DeckFlow: Specification Decomposition on a Multimodal Generative Canvas

Gregory Croisdale*, Emily Huang*, John Joon Young Chung†,
Anhong Guo*, Xu Wang*, Austin Z. Henley‡, Cyrus Omar*
*University of Michigan, Ann Arbor, USA
{gregtc, emihuang, anhong, xwanghci, comar}@umich.edu
†Midjourney, San Francisco, USA
jchung@midjourney.com
‡Carnegie Mellon University, Pittsburgh, USA
azhenley@cmu.edu

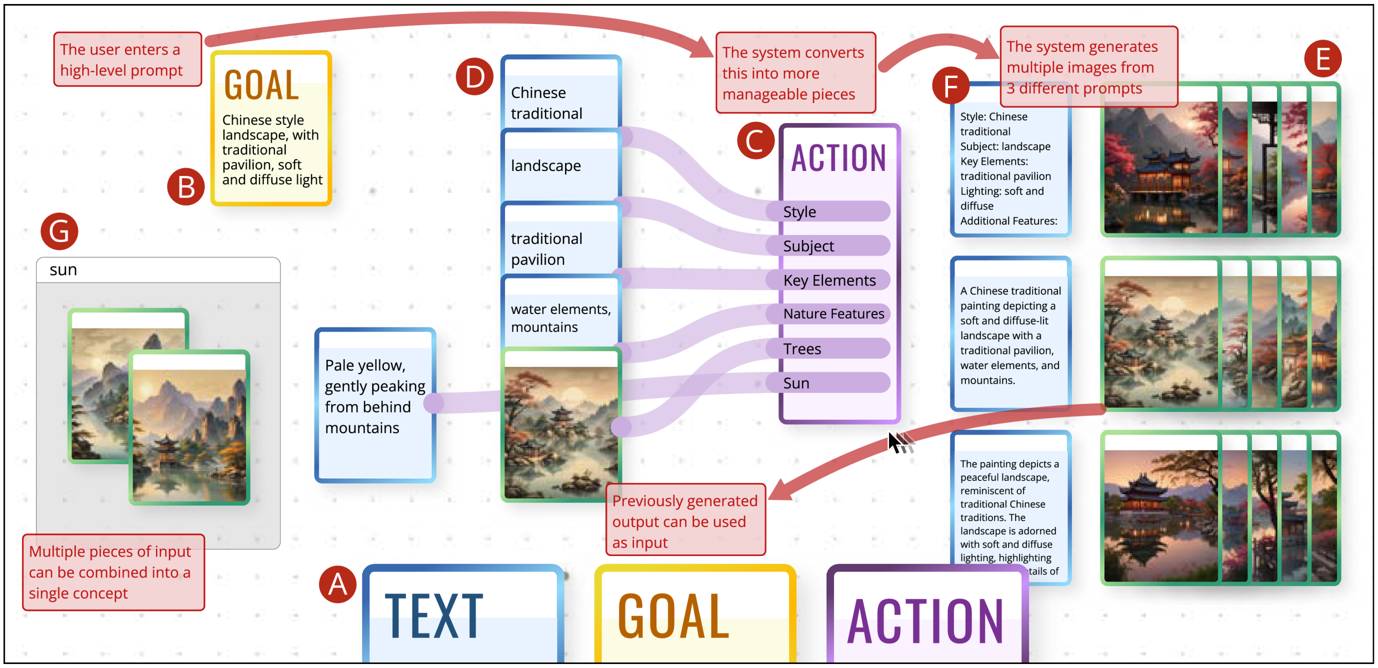

Fig. 1: DeckFlow is an infinite canvas for creating multimodal content. In this case, detailed in Section III, the user drags a Goal Card (b) from the Hand (a), which generates an Action Card (c) connected to several Text Cards (d) representing the decomposed specification. The Action Card spawns multiple Text Cards containing the constructed prompts (f), and images are generated using them (e). In a subsequent iteration of the task, the user moves some of them into a Cluster (g), and uses one as input to the Action Card.

