Imagination Tool: Accessible AI Image Generation Software to Support Child Ideation and Creative Expression
Paper type: System Demonstration

Melis Meriç[1], Seamus White[2], Alba Suárez Zapico[2],
Zelda Yanovich[2], Bethany Koby-Hirschmann[2], Rebecca Fiebrink[1]
[1] Creative Computing Institute, University of the Arts London, London UK. E-mail: r.fiebrink@arts.ac.uk
[2] Fam Studio, London UK. E-mail: seamus@famstudio.co

Abstract

Despite the rising popularity in image AI generation systems, key open questions remain as to both (1) how to design such systems for use by children, and (2) what roles generative AI may play in contexts in which media creation is not the end goal, but is rather used to support other human endeavours such as communication and ideation. We describe a new AI image generation tool, designed to support children engaging in creative ideation, expression, and visually communicating their ideas—in particular, their visions for the future. We describe key aspects of the system's implementation, including ensuring the system produces child-safe and task-appropriate images, and mechanisms for supporting children in ideation. This work can thus inform researchers in computational creativity who aim to address safety and usability of systems for children.

Introduction

Generative image AI tools such as Midjourney, Stable Diffusion, and DALL·E 2 can now create visually impressive images from text prompts. In principle, such tools offer new opportunities for creative expression and visual communication, including by people who lack skill in conventional image-making methods. However, they present challenges in real-world use, including the need for "prompt engineering" to achieve the desired output style, content, and quality. Further, models can produce images that are objectionable or illegal. These challenges are magnified when considering the needs of child users; not only do children present distinct safeguarding and usability requirements; children may come to a tool with little preexisting knowledge and understanding of AI. Nevertheless, we believe making AI image generation usable by and safe for children is a worthy goal, with the potential to facilitate new forms of creative engagement and to support children in other activities involving ideation and visual communication.

This paper describes the implementation of a new generative AI tool, built on Stable Diffusion, designed to support children engaging in creative ideation, expression, and visually communicating their ideas. We describe key aspects of the system's implementation, including those to ensure the system produces child-safe and task-appropriate images, and mechanisms for ensuring usability and supporting children in ideation. Our experiences in 3 workshops with children suggest that the tool generally satisfies these aims.

Related Work

Usability and Safety of AI Image Generation

Contemporary generative AI systems come with usability and safety challenges. For instance, "prompt engineering" is often necessary to produce images with the desired style and quality, but this can be difficult for novice users. Research is beginning to systematically explore strategies for producing good prompts (Oppenlaender 2023; Liu and Chilton 2022), but to our knowledge no prior work has examined prompt engineering approaches for making images appropriate to children, or for assisting children in constructing good prompts.

Additionally, popular generative AI models produce content inappropriate for children. Recent work (Qu et al. 2023) showed a high prevalence of unsafe (i.e., sexually explicit, violent, disturbing) images produced by all four popular text-to-image models they tested, finding that unsafe images could be generated even from prompts not containing inappropriate language or reference to unsafe content.

Many commercial systems include provisions for restricting the content that may be generated (e.g., (Open AI 2024), (Stability.AI 2022)). Yet such approaches are not designed with the safety requirements of child users in mind: it is easy to produce content that adheres to companies' filters and guidelines yet is inappropriate for children (e.g., portrayal of nudity-free women's bodies that are nevertheless sexualised; scenes that look disturbing even without violence).

Image Creation and AI Tools for Children

Most popular AI tools do not target child users, and some have terms of service prohibiting use by children under a given age. Two exceptions are Amazon's Create with Alexa (Landau 2022), which allows children to generate stories with illustrations and sound effects, and Kidgeni (2023), which allows users to make images from text prompts or line drawings. Neither system is transparent about approaches to safety. Research is beginning to explore how generative AI systems may be made for children; for instance NaCanva (Yan et al. 2023) allows children to make multi-modal nature collages, and Sun et al. (2022) explore how AI can be

used to support children learning to paint.

Supporting Child Co-Design and Ideation

Substantial research examines methods for eliciting children's ideas and voices in the design of new policies, services, or technologies. Our work is partly motivated by the fact that approaches to engaging children in co-design and futures thinking often incorporate drawing or image-making (Walsh et al. 2013). These include techniques like Big Paper (Guha et al. 2004), storyboarding (Truong, Hayes, and Abowd 2006), layered elaboration (Walsh et al. 2010), and Kid Reporter (Bekker et al. 2003). In comicboarding (Moraveji et al. 2007), an artist renders children's ideas for them, "allowing the child to focus on ideas rather than on translating their thoughts into drawings;" in principle, generative AI could play a similar role. AI has previously been used in tools for the general public to envision alternate futures (Rafner et al. 2023).

