3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows

Vivian Liu*
vivian@cs.columbia.edu
Autodesk Research
Toronto, Ontario, Canada

Jo Vermeulen
jo.vermeulen@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

George Fitzmaurice
george.fitzmaurice@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

Justin Matejka
justin.matejka@autodesk.com
Autodesk Research
Toronto, Ontario, Canada

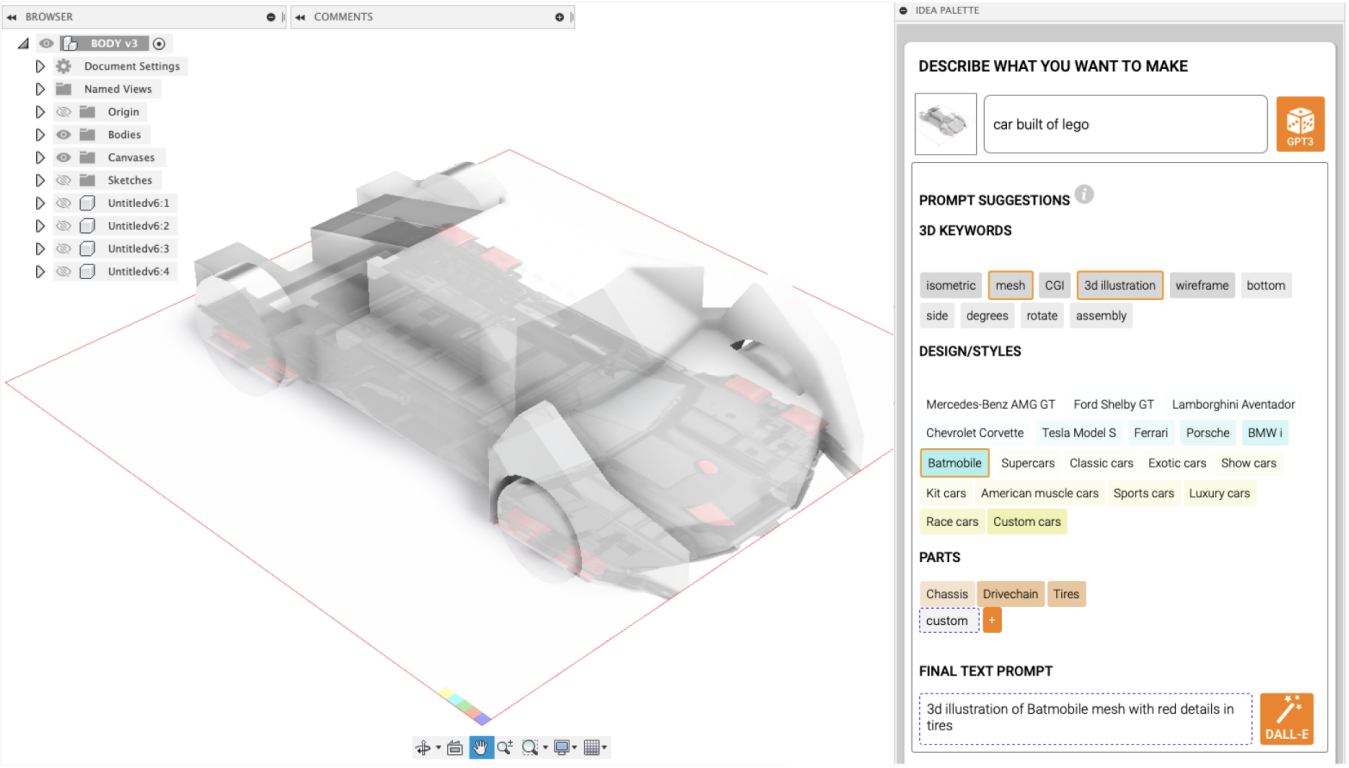


Figure 1: 3DALL-E integrates a state-of-the-art text-to-image AI (DALL-E) into 3D CAD software Fusion 360. This plugin generates 2D image inspiration for conceptual CAD and product design workflows. 3DALL-E helps users craft text prompts by providing 3D keywords, design/styles, and parts from GPT-3. Users can also generate from image prompts based on a render of their current workspace, letting users use their 3D modeling progress as a basis for text-to-image generations.

ABSTRACT

Text-to-image AI are capable of generating novel images for inspiration, but their applications for 3D design workflows and how designers can build 3D models using AI-provided inspiration have not yet been explored. To investigate this, we integrated DALL-E, GPT-3, and CLIP within a CAD software in *3DALL-E*, a plugin that generates 2D image inspiration for 3D design. 3DALL-E allows users to construct text and image prompts based on what they are modeling. In a study with 13 designers, we found that designers saw great potential in 3DALL-E within their workflows and could use text-to-image AI to produce reference images, prevent

*Also with Columbia University.

DIS '23, July 10–14, 2023, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9893-0/23/07…$15.00
https://doi.org/10.1145/3563657.3596098

design fixation, and inspire design considerations. We elaborate on prompting patterns observed across 3D modeling tasks and provide measures of prompt complexity observed across participants. From our findings, we discuss how 3DALL-E can merge with existing generative design workflows and propose prompt bibliographies as a form of human-AI design history.

CCS CONCEPTS

• **Applied computing** → **Media arts**; • **Human-centered computing** → **Interactive systems and tools**; • **Computing methodologies** → **Natural language generation**; *Shape modeling*.

KEYWORDS

creativity support tools, 3D design, DALL-E, GPT-3, CLIP, 3D modeling, CAD, co-creativity, creative copilot, ideation, prompt engineering, multimodal, text-to-image, AI applications, text-to-3D, workflow, diffusion

ACM Reference Format:

Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2023. 3DALL-E: Integrating Text-to-Image AI in 3D Design Workflows. In *Designing Interactive Systems Conference (DIS '23), July 10–14, 2023, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3563657.3596098

1 INTRODUCTION

Designing 3D models in CAD software is challenging—designers have to satisfy a number of objectives that can range from functional and aesthetic goals to feasibility constraints. Coming up with ideas takes a lot of exploration, even for experienced designers, so they often consult external resources for inspiration on how to define their geometry. They browse 3D model repositories [23], video tutorials, and image search engines to understand conventional designs and different aesthetics [48]. This process of conceptualizing CAD designs is pivotal to the product design process, yet few computational methods support it [35, 48].

A recent innovation that can more directly provide inspiration to designers is text-to-image AI. Tools such as DALL-E [17], Imagen [77], Parti [89], and Stable Diffusion [75] are AI tools that have the generative capacity to access and combine many visual concepts into novel images. Given text prompts as input, these tools can capture a wide variety of subjects and styles [52]. In online communities, users have already developed methods to elicit images with 3D qualities [52, 66] by including prompt keywords such as "3D render" or "CGI". Recent advancements have also allowed users to interact with text-to-image AI systems by passing in image prompts, where images are used as prompts in addition to text. Generations can now be varied or built off of previous generations. These innovative functions make the integration of text-to-image AI within existing creative authoring software more feasible.

However, how AI-provided image inspiration can contribute to CAD and product design workflows has not yet been fully explored. In this paper, we seek to understand how text-to-image AI can assist 3D designers with conceptual CAD and design inspiration and where in creative workflows designers can most benefit from AI assistance. Furthermore, we investigate how text-to-image tools respond to image prompts sent from 3D designers as they build up complexity in their designs. To do so, we integrated three large AI models—DALL-E, GPT-3, and CLIP—within Fusion 360, an industry standard software for computer-aided design (CAD). We implemented a plugin within the software which we call *3DALL-E*. This plugin helps translate a designer's goals into multimodal (text and image) prompts which can produce image inspiration for them. After a designer inputs their goals (i.e. to design a "truck"), the plugin provides a number of related parts, styles, and designs that help users craft text prompts. These suggestions are drawn from the world knowledge of GPT-3 [5] to help users familiarize themselves with relevant design language and 3D keywords that can better specify the text prompt. The plugin interactively updates an image preview from the software viewport that shows an image prompt which can be passed into DALL-E [72], giving users a direct bridge between their 3D design workspace and an AI model that can generate image inspiration. Additionally, having a lens on what the designer is actively working on allows the plugin to highlight what prompt suggestions may work best, which is implemented in the system by using CLIP [71] to approximate model knowledge. To evaluate 3DALL-E and how well it can integrate into 3D workflows, we conducted a user study with thirteen users of Fusion 360 who spanned a variety of backgrounds from industrial design to robotics. We found that 3DALL-E can benefit CAD designers as a system that supports conceptual CAD, helps prevent design fixation, produces reference images, and inspires design considerations.

We present the following contributions:

- 3DALL-E, a plugin that generates AI-provided image inspiration for CAD and product design by helping users craft text prompts with design language (different parts, styles, and designs for a 3D object) and image prompts connected to their work in progress.
- An exploratory user study (n=13) demonstrating text-to-image AI use cases in 3D design workflows and an analysis of prompting patterns and prompt complexity.

In our discussion, we propose prompt bibliographies, a concept of human-AI design history to track inspiration from text-to-image AI. We conclude on how text-to-image AI can integrate with existing design workflows and what can be best practices for generative design going forward.

2 RELATED WORK

2.1 Prompting

Prompting is a novel form of interaction that has come about as a consequence of large language models (LLMs) [5]. Prompts allow users to engage with AI using natural language. For example, a user can prompt an AI, "What are different parts of a car?" and receive a response such as the following, "Wheels, tires, and headlights". These prompts give LLMs context for what tasks they need to perform and help end users adapt the general pretraining of LLMs without further finetuning [4, 73]. By varying prompts, users can query LLMs for world knowledge, generative completions, summaries, translations, and so forth [5, 53]. Datasets around prompting are also beginning to emerge to benchmark generative AI abilities. PARTI [89] provides a schema and a set of prompts to investigate the visual language abilities of AI. Coauthor [50] provides a dataset of rich interactions between GPT-3 and writers. Audits of models

have also been performed by collecting generated outputs of AI models at scale and conducting annotation studies, as in [52] and [68]. As generative AI communities have gained momentum online, crowd-sourced efforts on Twitter and Discord have also organized to disseminate prompting guidance [66] that suggest experimentation with various style and medium keywords (e.g. "isometric", "3D render", "sculpture" etc.).

Recent research directions have begun to develop workflows around prompts. AI Chains [85] studied how complex tasks can be decomposed into smaller, prompt-addressable tasks. Promptchainer [84] unveiled an editor that helps users visually program chains of prompts. Prompt-based workflows were explored in [42] to make prototyping ML more accessible for industry practitioners. Other systems have tested pipelines that concatenate LLMs with text-to-image models. In Opal [53], a pipeline of GPT-3 initiated prompt suggestions generated galleries of text-to-image generations to help news illustrators explore design options in a structured manner. Similarly, a visual concept blending system in [27] used BERT [20] to surface shape analogies and prompt text-to-image AI for visual metaphors. A key finding from Opal and the visual blends system [27] that we apply in 3DALL-E is that LLMs can help generate prompts so end users can efficiently explore design outcomes.

New modes of prompting have also started to emerge. Users can now pass in image prompts and have AI models autocomplete images and canvases in methods called inpainting and outpainting [17, 65]. These functions have been implemented within state-of-the-art text-to-image AI systems [17]. 3DALL-E is the first to systematically generate image prompts from CAD software (Fusion 360) and help users incorporate their 3D design progress into text-to-image generations.

2.2 Generative Models

Generative AI models have long been excellent at image synthesis. However, many early models were class-conditional, meaning that they were only robust at generating images from the classes they were trained on [43, 44, 69, 70, 86, 88]. The most recent wave of generative AI models can now produce images from tens of thousands of visual concepts due to extensive pretraining. CLIP [71], a state-of-the-art multimodal embedding, was trained off of hundreds of millions of text and image pairs, giving it a broad understanding of both domains. The pretraining of CLIP has also helped it serve as an integral part of multiple generative workflows [15, 16, 22, 62] and training regimes [60, 78]. Large open-source efforts had previously paired CLIP with GAN models, using it as a discriminator to optimize generated images toward text prompts. The novelty of generating media through language has brought many text-to-image tools into production such as Midjourney, DALL-E, and Stable Diffusion. DALL-E [72] demonstrated how CLIP embeddings can help generate images with autoregressive and diffusion-based approaches. Diffusion is key within many of the aforementioned methods to increase the quality of text-to-image outputs [14, 17, 61]. New text-to-image approaches have led to more diverse methods of user interaction. Make-a-Scene [24] allows users to interact with generations by manipulating segmentation maps, and DALL-E gives users the ability to paint outside the edges of an image, allowing for unlimited canvases [65]. Textual inversion [75] gives users the ability to train and trade novel concepts learned by the AI [25] off of a few examples. These models have extraordinary generative capacity, but their ability to be used nefariously has also inspired new approaches to safeguarding AI outputs from redteaming [6] to large scale audits for social and gender biases [12].

Text-to-3D methods such as CLIP-Sculptor, DreamFusion, and Point-E [40, 67, 78, 79] also exist and are rapidly improving, but they have far longer inference times [40] and required computing power [40]. They are also often constrained to producing shapes that are limited in diversity [79], fidelity [78, 79], stylistic range [64], and capabilities for variable binding owing to the smaller volume of paired text-shape data online [79]. Advances using diffusion models as a prior have also made the generation of complex, textured 3D models possible [67]. However, text-to-3D approaches result in scene [67], voxel [78, 79], pointcloud [64], and mesh [67] representations that are medium or high fidelity from the get-go. This can start a designer off at an unfamiliar stage in their workflow (with a medium or high fidelity geometry they might not know how to edit) or with a representation they do not usually use for CAD. To support conceptual CAD from the earliest stages possible, we investigate text-to-image rather than text-to-3D in 3DALL-E as the most suitable starting point for AI-provided inspiration. We elaborate on how designers often start in 2D and build up to 3D forms using shape operations in Section 2.4.

2.3 Creativity Support Tools

Human-computer interaction research on creativity support tools has long showcased ways to facilitate text-based content creation. Early systems showed that users could iteratively define images based on chat and dialogue [21, 80]. AttriBit [9] allowed users to assemble 3D models out of parts matched on affective adjectives. Sceneseer [7] and Wordseye [13] allowed users to create scenes via sentences. However, since the advancement of AI tools, much of the momentum has now concentrated around human-AI co-piloted experiences. Systems such as Opal [53], Sparks [30], FashionQ [41], and the editors in [81] are examples of AI-assisted ideation. In tandem, many frameworks for computational creativity [54] and human-AI interaction [1] have cropped up to understand concerns such as ownership and agency when AI is involved in the creative process. Gero et al. [29] found that users can establish better mental models of what AI can and cannot do if they have a sense of its internal distribution of knowledge.