***Abstract*—Generative AI promises to allow people to create high-quality personalized media. Although powerful, we identify two fundamental design problems with existing tooling through a literature review. We introduce a multimodal generative AI tool, DeckFlow, to address these problems. DeckFlow supports a specification decomposition workflow where an initial goal is iteratively decomposed into smaller parts and combined using feature labels and clusters. DeckFlow supports generative space exploration by generating multiple prompt and output variations, presented in a grid, that can feed back recursively into the next design iteration. We evaluate DeckFlow for text-to-image generation against a state-of-practice conversational AI baseline for image generation tasks. We then add audio generation and investigate user behaviors in a more open-ended creative setting with text, image, and audio outputs.**

***Index Terms*—generative AI, prompt engineering, text generation, image generation, audio generation, infinite canvas**

I. Introduction

With a short text prompt, someone with minimal experience can generate a clever Shakespearean poem about a jaunt on a sunny day in ChatGPT, an image of a beautiful oil painting of a flowing wheat field in the style of Van Gogh in Midjourney, or a catchy pop-punk song pontificating about global warming in Suno. How can simple prompts lead to such complex and compelling output? Generative AI models rely on statistical patterns derived from massive amounts of training data. The model makes assumptions about the user's intent based on the most commonly observed patterns in the training data. Training data exhibits the same biases as online content in general [2], however, so this sort of one-shot prompting is limited in its ability to creatively support diverse users and niche use cases.

A longer version of this paper including supplemental figures is on arXiv [1].

In situations where the model's output is unsatisfying, the user may want to iteratively refine certain aspects of their specification and try again. However, common generative AI tools make it difficult to exert fine-grained control over what is generated (the **specification decomposition problem**). Users often resort to tweaking the prompt and re-running the model until the generated output is acceptable.

After providing an initial specification of their intent, users "roll the dice" by letting the generative AI model generate output, often multiple times. Generative AI is stochastic in nature, so each output differs slightly, forming a space of possible outputs for a given specification. Users may be interested in exploring this space, but many existing tools only present one output, forcing the user to sequentially ask the model to generate multiple outputs (the **generative space exploration problem**).

To address these problems, this paper introduces DeckFlow. DeckFlow is a **multimodal generative AI tool** designed to support a variety of creative activities. Our focus in this paper was on working with text, images, and audio, both as input to and as output from the tool. The teaser image shows a simplified example of a user workflow.

To address the **specification decomposition problem**, DeckFlow supports a specification workflow where an initial Goal Card, typically consisting of a text prompt, is decomposed into an Action Card with several textual labels that serve as "ports" in the dataflow diagram. Text, image, and audio cards can be connected to these ports.

To address the **generative space exploration problem**, users can click a button on an Action Card to generate three groups of three outputs from the underlying generative AI model. Each output appears directly on the infinite canvas next to the corresponding Action Card upon request. The user can freely delete, group, or rearrange these outputs to explore the design space. The user can also freely repurpose the output from one iteration of content generation as input to one or more other tasks, including future iterations of the same task.

In addition to our validation of DeckFlow as a whole, this paper contributes generalizable knowledge in the form of (1) our decomposition of the design space of generative AI tools around the two central problems that organize every section of this paper, (2) a set individual affordances in DeckFlow, validated by our study, that could be implemented in other visual generative AI tools, e.g. action cards and our lightweight generative space exploration affordances, and (3) insights about how humans engage in task and specification decomposition and generative space exploration in multimodal content generation tasks.

II. Background

A. Specification Decomposition

Users often need to specify several distinct aspects of a creative artifact, like its style, palette, tone, or rhythm. Conventional creative tools provide a variety of affordances specialized to each of these. However, many contemporary generative AI tools require expressing every aspect of the artifact using a natural language prompt. Practitioners therefore improvise, e.g. by including bullet-point lists or pasted reference images, but in some domains, this can limit their ability to specify their intent precisely.

