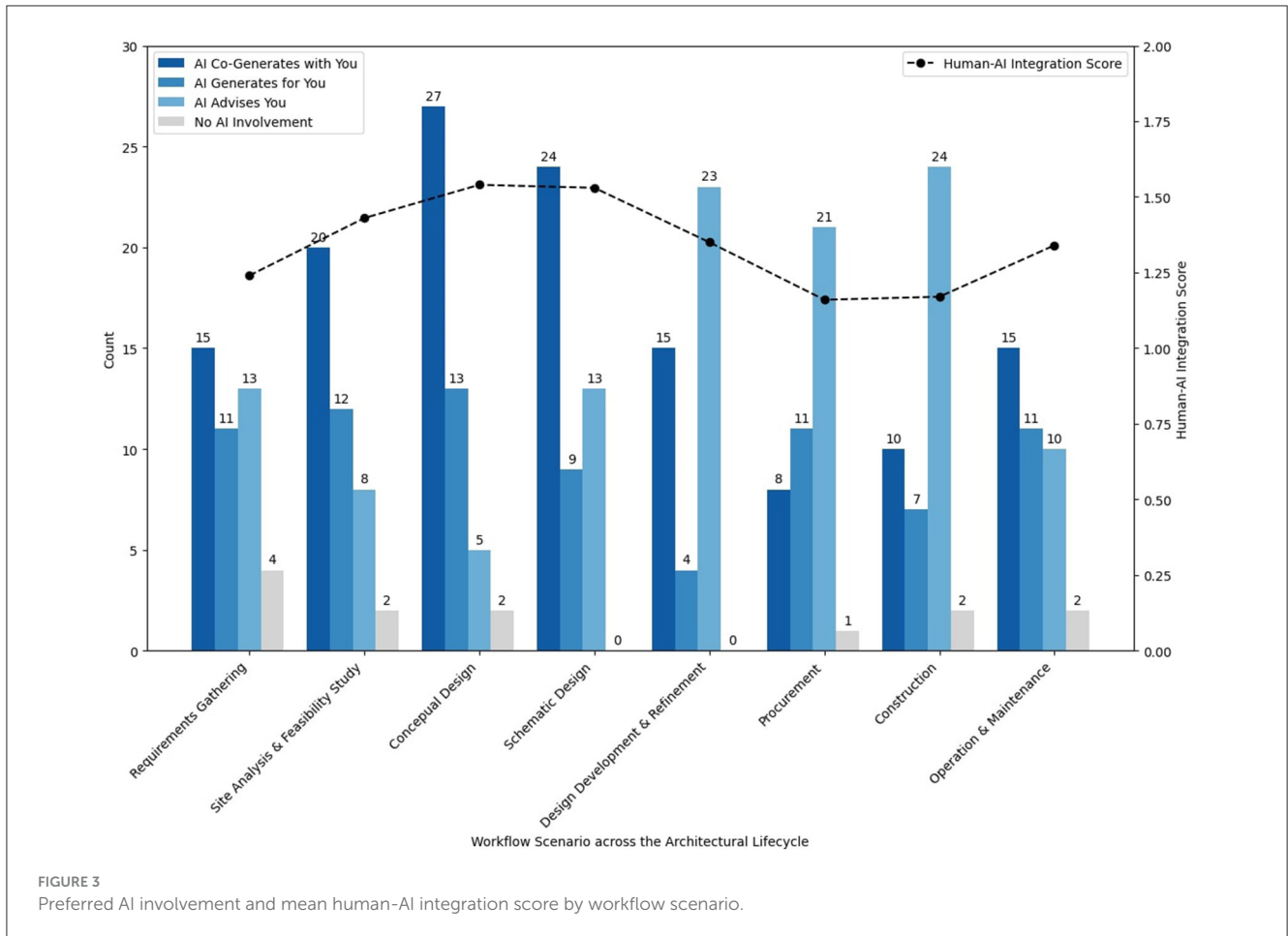
FIGURE 3
Preferred AI involvement and mean human-AI integration score by workflow scenario.

over time include properties and requirements of the final output (16, 62%), variations of output (16, 62%), examples of the final output (14, 54%), design fine-tuning (11, 42%), and interactive feedback (10, 38%). All 26 participants that completed part 2 brainstorming also reported a preference for maintaining and leveraging their past projects and outcomes as proprietary inputs to personalize their GD assistants to better align with their preferences. As measures of success in co-creation with GD assistants, top priorities include accuracy (24, 92%), efficiency (22, 85%), and time (20, 77%). These preferences are well-exemplified by one participant's approach in an early design scenario (Figure 4), where project requirements are imported, goals are specified, and interactions with the GD assistant involve selecting from generated options and providing feedback to refine and train the assistant to match individual tastes.

5 Study 2: case study on HI-TAM with a GD assistant prototype

In Study 2, we chose conceptual design as the workflow scenario where participants showed the strongest preference for human-AI integration observed from Study 1, allowing us to further investigate and evaluate the HI Approach.

5.1 Hybrid intelligence design approach

We now describe the 3 components of the *Hybrid Intelligence (HI) Approach* (see Figure 5 for an overview).

5.1.1 Programmable common language for representing human experts' design goals to GD assistant

In the realm of building layout design, a significant portion of the design requirements pertains to the types of spaces and their respective spatial and topological relationships. Inspired by knowledge representation languages used for ontology authoring and graph database, such as Resource Description Framework (RDF) (Arenas et al., 2009), we use sentences of the following syntax to represent design goals:

$$design\_goal \rightarrow subject \ \mathbf{unary\_relation} \mid subject$$
$$\mathbf{binary\_relation} \ object$$
$$subject \rightarrow \mathbf{space\_type}$$
$$object \rightarrow \mathbf{space\_type}$$

Similar to RDF, this programmable language focuses on characterizing relations between entities. In our case the entities are types of spaces, and the relations are
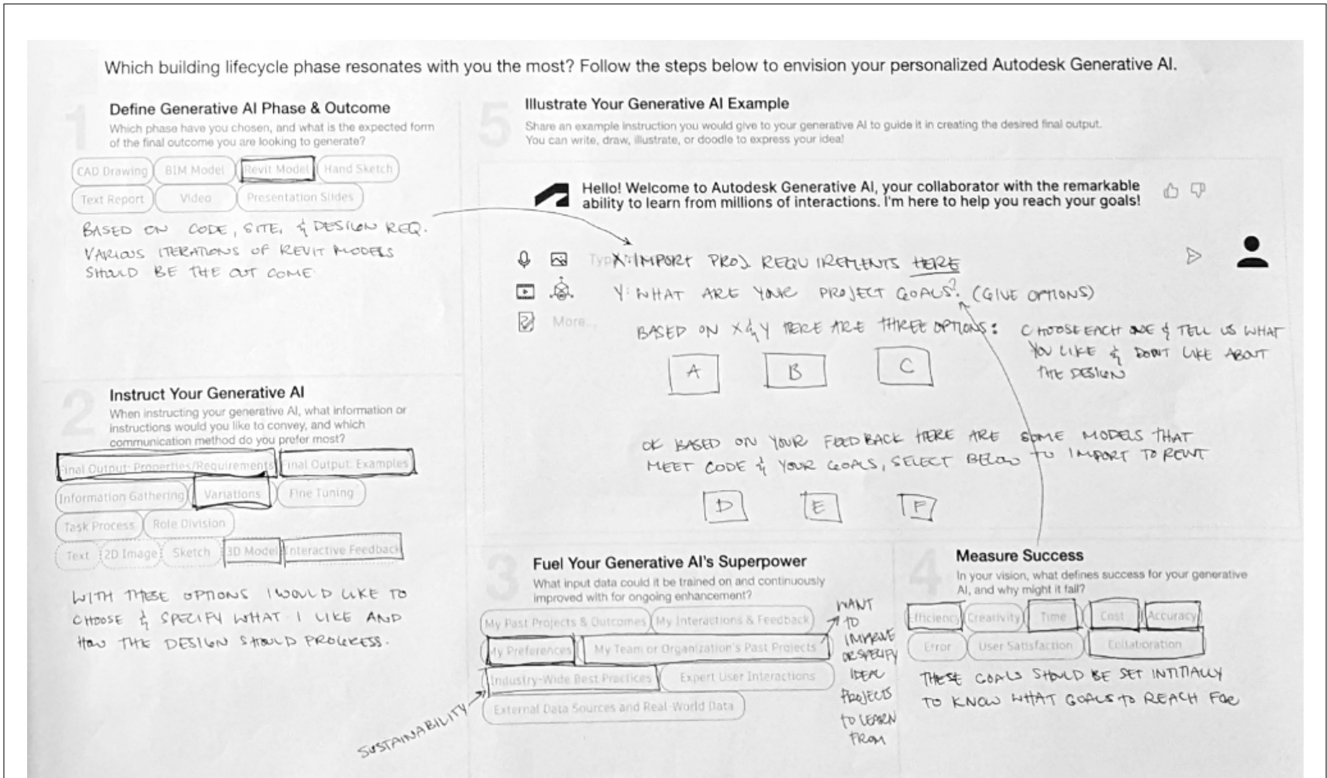
FIGURE 4
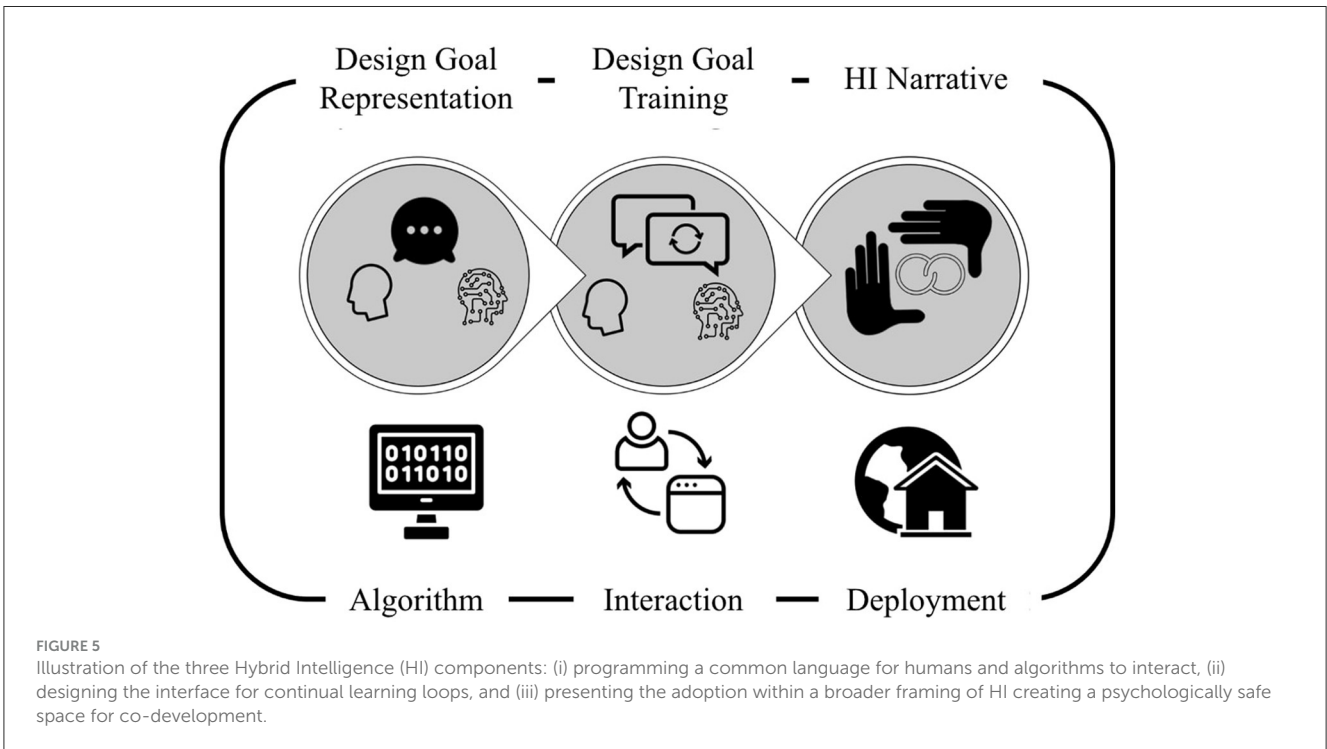An example of a participant's approach to personalizing GD assistant in co-creation.