Imagination Tool

Summary of Functionality

Our software, Imagination Tool, is a remotely-hosted Web app accessible via a browser. It supports children in generating images in their exploration of a design or ideation task. For instance, the version we will demo asks children to explore ideas around the question, "What will you design for London in 2050?"

A video walkthrough of the tool can be found at `https://www.youtube.com/watch?v=WT7yQIB89pk`. Upon launch, the app prompts the user to choose a design task (e.g., to design "a new way for people to get to school and around the city"). Next, the main screen shows a number of short questions accompanied by text boxes, asking questions such as "What is it?" and "What does it look like?" (Figure 1). After providing answers for at least one question, the user can click a button to "Make It!" Then (assuming no content moderation problems arise) the user is shown four images generated from their text. The user can select any of these to view a larger version.

The user can return to editing their text answers to modify the image. Or, if they are happy with the selected image, they can "submit" this image, with the option to add their image to a "gallery" showing images produced by all users. Clicking on an image in the gallery takes the user back to the main screen (Figure 1), showing the prompting questions and user answers used to make that image.

Ensuring Child-Appropriate Images

A high priority was to ensure the child safety of all generated images. Our tests revealed that even a "safe for work" version of Stable Diffusion (v1.6) still easily produced highly sexualised images of women for nonsexual prompts (e.g., "woman in swimsuit"), and often produced glitches in images that children could find disturbing (e.g., pictures of faces with one eyeball missing). Our approach to ensuring content safety can be summarised as follows:

Bad/Safe Words Lists: We began with a publicly available list of over 1300 "offensive/profane" English words list from von Ahn (n.d.). However, this list contained words that are not necessarily harmful (e.g., "fairy"), and words that would be unethical to prohibit (e.g., "black," "asian"). We therefore developed a custom "bad words" list by removing terms that we deemed non-harmful. We then augmented this to include any words *containing* words on the list (e.g., "murder" is on the list; we added "murders" and "murdered"). However, this proved too stringent, with words like "glass", "burgundy," and "hello" being prohibited (due to containing "ass," "gun," and "hell"). We therefore created a "safe words" list through (1) enumerating each word in the NLTK Python library's (Bird, Klein, and Loper 2009) English words dataset which included as a substring any word on the "bad words" list; (2) employing the alt-profanity-check library to check whether each enumerated word was harmful; and (3) saving non-harmful enumerated words to a "safe words" list. Our "bad words" and "safe words" lists are publicly available[1] for others to use and adapt.

Toxicity Detection: Filtering by words still does not preclude the possibility that text describes something inappropriate. (For instance, one can describe violent scenes using words like "blades," "cut," "ruin.") We therefore employed Perspective API (Jigsaw) to assess scores for "toxicity," "severe_toxicity," "identity_attack," "insult," "profanity," and "threat" for each user textbox.

Human Moderation: When any user text is flagged for having either (1) an input word that appears on our custom bad words list, or which contains a substring appearing on that list, and which does not also appear on our safe words list, or (2) a Perspective toxicity score above 70 (an experimentally determined threshold), a moderation pop-up appears. This allows an adult facilitator to approve or disapprove the users' submitted text. Individual flagged words are also highlighted in the interface directly.

Model Choice and Prompt Engineering: The steps above can be used with any text-to-image model. Certain models also provide further safety mechanisms. We chose to use Stable Diffusion v1.6 (Stability.AI 2023) because it has a built-in filter that attempts to prevent generation of explicit images (though this is far from foolproof), and it allows the use of **negative prompts** which specify types of image that are not desired. We constructed a set of negative prompts by including: (1) negative prompt words to improve image quality (e.g., "text," "blurry," "logo," "low-resolution," "ugly"); (2) terms we identified to improve safety (e.g., "nudity," "sexual content," "hate speech", "distorted faces", "distorted bodies"; (3) additional words suggested by ChatGPT as possibly useful (e.g., "blood," "surgeries.").