ChainForge can construct a prompt from a template string. Templated fields can be independently swept or frozen, enabling controlled A/B testing across a single dimension [3]. *CreativeConnect* lets users specify discrete keywords connected to specified regions of a sketch [4]. *CueFlik* frames specification decomposition as interactive concept learning: users label positive and negative image examples, and the system learns a weighted combination of visual features that can be re-applied across queries [5]. *PromptPaint* interpolates continuously between multiple prompts during the diffusion process, exposing a weighted blend rather than a concatenated string [6]. *PromptCharm* provides a mixed-initiative loop: an RL-based agent suggests refined prompts, while users can tweak token-level attention or in-paint masked regions, exposing prompt, attention, and pixel masks [7].

In the domain of strictly textual tools, *Sensecape* allows users to decompose prompts into individual parts, arranged spatially, and compare variants in parallel, merging the parts they like [8]. *Luminate* asks writers to tag sentences along qualitative dimensions (e.g., formality, concreteness) and then recombines those dimensions to generate tailored drafts [9].

B. Generative Space Exploration

Generative AI models are stochastic and can generate a wide variety of outputs for a given task specification. Long before modern generative AI models, visualization researchers argued that creative work benefits from design galleries: curated arrays of parameter variations that reveal structure in high-dimensional spaces [10]. Graphic-layout tools such as DesignScape revived the idea for automatic poster composition, presenting users with multiple exemplar layouts and allowing them to steer by favoring particular variants [11]. These gallery-based approaches exemplify a shift from producing a single "best" artifact to navigating a solution space. Modern generative models make that space substantially larger.

Rather than picking just one output from this space, many generative AI tools provide affordances for *generative space exploration*. *Sensecape* treats every generated response from part of a decomposed prompt as a movable card; users can request additional responses, duplicate or branch cards, build hierarchical concept maps, and thus form a multilevel mental model of the generative space [8]. *Promptify* offers a lighter abstraction: each successive revision of a single prompt appears on a zoomable canvas, preserving visual history and encouraging lateral comparison, though genuine branching still requires manually copying the prompt [12]. *ComfyUI*, aimed at experts, provides a direct manipulation approach which exposes seeds, schedulers, and CFG scales as node parameters so designers can sweep numeric ranges and cache intermediate latents, effectively turning low-level controls into gallery axes [13]. *Dreamsheets* gives a similar, but more accessible and manipulable interface, by image generation directly into a spreadsheet application [14].

III. DeckFlow

To address the design problems, we designed DeckFlow to support specification decomposition and exploration. To describe Deckflow, we will walk through an example DeckFlow usage scenario shown in the teaser image.

Chenxi has just moved into her new apartment and wants to decorate the dining room. Chenxi starts by dragging from the Hand (a), creating a Goal Card (b) in which she writes "Chinese style landscape, with traditional pavilion, soft and diffuse light." The Goal Card then creates an Action Card (c) with labels connected to discrete Text Cards (d), extracted from her high-level prompt: `Style`: "Chinese traditional", `Subject`: "landscape", `Key Elements`: "traditional pavilion", `Lighting`: "soft and diffuse", and `Natural Features`, but because these features weren't specified in the original prompt, this connection is empty. As a result, Chenxi thinks of natural features she is particularly interested in: "water elements, mountains." Satisfied with these settings, Chenxi asks the Action Card to generate some images from her specifications.

The Action Card begins by creating three prompts, using the labeled inputs as guidance. Chenxi is now presented with three rows of Image Cards (e), prompted using different Text Cards (f). The first row's Text Card is created by concatenating the inputs together: "Style: Chinese traditional, Subject: landscape...", resulting in images which resemble her input. The second row's Text Card is created by calling an LLM with those labeled inputs, creating a more coherent version of the prompt: "A Chinese traditional painting depicting a serene landscape..." The third row uses these inputs in a small local LLM which has been optimized to generate creative and appealing image prompts, yielding a less-precise, but more interesting row of Image Cards: "In the heart of a serene Chinese courtyard, a traditional Chinese painting unfolds..."

In a couple of the Image Cards from the second row, Chenxi likes the way the sun is peaking through the mountains in the background, but isn't quite sure how to describe it. She moves these Image Cards to a different region and forms a Cluster (g) around them, indicating to it that she wants to understand how to describe the phrase 'sun' from these Image Cards. After clicking the 'interpret' button, the Cluster generates a Text Card illustrating how to create a prompt which captures this essence: "Pale yellow, gently peaking from behind mountains."