FIGURE 5
Illustration of the three Hybrid Intelligence (HI) components: (i) programming a common language for humans and algorithms to interact, (ii) designing the interface for continual learning loops, and (iii) presenting the adoption within a broader framing of HI creating a psychologically safe space for co-development.

specifically spatial relations. For this study, **unary_relation** includes `at the center`, `on west`, `on east`, `on south`, `on north` and **binary_relation** includes `are close to`, `are away from`, `surrounded by`, `share the same door orientation with`. We also consider an office building setting for the layout design task, so **space_type**

include `meeting room`, `office`, `open_desk_space`, `lunch_space`, etc.

We compute an objective goal satisfaction score (as opposed to a subjective goal satisfaction score rated by participants) of each design goal by mapping each **unary_relation** and **binary_relation** to a hand-crafted heuristic function,[4] which takes a layout configuration as input and returns a real number in the range of $[0, 1]$ as output. The closer the value is to 1, the more satisfied the design goal. The *objective goal satisfaction* $\mathcal{GS}$ of a layout configuration $\mathcal{L}$ given a set of design goals G is computed as follows

$$\mathcal{GS}_{\mathcal{L}}(G) = \prod_{g \in G}(1 - \frac{1 - s(g)}{|G|})$$

where $s(g)$ is the satisfaction score of design goal g, and $|G|$ denotes the number of design goals.

The language is easily extendable or customizable by introducing more relations and space types. Each relation is associated with a function that returns a value in $[0, 1]$—in the sense the relations can be viewed as predicates under fuzzy logic semantics (Novák, 1987). It is consequently possible to extend this language with logical operators such as conjunction, disjunctions and implications. We leave this to future work for simplicity.

5.1.2 Continual learning loops for training the GD assistant with design goals

Leveraging the programmable design goals as a starting common language between the GD assistant and a human user, we created a training mechanism for users to iterate on their design goals and the tool-generated designs in feedback loops through the following steps (Figure 6): (i) A user reacts to a tool-generated design by marking spatial objects their like or dislike. (ii) The tool carries on the "conversation" by prompting the user with a popup window to select reason(s) for their likes or dislikes. The selected reasons are added into the tool as design goals following the programmable language. If the provided list does not capture their reason(s), they can choose "None of Them" and verbally describe their reasons, which are not included as input for the tool in the current task. (iii) The user repeatedly marks likes and dislikes and select reasons until they feel satisfied. They can also revise any added design goals if they detect any conflicts among them. (iv) Upon user request, the tool is invited to generate another round of designs, taking all the design goals including the updated ones from the previous rounds into consideration. (v) The user selects one preferred design and repeats steps (i) through (iv).

This training mechanism is based on two major sources of inspiration. One is the typical design critique process that architecture students would learn in their design studio and professional architects would practice at their daily work (Oh et al., 2013). The second is IML where the system is tightly coupled with the human in the loop of model training and thus results

in "more rapid, focused, and incremental model updates than in the traditional machine-learning process" (Fails and Olsen, 2003; Amershi et al., 2014; Dudley and Kristensson, 2018). Here are a few examples of how we translate these human-in-the-loop design opportunities into the training mechanism design: defining new constraints inspires expressing design goals, correcting errors in the training data inspires marking dislikes on the design, fine tuning parameters inspires adjusting previous design goals. These design considerations are also in line with the principles outlined in the *Guidelines for Human-AI Interaction* which seeks to guide interaction over time beyond one shot usage, including learning from user behavior, updating and adapting cautiously, encouraging granular feedback and conveying the consequences of user actions (Amershi et al., 2019).

5.1.3 HI narrative for nurturing partnership between human experts and the GD assistant

According to Rafner et al. (2022a), one critical issue that limits companies from successfully adopting AI solutions is employees' fear toward job automation and replacement, and thus requires a thoughtful deployment of these solutions into the professional work context. In their case study (Rafner et al., 2022a), a HI corporate narrative was created to onboard employees to an AI-supported editing tool and facilitate their adoption willingness through tool customization based on their preferences instead of following a standardized rigid workflow. We created our HI narrative (see the complete narrative in Supplementary Table S1) by following the proposed HI-TAM as the guiding design principles and taking inspirations from (Rafner et al., 2022a)'s narrative design. Specifically, our narrative introduces the GD assistant as a partner and emphasizes that the goal was to train the GD assistant sufficiently toward building a partnership instead of achieving the best quality design. Partnership was defined as "*the distribution of sub-tasks in an integrated and customizable workflow between you and the tool*". Participants were instructed to keep customizing the GD assistant "*by telling it about your preferences until you feel that you have trained it enough*".

5.2 System implementation

A prototype system is developed to implement the programmable language of design goals, training mechanism, and the expert's task workflow, with a focus on the task of building layout design. It is intended to represent a GD assistant that facilitates human designers to explore the design space more efficiently, and to iterate faster on both their design goals and the design itself.

For simplification, the layout design task is viewed as arranging a set of spatial objects (rooms, zones, etc.) within the building interior to satisfy user-provided design goals concerning spatial relationships between object types. Given the geometric definition of each spatial object, the goal is to find a transform (i.e., position and rotation) for each object within the space, maximizing the overall satisfaction of the provided design goals. This process adheres to a set of *general constraints*, including non-overlapping, containment conditions, and circulation validity.

---

4   For example, the design goal T_1 `are close to` T_2 is associated with the function $\prod_{x \in T_1, y \in T_2} IsCloseTo(x, y)$, where $IsCloseTo(x, y) = 1 - \frac{dist_{center}(x,y)}{D_R}$. x and y are spatial objects. $dist_{center}(x, y)$ is the Euclidean distance between the center of the bounding boxes of x and y. D_R is the largest possible distance between two points in the bounded 2D space.
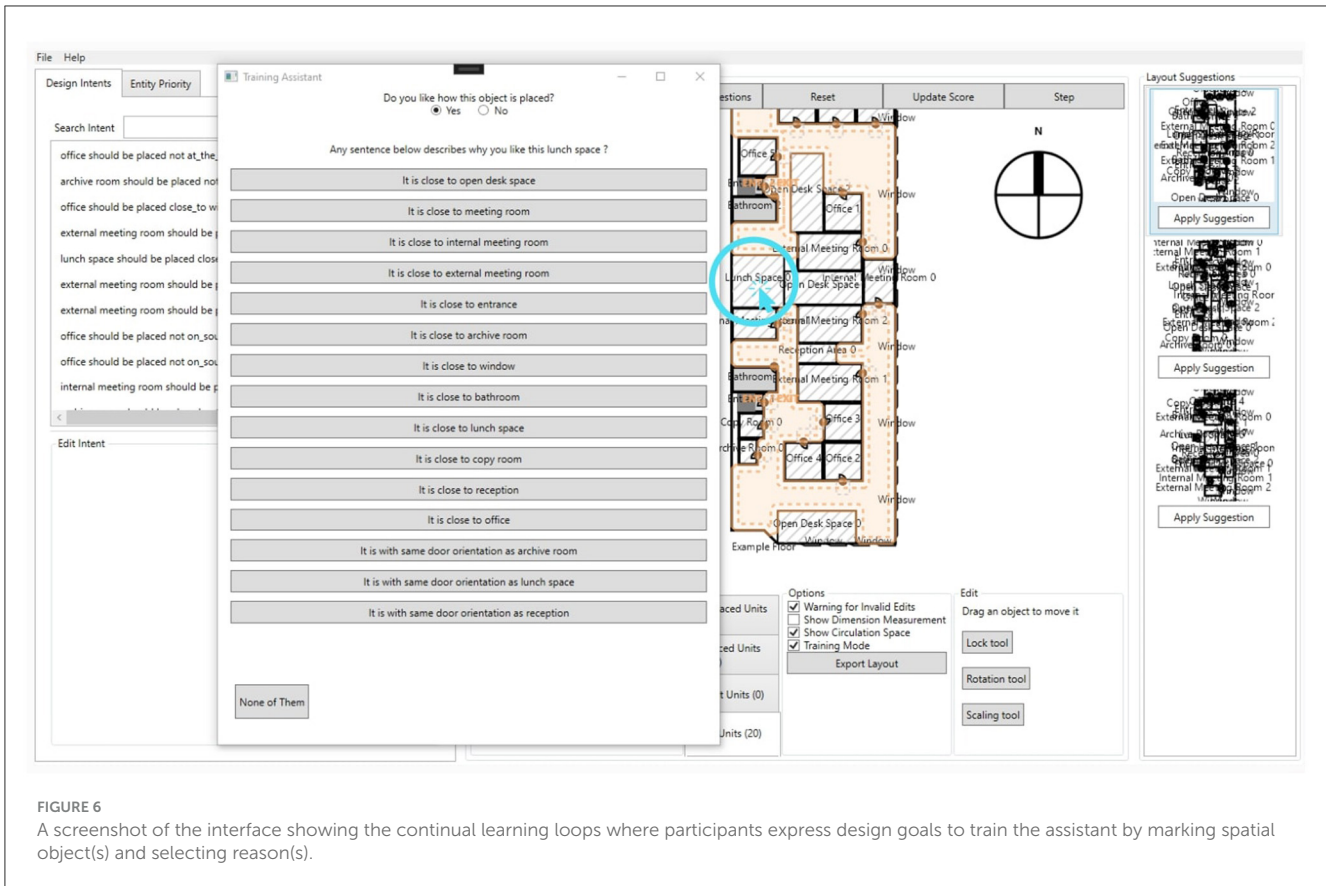
FIGURE 6
A screenshot of the interface showing the continual learning loops where participants express design goals to train the assistant by marking spatial object(s) and selecting reason(s).