Practices for creativity support tools that we revisit from an AI perspective include the idea of design galleries [56], timelines and design history [32], natural language exploration [26], and collaboration support [82]. DataTone [26] demonstrated how interactive prompting with widgets can help build specificity in a text-based interface. Suh et al. [82] demonstrated that AI-generated content could facilitate teamwork within groups by helping establish common ground between collaborators. While many systems have been built with generative AI capabilities [18, 31, 55] and even for text-to-image workflows [53]—none that we know of have applied text-to-image AI for 3D design workflows.

2.4 CAD Conceptual Design and Workflows

CAD is a highly complex design activity that usually involves a significant amount of conceptual design, as later stages of prototyping can incur material costs. Because CAD evolved in part from 2D drafting, CAD often relies on 2D representations such as freehand drawing and computer-assisted sketches [34, 45, 46]. In these early stages, designers are also gathering inspiration from external sources like 3D model repositories [23] (e.g. Onshape, Google Poly), video tutorials, and reference images [87] to inform their sketches. Users operate over these 2D representations (sketch profiles and planes) to apply constraints and dimensions and to take their models into 3D using operations such as extrusion, lofting, revolving and so on [34]. It has been found that early stage CAD and product design "tends to be ambiguous, incomplete, and expressive with high levels of uncertainties" [45], and there is less focus on constraints and parameters [46, 74]. Conceptual CAD also can involve text and image exploration; mechanical engineers perform system decomposition to understand model needs, and industrial designers collect moodboards and perform market research [35].

One direction within HCI work has focused on capturing and understanding CAD workflows. Screencast [2, 32] collects timelines of authoring operations from CAD help forums. From Screencast data, workflow graphs [8] have been proposed as a way to characterize 3D modeling workflows. These graphs have shown that users can arrive at 3D models through different paths. For example, to design a mug, a user can design in parts and in interchangeable sequences; they can first create the body of the cup, and then the handle, or vice versa. Examinations of CAD experts have also generalized CAD modeling as procedures of increasing detail, working from sketches to geometric forms to finishing features [34].

Prior work on applying generative models and AI for knowledge-based design in CAD and industrial engineering does exist [28, 49, 51, 58]. Liao et. al. note that parametric CAD tools do not offer "cognitive supports for search nor highlight new information a designer might not have thought of", which is where generative AI can assist by providing triggers for novel solutions [3]. The closest works to 3DALL-E would be DreamSketch [45] and Dream Lens [57], systems for generative design exploration. DreamSketch, helped explore 3D design ideas by passing in sketches, design variables, and constraints that retrieved generative designs from topology optimizers. Dream Lens helped users explore and visualize large-scale generative design datasets based on parameters. Rather than freehand sketches or parameters, 3DALL-E presents a method for supporting conceptual CAD through text-based exploration of design knowledge and text-to-image generations.

3 DESIGNING WITH 3DALL-E

3.1 Design Rationale

Engaging with text-to-image AI means coming up with many prompts. Users have to exhaustively experiment with AI to see what words it can understand and render well. To streamline prompt ideation for a CAD environment, 3DALL-E helps users efficiently assemble 3D design knowledge into prompts. For example, for a table, a user may know common designs like *"dining table"* or *"desk"* but may otherwise not know design vernacular (*"lift-top"*, *"drop-leaf"*, or *"nesting"* table) that 3DALL-E can efficiently supply.

3.2 The 3DALL-E interface

3DALL-E is provided as a panel on the right hand side of the 3D workspace (Fig. 1). Fig. 2 shows the steps users go through when designing with 3DALL-E inside their 3D workspace and presents the main interface components. 3DALL-E allows users to construct prompts relevant to their current 3D design, which can then be sent to DALL-E to retrieve AI-provided image inspiration. Once generations are received, users are able to download them, see a history of previous results, and create variations of generations that they want to explore more from. In what follows, we will present these different steps with a short walkthrough.

3.3 Constructing Text Prompts for AI-Provided Inspiration

Users begin at the starting state shown in Fig. 2-I, where they can describe what they want to make by typing in their goal (Fig. 2A). Once they do that, different prompt suggestions populate the sections with 3D keywords, designs/styles, and parts (Fig. 2-II). These suggestions help steer the generations toward results relevant to 3D modeling as well as provide design language a user might otherwise not be familiar with. For example, querying a chair could return a series of existing designs such as an egg chair, an Eames chair, or a Muskoka chair, helping familiarize the user with the design language befitting of chairs. Once users select a set of prompt suggestions (e.g. *"3d render, isometric, plant stool, wrought-iron"*), an automatically rephrased prompt appears in the final prompt box (e.g. *"isometric 3d render of a wrought-iron plant stool"*) as shown in Fig. 2-III. This prompt is still editable by the user, and a text box to add custom keywords is also available when clicking the orange '+' button in the parts section (Fig. 2F).

Prompt suggestions (Fig. 2C–E) are color-coded with a color for the group they belong to (blue for designs, green for styles, orange for parts) and varied in opacity to indicate how strongly their text aligns with the image prompt (see Fig. 4 for implementation details). For example, from a set of styles like "mid-century modern, contemporary, and art deco", if "art deco" was most strongly highlighted (i.e. more opaque – darker green), it meant that the image prompt had the greatest probability of being matched with "art deco". 3DALL-E suggests keywords to elicit 3D qualities particular to 3D models and renders, following design guidelines from related work [52, 66]. Styles are suggested to allow users to steer the aesthetic language of their generation and engage with inspiration spanning different time periods, traditions, and mediums (e.g. *"mid-century modern"*, *"Brutalist"*, or *"CGI"*). Using style keywords is also a recommended tip from prior work and existing AI systems [17, 52, 66]. 3DALL-E suggests parts as 3D models are often assemblies of parts, as established in work on part-based authoring systems [10] and part datasets [47, 59, 83]. Other dimensions like material and function could have been explored without loss of generality. However, we chose to focus on geometry-relevant suggestions instead of appearance (material) or abstract goals (function).

3.4 Crafting an Image Prompt

Users can also choose to include an image that is automatically extracted from their current 3D modeling workspace in addition to

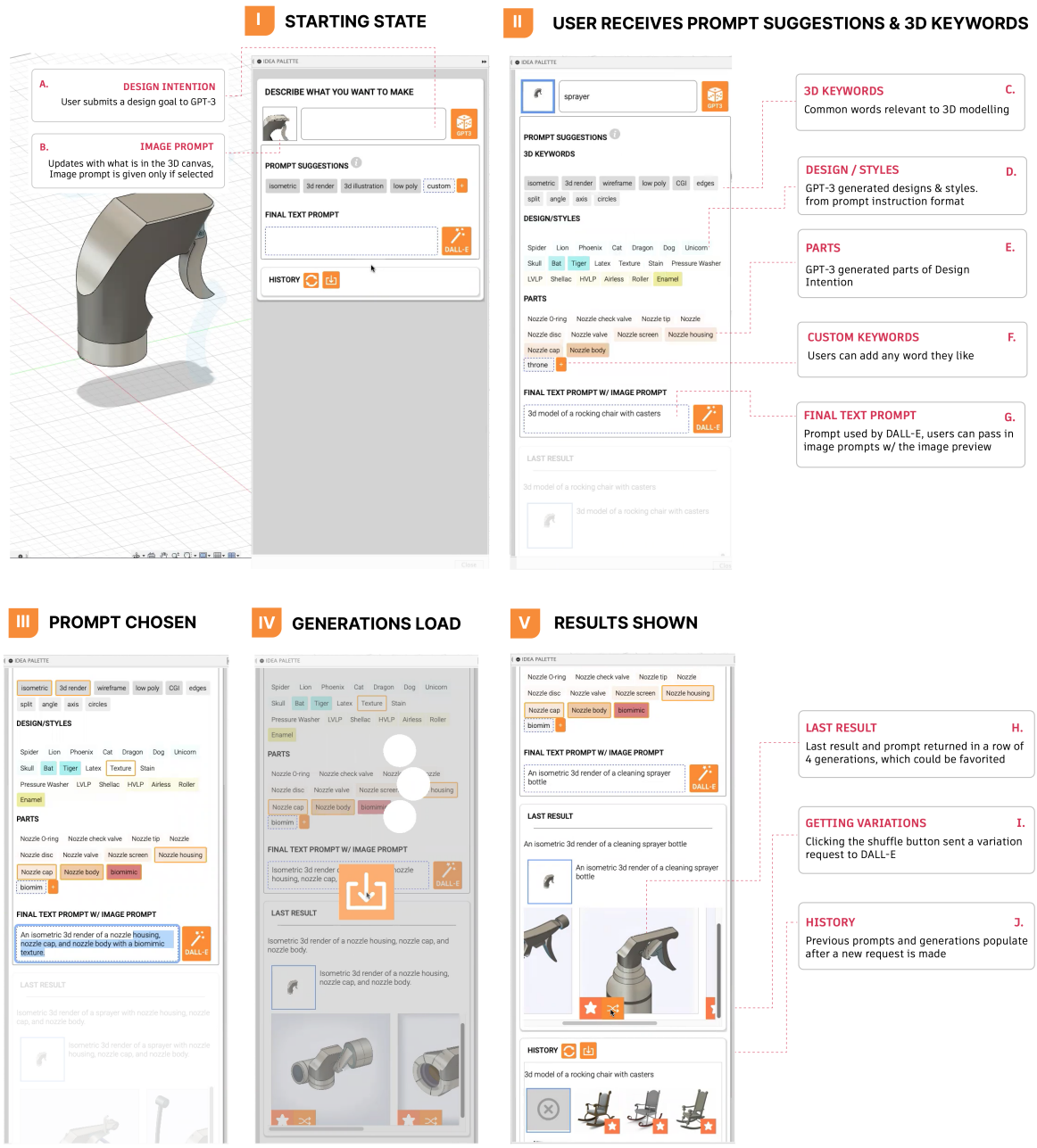


Figure 2: 3DALL-E walkthrough. Step I: Initial state, where users can type their design intentions. Step II: Users are presented with prompt suggestions from GPT-3. Step III: Selected suggestions are rephrased into an editable prompt. Step IV: Users wait as DALL-E generates. Step V: Results are shown. A cursor hovers over a shuffle icon, which is how users can launch variation requests from DALL-E.

their text prompt (image+text prompt) or choose to exclude it (text-only prompt). Image prompts are only passed in when users select the image preview (Fig. 2B), making it active. Using the 3D software to render the viewport allows 3DALL-E to programmatically deliver clean prompts without tasking the user with any erasing or masking. Users can easily toggle the visibility of certain parts of their model using the 3D software's built-in functionality and request for DALL-E to fill in the details for those hidden parts.

3.5 Receiving DALL-E Results and Retrieving Variations

Once the user is satisfied with the prompt, they click the DALL-E button next to the final text prompt (Fig. 2G) to generate either a text-only or image+text prompt (depending on whether the image preview is selected). While waiting for results (Fig. 2-IV), the user is shown a spinner animation. When the results are ready, the user can click the orange download button to pull the results from DALL-E into the 3DALL-E interface.

Results are returned in sets of four (Fig. 2-V). When the user hovers over a result, they are presented with a menu that allows them to 'star' their favorite results and click the 'shuffle' button to get more *variations* on that particular result (Fig. 2I). These are retrieved using DALL-E's built-in functions that generate similar images given an image input. Lastly, 3DALL-E also keeps a history of previous generations (Fig. 2J).

4 SYSTEM IMPLEMENTATION

3DALL-E was implemented within Autodesk Fusion 360 [37] as a plugin and written with the Fusion 360 API, Python, Javascript, Selenium, and Flask. Fig. 3 illustrates how we embedded DALL-E, GPT-3, and CLIP into one user interface. All actions in 3DALL-E were logged by the server to facilitate analysis of participant behavior in the study (Sect. 5). Note that 3DALL-E could be implemented generically in most 3D modeling tools. The needed functionality from Fusion 360 is relatively basic: a custom plugin system and ways to render the viewport as an image.

Prompt suggestions were populated by querying the GPT-3 API for the following: *"List 10 popular 3D designs for {QUERY}? 1.", "What are 10 popular styles of a {QUERY}? 1.", and "What are 10 different parts of a {QUERY}? 1."*. These queries were split using regular expressions such that each suggestion was one button on the interface. To rephrase chosen suggestions, GPT-3 was prompted: *"Put the following together: {SUGGESTIONS}"*.

Ten 3D keywords are sampled from a set of high frequency words (n=121) in a Fusion 360 Screencast dataset. Screencasts are videos used to communicate help and tutorials in forums [2, 32]. Automatic speech recognition (ASR) of these videos produced transcripts; these transcripts were processed with standard count vectorization using NLP modules from Sklearn, filtered out for general purpose words (words that were not specific to CAD), and sorted by frequency to get the final keywords set.

Text highlights were calculated by passing each of the prompt suggestions and the image prompt to CLIP, which was hosted on a remote server. CLIP produces softmaxed logit scores[1] that suggest how similar each text option was to the image, a value 3DALL-E renders as the opacity of each highlight. The stronger the highlight, the greater the probability a text option matched what a user had in their viewport. DALL-E was trained with CLIP text and image embeddings. By using CLIP's embedding in this way, users receive a computational guess for how well DALL-E might be able to interpret each prompt suggestion, while also dialing down the options they need to focus on (Fig. 4). The 3D keywords were by default gray, while designs, styles, and parts were matched to gradations of blue, green, and orange respectively.

[1] Applying softmax to logit scores yields normalized linear probabilities.

We used the Fusion 360 API to automatically save the viewport to a PNG image every 0.3 seconds. The workspace of Fusion 360 (the gridded background pictured in Fig. 1) was rendered transparently in the PNG image.

5 EVALUATION

Implementing 3DALL-E within Fusion 360 gave us a focused application context to evaluate text-to-image AI within a creative workflow. We set out to investigate the following research questions for 3DALL-E to understand in what ways text-to-image AI can be useful for 3D designers.

- *Generation Patterns within Workflows.* Are there certain patterns to how CAD designers use text-to-image generations within their workflows, and do these patterns differ depending upon the 3D modeling task?
- *Assisted Prompt Construction.* How helpful are different features (prompt suggestions, CLIP highlighting, automatically captured viewport images) for the construction of text and image prompts?
- *Prompt complexity.* How many concepts do people like to put within prompts?

To do so, we conducted an exploratory study with 3D CAD designers (n=13, 10 male, 3 female). Participants were recruited from internal channels within a 3D design software company as well as through a design institute mailing list at a local university. Participants were compensated with $50 dollars for 1.5 hours of their time. The average age of the participants was 28, and they had an average of 4.13 years of experience with Fusion 360 (min=1 year, max=8 years). Five had experience with the generative design environment within Fusion, and three had prior experience with AI / generative art systems. The participants spanned a range of disciplines from machining to automotive design. Domains of expertise, frequency of use, and years of experience with the 3D software are listed in Table 1. Based on the system implementation in a CAD software, we focused on CAD designers and product designers rather than 3D artists and 3D concept art more broadly.