Satisfied with this description, Chenxi decides to modify the existing Action Card to include this information. She begins by moving the old Image Cards to a region above the Action Card so that she can refer to them later. Chenxi adds the input 'sun' to the Action Card, and connects the Text Card from the Cluster to it. After regenerating, she realizes that the images are missing the pink cherry trees she liked from her previous set of images. Rather than using the Cluster this time, she decides to add a new input, 'trees', and connects an Image Card to it.

With these changes, Chenxi decides to re-generate a new batch using the Action Card. These next three rows of images provide different interpretations of her input. Chenxi finds one on the second row that she prefers.

IV. Evaluation

To evaluate the effectiveness of DeckFlow, understand patterns of user behavior, and extract generalizable insights relevant to the designers of other generative AI tools targeting end-users, we conducted two within-subjects studies recruiting from a Computer Science and Engineering email list. The first study compares DeckFlow with ChatFlow, a ChatGPT-like interface for text and image generation that we designed to use the same backend generative AI models (the **comparative study**), with 8 male, 7 female, and 1 non-binary participants. This is followed by a more in-depth study of user behavior using a version of DeckFlow modified to also support audio (the **multimodal behavioral study**), with 3 male and 4 female participants. Each session was scheduled to be 2 hours long, and participants were compensated with a $30 USD gift card.

1) Research Questions: These studies sought to understand how the novel interface, DeckFlow, impacted generative tasks: **RQ1:** How do users approach the Specification Decomposition problem? **RQ2:** How do users approach the Generative Space Exploration problem? **RQ3:** How are different modalities treated as input and output?

2) Procedures: In the comparative study, users were given two think-aloud tasks, shown in Table SI (in the supplementary material), for each tool with a flexible time limit of 10 minutes each. The tasks and the order of tools used were counterbalanced to avoid ordering effects. These tasks were a closed-ended task, where participants were asked to recreate a given image as closely as possible; these are designed to study situations in which a user's design space is narrowly defined; and an open-ended task, requiring participants to create an image that best satisfies a given text prompt; these are designed to study more exploratory, divergent design processes.

In the second, multimodal study, users were given two think-aloud tasks, shown in Table SII, counterbalanced to include 2 of the tasks from a pool of 3. These tasks were administered with a flexible time limit of 15 minutes each. After each task, state was reverted to the time in which a user used each of the primary input mechanisms: Text, Image, Audio, Cluster, as well as Action Card generation. If a user had not used one of the input mechanisms naturally, they were asked to perform another generation. During this retrospective think-aloud, users were asked questions relating to their expectations and their perception of the output.

We collected data through various means, including basic usage metrics (e.g., number of images and text inputs used), user ratings of different features, self-reported success in ChatFlow vs. DeckFlow, screen and voice recordings, and interviews at the beginning, after each task, and at the conclusion of each study.

V. Results

We detail the results of both user studies, comparing DeckFlow and the baseline ChatFlow using system logs, interview results, recordings of participant use, and survey results, in an effort to answer our research questions. To refer to participants, we use the format $P_A 2$, where the subscript indicates the study (A for Comparative Study, B for Multimodal Study), and the number is a unique identifier for that participant. To refer to participant counts, we use the format $n_B = 3$.

A. RQ1: How do users approach the Specification Decomposition problem in the interfaces?

1) Specification in ChatFlow: In the ChatFlow interface, all users but $P_A 2$ specified their requirements using a linear conversation. These messages were generally anthropomorphic, especially as users got more frustrated with ChatFlow's ability to follow instructions, such as these from $P_A 9$ in Close-Ended task A: "more closer and one branch!!!!", "I only see four birds?", and "I want one branch, why you give two branches again? also the birds are white-belle not yellow and red".

Some users ($n_A = 8$) began a task with a high-detail prompt, seeking to immediately create a potential final output. Some ($n_A = 10$), like $P_A 6$, began with a basic image, adding required details one after another. Some users ($n_A = 3$) downloaded some collection of favorable images, reloaded the page to reset the conversation, and uploaded the images to apply their context from a new perspective.