To support a real-time interactive workflow enabling the completion of the design task within 10–20 min, our study employs a simple greedy search algorithm rather than more complex alternatives such as genetic algorithms or machine learning-based generative models. We emphasize that the technical method is not the primary focus of this work; instead, the results from this study can be generalized to similar systems that may utilize different technical methods to reflect design goals in the generated artifacts. The greedy method assigns transforms to each spatial object based on the highest objective satisfaction score of relevant design goals and randomized the placement order to generate diverse design options. Our study confirmed the algorithm's effectiveness in generating layout configurations that fulfill the input design goals. The following procedure describes a detailed implementation of arranging each object on the floor plan (see more system implementation details in Supplementary material S5).

---

1. **Repeat** for *N* times to generate *N* layout options:

   (a) Generate a randomly ordered list *L* of all spatial objects
   (b) **For each** spatial object *obj* in *L*:

      (1) Use the architectural grid lines to obtain a finite set of possible transforms for *obj* with regard to the current object arrangement in the space;
      (2) **if** no candidate transform is available, put *obj* in a unfit list;
         otherwise, choose the candidate transform *tr* which would result in the highest goal satisfaction degree if assigned to *obj* (with the transform assignment to all other objects unchanged);
      (3) Assign *tr* to *obj*;

5.3 Materials and methods

For each HI-TAM construct, we included both a qualitative code that describes the participants' subjective experience and a quantitative measure to evaluate the construct with a numerical range, adapted from existing instruments (Table 1).

Study 2 took place online over Zoom and each participant was compensated $60 for their one-hour participation. Figure 7 illustrates the key steps in the HI task workflow. During the task, individual participants were encouraged to think aloud (Eccles and Arsal, 2017) and evaluate iterated designs by providing a numerical score of subjective design goal satisfaction. Additionally, the GD assistant logged objective design goal satisfaction and the number of design goals. Verbal and non-verbal behaviors were recorded and transcribed. Participants also completed a pre-task survey (e.g., demographics, design experience, AI experience and AI replacement fear), a post-task survey based on various constructs of the underlying model (e.g., willingness to adopt, willingness to train, providing as many suggestions as possible to improve the GD assistant) and a closing open-ended question probing AI's agency and role through visual characterization, adapted from Koch et al. (2019). Data collected included pre- and post-task surveys, GD assistant data logs, video recordings, and transcripts (see Table 1 for full survey questions and log data).

A total of eight participants were recruited through e-mail, with the following inclusion criteria: age between 18 and 65, working professionals with at least 3 years of architectural design experience, and fluency in English. They were between the ages

of 31 and 38, with 3–16 years of architectural design experience (mean = 8.6, median = 8), and three of them were female. All had previous experience working with GD tools such as Grasshopper or Dynamo and half of them had experience with machine learning or artificial intelligence for automated design. Almost all of them (seven out of eight) expressed that in 10 years from now, a moderate amount of their current tasks in architectural design could be done by a machine instead of themselves, such as designing room layout, collaborative design review, and checking building code compliance. All participants provided informed consent before the experiment. This study was approved by the Aarhus University IRB.

5.4 Analysis

In general, we combine and compare all the qualitative and quantitative results to seek complementary validation and explanations. We used Reflexive Thematic Analysis to examine user perceptions and interactions with the tool (Braun and Clarke, 2019). All authors participated in iterative coding rounds. Initial codes were identified from the transcripts, including known codes from survey questions such as "partnership" and "user control" and five constructs from the AI-TAM (Baroni et al., 2022). Collaborative and behavior intention codes were too vague to differentiate our data so they were replaced with willingness to train, adopt, and co-develop; codes for AI output trust, transparency, and perceived AI learning were also added to account for key aspects of HI. Relevant excerpts were coded with one or two of the most appropriate codes, and co-occurrences were mapped to determine linked constructs. Excerpts were labeled positive or negative. For example, P1 stated "*They are not quite consistent in a way I like*" referring both to the negative output quality of the output as well as negative AI output trust.

For quantitative measures, we first evaluated the reliability of any constructs with at least three questions using the same scale using Cronbach's alpha (Brown, 2002). We obtained a minimum alpha value above (α > 0.7), indicating a high internal consistency among items within the same construct and thus the survey measurement can be considered reliable. We also computed Pearson's correlation (Freedman et al., 2007) to explore relationships among these constructs in comparison to the qualitative co-occurrences. Furthermore, we conducted a hierarchical clustering (Yim and Ramdeen, 2015) of willingness to adopt, train and co-develop to explore the overall types of the resulted "partnership" profiles using Ward's method with square Euclidian distance as the distance or similarity between participants.

5.5 Results

5.5.1 HI-TAM

In this section, we present our results on the qualitative and quantitative links among constructs of the HI-TAM (see summarized codes and examples ranked by times of co-occurrences in Supplementary Table S2). In total, there were 106 relevant excerpts from the transcripts, ranging from 7 to 21 per participant.

The coding frequency enabled the identification of the strongest ties between the constructs. We report qualitative co-occurrences that have at least two instances across two participants. The quantitative results are based on significant statistical correlations.

5.5.1.1 Qualitative co-occurrences

The most common co-occurrence identified was between AI output quality and perceived AI learning (24 instances), observed by six out of eight participants. Co-occurrences between AI output trust and AI output quality (six), perceived AI learning and perceived usefulness (six), perceived usefulness and user control (five), and perceived usefulness and willingness to adopt (four) were also identified. Positive constructs were nearly twice as frequent as negative ones, with most co-occurrences being positive-positive or negative-negative, meaning that when construct one had a positive connotation, construct two also had a positive connotation. Exceptions included willingness to co-develop co-occurrences which were often positive-negative as negative comments were linked to suggestions for improvement.
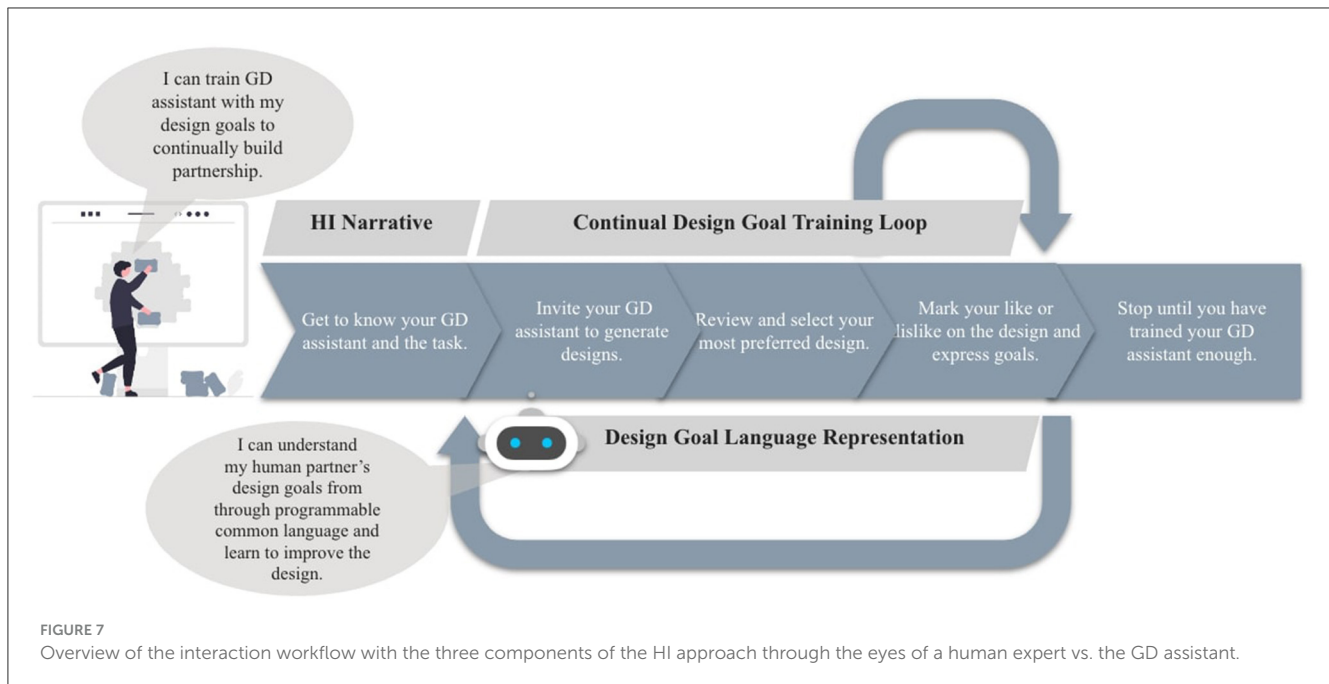
5.5.1.2 Quantitative correlations

From the quantitative analysis, all the statistical correlations were positive, meaning one construct increased with another. For example, user control was strongly associated with perceived AI learning ($r = 0.97$, $p < 0.001^{***}$) and AI output trust ($r = 0.76$, $p = 0.031^*$) suggesting users' control over the tool might play an important role in how they predict and trust the GD assistant and how they assess it as capable to learn and adapt through interaction. We also found correlations between previous constructs from AI-TAM and new constructs from HI-TAM. In particular, greater ease of use was associated with perceived partnership ($r = 0.75$, $p = 0.033^*$); the greater perceived usefulness was also associated with greater AI output quality ($r = 0.92$, $p = 0.001^{**}$), greater perceived AI learning ($r = 0.74$, $p = 0.037^*$), greater perceived partnership ($r = 0.74$, $p = 0.037^*$) as well as more willingness to train ($r = 0.81$, $p = 0.015^*$). Additionally, AI output quality was associated with willingness to train ($r = 0.88$, $p = 0.004^{**}$). Moreover, more willingness to train was associated with more willingness ($r = 0.72$, $p = 0.046^*$) to adopt, indicating participants' willingness to train and their willingness to adopt might have influenced or even reinforced each other when guided by the HI approach (asterisks indicate significance level: $^{***}p < 0.001$,$^{**}p < 0.01$,$^*p < 0.05$).