5.1 Experimental Design

Participants were given two different 3D modeling tasks: T_{edit} to edit an existing model and T_{create} to create a model from scratch. The intention of having these two tasks was to show how 3DALL-E might affect creative workflows at different stages of the 3D modeling process. The ordering of these tasks was counterbalanced to mitigate learning effects. This experimental design was approved by a relevant ethics board.

Before the study, participants were sent an email with DALL-E's content policy to disclose that they were going to use AI generative tools. During the study, participants were given a brief introduction to the different AI architectures involved (GPT-3, DALL-E) and given two general tips on prompting: 1) text prompts should include visual language, 2) text prompts are not highly sensitive to word ordering [52]. Participants were then given a walkthrough of the user interface and the different ways they could generate results from GPT-3 and DALL-E. The study was conducted virtually via Zoom and through remote control of the experimenter's Fusion 360 application and plugin.

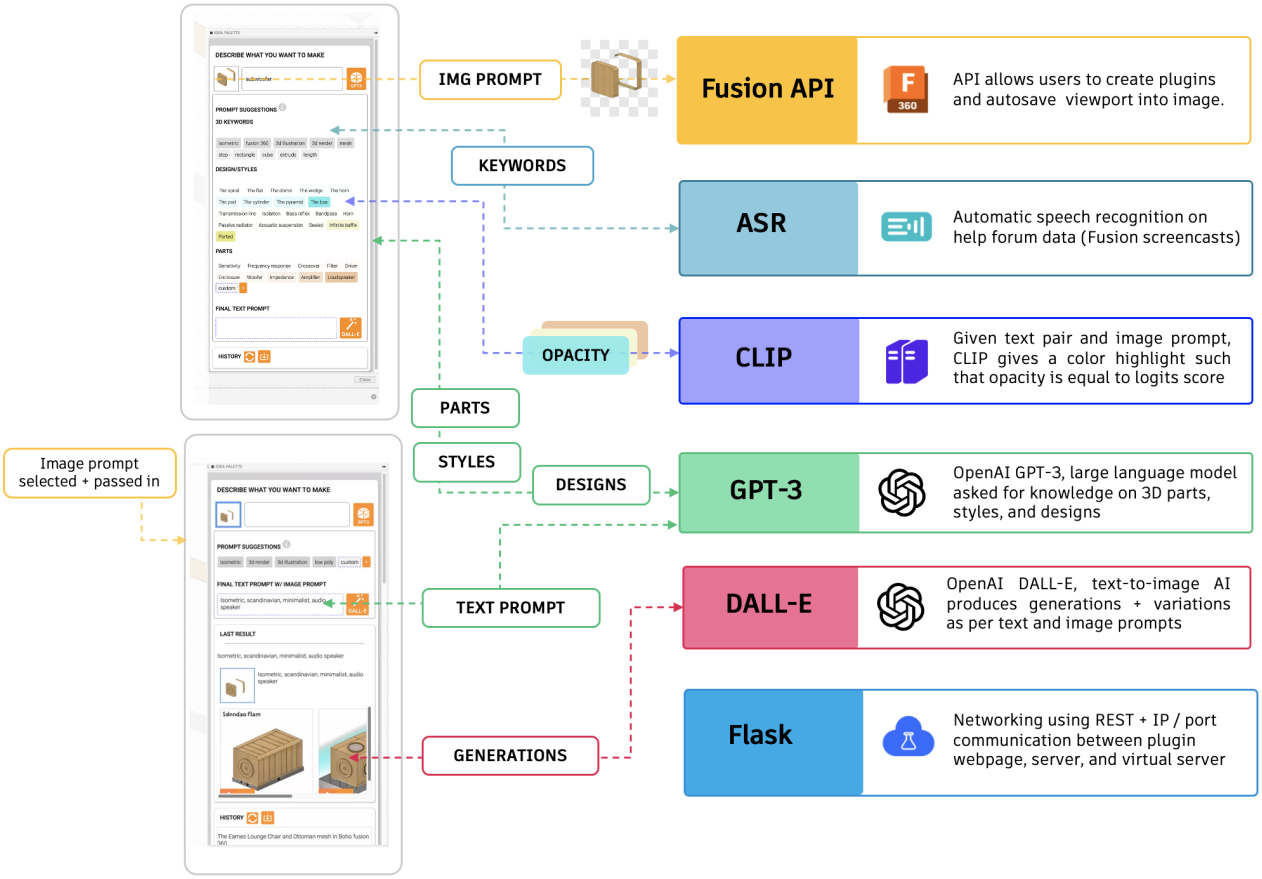


Figure 3: System design showing the architectures involved in 3DALL-E, which incorporates three large AI models into the workbench of an industry standard CAD software. In the top left panel, we show how text AI outputs are displayed in the UI. In the bottom left panel, we show how users could pass in image prompts and retrieve DALL-E generations within the plugin.

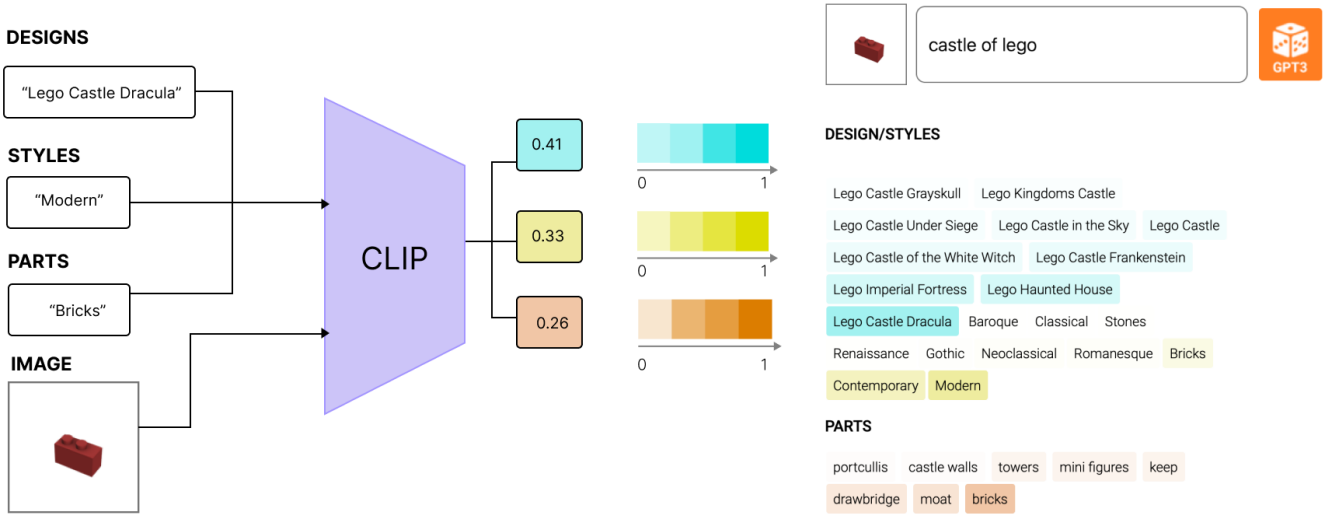


Figure 4: Diagram showing how text highlights were calculated using CLIP with image and text from the prompt suggestions as input. The CLIP logits score was set as the opacity of each prompt suggestion. Each type of suggestion was colored differently.

T_{edit} was to modify an existing 3D model that the participants had brought with them to the study. Participants were told to bring a non-sensitive model, meaning one that did not include corporate data. There were no constraints on what the model could have been. Examples of models brought in can be seen in Fig. 5. When a participant did not have a model to use, a random design was provided from the software's example library. This was the case for only one participant (P15).

For T_{create}, participants were allowed to pick whatever they wanted to design from scratch. For each task, participants had 30 minutes to work on their model with the assistance of 3DALL-E. We justify this duration of 30 minutes as a sufficient length of time based on prior work: DreamSketch [45] (30 to 60 minutes for 3D artifact creation) and Dream Lens [57] (25 minutes for generative design exploration). At the halfway point, participants were reminded of the time remaining and of any generation actions that they had not tried out yet from GPT-3 (prompt suggestions) or DALL-E (text-only prompts, image+text prompts, variations). Beyond this reminder, they were guided only if they needed assistance accomplishing something in the user interface. Examples of what participants created for T_{create} can be seen in Fig. 6. At the 30 minute mark, designers were told to wrap up their design.

After completing each task, participants marked generations in their history that they felt were inspiring and completed a post-task questionnaire, which included NASA-TLX [33], Creativity Support Index (CSI) [11, 55], and workflow-specific questions. These questions can be found in the supplementary material. A semi-structured interview was conducted to understand their experience.

5.2 Quantitative Feedback on 3DALL-E

5.2.1 Creativity Support and NASA-TLX Results. The metrics we measured showed that designers responded to 3DALL-E with enthusiasm. All responses were on a 7-point Likert scale. In terms of enjoyment, 12/13 participants rated their experience positively (≥ 5 out of 7) for T_{edit} (median: 6) and 11/13 for T_{create} (median: 6) . The majority of participants also responded positively that they were able to find at least one design to satisfy their goal: 10/13 respondents in T_{edit} (median: 6), 12/13 respondents in T_{create} (median: 7). Likewise, most participants reported that the system helped them fully explore the space of designs (9/13 responded positively for T_{edit} (median: 6) , 11/13 for T_{create} (median: 6)).

> *"I could spend ages in this." - P18*

In general, the post-task questionnaire results were similar for T_{edit} and T_{create}. However, on a few dimensions, participant responses were distributed slightly differently. For example for effort, responses for T_{edit} about tool performance *("How successful were you in accomplishing what you set out to do?")* were split across the spectrum, with 6/13 rating the tool positively (median: 4). For T_{create}, 10/13 participants rated the performance positively (median: 5). In terms of ease of prompting, while 13/13 respondents were positive that for T_{create} it was easy to come up with prompts (median: 7), 10/13 responded positively for T_{edit} (median: 5). We hypothesize that this could have been because for T_{edit} participants had to work under more constraints, bringing in 3D models that were often custom and near finished.

We note that frustration was low for both Tasks; 11/13 responded on the low side of the spectrum for T_{edit} (≤ 3) (median: 3), and 10/13 on the low side for T_{create}. For T_{edit} (median: 2), frustration was low in spite of the fact that 6/13 of participants disagreed to some degree (≤ 3) about having control over the generations.

> *"The amount of control you have with the system is very dependent upon how specific you get with the text. For example, if I make it super broad, you're obviously going to have less control because DALL-E is working off of less information. So it may provide its own information. It has to kind of fill in the gaps of what you're trying to say. But the more specific I got, the better results I got." - P1*

> *"It was a bit difficult to control. Some things I wasn't quite expecting. For example, with this one [generation of a watch] I expect that it would have more circular watch faces, but it came with ones that were more angular." - P8*

5.2.2 Usefulness of GPT-3, CLIP Highlights, Image Prompts . Lastly, to understand how helpful different features (prompt suggestions, CLIP highlighting, automatically captured viewport images) are in the construction of text and image prompts, we discuss workflow-specific questions about the prompting pipeline of 3DALL-E. Participants were asked about the usefulness of 3DALL-E for their usual workflow. For T_{edit}, 10/13 felt that it would be helpful (median: 5). For T_{create}, 10/13 also felt it would be helpful (median: 7).

In another question, we asked whether it was easy for participants to come up with new ways to prompt the system. Participants responded unilaterally positively for T_{create} (13/13 responded ≥ 5) and positively for T_{edit} (median: 6) (10/13 responded ≥ 5) (median: 6). Participants were also asked to rate how useful they found the GPT-3 suggestions. For T_{edit} and T_{create}, the responses were generally positive, at least 8/13 participants responded with 5 or higher for both tasks (T_{edit} median: 7, T_{create} median: 6).

> *"I'm looking for the right word and I think that's where this text [GPT-3] search can come in handy… I think it's helpful to know its language, to know what it finds." - P4*

> *"I think having the GPT-generated ones was useful. It allowed for some ideas I didn't consider… [ideas I] wouldn't have found the words for." - P13*

On whether or not the highlighting of prompt suggestions was useful, participants responded with more even distributions, though the distributions still skewed positive (8/13 in T_{edit} and 7/13 in T_{create} rated the statement at 5 or higher (median: 5, for both tasks). Participants tended to click on suggestions that were highlighted more strongly for text-image alignment, often choosing the most strongly highlighted suggestion within the category.

Lastly, we gauged participant response to image prompts, asking if they agreed that image prompts were incorporated well in their generations. For T_{edit}, 10/13 participants responded with a 5 or 6 for agreement (median: 6). For T_{create}, 8/13 participants responded with a 6 or 7 (median: 6).

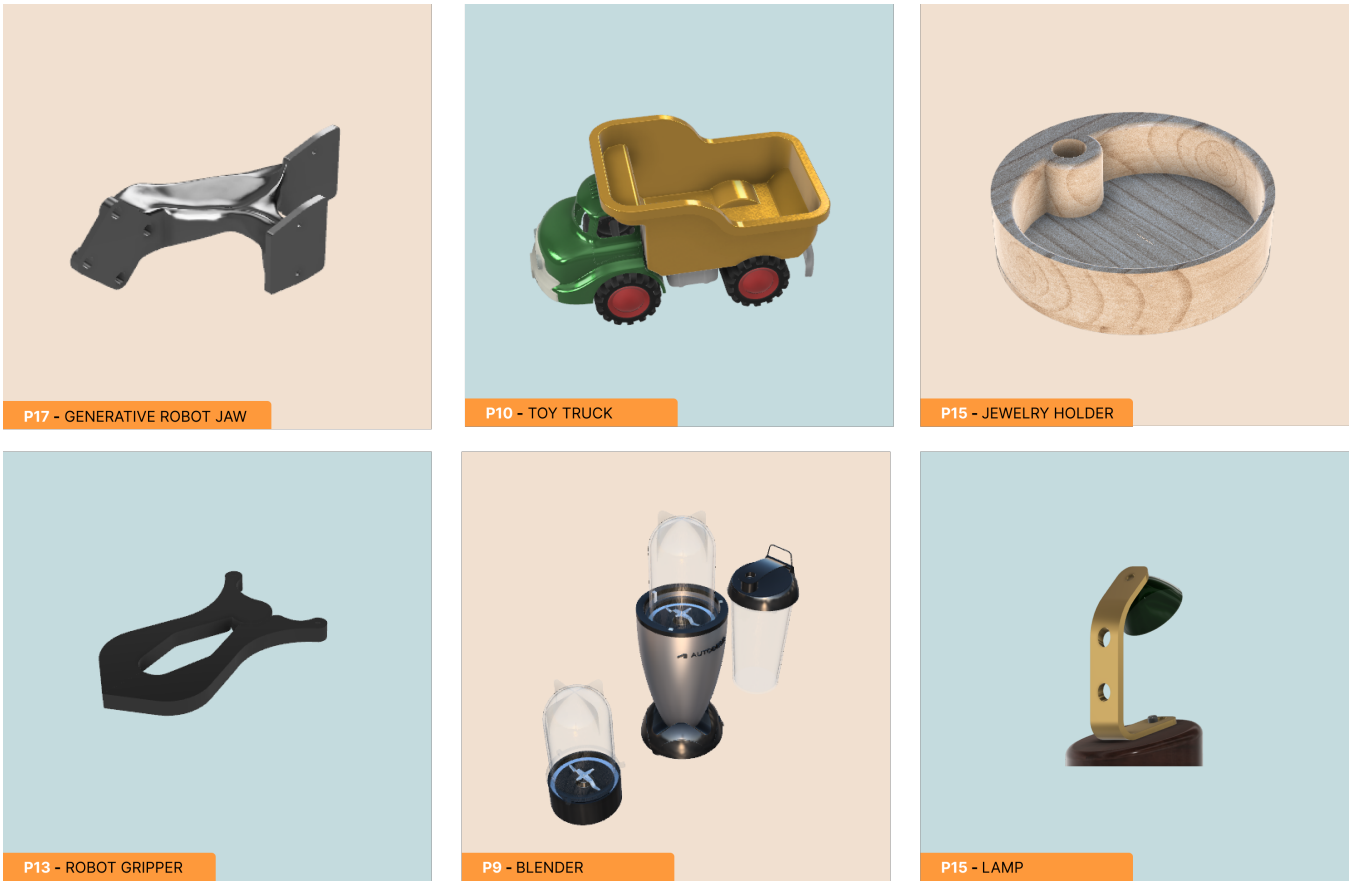


Figure 5: Examples of 3D designs participants brought in during T_{edit}, which was to edit an existing model.