$P_A 2$ used a unique technique in which they spent over 5 minutes crafting a detailed prompt in ChatFlow each time before generating their first output. After they received the output, they edited their original prompt.

2) Action Card Interactions: We observed three different types of labels for input in the Action Card used in each study:

1) Constraint: A specific, non-interpretive condition that the output should fulfill, such as $P_A 8$'s 'number of birds': '6'
2) Annotation: A label used to interpret or specify some input, such as $P_A 11$'s 'bird species': (image of a bird)
3) Instruction: A natural language instruction, such as $P_A 6$'s "More images like this" seen in Figure S4
4) Empty: An empty label

The 'instruction' label type was not anticipated in system design, but was utilized by a few different users ($n_A = 3, n_B = 2$). Counts of these label types used in the Multimodality study can seen in Table SIII. Some users ($n_A = 5, n_B = 3$) expressed difficulty in writing labels, as verbalized by $P_A 11$: "I had trouble coming up with my own annotations for the action cards... It was easier to use the goal card in the beginning when I didn't quite know what I wanted."

3) Clusters: Use of the Cluster led to discovery of concepts previously unknown by users in Close-Ended Tasks ($n_A = 4$). In $P_A 8$'s case, the Cluster even correctly identified the correct bird breed, the Zebra Finch, from Close-Ended Task A from only images previously generated in DeckFlow, without being prompted by the user. $P_A 11$ verbalized "The Cluster helped it feel less overwhelming, more like I was like getting to a point rather than diverging away from it."

Some users expressed that they did not remember to use a cluster ($n_A = 2, n_B = 3$), or avoided its use because they did not understand it ($n_A = 2$). Other users, however, did not find the cluster to be as useful as input; $P_B 4$ verbalized "I don't really understand how the clustering feature acts differently than if I were to use separate text prompts." $P_B 4$ later stated "Clustering works well to visually group the elements."

4) Goal Card Interactions: Goal Cards was popular among participants, primarily utilized to break down high-level prompts into more manageable components. $P_A 11$ articulated this benefit, stating, "The goal card helped a lot with breaking down large prompts into specific parameters, and I have the opportunity to choose which ones I want to implement."

During the comparative study, many users ($n_A = 4$) wanted multimodal input to the Goal Card. However, after this feature was implemented in the subsequent study, participants consistently used only text as input when creating an Action Card.

B. RQ2: How do users approach the Generative Space Exploration problem in the interfaces?

1) Similar performance in closed-ended tasks: In the rating scale results from the Comparative Study (Figure S5), there was no clear favorite for close-ended tasks; $P_A 11$ stated that "I prefer ChatFlow for [closed-ended tasks] because it's more suited for sentence prompts", but $P_A 9$ disagreed, saying "In the hard task, I definitely prefer DeckFlow because it gives me several options, and gives me an outline of what GPT expects".

2) DeckFlow was universally preferred in open-ended tasks: As seen in Figure S5, participants found DeckFlow preferable in usability in open-ended tasks, but similar in outcome. Notably, $P_A 6$ gave ChatFlow a 1 out of 7 when asked about ease of use for their open-ended task, stating "[ChatFlow] made more sense if I wanted to do one specific change to an image; something creative would be way easier in DeckFlow—you had such a wide array of images you got back so fast, and the prompt generation was really phenomenal.".

3) The role of divergence: The ability of an interface to diverge from past output was noted by many users ($n_A = 8, n_B = 4$). $P_B 7$ indicated that DeckFlow's design supports this: "I think at least the prompt part—generating image or audio... I don't need to write them on my own. The creation process is a very slow process if I do it by myself. For DeckFlow, I see more possibilities more quickly. *Although the image doesn't live up to my expectation, it helps me know what I want, because I know what I don't want*." Because each of the generated rows use different prompting methods, some users ($n_A = 5$) picked up on these differences, especially the less adherent but more creative third row: "For the third row—it's a higher quality picture, but it doesn't really reflect the prompt".