Notably, quantitative and qualitative results supported links between perceived usefulness and AI output quality, perceived AI learning, and perceived partnership as well as user control and AI output trust.

5.5.2 Design generation process

In the proposed HI-TAM model, we identified links between participants' willingness to train and AI output quality, perceived usefulness, and perceived AI learning, respectively. To further examine such relationships, we describe an example of the design generation process and then present individual processes contributing to participants' rated willingness to train including participants' training efforts over time, underlying motivations to train/stop training as well as the GD assistant's learning curve in response to user training.

FIGURE 7
Overview of the interaction workflow with the three components of the HI approach through the eyes of a human expert vs. the GD assistant.

5.5.2.1 Individual participant process reflections

Table 2 and Figure 8 summarize the full design generation process for P1. We list some interesting observations below:

- P1 initially viewed the design generated in the first round as "*completely random*", but only provided three design goals, possibly due to a lack of trust and familiarity with the GD assistant. P1 chose to start with simpler feedback to gauge the GD assistant's response.

- In the second round, P1 noticed a small improvement - "*I see some offices are close to the windows now*", aligning with their previous design goals. This significantly increased P1's engagement, leading to a substantial increase in the number of design goals compared to the first round.

- In Round 3, P1 found the design mostly acceptable with only a few minor issues, commenting "*they definitely start to make sense. I'm actually a bit impressed by how this is going. Still have some weird things like office 6.*" This resulted in a decrease in the number of design goals for this round.

- P1 stopped reviewing the design at round 4 due to the perceived lack of significant improvement compared to the previous round. They expressed a strong belief that additional rounds would not make a substantial difference, stating, "*It's kinda similar to the last round. It's getting there but not perfect. I don't think it will get perfect. It's getting there but never perfect. That's one reason why I don't trust AI.*".

- When presented with a list of potential reasons for likes and dislikes as design goals, P1 actively explored alternative options to determine their validity, rather than simply selecting the one that aligns with their initial thoughts. They expressed this mindset by stating in the second round, "*Let's see if there is any other option that makes sense*".

5.5.2.2 User efforts in training the GD assistant

Figure 9 illustrates the progression of design goals over time for each participant and indicates when they ceased training the GD assistant, signifying the end of their involvement in the design generation process.

We categorize the participants' training efforts into three categories, emphasizing their qualitative narratives regarding the reasons for discontinuing the training process of the GD assistant:

5.5.2.2.1 AI output quality: reaching a sufficiently good design

P2 discovered a highly effective solution by the end of the third round, expressing satisfaction with the outcome, stating, "*This third option is really strong...It's a really good result.*"

Both P4 and P8 expressed low expectations for the GD assistant. P4 mentioned additional requirements regarding circulation and building codes but did not attribute them to the GD assistant's responsibility. On the other hand, P8 simply desired a design that served as a good starting point, stating, "*I think I'm pretty happy with it. I would take this layout and massage it.*"

5.5.2.2.2 Perceived AI learning: anticipating no further design improvements

P1 expressed a high level of satisfaction with the last round of design, acknowledging the unlikelihood of achieving perfection, and considered it a good starting point for further refinement by a human designer. In their comments, P1 stated, "*It's kinda similar to last round*", and expressed, "*This is a good starting point, and I would assume if I keep generating new designs it'll keep giving me 4.5*"—referring to the same subjective satisfaction rating. "*It could get close, but never perfect...like 4.5 (out of 5).*"

TABLE 2 P1's design generation process: each column represents a design generation round displaying the chosen design, participant quotes, and reasons for likes/dislikes submitted by participants as design goal input for the GD assistant.

Round 1	Round 2	Round 3	Round 4
"Looks completely random."	*"I see some offices are close to the windows now."*	*"They definitely start to make sense. I'm actually a bit impressed by how this is going. Still have some weird things like office 6."*	*"It's kinda similar to the last round. It's getting there but not perfect."*
(-) Office 4 is at the center **(−) The reception area is away from the entrance (−) The archive room is at the center	(+) Office 1 is close to the window (+) External meeting room 2 is close to some internal meeting room (+) The lunch space is close to some meeting room (+) External meeting room 0 is close to internal meeting room (+) External meeting room 0 has the same door orientation as the entrance (−) Office 2 is on south (−) Office 6 is on south (−) Internal meeting room 1 is close to the entrance **(−) The reception area is away from the entrance	(−) Office 6 is at the center (−) The archive room is close to the window (+) Open desk space 2 is at the center (−) The lunch space is away from the bathrooms	

"+" and "−" indicate aspects of the design they like and dislike, respectively, while "**" represents additional reasons not listed in the popup window but recorded out side of the GD assistant.
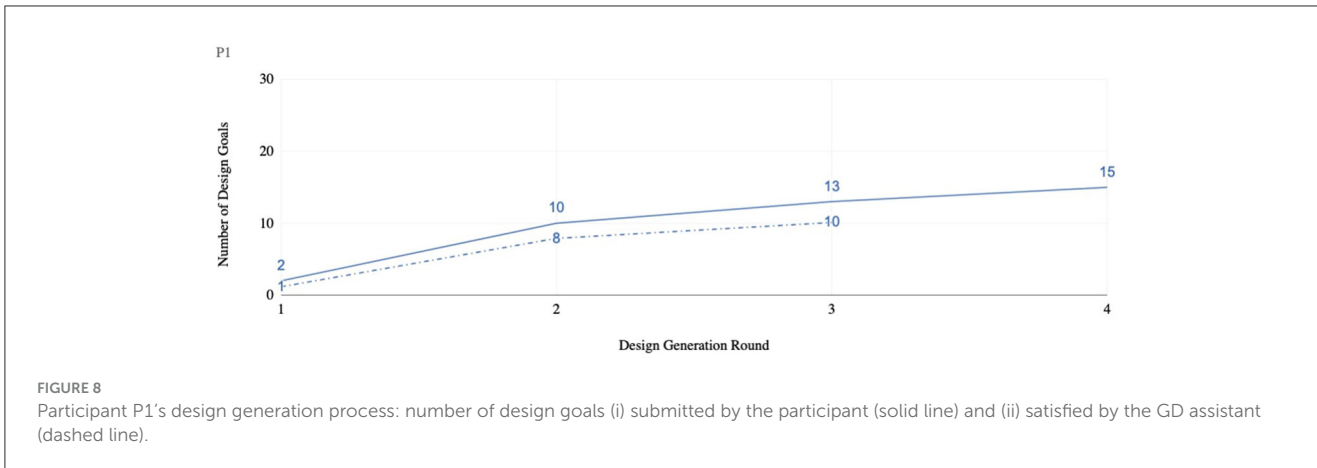
FIGURE 8
Participant P1's design generation process: number of design goals (i) submitted by the participant (solid line) and (ii) satisfied by the GD assistant (dashed line).
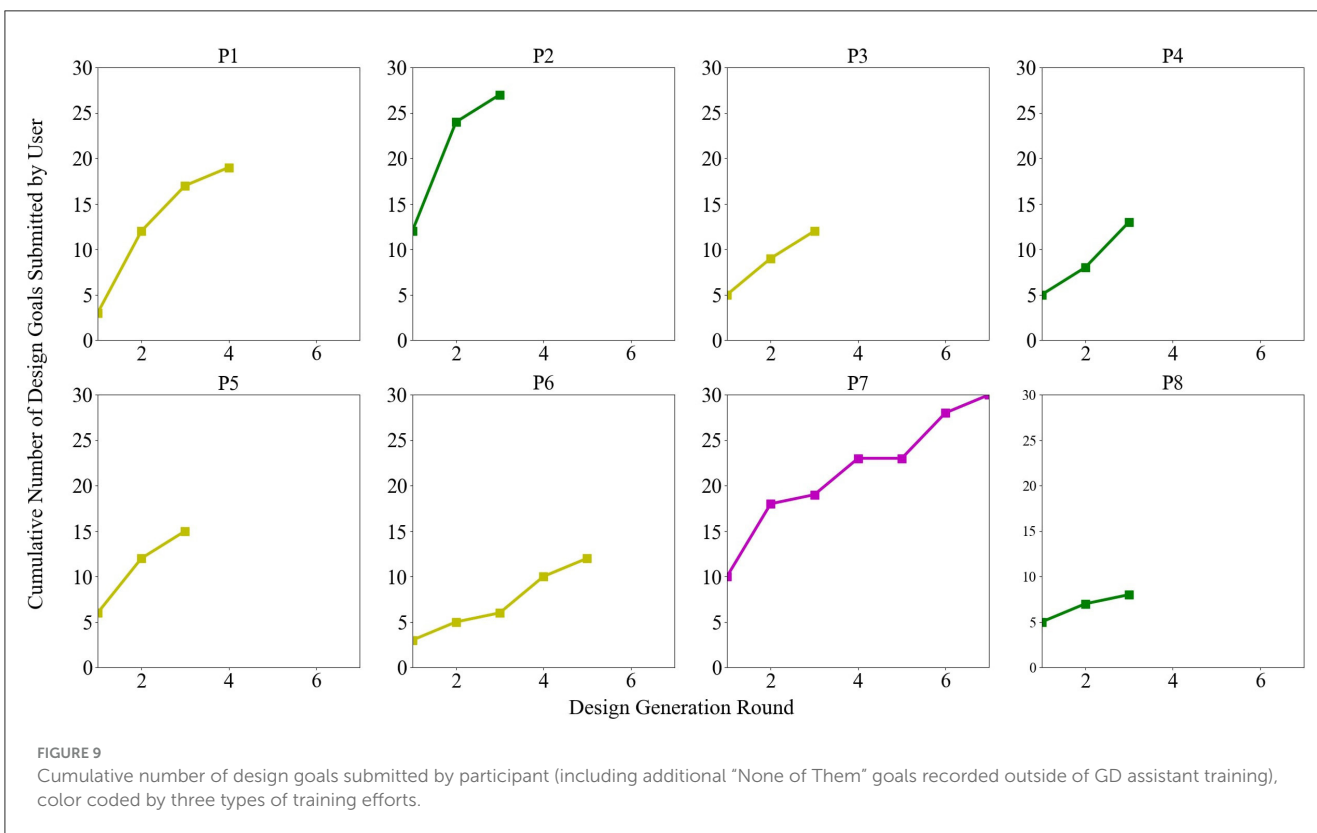

FIGURE 9
Cumulative number of design goals submitted by participant (including additional "None of Them" goals recorded outside of GD assistant training), color coded by three types of training efforts.

P3 believed that the egress design[5] could be further improved, but lacked a way to express their related design goals. However, they found the design acceptable when disregarding the egress design aspect.