Table 1: Table of participant details, with discipline, Fusion360 usage frequency, and years of experience. We list labels for the model they designed during T_{create} and labels for the model they brought in (T_{edit}).

ID	Discipline	Fusion360 Freq.	Exp.	T_{edit}	T_{create}
P1	Mech. engineering, CAD for robotics competitions	Few times /week	4 yrs	robot	prosthetic hand
P2	Design grad student, CAD + drone design instructor	Daily	4 yrs	drone	airplane
P3	Mech. engineering + design student, CAD hobbyist	Few times /year	1 yr	ring	iPhone
P4	Technical CAD software demos and sales	Daily	7 yrs	machined part	Bluetooth ear gauge
P5	Mech. engineering student, CAD hobbyist	Few times /month	2 yrs	jewelry holder base	outdoor 3D scene
P8	Mechanical engineer	Few times /month	2.5 yrs	table top	table
P9	Mech. engineering student, CAD hobbyist	Few times /year	2 yrs	spray bottle	mittens
P10	Technical accounts executive for CAD (demos)	Few times /week	8 yrs	truck	shelf
P11	CAD technical support for machining	Daily	1.3 yrs	blender	bottle
P13	CAD software engineer, prev. industrial designer	Daily	8 yrs	gripper	speakers
P15	Technical product manager at car company	Few times /year	5 yrs	lamp	bookshelf
P16	Mechanical engineer	Few times /year	1 yr	sensor mount	screwdriver
P18	Technical sales (CAD demos), industrial designer	Daily	8 yrs	microphone stand	car

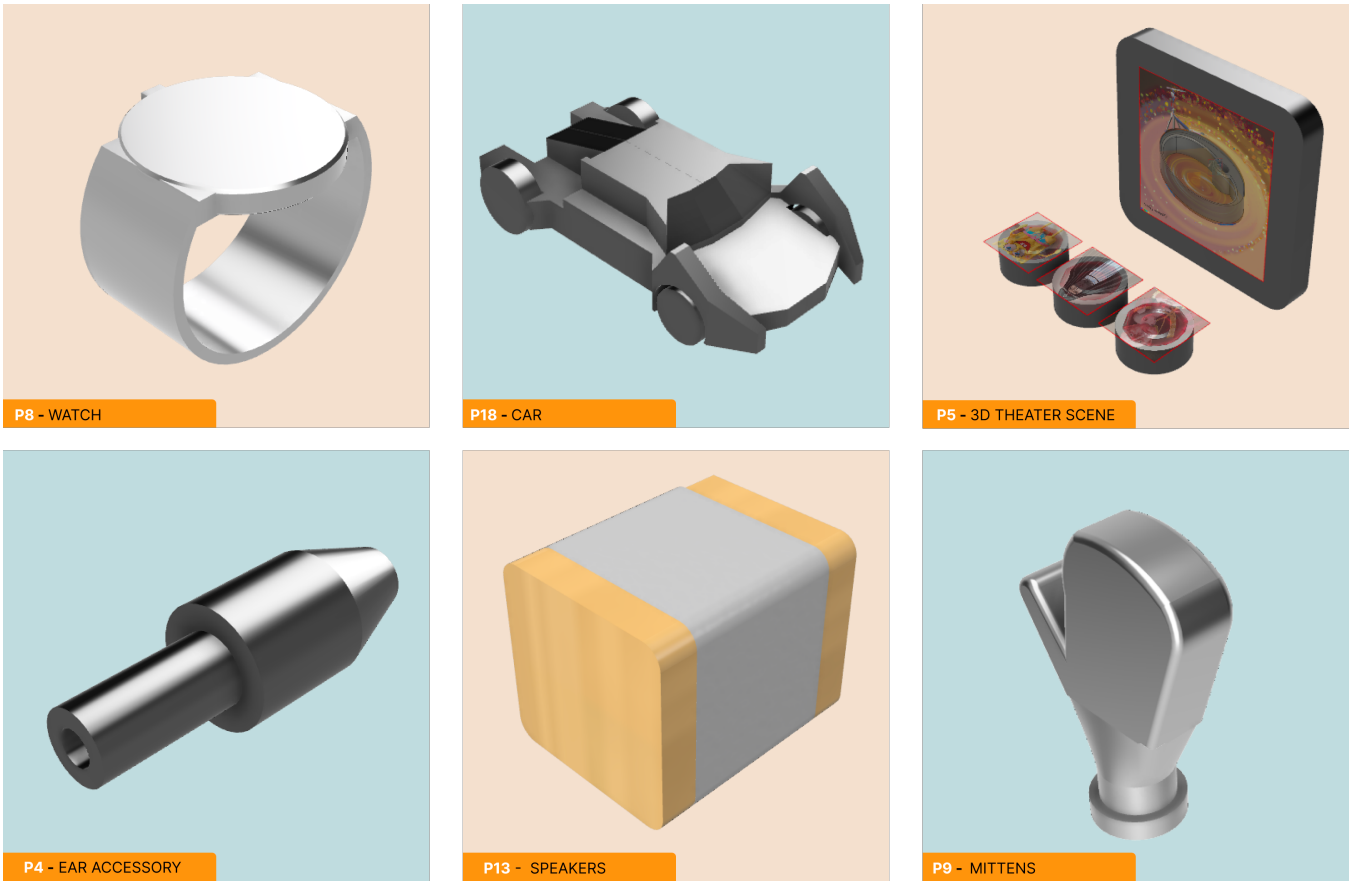


Figure 6: Example of 3D designs participants came up with during T_{create}, which was to create a model from scratch.

> *"Image prompts definitely allowed me to tailor the outcomes towards what I was hoping for or expecting maybe…I'd have struggled to replicate [the render type] if I hadn't done the click on the image [sent in an image prompt] and create some variations. I think once I found something I liked, using those variations made it much easier to stuck to that design theme." - P13*

> *"This middle one is pretty insane…it has integrated my design into the image properly…even as an assembly, I think that's completely nuts…[An image prompt] connects what I'm working on with it [DALL-E]…otherwise it might be giving some random results, and after a while it might become redundant for me."] - P18*

We analyzed participant prompt logs to quantify how often participants used 3DALL-E-provided prompt suggestions. For both T_{edit} and T_{create}, we counted how many times participants used the 3DALL-E-provided prompt suggestions (3D keywords, designs, parts, and styles) and how many times participants provided a custom keyword. Collectively, these represented all the keywords within prompts. Across both tasks and all participants, we found that 3DALL-E-provided prompt suggestions accounted for 63.61% of all prompt keywords, showing that participants heavily used the GPT-3 function of 3DALL-E. We also see in Fig. 7 that 3DALL-E provided the majority of prompt keywords (at least half) for 9/13 participants in T_{create} and 9/13 participants in T_{edit}. These results are summarized in Table 2.

Table 2: Source of prompt keywords across tasks, comparing the frequency of prompt keywords supplied by participants versus by 3DALL-E. 3DALL-E provided the majority of prompt keywords in both tasks.

	Participant-provided	3DALL-E provided
T_{edit}	34.95%	65.05%
T_{create}	38.64%	61.36%
Both tasks	36.39%	63.61%

5.3 Prompting Behavior

We were able to observe certain patterns of prompting with 3DALL-E as each generation action was logged by our interface. From these logs for both GPT-3 and DALL-E, we were able to provide timelines of generation activity in Fig. 10 (T_{edit}) and Fig. 11 (T_{create}).

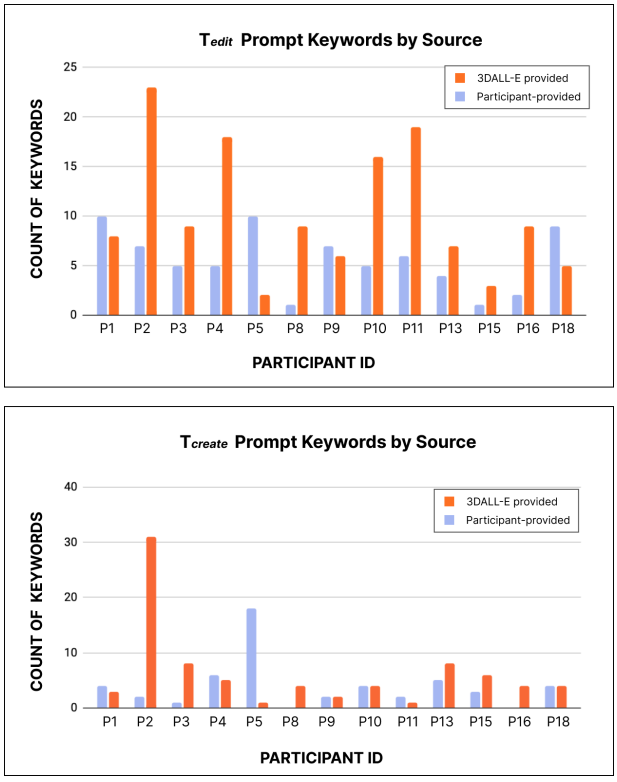


Figure 7: Count of prompt keywords by source (3DALL-E- or participant-provided) for each participant during T_{edit} (top) and T_{create} (bottom). 3DALL-E provides at least half of prompt keywords for 9/13 participants in both tasks.

5.3.1 AI-first, AI-throughout, or AI-last. One of the most salient ways to distinguish participants was at which points in their workflow they took to 3DALL-E and at which points they focused on Fusion 360. Some participants were *AI-first*, meaning they tended to sift through AI generations first until they had a better grasp of its abilities or until they found a design that they liked before taking any significant 3D design actions. For example, P18 (top row of Fig. 11), a technical software specialist with an industrial design background, was trying to make a car. They first began looking for inspiration for a matchbox car, before diving into prompt suggestions like "sports car". Text prompts that P18 tried included *"a single sports car built like a Lego building block, view from the top."* and *"The Dark Knight Rises: the body of a car as a Lego building set"*. They added perspective ("view from the top") and a number word ("single") to specify the composition of their generation and tried "The Dark Knight Rises" as a style suggested by 3DALL-E for the query *"matchbox car"*. After liking one of the resulting generations (Fig. 13), P18 used the result as a reference image. For the rest of the duration of the task, P18 modelled within Fusion 360. P18 first traced over half of the generation like a blueprint before extruding faces to varying heights. They then beveled and chamfered these starting blocks of a car to add ridges and windshields and subtracted material to make room for wheels. They ended by mirroring the half of the car they modeled to create a full symmetrical car. P18's prompting and modeling workflow for T_{create} is shown in Fig. 9.

The *AI-last* pattern occurred when participants jumped straight into their existing workflows for 3D design and tried 3DALL-E later. We see this in the rows of Fig. 11 that start off with orange bars, which indicate that participants started modeling from the get-go of the task. P11, for example, was trying to make a bottle. They began by sketching the cross-section of a bottle and revolving it 360 degrees to create a form. After filleting the base to round it and hollowing it out with a hole, they found prompt suggestions from 3DALL-E like *"fusion 360"* and *"Coca-Cola"*. Using a generation prompted from *"Front view Coca-Cola Bottle"*, they edited their bottle cross-section to match that of the generation. Only after they had created this basic bottle did they start looking for inspiration; seeing generations of Coca-Cola bottles *later* helped P11 figure out how to bring complexity into the cross-section of their design. P16 (second row in Fig. 11) was another AI-last participant. They already had an existing screwdriver concept in their mind. They began by sketching and extruding a rounded rectangle for the grip of the screwdriver, dimensioning accordingly. They worked on the flat-head tip by extruding a narrow cylinder and lofting the face out to a point. After making a rough model, they tried 3DALL-E with prompts specific to flat-head screwdrivers and used their existing modeling progress as an image prompt. P16 commented that 3DALL-E inspired them to consider different handle cross-sections (e.g. hexagonal, square) and grooved grips. Note that the AI-last pattern, jumping into a participant's existing workflow with Fusion 360, was more prevalent in T_{create}.

However, there were also participants who queried *AI-throughout*. Many participants (P13, P1, P8, P10) would intermittently craft an image prompt by briefly working within Fusion 360 and then start generating. We see these actions whenever participants would have a short window of Fusion time that led up to image+text generation (medium blue dots in Fig. 10 and Fig. 11). During these short windows, participants were generally changing their camera perspective or the visibility of different parts in their assemblies. For example, P10 hid the hopper of a toy truck they had brought in and tried to generate different semi-trailers using prompts such as *"Jeep Gladiator snow plow truck"*. P13 (during T_{create}) was another *AI-throughout* participant. They first built up a base for an audio speaker they wanted to design and applied wood and chrome finishes for a Scandinavian design aesthetic. They then tried prompts with lighting elements (e.g. *"Isometric Scandinavian minimalism audio speaker with built-in lights"*). They built towards a generation they liked for a while, adding details of a speaker cone and applying tessellation and reducing operations to give the speaker body structural texture. Then they began to create image prompts for 3DALL-E to fill in—deleting faces and extrusions or hiding bodies in their geometry. They wanted to see the different ways the middle section of their speaker could be autocompleted. We see P13's work in T_{create} and the way they utilized AI-throughout their workflow illustrated in Fig. 9.