Evaluation: We tested the safety measures using informal "red teaming" in which we attempted to break the tool, as well as using ChatGPT to generate (1) a list of possible harmful topics, and (2) image generation prompts for each

[1] `https://github.com/fiebrink1/ImaginationTool_Public`

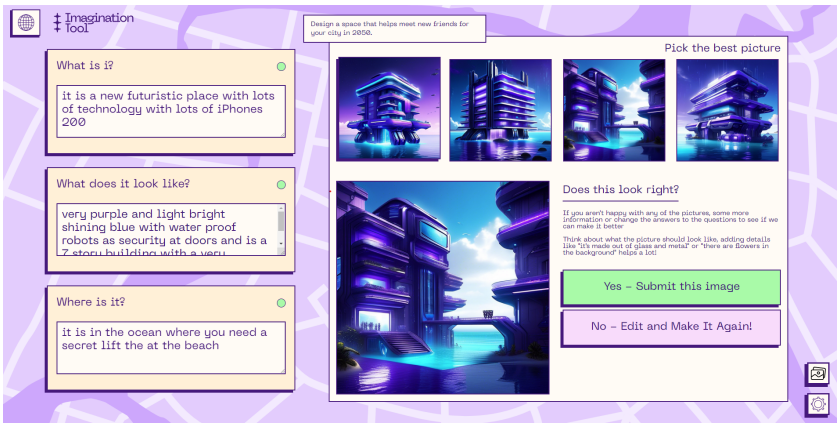

Figure 1: The main screen of Imagination Tool provides several text input boxes. Once the user clicks a button to "Make it!", the right side of the screen shows four generated images, with one user-selected image enlarged.

topic. With the above safety measures in place, the system failed to produce any images we judged to be objectionable.

Generating Satisfying, Task-Appropriate Images

An early experiment with children provided a single text input box for them to describe an imagined concept (similar to most existing text-to-image interfaces). Children tended to describe their ideas abstractly and conceptually, rather than visually, and they were often disappointed with images generated from their descriptions.

We therefore implemented a guided input approach with multiple questions and text boxes, such as "What is it?", "What does it look like?", and "Where can you find it?" The text responses to these questions were concatenated and inserted into a longer text string with additional information (see below) to form the prompt sent to Stable Diffusion. We observed that these questions encouraged children to elaborate on the visual appearance of their imagined concept, supported creative ideation by incrementally building up an idea, and gave insight into how changing their responses could alter the generated image.

It is common practice in text-to-image generation to include further prompt text to enforce a desired visual style and promote high visual quality (LetsEnhance.io 2023; Liu and Chilton 2022). We experimentally identified prompt modifier terms that specified visual styles, medium, lighting, and colour. The Imagination Tool allows users to select amongst these using a drop-down box, with the defaults set to "hopeful digital concept art," "photo," "bright colours," and "volumetric lighting," respectively. We also add "quality booster" terms, some of which (e.g., "Pixar and Disney animation") appeared to further enforce the child-safety of the generated images (e.g., even test prompts describing violence were rendered as peaceful and child-friendly).

Supporting Ideation

Imagination Tool employs three main mechanisms to support children's ideation, specifically:

Asking for Answers to Specific Descriptive Questions: As described above, the main interface employs several prompting questions for users. This encourages children to consider multiple facets of their ideas, and ensures these are reflected in the generated image.

Gallery: The gallery of images produced by other users illustrates what the tool is capable of, and clicking on an image in the gallery also allows children to see the answers to the prompting questions that resulted in a gallery image.

Imagination Bot: Inspired by previous work (Baek et al. 2023) suggesting that a language model chatbot could help craft better text-to-image prompts, we implemented an "Imagination Bot" chatbot assistant (Figure 2) to provide optional support. Users can use conversations with Imagination Bot about each on-screen question in Figure 1 to begin and deepen idea generation. Imagination Bot uses the OpenAI Assistants API with the GPT-4 model (Open AI 2023). The instructions and opening message for the Assistant were tailored to each question, design prompt, and existing user response text (if any). A user can trigger Imagination Bot for an individual question at any time.

The OpenAI Assistant instructions specify the bot's output length and language level to suit the needs of the age group. Crucially, it always concludes responses with a question. The instructions also include hard-coded specific guidance for each design task and question to allow further framing of the bot's conversation. Also, by differentiating the conversation trigger based on whether the user has currently provided any text for the question associated with the bot, the bot can encourage users to deepen unfinished concepts or propose relevant starting points.

When users decide to finish with Imagination Bot, they can request that the bot directly "auto-fills" a short text segment and inserts it into the corresponding user textbox automatically. The Assistant instructions used for this task are: "You are an assistant that creates one sentence simple image prompts for children to use in a text to image generation model. The prompt should be about [design prompt, user input]." Alternatively, users can enter their own text.