P5 expressed that their design goals were being captured by the GD assistant. However, they found the resulting layout to be unsatisfactory, which led to their reluctance to add more design goals to improve it. In their own words, *"I feel like on the one hand, it is following the things I've said, but the plans are all still crazy."*

P6 expressed satisfaction with the design in terms of adjacency, but desired to improve circulation, although they faced difficulty in

expressing this particular design goal. They stated, *"Function wise, it works. But my comments on circulation cannot be reflected."*

5.5.2.2.3 Perceived usefulness: balancing precision and time constraints

P7 indicated a strong desire to train the GD Assistant to precisely replicate their specific plan. This resulted in numerous iterations even after the layout was already deemed acceptable. However, due to time constraints, they had to be stopped, *"This was pretty good. I just want the open desk space to be a little bit close to the office but this is just because I had designed the office before (in this way)... I'm biased... that background may not actually be helping here. I think this could be workable."*

5   "Egress design" refers to the arrangement of exit routes in a building.

5.5.2.3 Learning characteristics of the GD assistant

Figure 10 depicts the learning characteristics of the GD Assistant, where steeper lines indicate a greater level of goal satisfaction, with a maximum of 45 degrees from the X-axis. Several participants (P5, P7, P8) show decreasing steepness toward the end, suggesting varying degrees of saturation.[6] Participant P8 even exhibited a slight negative curvature, indicating conflicting design goals. P1, P2, P3, P4, and P6 stopped before the lines flattened, either due to finding a satisfactory design (P2 and P4) or feeling that their remaining design goals couldn't be adequately expressed in the system (P1, P3 and P6).

5.5.2.3.1 Design goal conflicts restrict the GD assistant's learning

P5 had a specific requirement to separate objects of the same type, such as external meeting rooms not being close to each other. However, finding suitable areas to satisfy this requirement within the limited floor plan space proved challenging, as concentrating objects in those areas could have violated the separation goal.

5.5.2.3.2 Divergent training enhances the GD assistant's longer-term learning compared to convergent training

As users add more design goals, there is a tendency for a decline in the GD assistant's capacity to satisfy all goals due to an increased likelihood of goal conflict within the limited floor plan space. However, there are significant individual differences in participants' training approaches. In particular, P2 and P7 achieved significantly different overall design goal satisfaction by the GD assistant, with P2 exhibiting much higher satisfaction compared to P7, despite submitting a similar total number of design goals. Specifically, P2 had 70% of 27 goals satisfied while P7 had 57% of 30 goals satisfied.

P2 took an explorative, divergent approach to evaluating the generated designs, which involved assessing their inherent quality and providing specific comments on whether they were good or bad. As evidence of this approach, P2 made a comment about one of the designs, saying, "*This one is interesting because it forces you to go through the reception area, which is good,*" indicating a willingness to accept unexpected features and explore novel directions. In the contrary, P7 seemed to compare the generated designs with an ideal concept they had in mind. They provided feedback based on how closely the designs matched their ideal, stating, "*This was pretty good. I just want the open desk space to be a little bit closer to the office, but this is just because I have designed an office before in this way... I'm biased...that background may not actually be helping here. I think this could be workable.*" As a result, P7's design goals were more specific and more likely to exhibit early saturation and generate conflicts with continued training. In contrast, the more explorative approach of P2 resulted in many fewer signs of saturation and thus a potential for further training or personalization. This demonstrates the different design considerations and success criteria related to training a virtual assistant in settings ranging from wanting

to realize preconceived ideas to engaging in more open-ended problem and solution exploration.
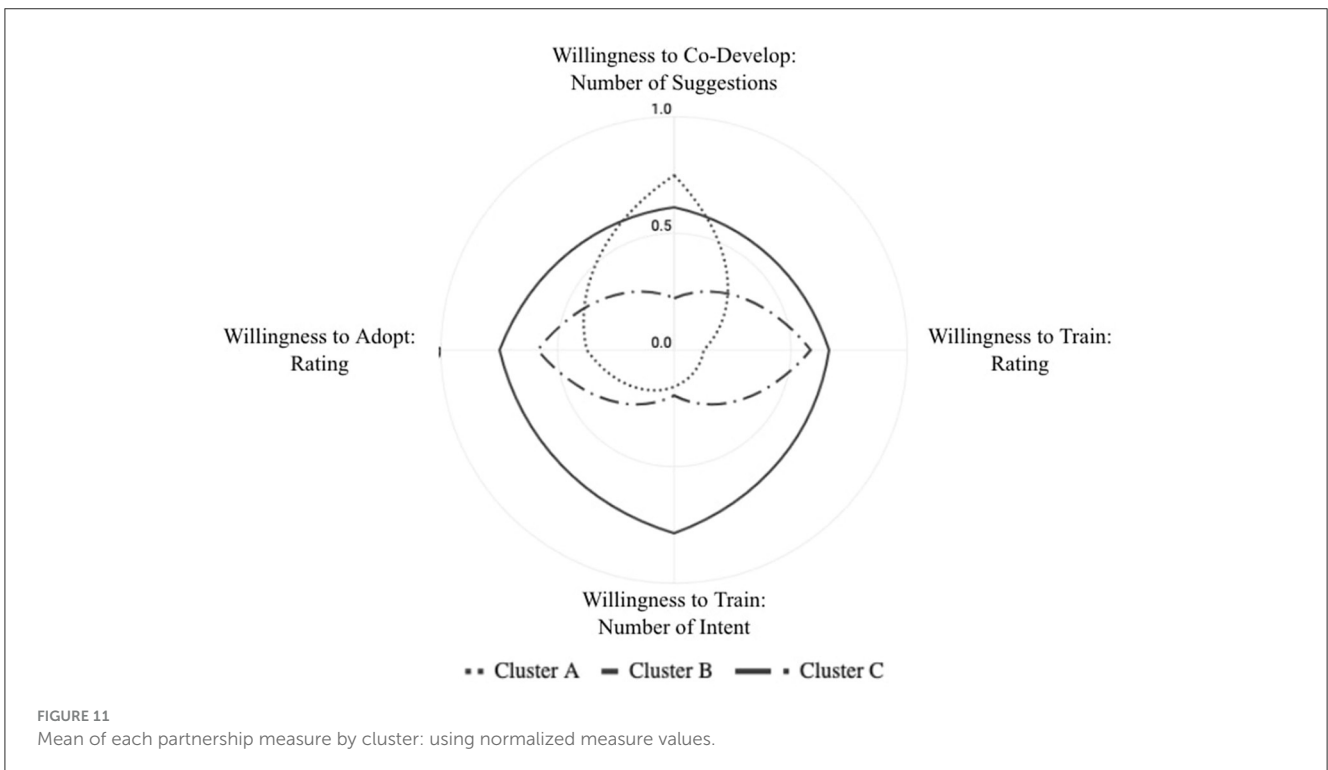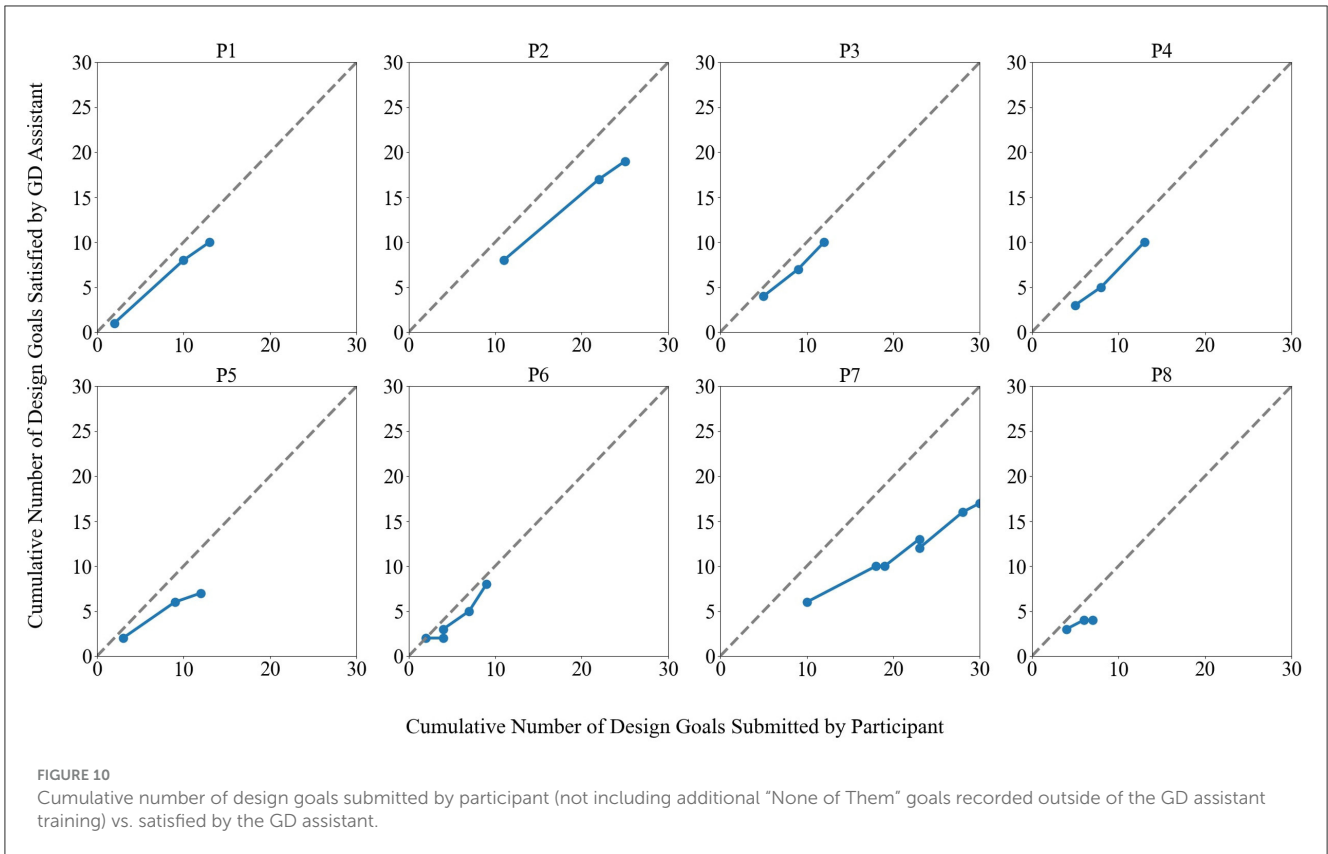
5.5.3 Three types of partnership

The hierarchical clustering resulted in three distinct clusters of participants, differentiated by four quantitative measures across three partnership constructs with the GD assistant: (1) willingness to adopt, measured by post-task willingness to use it in practice, (2) willingness to train, measured by the number of design goals contributed during interaction and the post-task willingness to invest time in training, and (3) willingness to co-develop, measured by the number of suggestions proposed to improve the assistant. The hierarchical clustering resulted in 3 distinct clusters of participants that vary according to four quantitative measures of the three partnership constructs regarding GD assistant, (1) willingness to adopt measured by the post-task willingness to use it in work practice, (2) willingness to train measured by the number of design goals participants contributed during interaction and the post-task willingness to invest time in training to use it in work practice, (3) willingness to co-develop measured by the number of suggestions proposed to improve it after the task.