C. RQ3: How are different modalities treated as input and output?

1) Text dominates as specification input: Users preferred using text to specify their intent, using it for 89.5% of their Action Card inputs, and rating text much higher than images or audio for each of the modified Creativity Support Index questions [15]: "Sometimes, I'm not sure to what extent do the audio or images affect the output, but I can always resort to text as the input, and it's very clear to me" ($P_B 3$).

Some users ($n_B = 3$) mentioned that it felt like the game "telephone" due to context being revealed through text and the model's gaps in modal understanding: "I think they all go through text; it reminds me of the old Google Translate, how if you translate to different languages, especially obscure ones... it would go through English, and you would see the English words that they couldn't translate. *The image and audio cannot directly translate to each other, but they get described by text first, which is more efficient since you don't need a separate model to communicate between that audio and the image*. At least 'telephone game' results were all pretty. It was positive for entertainment since you get slightly different things, but negative for trying to make something precise" (P_B 1).

One notable exception is P_B 6, who gave Audio as Input a 7 out of 7 when asked about its Precision, stating "It's always a problem to get everything in my mind to the model. No matter with text, image, or audio... being able to express myself in different ways helped me. Like to say, if someone doesn't know Picasso, I can at least show them an image."

2) Text was the least interrogated output: Despite being the dominant choice for input, text only accounted for 16.8% of the output generated. Additionally, of the *final text outputs* in the Multimodality Study, 6/12 ended in incomplete sentences due to the model reaching its token limit, such as P_B 5's chosen text output for the Creature task: "..., and small invertebrates. The creature often [sic]".

3) Strong reactions to Audio: Despite only averaging 2 bulk generations per task, some participants (n_B=3) had strong reactions to generated audio. For example, as seen in Figure S9, P_B 5 was unhappy with their generated audio content which sounded vaguely like giggling and crackles: "That was nightmarish... I'm not gonna hold it against the model. I think if it were a better model, this would be closer to what I want." P_B 7, when generating audio for the Creature task, indicated they were somewhat fearful of listening to the audio, worried that it would be "creepy" due to the image used as input. However, after one generation, they stated "*I think the sound is what makes a creature imaginative most.* The visual part, although it may look like a different creature than what exists on Earth, maybe I can imagine it from the descriptions, but the audio, I totally cannot expect... I think audio is the most important part to make this creature like an alien. The audio makes me feel like the creature exists."

VI. Discussion

A. Conversational Expectations

Though DeckFlow addressed specification decomposition through structured components, users frequently approached it conversationally, potentially revealing ingrained mental models from interfaces like ChatGPT. Users created instruction-style labels in Action Cards (V-A2) rather than using the intended annotation types, and used Goal Cards as if they were beginning a conversation (V-A4). Combined with the frustration of DeckFlow lacking a centralized memory model, it seems that user expectations of natural language have begun to include features found in Chat interfaces. Future interface designers must either accommodate these conversational patterns, or provide clearer scaffolding for alternative interaction models.

B. Multimodal Inputs and Generative Space Exploration

DeckFlow's approach to generative space exploration through multimodal inputs showed mixed results. While supporting iterative refinement (V-A), it sometimes failed to maintain consistency in direct iteration (V-B1), and created disconnection between modalities (V-C1). Users strongly preferred text for specification (89.5% of inputs) but spent most time with images (61% of generations), and showed strongest emotional responses to audio (V-C3).

These findings suggests that modalities can serve different roles in generation: text for precise specification, images for quickly understood output, and audio for emotional engagement. Future work could enhance exploration by giving direct control, such as attention masking [7] or model adaptation [16], [17].

C. Threats to Validity

Our lab studies had several limitations: short tasks (10–15 min), a homogeneous participant pool (18–25 year-old CS/EE students), unfamiliar equipment, awareness of the researchers' roles, and asymmetric editing capabilities across modalities. The ChatGPT-like baseline represents just one possible comparison point in a rapidly evolving landscape. These factors limit generalizability to other tools, populations, and real-world creative workflows.