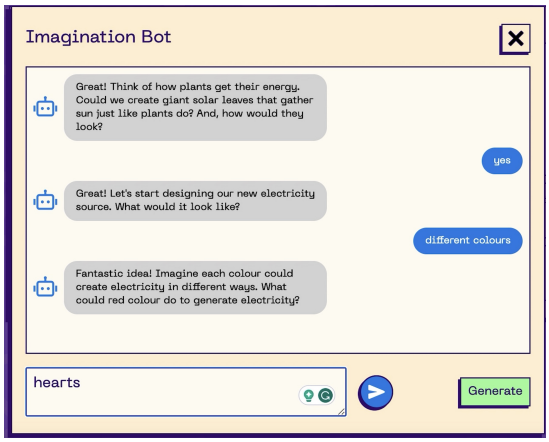

Figure 2: A screenshot of Imagination Bot in use, triggered for the "What is it?" question.

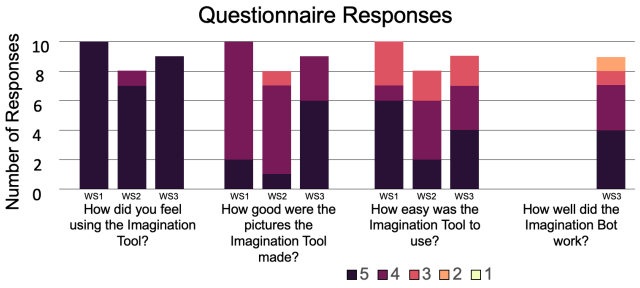

Figure 3: Responses from individual questionnaires from workshops (WS) 1, 2, and 3. (Note that Imagination Bot was only used in Workshop 3.)

Other Usability Features to Support Children

We use a Grammarly browser plug-in to automatically suggest corrections to misspelled words. Further, we use the Chrome browser's native Web Speech API to enable optional live speech-to-text input for the prompt boxes in the main interface, removing the need to type.

Children's Use of the Imagination Tool

The Imagination Tool design was iteratively refined through three workshops with 27 (total) children aged 8–11. Following an introduction discussing AI and AI safety, we invited children to imagine they'd traveled to the future. Children worked in groups of 2–3. Each group had one laptop and was accompanied by one researcher who recorded observations and provided assistance as needed. While a full description of these workshops and their outcomes is beyond the scope of this system demonstration paper, we present some key outcomes below to provide readers with additional insight into the safety and usability of the system.

The safety mechanisms described above worked; we did not observe any instances of harmful or disturbing images generated by the models. Seven false-positive flags occurred in the workshops, all due to words being flagged by the "bad words" list and not being present on the "safe words" list (e.g., "cats and dogs rolling *balls* to each other").

Children generally enjoyed their experiences with the Imagination Tool, felt it made good pictures, and found it easy to use (Figure 3). The Chrome browser's Web Speech API usually transcribed children's speech correctly and quickly, even in a noisy classroom. However, some friction remained due to transcribing errors. Children frequently spent time browsing the gallery, and viewing the specific text used to create a gallery image seemed to boost children's confidence in crafting effective prompts themselves.

We observed that more confident children enjoyed playfully engaging the Imagination Bot or ignored it entirely. However, some less confident users heavily utilised it for initial idea generation and ideation support, eventually transitioning to greater independence. Children generally utilised the bot more in the earlier questions, e.g., with one group exchanging 13 messages for the first question, 7 for the second, and none for the final question. Some Imagination Bot chats suggest the bot could restrict creativity by prematurely steering children down particular pathways (e.g., immediately suggesting that children exploring "a new way to make energy for the city" consider how plants get their energy).

Conclusions and Future Work

We have provided a description of a new software system, Imagination Tool. It is our hope that this work can provide concrete guidance and inspiration to others designing generative AI systems for children, as well as informing future work that explores how AI image creation can play a role in human tasks such as ideation and co-design.

Our approach to ensuring safety seems to work, particularly when an adult facilitator is present. However, it is not ideal for every development or research team to address child safety in an ad hoc way, from scratch. Implementing safety mechanisms is a time- and labour-intensive process, and evaluating them confidently is a daunting task—one for which most developers and researchers are not trained. We hope that in the future, researchers might therefore work to establish and share community-driven approaches to ensuring safety of generative AI systems. Ideally such systems would be usable by small projects such as ours, implemented as open-source for full transparency, and configurable for different ages and contexts.

While Imagination Tool encourages children to explore ideas and communicate them visually, further work is also needed to bridge the gap between children's activities with the tool and helping stakeholders (e.g., policymakers or designers) interpret children's ideas and inform action.

Author Contributions

Implementation: MM, SW. Iterative system design, workshops: All. Writing: MM, SW, ASZ, RF.

Acknowledgments

This work was funded by Innovate UK Feasibility Studies for Artificial Intelligence Solutions grant 10079999.

References

[alt-profanity-check] alt-profanity-check. https://pypi.org/project/alt-profanity-check/.