Participants would also use text-only prompts to take them towards new directions. P9 used text prompts to pivot their design multiple times and better scope their 3D design. Originally, P9 intended on creating a prosthetic hand and tried generating *"A 3D model of a robotic hand with two fingers"*. After finding modeling a

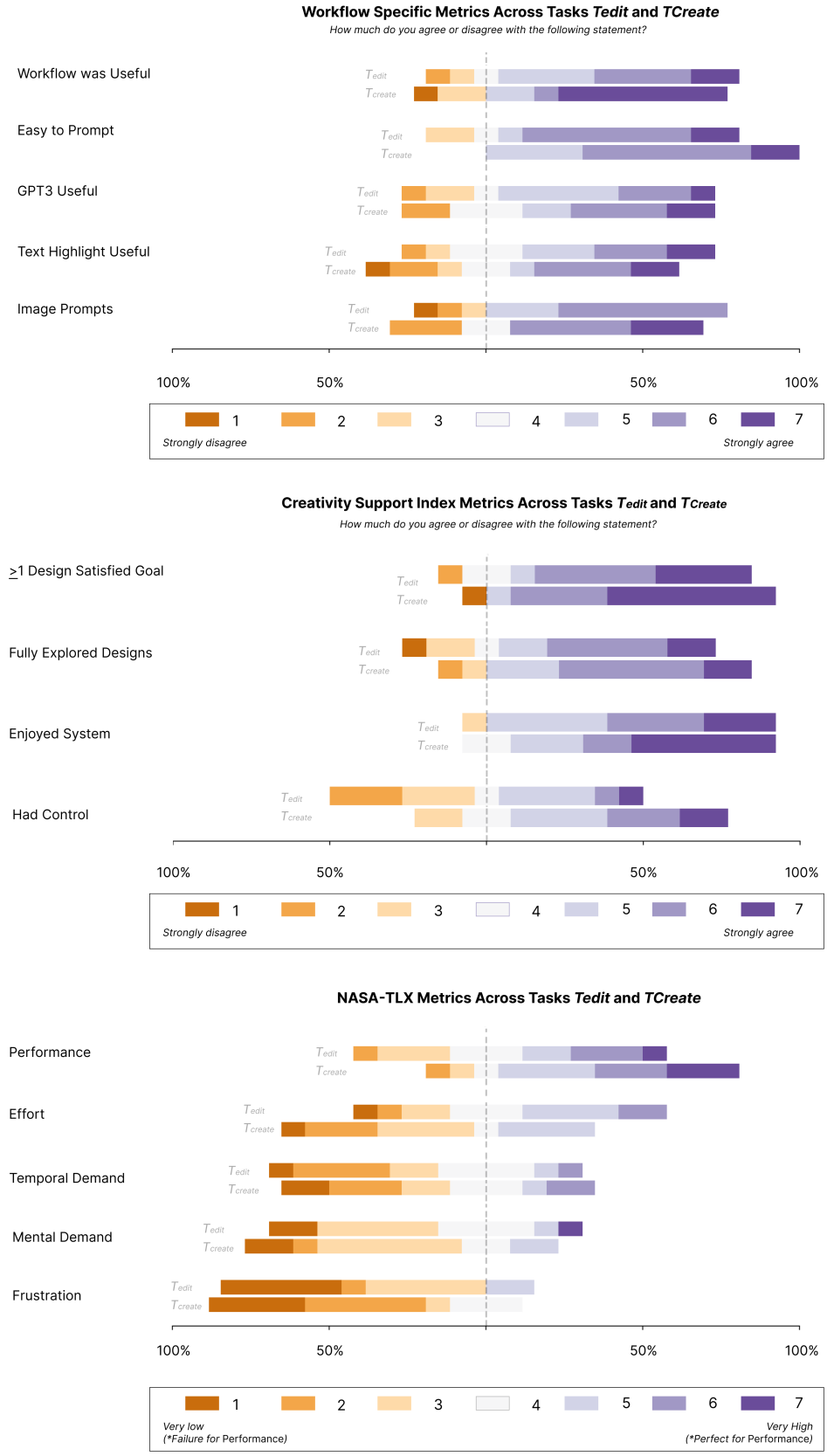


Figure 8: Distribution of Likert scale responses on NASA-TLX, creativity support index, and workflow-specific questions across all participants for both T_{edit} and T_{create}. Full questions are in the Appendix.

Figure 9: Prompting and 3D modeling workflows of design process of three participants (P18, P13, and P1). P18 created a car, P13 created an audio speaker, and P1 edited a robot. Timelines are vertical with the markers representing different generation requests and yellow intervals representing CAD time. The markers preserve order but the time stamps across participants are not aligned / to scale.

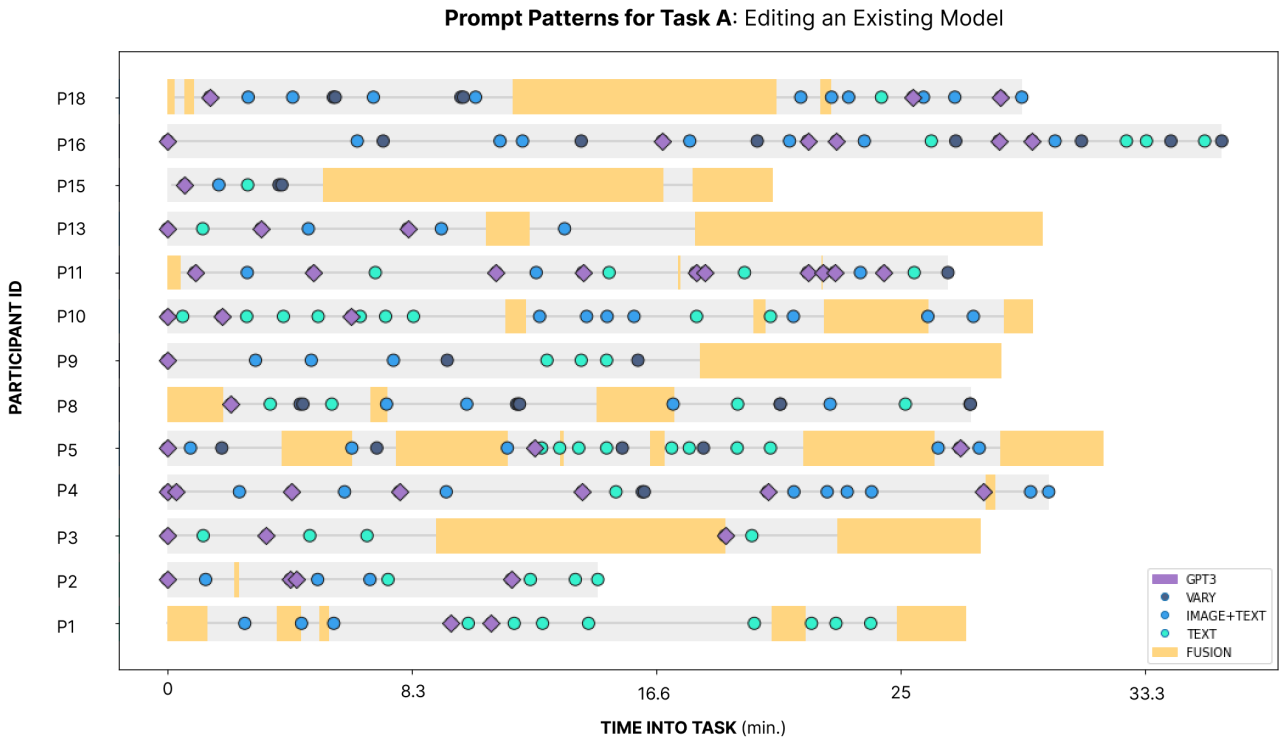


Figure 10: Pattern of generation activity for T_{edit}, when participants edited an existing model.

hand to be too complex because of how articulated they are, they tried text-only prompts *"3d model of a human fist"* and *"3d model of mittens"* to explore what they could more feasibly model, exploring divergently. Deciding on mittens, they imported a generation as a reference and sketched over it. After extruding the sketch and applying fillet operations to round out the mittens, P9 added cuff sleeves, a detail inspired by the generation.

In terms of generation patterns for GPT-3, nearly everyone started with generating from GPT-3 (though this could be because of the organization of the user interface). Many continued to use GPT-3 throughout each task, and we can see this reflected in the fact that there are purple diamonds (GPT-3 actions) at the early, middle, and late stages of workflows for both T_{edit} and T_{create}.

5.3.2 Switches between Types of Prompting. Eight participants passed in *an image prompt* as their first action in T_{edit}, and eight participants passed in *text prompts* as their first generation action for T_{create}. This suggests that participants may be more likely to pass in an image prompt if they already have work on their page. Aggregating across all the different generations across T_{edit} and T_{create}, we did not see that any mode of prompting was favored more than the rest. Preferences in prompting were highly dependent upon the participant and also how well the participant felt like the generations incorporated their image prompts. For example, even though P13 found image prompts useful, they felt like image prompts were incorporated in an "awkward" way, as they had more glaring visual artifacts than text-only generations.

In certain rows in Fig. 10 and Fig. 11, we could see that some participants would shift away from using image prompts and focus on text-only prompts. A case in point of this was when P1 worked on a tank-drive robot that they had built for a FIRST [39] robotics competition during T_{edit} (pictured in Fig. 9). To craft image prompts, they played around with different angles of their models and toggled the visibility of parts like the wheels and ground plane of their model. The robot was a highly convoluted assembly, and while they found that 3DALL-E could generate decently even on these visually complex image prompts, they ended up passing in a series of text-only prompts like *"3D illustration of a Roomba with four wheels powered by motors"* and *"flat image of a toy wheel"* (focusing in on a specific part rather than trying to get 3DALL-E to work with the full assembly was also a common strategy of participants). In this situation, the text-only generations were easier for P1 to parse and make sense of. P5 was another example of someone who pivoted away from passing in image prompts to use text-only prompts after receiving sets of unsatisfying generations during T_{edit}. The image prompt that they passed in was a mechanical base, so the generations building off of that were all visually indeterminate (not recognizable as any particular object). P5 instead decided to generate textures of water and maple syrup to project onto their original model (as seen in Fig. 6), finding this to be an easier way to make use of their part and 3DALL-E.

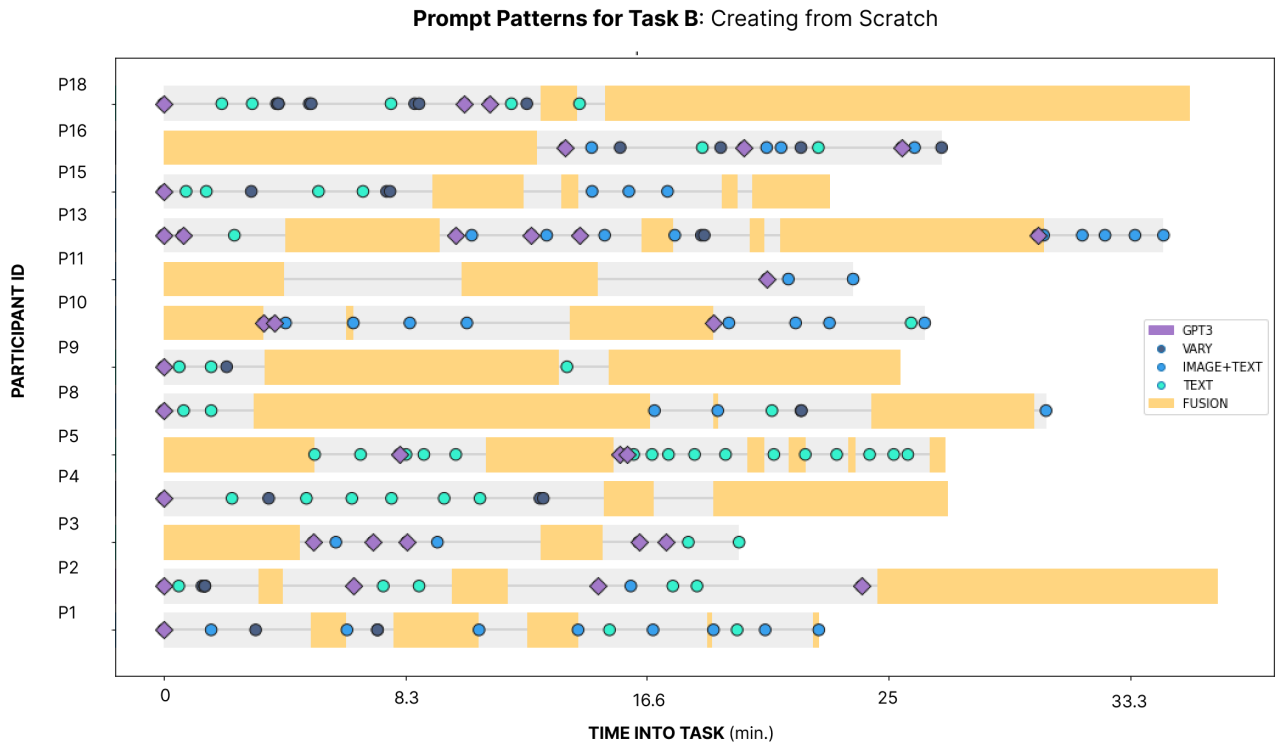


Figure 11: Pattern of generation activity for T_{create}, when participants created a model from scratch.

6 PROMPT COMPLEXITY

It can be challenging for an end user to understand how lengthy or detailed a text-to-image prompt should generally be, which is why we studied prompt complexity with 3DALL-E. In 3DALL-E, GPT-3 would automatically rephrase selected prompt suggestions while adding a small amount of connecting words. Based on this design, we could measure complexity as the number of concepts forming the basis of a prompt. For example, if *"3d render, minimalist, chair"* was rephrased as *"3d render of a minimalist chair"*, we gave the prompt a count of 3 concepts.

However, participants also had the ability to edit the final prompt and to add or subtract concepts of their own. In cases where the text prompt mostly came from the participant rather than GPT-3, we counted the number of concepts based on rules from linguistics and natural language processing. The prompt complexity was then the number of noun phrases and verbs in a prompt, ignoring prepositions, function words, and stop words. Count words were ignored; they were considered modifiers for the noun phrases they were a part of (e.g. "five fingers" was one concept).

We annotated text-only and image+text prompts with the number of concepts. We did not annotate variations for complexity because the generation of those images were not directly informed by text prompts. From these annotations, we charted prompt complexity across participants in Fig. 12. We found that participants tended to explore between two to six prompts, which is where most of the density of points concentrates in Fig. 12. We see that participants were also willing to try a range of concepts, as we can see in the wide spread of P2, P9, and P10. Fig. 12 also shows that participants could easily assemble prompts of over six concepts with this workflow.