Figure 11 and Supplementary Table S3 display the average statistics for each cluster profile contour. For each cluster, we describe their profile statistics and the most common qualitative code patterns shared across participants including their suggestions to co-develop the GD assistant proposed after the design generation process (Figure 12, see suggestion examples in Supplementary Table S4). By relating qualitative and quantitative results in this section, we explore and reason further hypotheses about individual differences when experiencing the HI approach, and point to future directions for researching and improving the HI approach.

Cluster A (P5, P8) represents those who are the least willing to adopt, the least willing to train but the most willing to co-develop the tool. Both participants had notable occurrences of user control (3, 5 times) and perceived usefulness (4, 7 times), mostly with a negative sentiment. In particular, P8 who had the most mentions of willingness to co-develop (6 times) and linked willingness to co-develop with perceived usefulness and user control. He provided a suggestion of how the system could be changed to increase control: "*Have you used the feature in DALL-E where you can lock part of the image and regenerate? Yeah, I think that would be useful. That way you feel more in control of what you are doing.*" They both found it more pressing to improve user control of the GD assistant in their suggestions before they could invest in training or adopting it.
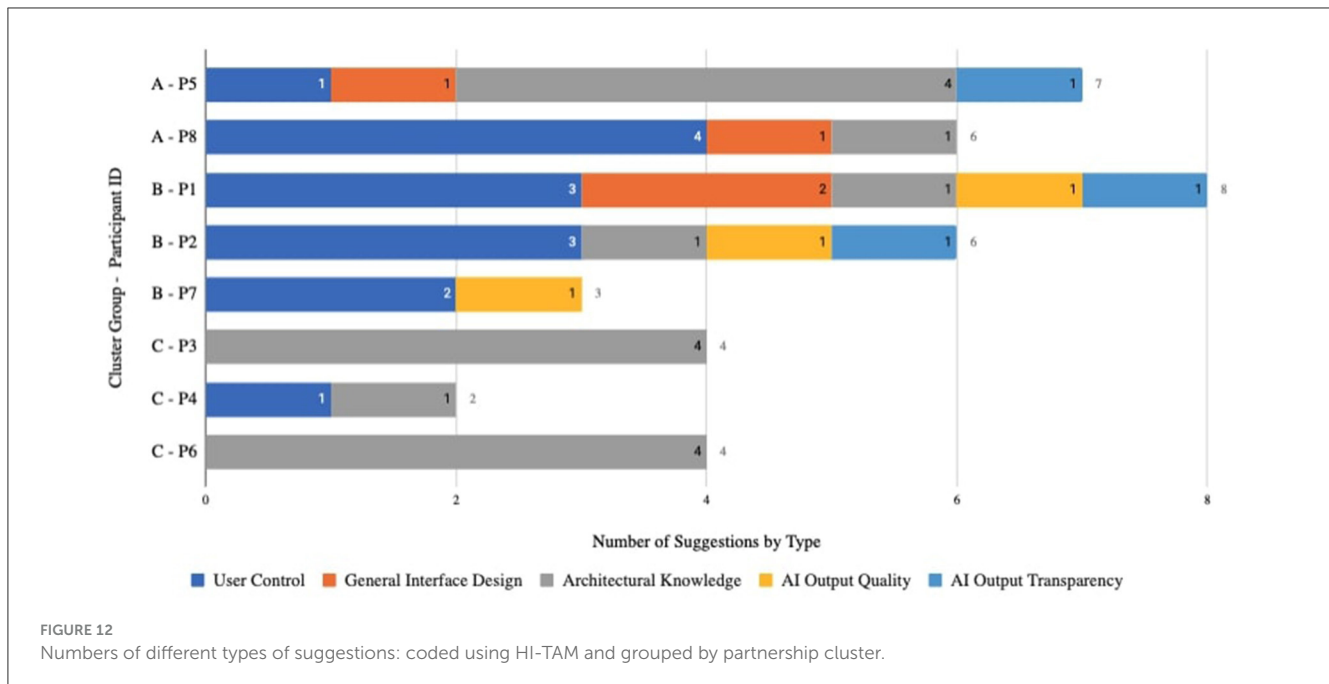
Cluster B (P1, P2, P7) represents those who are most willing to adopt, most willing to train and also highly interested in co-developing the GD assistant. They shared the most of perceived AI learning (10, 9, 9 times) with a strong positive sentiment and AI output quality (18, 9, 3 times) with a mixed sentiment. As an example, P7 held his conflicted input to the GD assistant responsible for the resulting negative output and expressed his duty to train the GD assistant with a strong belief in its learning capability: "*I don't know if my rules are all following each other. Some of them might be contradictory. But it seems like it is learning...So I guess I should tell it. Tell the tool which one is my favorite even*

---

6   In the graphs, there were instances where the same number of design goals submitted corresponded to multiple numbers of design goals satisfied (e.g., P6, P7). This occurred when participants generated multiple rounds of new designs without submitting additional design goals to explore different design options.

FIGURE 10
Cumulative number of design goals submitted by participant (not including additional "None of Them" goals recorded outside of the GD assistant training) vs. satisfied by the GD assistant.

FIGURE 11
Mean of each partnership measure by cluster: using normalized measure values.

out of these." All of them speculated GD assistant's learning and reflected on how their training input can be improved with more user control and AI output transparency, and consequently such training can improve the output quality, which might in return held them optimistic and responsible for both adopting and improving the tool.

FIGURE 12
Numbers of different types of suggestions: coded using HI-TAM and grouped by partnership cluster.

Cluster C (P3, P4, P6) represents those with a moderate level of willingness to adopt or train with the least willingness to co-develop. They shared the most mentions of perceived AI learning (2, 5, 6 times) with a mostly positive sentiment and perceived usefulness (4, 7, 2 times) with a mixed sentiment. Moreover, they all hoped to appropriate GD assistant for alternative uses other than creative design. In particular, P6 was mindful of how GD assistant was learning from his behavioral input but expressed his concerns about its artistic expression limitation: "There are two aspects in architectural design: the scientific part also the functional part can be definitely done with machine learning, with the program. It's just the artistic part...It's kind of tricky because it's just like a poem. It's so personal." Therefore he suggested GD assistant could be better used in its strength, "the scientific part also the functional part", for producing or checking designs in compliance with building code: "So a lot of those rules can be applied to the master plan, and can probably be used to generate like a basic layout without concept. A messy analysis can all be done by a machine and you work on the more artistic part". In summary, all of them acknowledged GD assistant's learning capability to some extent but had strong opinions on how human experts and AI should handle different design tasks. To them, GD assistant can never design or learn like humans do regardless of the amount of training investment. Thus, it made sense to them to appropriate it for other contexts rather than improving it further. Additionally, they suggested that GD assistant could be improved in performance by incorporating specialized architectural knowledge to have a built-in sense of high-level design goals such as circulation, privacy, lighting and building code.

5.5.4 User characterization of the GD assistant

Visual characterization and analogy have been utilized in previous research to probe participants' perception of AI agency

and role (e.g., Koch et al., 2019). Building upon this approach, we employed visual characterizations (typically asking "*If you have to describe it in a visual image, how would you characterize the role of the tool within the design generation process?*") to gain deeper insights into how participants perceived the roles of the GD assistant, particularly in terms of their varied senses of partnership. Table 3 summarizes participants' visual characterizations of the GD assistant after completing the design generation task, covering a wide range of animate (P1, P5, P8) and inanimate entities (P2, P4, P6, P7, P8).

The analysis of these characterizations revealed several notable results. Positive mentions were primarily related to AI output transparency, perceived usefulness, and perceived AI learning. Participants (P2, P7, P8) appreciated the transparency of the AI's output, finding it beneficial for understanding the design process. Additionally, participants (P4, P5, P6) recognized the usefulness of the GD assistant in performing specific tasks. On the other hand, criticism revolved around the perceived ease of use, particularly concerning the repetitiveness of communication, as expressed by participants (P1, P2, P8).

Additionally, user control seems to have a mix of positive and negative perceptions. Some participants (P6, P8) appreciated the control they had over the design process. However, P2 expressed negative perceptions of user control, indicating the need for further exploration and improvement.

It is noteworthy that participants' visual characterizations highlighted the importance of Perceived AI learning, AI output transparency and trust. Perceived AI learning is linked with both AI output transparency and AI trust as seen in the HI-TAM model (see Figure 1). P7 valued the consistency and predictability of the GD assistant's performance while acknowledging occasional surprises in its capabilities. This suggests that a balance between reliability and unexpected outcomes can contribute to a positive user experience.

TABLE 3 Participants' visual characterizations of the GD assistant: summarized from original quotes.

	Summarized characterizations	HI-TAM codes (+ positive, − negative)
P1	A curious child constantly asks for parents' guidance in learning cooking. While improvement occurs over time, there are concerns about the trustworthiness of the initial output quality and the burdens of repetitive questions.	●(−) AI output quality ●(−) Perceived ease of use ●(+) Perceived AI learning
P2	A regular dice creates a sense of partnership with some specialized patterns but lacks full control, requiring iteration and frustration when communicating through obvious questions.	●(−) User control ●(+) AI output transparency ●(−) Perceived ease of use ●(+) Perceived partnership
P3	NA (P3 skipped this question.)	
P4	A hairbrush simple in its capabilities but helpful in performing specific tasks such as improving blood flow on the skin and grooming hair, serving multiple distinct functionalities.	●(+) Perceived usefulness
P5	A less annoying and unhelpful Microsoft paper clip that doesn't pretend to have a personality or befriend the user.	●(+) Perceived usefulness
P6	Magnetic blocks of physical models that users can move around and experiment collaboratively for early concept design.	●(+)User control ●(+)Perceived usefulness
P7	A weighted dice that is consistent, predictable and capable of surprising by performing unexpected tasks.	●(+) AI output transparency ●(+) AI output trust
P8	A young boy wearing a button-up shirt, intelligent but not straightforward or playful, lacking refinement, yet still reliable and has the potential for significant value after prolonged training. A blocky Casio scientific calculator, lacking coolness and sleekness while reliable and transparent, parallel to the control and precision offered by Revit software.	●(+) User control ●(+) AI output transparency ●(+) AI output trust ●(−) Perceived ease of use ●(+) Perceived AI Learning

6 Discussion

Firstly, it is important to underscore that in Study 1 no more than four participants (10% of 41) expressed a preference to exclude generative AI entirely at any phase of the building lifecycle. In Study 2, seven out of eight participants expressed that 10 years from now, a moderate amount of their current tasks in architectural design could be done by a machine instead of themselves. These observations, consistent with recent studies of generative AI-assisted writing (Noy and Zhang, 2023), underscore the importance of investigating effective human-AI approaches within GD. In this paper, we set out to investigate to what extent the HI approach helps human experts build a sense of partnership in design co-creation. We define the concept of partnership as a human expert's willingness to contribute to the creative tool, a personal GD assistant, during and after co-creation. We here discuss the results from our studies and implications for future work. Given the small scale of our studies within one domain, the generalizability of all results should be empirically tested further.