[Baek et al. 2023] Baek, S.; Im, H.; Ryu, J.; Park, J.; and Lee, T. 2023. PromptCrafter: Crafting text-to-image prompt through mixed-initiative dialogue with LLM. *arXiv preprint arXiv:2307.08985*.

[Bekker et al. 2003] Bekker, M.; Beusmans, J.; Keyson, D.; and Lloyd, P. 2003. Kidreporter: A user requirements gathering technique for designing with children. *Interacting with computers* 15(2):187–202.

[Bird, Klein, and Loper 2009] Bird, S.; Klein, E.; and Loper, E. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.

[Guha et al. 2004] Guha, M. L.; Druin, A.; Chipman, G.; Fails, J. A.; Simms, S.; and Farber, A. 2004. Mixing ideas: A new technique for working with young children as design partners. In *Proceedings of the 2004 conference on Interaction design and children: building a community*, 35–42.

[Jigsaw] Jigsaw. Perspective API. https://perspectiveapi.com/.

[Kidgeni 2023] Kidgeni. 2023. https://kidgeni.com/.

[Landau 2022] Landau, D. 2022. 'Alexa, let's make a story': Amazon introduces a new feature that turns your Echo device into a storytelling companion. https://www.aboutamazon.com/news/devices/what-is-create-with-alexa.

[LetsEnhance.io 2023] LetsEnhance.io. 2023. AI Image Guide: Learn the Nuances of Prompt Writing in 5 Minutes. https://letsenhance.io/blog/article/ai-text-prompt-guide/.

[Liu and Chilton 2022] Liu, V., and Chilton, L. B. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–23.

[Moraveji et al. 2007] Moraveji, N.; Li, J.; Ding, J.; O'Kelley, P.; and Woolf, S. 2007. Comicboarding: Using comics as proxies for participatory design with children. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1371–1374.

[Open AI 2023] Open AI. 2023. Assistants API (Beta). https://platform.openai.com/docs/assistants/overview.

[Open AI 2024] Open AI. 2024. OpenAI Platform: Moderation. https://platform.openai.com/docs/guides/moderation/overview.

[Oppenlaender 2023] Oppenlaender, J. 2023. A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology* 1–14.

[Qu et al. 2023] Qu, Y.; Shen, X.; He, X.; Backes, M.; Zannettou, S.; and Zhang, Y. 2023. Unsafe diffusion: On the generation of unsafe images and hateful memes from text-to-image models. *arXiv preprint arXiv:2305.13873*.

[Rafner et al. 2023] Rafner, J.; Zana, B.; Beolet, T.; Buyukguzel, S.; Michel, E.; Maiden, N.; Risi, S.; and Sherson, J. 2023. Crea.visions: A platform for casual co-creation with a purpose envisioning the future through human-AI collaboration with multiple stakeholders. In *Proceedings of the 14th International Conference on Computational Creativity*.

[Stability.AI 2022] Stability.AI. 2022. Stable diffusion public release. https://stability.ai/news/stable-diffusion-public-release.

[Stability.AI 2023] Stability.AI. 2023. Stable diffusion v1.6 release. https://platform.stability.ai/docs/release-notes.

[Sun et al. 2022] Sun, J.; Gu, C.; Chen, J.; Wei, W.; Yang, C.; and Jiang, Q. 2022. A study of the effects of interactive ai image processing functions on children's painting education. In *International Conference on Human-Computer Interaction*, 93–108. Springer.

[Truong, Hayes, and Abowd 2006] Truong, K. N.; Hayes, G. R.; and Abowd, G. D. 2006. Storyboarding: An empirical determination of best practices and effective guidelines. In *Proceedings of the 6th conference on Designing Interactive systems*, 12–21.

[von Ahn] von Ahn, L. Offensive/profane word list. https://www.cs.cmu.edu/∼biglou/resources/.

[Walsh et al. 2010] Walsh, G.; Druin, A.; Guha, M. L.; Foss, E.; Golub, E.; Hatley, L.; Bonsignore, E.; and Franckel, S. 2010. Layered elaboration: A new technique for co-design with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1237–1240.

[Walsh et al. 2013] Walsh, G.; Foss, E.; Yip, J.; and Druin, A. 2013. FACIT PD: A framework for analysis and creation of intergenerational techniques for participatory design. In *proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2893–2902.

[Yan et al. 2023] Yan, Z.; Wu, Y.; Luo, D.; Zhang, C.; Jin, Q.; Chen, W.; Wu, Y.; Chen, X.; Wang, G.; and Mi, H. 2023. Nacanva: Exploring and enabling the nature-inspired creativity for children. *Proceedings of the ACM on Human-Computer Interaction* 7(MHCI):1–25.