We note that even when the prompts were filled with concepts: *"V-shape,Y, Tricopter, Sports, Abstract, Landscape, Aerial, Gimbal, Camera, Transmitter, Flight controller, Receiver"* , 3DALL-E could still return legible images. For this prompt, P2 received generations that had laid out displays of product components. P2 was an obvious outlier in the complexity of the prompts that they provided. They were keen on trying to "break the system" and passed prompts averaging 10 concepts. We did not discern a difference between complexity observed for T_{edit} and T_{create}.

7 QUALITATIVE FEEDBACK

7.1 3DALL-E Use Cases for CAD Design

7.1.1 Use Case: Preventing Design Fixation. Participants demonstrated different use cases of 3DALL-E as they progressed through the tasks. The most commonly acknowledged use case was using the system for inspiration, particularly in the early stages of a design workflow. P10 contextualized some of the challenges that 3D designers face on the job, such as design fixation and time constraints. *"A lot of times designers get stuck, they get tunnel vision...the folks at [toy design company] used to say to me, "We can't come up with enough designs...it takes too long to come up with a design, so then we only get two or three...we would like to see thousands of design*

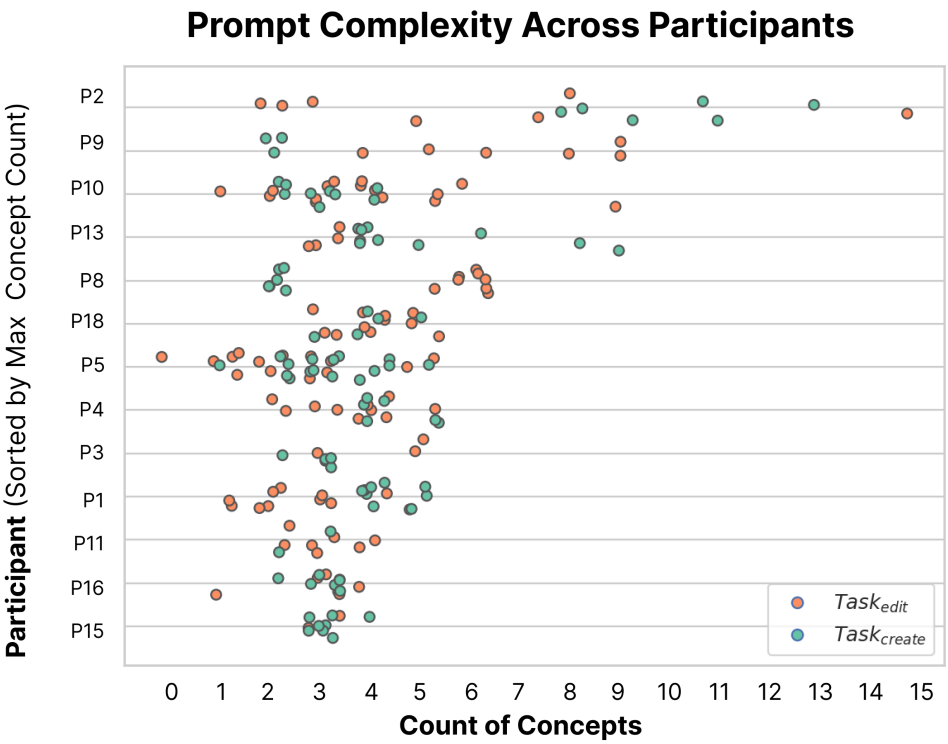


Figure 12: Prompt complexity measured across participants, where complexity is the count of concepts in each text-only and image+text prompt. Participants span the X-axis, sorted by the count of their most complex prompt. The values are jittered to show multiplicity; many prompts mapped to the same number of concepts. Complexity tended to concentrate between two to six concepts, as seen by the density of prompts within that interval. Each datapoint was colored based on prompt task.

options and variations...the [designer's] goal is to start throwing as many designs out there as they can."

Participants felt like the tool could be *"game-changing"* (P11) for certain industries such as consumer products, automobiles, and game assets (P10, P11, P15, P17, and P18). They likened it to existing search and intelligent suggestion tools like stock photography websites (P15) and Google Images, but noted that with 3DALL-E, it was better in that users could access inspiration without leaving their workbench (P11). For example, P8 was building a table and explored many different design styles from *"industrial minimalist"* to *"nature-inspired Scandinavian"* by using text-only prompts in quick succession. They also passed in the table top they had already modelled to see how it could be completed by 3DALL-E as a *"CGI traditional farmhouse table with centerpiece drawers"*. They did all of this without having to switch applications, which is important for 3D design software, as it requires focused modeling time.

7.1.2 Use Case: Reference Images for 3D Geometry. Many participants (P18, P3, P13, P11) imported generations into the 3D software as reference images to model off of. P18 and P13, both of whom had backgrounds in industrial design, described how designers traditionally gather reference images to build their models as part of their CAD workflows. These images generally aligned with specific views: front, side, top-down, perspective, or isometric. P18 said, *"I would probably need at least three images: top, side, and front view to even understand it three-dimensionally...that's what a designer would pass to an engineer to then build it. I would try to force it [3DALL-E] to create a top view, side, front view that are somehow matching."*

Figure 13: Three DALL-E generations participants (P18, P15, P9) found inspirational from the prompts: *"The Dark Knight Rises: the body of a car as a Lego building set top view"*, *"3D render of a desk lamp Victorian"*, and *"isometric 3d renders of a cleaning sprayer bottle"*.

P18 used a top view generation as a basis for their model as shown in Fig. 9. We note that most generations came back angled and at perspectives unless the prompt explicitly specified viewpoints like "top view" or "flat", and that 3DALL-E did not always

capture "isometric" and "perspective" views in the technically accurate sense of those words. Nonetheless, even if generations were not drawn to perspective or as clean as technical drawings and renders usually are, participants still found them useful as reference images.

Other participants used the generations as references albeit more loosely. P15, liking a *"3D render of a desk lamp Victorian"* (Fig. 13), made the arm of their lamp skinnier as per the generation. P9, observing generations from prompts such as *"isometric 3d renders of a cleaning sprayer bottle"* (Fig. 13), noted that they could subtract volume from the outer contours of their model and reduce the amount of material used, which was part of their goal to design a more sustainable spray bottle top.

7.1.3 Use Case: Textures and Renders for Editing Appearance. Participants would also edit their model appearance towards the look of generations (P13, P15, P10, P3). They could do this by applying textures within the software and dragging and dropping materials from the software's material library onto surfaces. P5, innovatively used generations as textures to help build a 3D outdoor movie theater scene. Their scene was built out of simple geometries, and atop these geometries, they placed generations of a *"jello bed"* and generated portraits of pop culture characters (pictured in Fig. 6).

P1 mentioned that 3DALL-E could be useful for product design presentations to show the function or interaction of things being designed. As P1 made a prosthetic hand, they imported a generation and started to model atop it. Curious about how a text+image prompt would fare if it included a generation transparently overlaid over their geometry, they generated and found compelling images that could visually situate their designs with their use cases in product design presentations.

7.1.4 Use Case: Inspiring Collaboration. Design in industry is a team effort, and while 3DALL-E was evaluated in the context of a single user, many participants acknowledged that 3DALL-E could be beneficial in teams. P16 mentioned that from their industry experience, 3DALL-E would be excellent for establishing communication between mechanical engineers and industrial designers. Mechanical engineers focus on function, while industrial designers focus on aesthetics. P16 felt that 3DALL-E could help both sides pass around design materials for discussion and common ground.

P13, who was an industrial designer, noted that teams could also do multi-pronged exploration with 3DALL-E. Because each team member would have individual prompting trajectories, a team could easily produce diverse searches and more variety during brainstorming. P3 mentioned that there are already points within their industry (automotives) where there are hand-offs between the people who generate design ideas and the people who execute them. Technical sales specialist P4 also mentioned that they could instantly see 3DALL-E being useful for their clients, many of whom have bespoke requests such as organic fixtures for restaurants and museums or optimized shapes for certain materials.

7.1.5 Use Case: Inspiring Design Considerations. 3DALL-E also inspired design considerations by making participants think about different aspects such as functionality or manufacturability. For example, P1 was looking for a wheeled robot. Seeing generations where robot bodies were varied in the number of wheels they had or how far off the ground they were made P1 think about the different amounts of motor power these robots would require. While 3DALL-E could not guarantee the feasibility of every generated design, some participants (P1, P8) liked that 3DALL-E inspired them to think through details such as how manufacturable a design was.

Participants also felt like they could elicit unique, out-of-the-norm designs from 3DALL-E and use it to let them gauge the uniqueness of their own designs. P4 wanted to design a product that did not exist in the real world yet: an ear gauge electronic for their son. They treated the model's inability to come up with their exact vision in generations as a good thing, interpreting it to mean that the product did not exist yet and therefore had patentable value. *"We [DALL-E] started to lose a little bit when we started putting in the 'Bluetooth ring', which is good because that tells me... probably out there in the real world, nobody's actually doing this... that made me feel good about the fact that I might have a predicate design in my head."* P2, who had taught drone design classes, also felt like right off the the bat, 3DALL-E was able to produce unique aesthetics beyond what is typically seen in drones, something their students generally struggled to do. P15 also felt like 3DALL-E could have educational value as they looked around for ways to accomplish something they saw in a lamp generation: *"being able to reverse engineer... that is a cool learning aspect."* 3DALL-E could not guarantee the educational or patentable value of a generation, but it inspired participants (P4, P2, P15) to think about design considerations such as design conventions, uniqueness, and plausibility.

7.1.6 Weaknesses in terms of CAD. Some participants did comment that text-to-image AI may have weaknesses in applications like machining and simulation or the construction of internal components and other function-focused parts. P9 pointed out that it would be difficult to generate geometries that enclose parts, because if a user was to pass in an image prompt of that part, 3DALL-E would be unable to draw housing over it. Likewise, a participant mentioned that they could imagine 3DALL-E being used to design the facade of the car, but they did not believe that it could design a more internal component not easily describable in layman's terms.

7.2 Comparing with Traditional Workflows

Our exploratory study invited designers to stress test 3DALL-E across the settings of a wide range of disciplines. Participants were impressed with the ability of the model to generate even when they passed prompts filled with technical jargon like "CNC machines", "L-brackets", or "drone landing gear". Still, prompting remains very distinct from the workflows participants usually go through. Many participants described their regular design process as multiple phase progressions from low fidelity to high fidelity. They mentioned roughing out designs first, putting placeholders within robotic assemblies (P1), box blocking up to complexity (P13), and redesigning from the ground up again and again (P18). Even though 3DALL-E only provided images of 3D designs, these designs could have high fidelity details that could shortcut participants to later stages of the design process.

7.2.1 Text Interactions in 3D Workflow. The most distinct difference in workflows is that 3DALL-E is text-focused, but text is not central to 3D design workflows, which are usually based on the direct manipulation of the geometry. P13 mentioned that designers

primarily operate visually. *"The only reason I really use text in an industrial design context is [for] making notations on a design...to explain what a feature is...to write a design specification...but the majority of the time is image focused."* Because of this, P13 preferred the "image-based approach" within 3DALL-E where they could "provide it with a starting image and get variants of that". P4, however, thought that in some respects designers *are* often engaging with text, but in the form of numbers, properties, parameters, equations, and configurations. *"[We] do it in a smart way...[we] drive it with the math equation. This is something we can do in parameters, and it is very text-based."*

7.2.2 Problem Solving with 3DALL-E. P10 and P4 described their day-to-day job tasks as customer-facing CAD specialists as problem solving and finding design solutions. P10 began the study wondering if 3DALL-E could solve a problem they were facing in their job: packaging a toy truck. To do so, they like many of the participants, tried employing 3DALL-E as a problem solver. P10 tested prompts such as *"create a toy dump truck and fire truck with plastic material"* and *"protect a sphere with foam"* to see if 3DALL-E could help encase a 3D model. From the results they saw, they concluded that 3DALL-E *"was not intended to be a problem solver type of tool"*.

P13 set up image prompts as autocompletion problems. As they built an audio speaker for T_{create}, they commented that they were *"creating two pieces of geometry and using it [3DALL-E] as a connection between the two… kind of like the automated modeling command"* [36]. They also tried other innovative ways of creating image prompts: *"a hacky approach, trying to keep preserved geometries with the faces and using 3DALL-E to fill in the gaps"*.

7.2.3 Driving the Design. When AI input is added into a workflow, questions of who drives the design process and who owns the final design can arise. While P9 liked that 3DALL-E augmented their workflow with what they called dynamic feedback, they felt as though their design was being driven by the generations. *"Initially, the image did not really meet my expectation… but eventually I was also trying to not imagine anything and just depend upon what it was suggesting."* P3 mentioned that they felt as if they were driven by 3DALL-E, while P15 mentioned that sometimes in the midst of exploring, they felt they were not gravitating towards building.

As for ownership, many participants felt like the designs they created with 3DALL-E would still be their own. P1 stated on ownership, *"A lot of 3D modeling is stealing...borrowing premade files online, and then assembling it together into a new thing. For this robot, we borrowed these assemblies from already premade files that were sold by the company. We modelled based off of that, but the majority of this robot can be considered ours because we determined the placement."* P13 was also not worried about ownership concerns, stating that even now, anyone can recreate any model found online, but that *"it's about the steps you go through to get there."*

P18 mentioned that for an AI to be applied to the real world, it still takes an expert designer's understanding of the market and customer needs. *"I would use my know-how of manufacturing processes and the market or style. My service would adopt AI as a source of inspiration rather than as the solution."* Reflecting on if AI inspiration became mainstream without designers in the loop, they expressed concerns that *"if everyone would converge on the same designs [because] it only learns from the input it gets from people… we might lose creativity."*

7.3 Comparison with Existing Generative CAD Tools

Five of 13 participants had experience with the existing generative design mode within the 3D CAD software [38]. Generative design (GD) is an environment in Fusion 360 in which the completion of a 3D design is set up like a problem: users define physical constraints and geometric filters that allow a model to be autocompleted. We did not directly compare with GD, because hardware constraints made 3DALL-E incompatible with GD. However, we did ask participants with experience in GD to compare and contrast the two.

A primary difference was that GD allows users to directly manipulate the model geometry, which differs from the text-based interaction of 3DALL-E. GD results therefore free the user from doing more modeling work. What one participant liked about GD was that *"once they set up the problem, they could just hit go… don't have to actually worry about lofting and modeling"*. However, participants mentioned that GD has a higher barrier of entry; users are burdened with calculating loads and non-conflicting constraints, which requires some understanding of physics and engineering.