VII. Conclusion

We have presented DeckFlow, a novel interface for iterative human-AI co-design in generative content creation. DeckFlow addresses the **specification decomposition problem** via Goal Cards that break into labeled Action Cards supporting multimodal inputs, and the **generative space exploration problem** by presenting structured output groups that represent different creative directions directly on the canvas. Our evaluations demonstrated DeckFlow supports diverse workflows and improves outcomes for open-ended creative tasks. As generative AI evolves, interfaces like DeckFlow will be crucial in empowering users while maintaining creative control. Future work could extend the specification design space, investigate long-term impacts on creative processes, and develop tools supporting diverse needs for content creation.

VIII. Acknowledgements

We would like to thank those who contributed to the project by providing valuable feedback, including but not limited to the anonymous reviewers of this and previous submissions, our anonymous study participants, Dhruv Jain, Alanson Sample, Farnaz Jahanbakhsh, Toby Li, Cade Brown, and the 2024 VL/HCC Graduate Consortium.

Additionally, we would like to thank the facilities and administrative staff at the University of Michigan Department of Computer Science and Engineering, particularly Stephen Reger and Jasmin Stubblefield.

References

[1] G. Croisdale, E. Huang, J. J. Y. Chung, A. Guo, X. Wang, A. Z. Henley, and C. Omar, "Deckflow: Iterative specification on a multimodal generative canvas," 2025.

[2] M. Zhou, V. Abhishek, T. Derdenger, J. Kim, and K. Srinivasan, "Bias in generative ai," 2024.

[3] I. Arawjo, C. Swoopes, P. Vaithilingam, M. Wattenberg, and E. Glassman, "Chainforge: A visual toolkit for prompt engineering and llm hypothesis testing," 2023.

[4] D. Choi, S. Hong, J. Park, J. J. Y. Chung, and J. Kim, "Creativeconnect: Supporting reference recombination for graphic design ideation with generative ai," *arXiv preprint arXiv:2312.11949*, 2023.

[5] J. Fogarty, D. Tan, A. Kapoor, and S. Winder, "Cueflik: interactive concept learning in image search," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (New York, NY, USA), p. 29–38, Association for Computing Machinery, 2008.

[6] J. J. Y. Chung and E. Adar, "Promptpaint: Steering text-to-image generation through paint medium-like interactions," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, (New York, NY, USA), Association for Computing Machinery, 2023.

[7] Z. Wang, Y. Huang, D. Song, L. Ma, and T. Zhang, "Promptcharm: Text-to-image generation through multi-modal prompting and refinement," *arXiv preprint arXiv:2403.04014*, 2024.

[8] S. Suh, B. Min, S. Palani, and H. Xia, "Sensecape: Enabling multilevel exploration and sensemaking with large language models," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, p. 1–18, ACM, Oct. 2023.

[9] S. Suh, M. Chen, B. Min, T. J.-J. Li, and H. Xia, "Structured generation and exploration of design space with large language models for human-ai co-creation," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2024.

[10] J. Marks, B. Andalman, P. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, and W. Ruml, "Design galleries: A general approach to setting parameters for computer graphics and animation," in *SIGGRAPH*, pp. 389–400, 1997.

[11] P. O'Donovan, A. Agarwala, and A. Hertzmann, "Designscape: Design with interactive layout suggestions," in *CHI*, pp. 1221–1224, 2015.

[12] S. Brade, B. Wang, M. Sousa, S. Oore, and T. Grossman, "Promptify: Text-to-image generation through interactive prompt exploration with large language models," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, (New York, NY, USA), Association for Computing Machinery, 2023.

[13] Comfy Org, "Comfyui," 2024.

[14] S. G. Almeda, J. Zamfirescu-Pereira, K. W. Kim, P. Mani Rathnam, and B. Hartmann, "Prompting for discovery: Flexible sense-making for ai art-making with dreamsheets," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, (New York, NY, USA), Association for Computing Machinery, 2024.

[15] E. Cherry and C. Latulipe, "Quantifying the creativity support of digital tools through the creativity support index," *ACM Trans. Comput.-Hum. Interact.*, vol. 21, jun 2014.

[16] R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. H. Bermano, G. Chechik, and D. Cohen-Or, "An image is worth one word: Personalizing text-to-image generation using textual inversion," 2022.

[17] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," 2021.