6.1 HI-TAM and the adoption of the GD assistant

Study 1 highlighted professionals' readiness to adopt the HI approach, characterized by relatively higher human-AI integration, and their preferences for personalizing the GD assistant. These results shed light on the suitability of the HI approach as a foundation for real-world application design. Notably, higher human-AI integration was observed in the early phases of the building lifecycle. For effective real-world application design, it is crucial to break down processes or tasks into manageable stages.

This approach allows for understanding where end-users desire integration of the GD assistant, ensuring alignment with their workflow needs.

Given that the participants in Study 2 were asked to report their goal satisfaction with the generated layouts after each round, it is unsurprising that AI output quality and perceived AI learning were the most frequently identified constructs. This co-occurrence suggests that human experts in architectural design highly value the ability of the GD assistant to generate high-quality design solutions while also rapidly adapting to their personal preferences. The link between AI output quality and AI output trust can be explained by the fact that the participants were generally happier with the outcome of AI output when it consistently reflected their design goals. Future iterations of this GD assistant could allow users to see the factors that are contributing to its output, such as the specific design preferences and goals that have been learned over time and which design goals it is weighting most when generating the design. In an effort to combat algorithmic overconfidence (Lacroux and Martin-Lacroux, 2022), future designs could display the GD assistant's level of confidence in the generated design as a means of the algorithm communicating to the end user. Bi-directional communication has also been shown to increase perceived partnership (Rezwana and Maher, 2022).

From the quantitative analysis in Study 2, the high and positive correlation found between willingness to train and willingness to adopt in this study could be a promising indicator of a "pathway to adoption". In other words, positive experiences in training a virtual GD assistant could lead to a greater likelihood of adoption.

Both our quantitative and qualitative results from Study 2 supported a link between perceived usefulness and AI output quality, perceived AI learning, and perceived partnership. This is supported by the AI-TAM (Baroni et al., 2022) which combined

constructs of AI Output Trust and AI Output Quality into the "super construct", explainable AI (XAI), correlating with perceived usefulness (0.74). While the link between AI output trust and perceived usefulness was not supported in our analysis, perceived usefulness was linked to AI output transparency and AI perceived learning. As the addition of AI output quality and perceived AI learning allowed the participants' feedback to be more granular, it is unsurprising that there were fewer instances of perceived AI trust as the codes are conceptually related.

Our study's results suggest that perceived ease of use, as described in the AI-TAM, may have been absorbed into more fine-grained constructs such as AI output quality, AI output transparency, AI output trust, and perceived AI learning in our HI-TAM. Similarly, perceived usefulness may be broken down into more granular metrics to evaluate and inform AI design.

Perceived partnership is one of the new constructs we added to the HI-TAM given its importance in human-AI co-creative systems (Rezwana and Maher, 2022). Our HI-TAM shows quantitative correlations and qualitative co-occurrences between perceived partnership and perceived usefulness. Interestingly, there were no significant correlations or strong co-occurrences between perceived partnership and willingness to co-develop. We maintain that the construct of partnership is important to the HI-TAM and postulate the missing link between perceived partnership and willingness to adopt may be due to the fact that the tool is still in prototype form and may not exhibit enough qualities to warrant partnership. In our analysis, we also identify three distinct clusters of participants that vary according to 4 quantitative measures of the three partnership constructs (willingness to adopt, willingness to train and willingness to co-develop). While our sample is small, these initial results shed light on the possibility that different modes of training may be necessary for different types of users.

6.2 Training of the GD assistant

In Study 1, a majority of participants (62%) indicated that they preferred using properties and requirements of their final output, along with variations in output, as methods for training and improving GD assistants. In contrast, none of the participants in Study 2 trained the GD assistant long enough to reach the saturation regime. This implies that by incorporating additional design goals, the GD assistant's outputs could have been further enhanced. This highlights the potential of training the GD assistant with a broader range of design goals, which can lead to a more personalized GD assistant capable of better aligning with multiple design objectives.

Additionally, motivating participants to submit design goals for GD assistant training can be influenced by various factors such as the expressivity of the common language used between the user and the GD assistant, the positive correlation between the quality of the output and the number of design goals submitted, and the responsiveness of the GD assistant to the submitted design goals. It is essential to understand these promoting factors as well as those that act as detractors. One significant detractor is the presence of design goal conflicts, where participants were unclear if their design goals were "contradictory" and they "confused" the assistant. P7 exemplifies this with the following excerpt "*I have said*

this before. That [design feature] should be closer to the window. And now I'm saying...I think I'm contradicting myself." This can discourage participants from continuing the training. Participants in our study suggested that future iterations of the GD assistant could benefit from features that help users track and manage these conflicts, thereby facilitating continuous training and improving the perceived AI transparency. By considering these factors, we can create an environment that fosters motivation and enables the GD assistant to reach its full potential. Also, it would be interesting to use gamification features to enhance willingness to engage in long-term virtual assistant training (Afyouni et al., 2019).

Lastly, the training of a virtual assistant should be tailored based on the desired outcome and the level of openness in problem-solving. When users aim to realize preconceived ideas, the training process should prioritize capturing and reproducing those specific ideas effectively. On the other hand, when users engage in open-ended exploration, the training approach should emphasize adaptability, novelty, and the generation of diverse solutions. Further research can explore ways to incorporate users' goals, preferences and intentions into the training of AI systems, enhancing their ability to align with individual user needs.

6.3 Design implications

The findings from this study have notable practical implications for AI tool designers and developers. First, the participants' overall willingness to engage with and adopt generative AI assistants underscores the importance of designing tools that are flexible, user-friendly, and capable of learning from users' inputs. One key insight is the need for AI systems to not only deliver high-quality output but also transparently communicate how they adapt to user preferences over time. By offering features that allow users to track the AI's learning process and better understand how their feedback influences the system's decisions, developers can foster trust and partnership between humans and AI, crucial for long-term adoption.

Furthermore, the research highlights that a personalized, adaptable AI assistant aligned with users' evolving design goals can significantly enhance workflow integration. Designers of AI tools should focus on embedding customization features that allow users to personalize their interaction with AI systems, catering to specific phases of a task or lifecycle. Importantly, this includes providing feedback mechanisms that display the AI's confidence and uncertainty, promoting transparency and mitigating overconfidence, a common issue in AI systems.

Beyond the architectural domain, the implications of this research extend to industries such as product design, engineering, and media, where co-creative processes are central. The HI (Hybrid Intelligence) approach offers a framework for developing AI systems that complement human creativity rather than replace it. The integration of AI in design and creative industries could potentially reshape workflows, enabling professionals to focus on more strategic, high-level tasks while delegating routine, repetitive work to AI. This partnership model may also apply to other fields, such as healthcare and education, where AI systems can serve as collaborative agents, enhancing human expertise and decision-making capabilities.

6.4 Limitations and future research

The limited sample sizes in Study 1 ($N = 41$) and Study 2 ($N = 8$) from the architectural industry constrain the generalizability of our findings. To strengthen the validity of these preliminary results, future studies should involve larger, stratified samples, taking into account participants' varying levels of industry experience, exposure to generative AI and their levels of psychological safety with AI assistant tools at work. This would allow for a more robust analysis of how these factors influence the effectiveness of AI adoption in architectural design practices.

One key area to explore is whether positive experiences in training AI assistants directly lead to higher adoption rates or if this effect is skewed by selection bias - AI-enthusiasts who are naturally more willing to engage with new tools. To further investigate this, future studies could manipulate the training experience (e.g., hybrid intelligence-induced positive experiences vs. control conditions without such induction) and measure adoption differences. Structural equation modeling (SEM) could be applied to analyze the stepwise effects of hybrid intelligence (HI) on user experience, leading to adoption. Hypotheses for this approach could include: *"HI interactions will result in improved user experience and increased likelihood of long-term adoption compared to non-HI interactions."* This approach would help isolate the causal relationship between training experiences and adoption, providing more robust insights into the mechanisms driving AI integration in professional workflows.

Future research should also explore ways to further refine the interaction between humans and AI by experimenting with HI across different domains and investigating long-term impacts of AI transparency and feedback mechanisms on user trust and AI tool effectiveness. Combining quantitative surveys with in-depth interviews could yield deeper insights into users' cognitive and emotional responses to working with AI, further guiding the design of tools that promote productive human-AI partnerships.

7 Conclusion and future work

In conclusion, our study sheds light on the potential of a Hybrid Intelligence Technology Acceptance Model (HI-TAM) to inform the design of generative design assistants that facilitate a co-creative partnership between human experts and algorithms. Opportunities for future work include improving the functionality, user experience, and integration of the current GD assistant prototype with existing design workflows and processes. Another avenue is to validate the effectiveness of the HI narrative in creating a psychologically safe learning and training environment by comparing it to a control group with a more neutral algorithmic description. The HI narrative could be further improved to address human experts' concerns about AI adoption and job displacement. Furthermore, we plan to validate the HI-TAM model on other HI systems to verify its applicability across diverse contexts. Additionally, scalability and generalizability of HI-empowered interactions with GD assistants and AI systems in other domains and contexts represent a promising line of future research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Aarhus University IRB; Autodesk. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing, Project administration. JR: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. YW: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing, Resources, Software, Visualization. JS: Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing, Supervision.

Conflict of interest

YM was employed by Autodesk Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2024.1460381/full#supplementary-material

References

Afyouni, I., Einea, A., and Murad, A. (2019). "RehaBot: gamified virtual assistants towards adaptive telerehabilitation," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, 21–26.

Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., et al. (2020). A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer* 53, 18–28. doi: 10.1109/MC.2020.2996587

Alcaide-Marzal, J., Diego-Mas, J. A., and Acosta-Zazueta, G. (2020). A 3D shape generative method for aesthetic product design. *Design Stud.* 66, 144–176. doi: 10.1016/j.destud.2019.11.003

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. (2014). Power to the people: the role of humans in interactive machine learning. *Ai Mag.* 35, 105–120. doi: 10.1609/aimag.v35i4.2513

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., et al. (2019). "Guidelines for human-AI interaction," in *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, 1–13.

Arenas, M., Gutierrez, C., and Pérez, J. (2009). "Foundations of RDF databases," in *Reasoning Web International Summer School*, eds. S. Tessaris et al. (Heidelberg: Springer Berlin Heidelberg), 158–204.

Baroni, I., Calegari, G. R., Scandolari, D., and Celino, I. (2022). AI-TAM: a model to investigate user acceptance and collaborative intention in human-in-the-loop ai applications. *Hum. Comp.* 9, 1–21. doi: 10.15346/hc.v9i1.134

Berditchevskaia, A., and Baeck, P. (2020). *The Future of Minds and Machines: How Artificial Intelligence Can Enhance Collective Intelligence*. London: Nesta.

Borji, A. (2023). A categorical archive of chatgpt failures. *arXiv* [preprint]. doi: 10.21203/rs.3.rs-2895792/v1

Boston Consulting Group (2020). *Study Finds Significant Financial Benefits With AI*.

Braun, V., and Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qual. Res. Sport Exerc. Health* 11, 589–597. doi: 10.1080/2159676X.2019.1628806

Brown, J. D. (2002). The cronbach alpha reliability estimate. *JALT Test. Eval. SIG Newslett.* 6, 17–19.

Caetano, I., Santos, L., and Leitã, A. (2020). Computational design in architecture: defining parametric, generative, and algorithmic design. *Front. Arch. Res.* 9, 287–300. doi: 10.1016/j.foar.2019.12.008

Creswell, J. W., and Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 319–340. doi: 10.2307/249008

Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. (2019). Hybrid intelligence. *Bus. Inf. Syst. Eng.* 61, 637–643. doi: 10.1007/s12599-019-00595-2

Demirel, H. O., Goldstein, M. H., Li, X., and Sha, Z. (2023). Human-centered generative design framework: an early design framework to support concept creation and evaluation. *Int. J. Hum. Comp. Interact.* 40, 933–944. doi: 10.1080/10447318.2023.2171489

Diamond, L., Busch, M., Jilch, V., and Tscheligi, M. (2018). "Using technology acceptance models for product development: case study of a smart payment card," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 400–409.

Dubiel, M., Halvey, M., and Azzopardi, L. (2018). A survey investigating usage of virtual personal assistants. *arXiv* [preprint]. doi: 10.48550/arXiv.1807.04606

Dudley, J. J., and Kristensson, P. O. (2018). A review of user interface design for interactive machine learning. *ACM Transact. Interact. Intell. Syst.* 8, 1–37. doi: 10.1145/3185517

Eccles, D. W., and Arsal, G. (2017). The think aloud method: what is it and how do i use it? *Qual. Res. Sport Exerc. Health* 9, 514–531. doi: 10.1080/2159676X.2017.1331501

Fails, J. A., and Olsen Jr, D. R. (2003). "Interactive machine learning," in *Proceedings of the 8th International Conference on Intelligent User Interfaces*, 39–45.

Freedman, D., Pisani, R., and Purves, R. (2007). *Statistics Fourth Edition*. New York, NY: W. W. Norton & Company.

Gunadi, D., Sanjaya, R., and Harnadi, B. (2019). "Examining the acceptance of virtual assistant-vanika for university students," in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)* (Semarang: IEEE), 1–4.

Hartmann, M. (2009). "Challenges in developing user-adaptive intelligent user interfaces," in *LWA, ABIS-6* (Citeseer).

Inie, N., Falk, J., and Tanimoto, S. (2023). "Designing participatory AI: creative professionals' worries and expectations about generative AI," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, CHI EA '23* (New York, NY: Association for Computing Machinery).

Jameson, A. D. (2009). Understanding and dealing with usability side effects of intelligent processing. *AI Mag.* 30, 23–23. doi: 10.1609/aimag.v30i4.2274

Jian, J.-Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergonomics* 4, 53–71.

Kazi, R. H., Grossman, T., Cheong, H., Hashemi, A., and Fitzmaurice, G. W. (2017). Dreamsketch: early stage 3D design explorations with sketching and generative design. *UIST* 14, 401–414. doi: 10.1145/3126594.3126662

Kepuska, V., and Bohouta, G. (2018). "Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (Las Vegas, NV: IEEE), 99–103.

Keshavarzi, M., Hotson, C., Cheng, C.-Y., Nourbakhsh, M., Bergin, M., and Rahmani Asl, M. (2021). "Sketchopt: sketch-based parametric model retrieval for generative design," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6.

Koch, J., Lucero, A., Hegemann, L., and Oulasvirta, A. (2019). "May AI? Design ideation with cooperative contextual bandits," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Lacroux, A., and Martin-Lacroux, C. (2022). Should i trust the artificial intelligence to recruit? Recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during resume screening. *Front. Psychol.* 13:895997. doi: 10.3389/fpsyg.2022.895997

Maedche, A., Morana, S., Schacht, S., Werth, D., and Krumeich, J. (2016). Advanced user assistance systems. *Bus. Inf. Syst. Eng.* 58, 367–370. doi: 10.1007/s12599-016-0444-2

Mao, Y., Rafner, J., Wang, Y., and Sherson, J. (2023). "A hybrid intelligence approach to training generative design assistants: partnership between human experts and ai enhanced co-creative tools," in *HHAI 2023: Augmenting Human Intellect*, eds. P. Lukowicz et al. (IOS Press), 108–123.

Mayer, R. C., and Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: a field quasi-experiment. *J. Appl. Psychol.* 84:123. doi: 10.1037/0021-9010.84.1.123

Nagy, D., Villaggi, L., and Benjamin, D. (2018). "Generative urban design: integration of financial and energy design goals in a generative design workflow for residential neighborhood layout," in *Symposium on Simulation for Architecture and Urban Design, Vol. 3* (Delft).

Novák, V. (1987). First-order fuzzy logic. *Stud. Logica* 46, 87–109. doi: 10.1007/BF00396907

Noy, S., and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381, 187–192.

Oh, C., Song, J., Choi, J., Kim, S., Lee, S., and Suh, B. (2018). "I lead, you help but only with enough details: understanding user experience of co-creation with artificial intelligence," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

Oh, S., Jung, Y., Kim, S., Lee, I., and Kang, N. (2019). Deep generative design: integration of topology optimization and generative models. *J. Mech. Design* 141:229. doi: 10.1115/1.4044229

Oh, Y., Ishizaki, S., Gross, M. D., and Do, E. Y.-L. (2013). A theoretical framework of design critiquing in architecture studios. *Design Stud.* 34, 302–325. doi: 10.1016/j.destud.2012.08.004

Prakash, N., and Mathewson, K. W. (2020). Conceptualization and framework of hybrid intelligence systems. *arXiv* [preprint]. doi: 10.48550/arXiv.2012.06161

Rafner, J., Bantle, C., Dellermann, D., Söllner, M., Zaggl, M. A., and Sherson, J. (2022a). "Towards hybrid intelligence workflows: Integrating interface design and scalable deployment," in *HHAI2022: Augmenting Human Intellect*, eds. F. Lorig et al. (IOS Press), 310–313.

Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T., and Sherson, J. (2023). Creativity in the age of generative ai. *Nat. Hum. Behav.* 7, 1836–1838. doi: 10.1038/s41562-023-01751-1

Rafner, J., Dellermann, D., Hjorth, A., Verasztó, D., Kampf, C., Mackay, W., et al. (2022b). Deskilling, upskilling, and reskilling: a case for hybrid intelligence. *Morals Mach.* 1, 24–39. doi: 10.5771/2747-5174-2021-2-24

Rezwana, J., and Maher, M. L. (2022). Designing creative ai partners with cofi: a framework for modeling interaction in human-AI co-creative systems. *ACM Trans. Comput. Hum. Interact.* 30, 1–28. doi: 10.1145/3519026

Salovaara, A., and Tamminen, S. (2009). Acceptance or appropriation? A design-oriented critique of technology acceptance models. *Fut. Interact. Des. II* 157–173. doi: 10.1007/978-1-84800-385-9_8

Shea, K., Aish, R., and Gourtovaia, M. (2005). Towards integrated performance-driven generative design tools. *Automat. Construct.* 14, 253–264. doi: 10.1016/j.autcon.2004.07.002

Sherson, J. (2024). *A Hybrid Intelligent Change Management Approach to Generative Ai Adoption. HHAI 2024: Hybrid Human AI Systems for the Social Good*, 472.

Sherson, J., Rabecq, B., Dellermann, D., and Rafner, J. (2023). "A multi-dimensional development and deployment framework for hybrid intelligence," in *HHAI 2023: Augmenting Human Intellect* (IOS Press), 429–432.

Sherson, J., Rafner, J., and Büyükgüzel, S. (2024). "Operational criteria of hybrid intelligence for generative ai virtual assistants," in *HHAI 2024: Hybrid Human AI Systems for the Social Good* (IOS Press), 475–477.

Sherson, J., and Vinchon, F. (2024). "Facilitating human feedback for GENAI prompt optimization," in *HHAI 2024: Hybrid Human AI Systems for the Social Good*.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Transact. Interact. Intell. Syst.* 10, 1–31. doi: 10.1145/3419764

Shneiderman, B., and Maes, P. (1997). Direct manipulation vs. interface agents. *Interactions* 4, 42–61. doi: 10.1145/267505.267514

Song, Y. W. (2019). *User Acceptance of an Artificial Intelligence (AI) Virtual Assistant: An Extension of the Technology Acceptance Model* (PhD thesis). UT Electronic Theses and Dissertations.

Van Turnhout, K., Bennis, A., Craenmehr, S., Holwerda, R., Jacobs, M., Niels, R., et al. (2014). "Design patterns for mixed-method research in HCI," in *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, 361–370.

Yim, O., and Ramdeen, K. T. (2015). Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *Quant. Methods Psychol.* 11, 8–21. doi: 10.20982/tqmp.11.1.p008

Yuan, X., Chen, Y., Zhao, Y., Long, Y., Liu, X., Chen, K., et al. (2018). "Commandersong: a systematic approach for practical adversarial voice recognition," in *27th {USENIX} Security Symposium ({USENIX} Security 18)*, 49–64.

Zanzotto, F. M. (2019). Human-in-the-loop artificial intelligence. *J. Artif. Intell. Res.* 64, 243–252. doi: 10.1613/jair.1.11345