> *"You're [GD] focused on strength, endurability of the model itself, really driven as a manufacturing task… your end result is something that's makeable… whereas this process [3DALL-E] is more on the creative side."*

P2 mentioned that 3DALL-E allowed users to come up with outcomes far more efficiently than GD. In the span of a 30-minute task, users were able to browse hundreds of results, with the first results coming in a matter of seconds, whereas P2 has previously had to wait multiple hours or even days for GD. P2 and P18 were enthusiastic that GD and 3DALL-E could merge. P10 suggested that one way these two tools could complement each other is if *"this tool [3DALL-E] could be used to generate shapes… pass it off to the generative design [GD] to optimize"*.

8 DISCUSSION

Our results demonstrate high enthusiasm for text-to-image tools within 3D workflows. With 3DALL-E, participants had a tool for conceptual CAD that could help them combat design fixation and get a variety of reference images and inspiration. Furthermore, we elaborated prompting patterns that can help understand when and what types of text-to-image generation can be most helpful. In measuring prompt complexity, we showed that many prompts fall within a range of two to six concepts, providing a heuristic that can be implemented in text-to-image prompt interfaces. The following discussion focuses on best practices for helping 3D designers bring their own work into AI-assisted design workflows and the implications of these workflows.

8.1 Prompt Bibliographies

A strength of studying 3D workflows was that there was no conflict between the AI and human on the canvas, as the AI had no part in the physical realization of the design. We believe this helps mitigate

ownership concerns and makes text-to-image AI very promising for 3D design tools. Currently, AI-generated content is a gray area due to concerns of attribution and intellectual property [76]. Currently, there is no way to tell how heavily an AI-generated image borrows from existing materials. As generated content becomes more prevalent on platforms, it is important to develop practices of data provenance [19]. We propose the notion of *prompt bibliographies* to provide information on what informed designs and to separate out which contributions were human and which were AI. These can work to clarify ownership and intellectual property concerns.

Prompt bibliographies, illustrated in Fig. 14, could likewise help track designer intentions and enrich the design histories that software tools provide, which generally capture commands and actions (but not intentions). The bibliographies can be merged within the history timeline features that are present in tools like Fusion 360 and Photoshop, helping prompting integrate better with the traditional workspaces of creative tools.

Sharing prompt bibliographies with their outcomes (i.e. 3D models) can also help respect all the parties that are behind these AI systems. End users can easily query for the styles of artists (as they already do) and create derivative works that dilute the pool of images attributed to artists. Prompt bibliographies may be especially relevant for CAD designers as CAD is highly intertwined with patents, manufacturing, and consumer products.

8.2 Enriching creative workflows with text

The advancements in prompting may push text prompting as a type of interaction into creative tools, even if creative workflows have traditionally not revolved around text. In 3DALL-E, we show the benefit of having a language model scaffold the prompting process. By giving the user fast ways to query and gesture towards what an AI is most likely to understand (as 3DALL-E did with the highlighted text options), we enable users to have more opportunities to understand what language may work best with an AI. At the same time, 3DALL-E helped users easily reach the design language of their domain, be it robotics or furniture design. In the quantitative survey results, participants felt it was easy to come up with prompts near unanimously for T_{edit} and unanimously for T_{create}.

It is important to understand where in a workflow assistance can be of most use. Our survey results reflect that 3DALL-E produced a slightly more positive experience when it was introduced earlier on in the process. This was corroborated by many participants who said they saw this tool being most helpful in the early stages of design. Well-placed AI assistance, such as early stage ideation with GPT-3, trying a text-only prompt to pivot directions, or carefully setting up an image prompt for 3DALL-E to fill in—can be greatly constructive and address painpoints like design fixation that CAD and 3D designers in general feel today. Furthermore, if we understand the scope of the tasks we want AI to handle within a workflow, such as having GPT-3 suggest different parts of a model or having DALL-E generate reference images from front, side, and top views, we can better fit general purpose models to their task. We can have stronger checks on the prompt inputs and generation outputs if we understand what is within scope of the task. For example, when P16 wanted a "flat head" screwdriver, they were returned results about a medical syndrome—something that could be avoided with content filtering guards checking for relevance to 3D design. AI models may not have to bear the full burden of providing good and ethical answers if we can have multiple checkpoints for propriety.

8.3 Generalizability

The design workflow posed in 3DALL-E is generalizable and can easily be used as a blueprint for text-to-image AI integration with different design software. The idea behind surfacing 3D keywords from application related data (as we did with Fusion 360 Screencast data) also introduces ideas for how prompts can be tailored towards the technical vocabulary of a software. The idea of passing in image prompts is also easily extendable to different creative tools, even those outside of the 3D space. For example, graphic editing tools can pass in image prompts based on active layers chosen by a user. Animation software and video editors can send in choice frames for anchored animations and video stylization. A takeaway of this paper is to take advantage of the complex hierarchies that users build up as they design, such as the way 3DALL-E takes advantage of the fact that 3D models are generally assemblies of parts. With 3DALL-E, users could isolate parts and send clean image prompts without the burden of erasing or masking anything themselves.

8.4 Benefits of Text-to-Image for CAD

Few tools currently explicitly support conceptual CAD [35, 48]. 3DALL-E supports conceptual CAD not only at the beginning of the design process, but also throughout their workflow, as evidenced by the different usage patterns. It provides visual assets for CAD / product design as well as design knowledge that is otherwise difficult to collect (e.g. standard designs, specific part terminology). These visual assets can be utilized for detailed sketching within CAD, for appearance editing through materials, or for the inspiration of design considerations. 3DALL-E also presented directions that can solve weaknesses of existing generative tools (GD) for CAD. By having 3DALL-E define shapes and then having the GD environment optimize them, existing generative tools could better align with what designers visually want, and go beyond physical constraints like loads and forces.

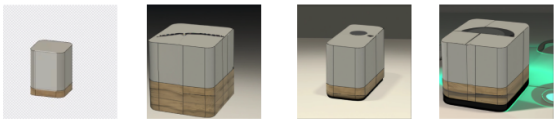
We demonstrated the efficacy of 3DALL-E at supporting a diverse set of potential CAD end users: mechanical engineers, industrial designers, roboticists, machining specialists, and hobbyist makers. 3DALL-E's interdisciplinary design knowledge is both a strength of AI pretraining as well as the ability of designers to make integrative leaps to meet the AI halfway [81]. Additionally, the modular nature of 3DALL-E in Fusion 360 demonstrates an idea of separating out AI assistance from traditional non-AI direct manipulation features. Lastly, the text-based nature of the tool and its ready acceptance with designers demonstrates how text interactions can facilitate a low threshold, high ceiling design tool for CAD [63].

8.5 Future Work and Limitations

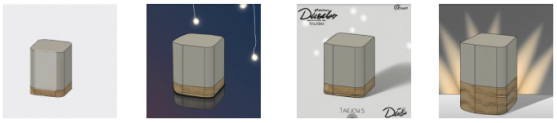
A necessary line of future work to make text-to-image AI more usable for CAD will be to integrate it with sketch-based modeling. Sketching is fundamental to CAD and reliant on the creation and manipulation of clean primitives (splines, lines, etc.), and controlling the composition of text-to-image generations based on sketches would be highly useful.

PROMPT BIBLIOGRAPHY: Design of Audio Speakers (P13-B)

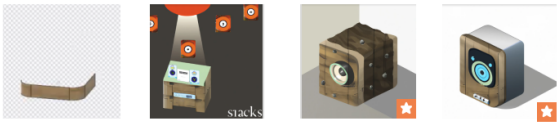
IMAGE+TEXT: Scandinavian, 3D illustration, I'm looking for a good pair of portable Bluetooth speakers



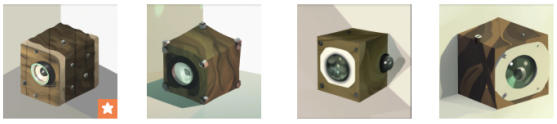
IMAGE+TEXT: Scandinavian Minimalism audio speaker with lights



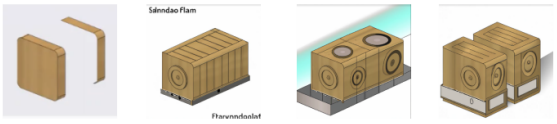
IMAGE+TEXT: Scandinavian Minimalism audio speaker with lights



VARIATION



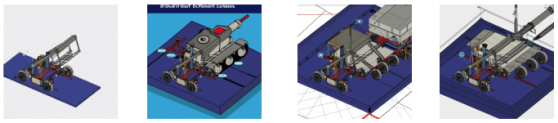
IMAGE+TEXT: Isometric Scandinavian minimalism, audio speaker with built-in light



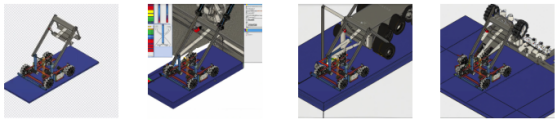
IMAGE+TEXT: Isometric, scandinavian, minimalist, audio speaker

PROMPT BIBLIOGRAPHY: Design of Tank Drive Robot (P1-A)

IMAGE+TEXT: Robot with tank drive system



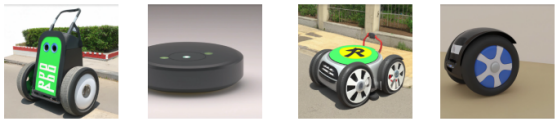
IMAGE+TEXT: Robot with tank drive system



IMAGE+TEXT: Robot with wheels

TEXT: 3d illustration of a Roomba with wheels

TEXT: 3d illustration of a Roomba with four wheels



TEXT: 3d illustration of a Roomba with four wheels powered by motors

Figure 14: Prompt bibliographies, a design concept we propose for tracking human-AI design history. As prompts become a part of creative workflows, they may be integrated into the design histories already kept by creative authoring software. This bibliography tracks text and image prompts, as well as which generations inspired users during the tasks.

In terms of limitations, 3DALL-E, owing to its implementation in a CAD software, is object-oriented and intended to support CAD product designs (and not 3D art more broadly). It can also occasionally return prompt suggestions that are imperfect or irrelevant to their category (e.g. a "cylinder" suggestion could be categorized as both part or design). There were also times during the study when we experienced technical difficulties. For example, some participants had their DALL-E results cancelled. Moreover, when participants tried to compare the generative design environment with 3DALL-E, the software crashed, so we were unable to directly compare 3DALL-E with GD. However, the existence of GD, a cloud-based generative design tool for Fusion 360, shows that there are already CAD designers who utilize generative assistance, and our interviews illustrate that they are open to it improving further. As such, future work can explore how these tools could merge, as 3DALL-E has the potential to help with text-based exploration of GD outcomes. Text-to-3D methods will also be meaningful to explore as they mature in capability of expression, become faster to run at inference, and become more widely available.

Data privacy will also be a key concern in the future. Design know-how and details are the intellectual property of companies and is safeguarded by high-value product industries (i.e. cars). We asked participants to use non-sensitive files, but in the future it will become important to understand how intellectual property passed to AI systems can be protected and not given as free training data. While each prompt needs to be examined for content policy and ethics adherence, there are looming trade-offs to be made in data privacy and AI regulation.

8.6 Broader Impact

Text-to-image methods have entered the mainstream conversation as a tool that has the potential to impact creative jobs and livelihoods. People have begun to utilize these methods to generate logos, vector illustrations, fashion designs, and so on. This paper is a case study for how generative AI tools can be integrated within the conceptual CAD design stages and how CAD design processes can be augmented rather than automated away. There are ongoing discussions about copyright and existing artist work being leveraged as training data that are rightfully merited. However, we believe that the positive response from participants to 3DALL-E illustrates the utility that these tools can present to creatives. Key aspects we think are important for these tools to be successful are that they are narrowed in scope, introduced at early stages of the design process, and still leave room for the creative to exercise their artistic license.

9 CONCLUSION

3DALL-E introduced text-to-image AI into 3D workflows and was evaluated in an exploratory study with 13 designers. This study elaborated a number of use cases for text-to-image AI from providing reference images to facilitating collaboration to inspiring design considerations. From participant prompts, we observed different types of prompting patterns depending on whether the user engaged with 3DALL-E first, last, or throughout their process. Furthermore, we provided measures of prompt complexity across participants and propose a concept for tracking human-AI design history through prompt bibliographies.

ACKNOWLEDGMENTS

We thank the AI Lab at Autodesk Research and OpenAI.

REFERENCES

[1] Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul Bennett. 2019. Guidelines for Human-AI Interaction. 1–13. https://doi.org/10.1145/3290605.3300233

[2] Autodesk. 2022. Autodesk Screencast. https://knowledge.autodesk.com/community/screencast Retrieved September 15, 2022.

[3] Marcelo Bernal, John R. Haymaker, and Charles Eastman. 2015. On the role of computational support for designers in action. *Design Studies* 41 (2015), 163–182. https://doi.org/10.1016/j.destud.2015.08.001

[4] Gwern Branwen. 2020. Gpt-3 creative fiction. https://www.gwern.net/GPT-3

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. https://doi.org/10.48550/ARXIV.2005.14165

[6] Miles Brundage, Shahar Avin, Jasmine Wang, Haydn Belfield, Gretchen Krueger, Gillian Hadfield, Heidy Khlaaf, Jingying Yang, Helen Toner, Ruth Fong, Tegan Maharaj, Pang Wei Koh, Sara Hooker, Jade Leung, Andrew Trask, Emma Bluemke, Jonathan Lebensold, Cullen O'Keefe, Mark Koren, Théo Ryffel, JB Rubinovitz, Tamay Besiroglu, Federica Carugati, Jack Clark, Peter Eckersley, Sarah de Haas, Maritza Johnson, Ben Laurie, Alex Ingerman, Igor Krawczuk, Amanda Askell, Rosario Cammarota, Andrew Lohn, David Krueger, Charlotte Stix, Peter Henderson, Logan Graham, Carina Prunkl, Bianca Martin, Elizabeth Seger, Noa Zilberman, Seán Ó hÉigeartaigh, Frens Kroeger, Girish Sastry, Rebecca Kagan, Adrian Weller, Brian Tse, Elizabeth Barnes, Allan Dafoe, Paul Scharre, Ariel Herbert-Voss, Martijn Rasser, Shagun Sodhani, Carrick Flynn, Thomas Krendl Gilbert, Lisa Dyer, Saif Khan, Yoshua Bengio, and Markus Anderljung. 2020. Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims. https://doi.org/10.48550/ARXIV.2004.07213

[7] Angel X. Chang, Mihail Eric, Manolis Savva, and Christopher D. Manning. 2017. SceneSeer: 3D Scene Design with Natural Language. https://doi.org/10.48550/ARXIV.1703.00050

[8] Minsuk Chang, Ben Lafreniere, Juho Kim, George Fitzmaurice, and Tovi Grossman. 2020. Workflow Graphs: A Computational Model of Collective Task Strategies for 3D Design Software. In *Graphics Interface 2020*. https://openreview.net/forum?id=qXEzq5agzIN

[9] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. 2013. AttribIt: Content Creation with Semantic Attributes. *ACM Symposium on User Interface Software and Technology (UIST)* (Oct. 2013).

[10] Siddhartha Chaudhuri, Evangelos Kalogerakis, Stephen Giguere, and Thomas Funkhouser. 2013. AttribIt: Content Creation with Semantic Attributes. In *Proc. UIST*. ACM.

[11] Erin Cherry and Celine Latulipe. 2009. The creativity support index. 4009–4014. https://doi.org/10.1145/1520340.1520609

[12] Jaemin Cho, Abhay Zala, and Mohit Bansal. 2022. DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generative Transformers. https://doi.org/10.48550/ARXIV.2202.04053

[13] Bob Coyne and Richard Sproat. 2022. WordsEye: an automatic text-to-scene conversion system. https://doi.org/10.1145/383259.383316

[14] Katherine Crowson. 2021. afiaka87/clip-guided-diffusion: A CLI tool/python module for generating images from text using guided diffusion and CLIP from OpenAI. https://github.com/afiaka87/clip-guided-diffusion

[15] Katherine Crowson. 2021. Rivers Have Wings. https://twitter.com/RiversHaveWings

[16] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato, and Edward Raff. 2022. VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. *arXiv preprint arXiv:2204.08583* (2022).

[17] Boris Dayma, Suraj Patil, Pedro Cuenca, Khalid Saifullah, Tanishq Abraham, Phúc Le Khac, Luke Melas, and Ritobrata Ghosh. 2021. DALLE Mini. https://doi.org/10.5281/zenodo.1234

[18] Manoj Deshpande, Eric Sauda, and Mary Lou Maher. 2020. Towards Co-Build: An Architecture Machine for Co-Creative Form-Making.

[19] Daniel Deutch, Tanu Malik, and Adriane Chapman. 2022. Theory and Practice of Provenance. In *Proceedings of the 2022 International Conference on Management of Data* (Philadelphia, PA, USA) *(SIGMOD '22)*. Association for Computing Machinery, New York, NY, USA, 2544–2545. https://doi.org/10.1145/3514221.3524073

[20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

arXiv:1810.04805 [cs.CL]
[21] Alaa El-Nouby, Shikhar Sharma, Hannes Schulz, R Devon Hjelm, Layla El Asri, Samira Ebrahimi Kahou, Y. Bengio, and Graham Taylor. 2018. Keep Drawing It: Iterative language-based image generation and editing.
[22] Patrick Esser, Robin Rombach, and Björn Ommer. 2020. Taming Transformers for High-Resolution Image Synthesis. https://doi.org/10.48550/ARXIV.2012.09841
[23] Thomas Funkhouser, Michael Kazhdan, Philip Shilane, Patrick Min, William Kiefer, Ayellet Tal, Szymon Rusinkiewicz, and David Dobkin. 2004. Modeling by Example. *ACM Trans. Graph.* 23, 3 (aug 2004), 652–663. https://doi.org/10.1145/1015706.1015775
[24] Oran Gafni, Adam Polyak, Oron Ashual, Shelly Sheynin, Devi Parikh, and Yaniv Taigman. 2022. Greater Creative Control for AI image generation. https://ai.facebook.com/blog/greater-creative-control-for-ai-image-generation/
[25] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. 2022. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. https://doi.org/10.48550/ARXIV.2208.01618
[26] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 489–500. https://doi.org/10.1145/2807442.2807478
[27] Songwei Ge and Devi Parikh. 2021. Visual Conceptual Blending with Large-scale Language and Vision Models. arXiv:2106.14127 [cs.CL]
[28] John S. Gero and Mary Lou Maher. 1993. Modeling Creativity and Knowledge-Based Creative Design.
[29] Katy Ilonka Gero, Zahra Ashktorab, Casey Dugan, Qian Pan, James Johnson, Werner Geyer, Maria Ruiz, Sarah Miller, David R. Millen, Murray Campbell, Sadhana Kumaravel, and Wei Zhang. 2020. *Mental Models of AI Agents in a Cooperative Game Setting*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376316
[30] Katy Ilonka Gero, Vivian Liu, and Lydia B. Chilton. 2021. Sparks: Inspiration for Science Writing using Language Models. https://doi.org/10.48550/ARXIV.2110.07640
[31] Arnab Ghosh, Richard Zhang, Puneet K. Dokania, Oliver Wang, Alexei A. Efros, Philip H. S. Torr, and Eli Shechtman. 2019. Interactive Sketch and Fill: Multiclass Sketch-to-Image Translation. arXiv:1909.11081 [cs.CV]
[32] Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2010. Chronicle: Capture, Exploration, and Playback of Document Workflow Histories. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 143–152. https://doi.org/10.1145/1866029.1866054
[33] S. G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology* 52 (1988), 139–183.
[34] Nathan W. Hartman. 2004. Defining Expertise in the Use of Constraint-based CAD Tools by Examining Practicing Professionals. *Engineering Design Graphics Journal* 69 (2004).
[35] Fariz Muharram Hasby and Dradjad Irianto. 2022. Conceptual Design Assessment Method for Collaborative CAD System. In *4th Asia Pacific Conference on Research in Industrial and Systems Engineering 2021* (Depok, Indonesia) *(APCORISE 2021)*. Association for Computing Machinery, New York, NY, USA, 254–261. https://doi.org/10.1145/3468013.3468340
[36] Autodesk Inc. 2022. Autodesk Fusion 360 faster performance and quality of life updates. https://www.autodesk.com/products/fusion-360/blog/usability-and-performance-improvements-fusion-360/
[37] Autodesk Inc. 2022. Fusion 360. https://help.autodesk.com/view/fusion360/ENU/?guid=GUID-7B5A90C8-E94C-48DA-B16B-430729B734DC
[38] Autodesk Inc. 2022. Generative design for manufacturing with Fusion 360. https://www.autodesk.com/solutions/generative-design/manufacturing
[39] For Inspiration, Recognition of Science, and Technology (FIRST). 2022. FIRST Robotics Competition. https://www.firstinspires.org/robotics/frc
[40] Ajay Jain, Ben Mildenhall, Jonathan T. Barron, Pieter Abbeel, and Ben Poole. 2022. Zero-Shot Text-Guided Object Generation with Dream Fields. (2022).
[41] Youngseung Jeon, Seungwan Jin, Patrick C. Shih, and Kyungsik Han. 2021. *FashionQ: An AI-Driven Creativity Support Tool for Facilitating Ideation in Fashion Design*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445093
[42] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. PromptMaker: Prompt-based Prototyping with Large Language Models. https://doi.org/10.1145/3491101.3503564
[43] Tero Karras, Samuli Laine, and Timo Aila. 2018. A Style-Based Generator Architecture for Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1812.04948
[44] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. arXiv:1912.04958 [cs.CV]
[45] Rubaiat Habib Kazi, Tovi Grossman, Hyunmin Cheong, Ali Hashemi, and George W. Fitzmaurice. 2017. DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (2017).
[46] Sumbul Khan and Bige Tunçer. 2019. Gesture and speech elicitation for 3D CAD modeling in conceptual design. *Automation in Construction* (2019).
[47] Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. 2018. ABC: A Big CAD Model Dataset For Geometric Deep Learning. https://doi.org/10.48550/ARXIV.1812.06216
[48] Hitoshi Komoto and Tetsuo Tomiyama. 2012. A Framework for Computer-Aided Conceptual Design and Its Application to System Architecting of Mechatronics Products. *Comput. Aided Des.* 44, 10 (oct 2012), 931–946. https://doi.org/10.1016/j.cad.2012.02.004
[49] Carmen Krahe, Maksym Kalaidov, Markus Doellken, Thomas Gwosch, Andreas Kuhnle, Gisela Lanza, and Sven Matthiesen. 2021. AI-Based knowledge extraction for automatic design proposals using design-related patterns. *Procedia CIRP* 100 (2021), 397–402. https://doi.org/10.1016/j.procir.2021.05.093 31st CIRP Design Conference 2021 (CIRP Design 2021).
[50] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3491102.3502030
[51] Jing Liao, Preben Hansen, and Chunlei Chai. 2020. A framework of artificial intelligence augmented design support. *Human–Computer Interaction* 35, 5-6 (2020), 511–544. https://doi.org/10.1080/07370024.2020.1733576 arXiv:https://doi.org/10.1080/07370024.2020.1733576
[52] Vivian Liu and Lydia B. Chilton. 2021. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. arXiv:2109.06977 [cs.HC]
[53] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. https://doi.org/10.48550/ARXIV.2204.09007
[54] Maria Teresa Llano, Mark d'Inverno, Matthew Yee-King, Jon McCormack, Alon Ilsar, Alison Pease, and Simon Colton. 2022. Explainable Computational Creativity. (2022). https://doi.org/10.48550/ARXIV.2205.05682
[55] Ryan Louie, Any Cohen, Cheng-Zhi Anna Huang, Michael Terry, and Carrie J. Cai. 2020. Cococo: AI-Steering Tools for Music Novices Co-Creating with Generative Models. In *HAI-GEN+user2agent@IUI*.
[56] J. Marks, B. Andalman, P. A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber. 1997. Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., USA, 389–400. https://doi.org/10.1145/258734.258887
[57] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. 2018. Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173943
[58] Gabriele Mirra and Alberto Pugnale. 2022. Expertise, playfulness and analogical reasoning: three strategies to train Artificial Intelligence for design applications. *Architecture, Structures and Construction* 2 (03 2022). https://doi.org/10.1007/s44150-022-00035-y
[59] Kaichun Mo, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. 2019. PartNet: A Large-Scale Benchmark for Fine-Grained and Hierarchical Part-Level 3D Object Understanding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society. https://doi.org/10.1109/CVPR.2019.00100
[60] Ron Mokady, Amir Hertz, and Amit H. Bermano. 2021. ClipCap: CLIP Prefix for Image Captioning. https://doi.org/10.48550/ARXIV.2111.09734
[61] Ryan Murdock. 2022. lucidrains/big-sleep: A simple command line tool for text to image generation, using OpenAI's CLIP and a BigGAN. Technique was originally created by https://twitter.com/advadnoun. https://github.com/lucidrains/big-sleep
[62] Ryan Murdock and Phil Wang. 2021. Big Sleep.
[63] Brad Myers, Scott E. Hudson, and Randy Pausch. 2000. Past, Present, and Future of User Interface Software Tools. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (mar 2000), 3–28. https://doi.org/10.1145/344949.344959
[64] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. 2022. Point-E: A System for Generating 3D Point Clouds from Complex Prompts. https://doi.org/10.48550/ARXIV.2212.08751
[65] OpenAI. 2022. Dall·E: Introducing outpainting. https://openai.com/blog/dall-e-introducing-outpainting/
[66] Guy Parsons. 2022. The DALL-E 2 prompt book. https://dallery.gallery/the-dalle-2-prompt-book/
[67] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. *arXiv* (2022).

[68] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art. In *Creativity and Cognition* (Venice, Italy) *(C&C '22)*. Association for Computing Machinery, New York, NY, USA, 15–28. https://doi.org/10.1145/3527927.3532792

[69] Tingting Qiao, Jing Zhang, Duanqing Xu, and Dacheng Tao. 2019. Learn, Imagine and Create: Text-to-Image Generation from Prior Knowledge. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/d18f655c3fce66ca401d5f38b48c89af-Paper.pdf

[70] Tingting Qiao, Jing Zhang, Duanqing Xu, and Dacheng Tao. 2019. MirrorGAN: Learning Text-to-image Generation by Redescription. https://doi.org/10.48550/ARXIV.1903.05854

[71] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]

[72] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. https://doi.org/10.48550/ARXIV.2204.06125

[73] Laria Reynolds and Kyle McDonell. 2021. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. arXiv:2102.07350 [cs.CL]

[74] B. F. Robertson and D. F. Radcliffe. 2009. Impact of CAD Tools on Creative Problem Solving in Engineering Design. *Comput. Aided Des.* 41, 3 (mar 2009), 136–146. https://doi.org/10.1016/j.cad.2008.06.007

[75] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 10684–10695.

[76] Kevin Roose. 2022. An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

[77] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. https://doi.org/10.48550/ARXIV.2205.11487

[78] Aditya Sanghi, Hang Chu, Joseph G. Lambourne, Ye Wang, Chin-Yi Cheng, Marco Fumero, and Kamal Rahimi Malekshan. 2021. CLIP-Forge: Towards Zero-Shot Text-to-Shape Generation. https://doi.org/10.48550/ARXIV.2110.02624

[79] Aditya Sanghi, Rao Fu, Vivian Liu, Karl Willis, Hooman Shayani, Amir Hosein Khasahmadi, Srinath Sridhar, and Daniel Ritchie. 2022. TextCraft: Zero-Shot Generation of High-Fidelity and Diverse Shapes from Text. https://doi.org/10.48550/ARXIV.2211.01427

[80] Shikhar Sharma, Dendi Suhubdy, Vincent Michalski, Samira Ebrahimi Kahou, and Yoshua Bengio. 2018. ChatPainter: Improving Text to Image Generation using Dialogue. https://doi.org/10.48550/ARXIV.1802.08216

[81] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* (jan 2022). https://doi.org/10.1145/3511599 Just Accepted.

[82] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 582, 11 pages. https://doi.org/10.1145/3411764.3445219

[83] Karl D. D. Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G. Lambourne, Armando Solar-Lezama, and Wojciech Matusik. 2021. Fusion 360 Gallery: A Dataset and Environment for Programmatic CAD Construction from Human Design Sequences. *ACM Transactions on Graphics (TOG)* 40, 4 (2021).

[84] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. 2022. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. https://doi.org/10.48550/ARXIV.2203.06566

[85] Tongshuang Wu, Michael Terry, and Carrie J Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. https://doi.org/10.1145/3491102.3517582

[86] Weihao Xia, Yujiu Yang, Jing-Hao Xue, and Baoyuan Wu. 2020. TediGAN: Text-Guided Diverse Face Image Generation and Manipulation. https://doi.org/10.48550/ARXIV.2012.03308

[87] Kai Xu, Hanlin Zheng, Hao Zhang, Daniel Cohen-Or, Ligang Liu, and Yueshan Xiong. 2011. Photo-inspired model-driven 3D object modeling. In *ACM SIGGRAPH 2011 papers on - SIGGRAPH '11*. ACM Press, Vancouver, British Columbia, Canada, 1. https://doi.org/10.1145/1964921.1964975

[88] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. 2017. AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. https://doi.org/10.48550/ARXIV.1711.10485

[89] Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, Ben Hutchinson, Wei Han, Zarana Parekh, Xin Li, Han Zhang, Jason Baldridge, and Yonghui Wu. 2022. Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. https://doi.org/10.48550/ARXIV.2206.10789