

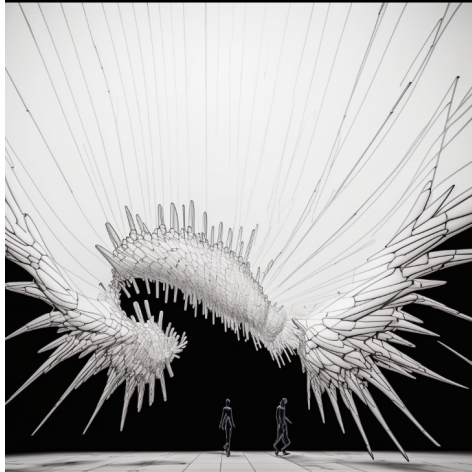
Giving Up Control

Hybrid AI-Augmented Workflows for Image-Making

By Joshua Vermillion, UNLV School of Architecture



Midjourney, a text-to-image latent diffusion model, was publicly released on July 12, 2022. Since that summer, the technological space dealing with AI and machine learning, and specifically latent diffusion models, has moved at a rapid pace (Figure 2). Image-making is now easier than ever—input text and, in a very short amount of time, get images in return. And while much has been made about AI-generated images in the general public as well as in the press, designers should focus on the actual creative processes that integrate latent diffusion models for image-making.



This paper argues that if we are to come to creative terms with AI then we must critically interrogate these new tools in the production of creative works. In doing so, designers can start to chart new creative workflows that integrate AI while augmenting and then evolving our current ways of designing.

Figure 1: Hybrid human/machine dynamic to augment the creative design process¹

¹A graphite hand drawing of a crab claw (top), a rendering of a spiky spatial installation (middle), an AI-rendered blend of the previous two images combining outputs from my traditional design process into a human/AI hybrid drawing (bottom).

Rather than seeing AI as something that will replace human creativity, we discover ways that AI models will augment human creativity in hybrid (human and AI) creative endeavors.



Figure 2: Outputs from different Midjourney models using the same prompt: "a monumental undulating space articulated by twisted wood strips".²

Background: Interlocutors, not Usurpers

As the subtitle implies, I'm here to argue that AI tools, while appearing to do creative things, aren't replacing human creativity. They aren't usurping our tasks, but rather they are enabling us to be more creative while also more productive. But *how* do these systems augment creativity and creative processes? While others have described how these latent diffusion models are scientifically trained and then how they perform (Dhariwal and Nichol 2021, Radford et al 2021), we should closely examine how generative-AI is used creatively by experienced designers and artists.

If we try to understand diffusion models as creative mediums or tools, then we need to better understand these models' affordances, or simply, what tasks they're good at performing (Gibson 1977), as well as how these capabilities and biases re-shape our creative processes. There are two higher-level ways that machine learning and AI can augment what we, as humans, already do. The first seems obvious and falls under the category of optimization. AI models can be trained to perform large or complex tasks for us, and can resolve problems that are too complicated or messy for human brains to solve. Often these sorts of problems

² Outputs from different Midjourney models using the same prompt: "a monumental undulating space articulated by twisted wood strips". In comparing the results, we can see just how far the technology has evolved in just over one year.

rely on enormous data sets and competing sets of constraints, which makes deducing an optimum outcome or scenario too complex for the human brain but suitable for computation.

But the second “classification” of AI that makes it very useful is one of prediction. This generative aspect of AI, where the human gives the model some sort of prompt (words, images, etc) and then the model, in turn, generates results, is the reason why many creatives are very interested in AI and why other creatives are very nervous about it. Coming to creative terms with new technologies can be daunting enough, and certainly giving up some direct control of the creative process to a computer can be equally daunting. Nothing speaks better to this new way of working, augmenting creativity with AI, than large language models and image-making.

But what is creativity? The cognitive scientist, Margaret Boden, defined creativity in terms of outputs that are new, surprising, and valuable (1990). But architectural theorist Neil Leach points out the shortcomings of Boden's definition of creativity, namely, that it fails to account for the creative process and only focuses on the results (2022). The argument that creativity is dependent on process shifts attention away from judging the results of creative partnership, and instead, spotlights the creative roles and burdens of humans versus machines. In a hybrid human-AI creative partnership, it seems as though humans give up control of some of the creative process, and in particular, significant parts of creating.

Giving up Control: Image-making to Image-finding

As mentioned before, any creative medium has a set of capabilities, affordances, uses, and limits. One of the diffusion model's qualities that make it useful (in a bit of counter-intuitive logic) is the relative lack of control over its generative outputs. The model wields direct control over the generation of graphics which means that the designer only has indirect control on the results. But I would argue that this relative lack of control is a feature rather than a bug.

For AI to truly be of use to the creative process, it needs to be both instrument to and interlocutor with the human designer or artist in creating, exploring, and reacting to ideas.

Unlike many digital tools before it, AI is most productive in suggesting ideas and responses to problems that are messy and not initially well-defined, and therefore useful in earlier stages of the design process, almost like sketches—well, generative sketches.

Working with AI diffusion models is a lot like working with a large creative team ready to visualize a varied spectrum of idea responses—from the obvious to the bizarre—almost as quickly as you can write down ideas (even very open-ended ideas). The results are merely two-dimensional images, and admittedly there is a lot more to designing objects, spaces, and architecture than just making flat images, as elegantly pointed out in an essay by designer and scholar, Andrew Kudless (2022). However, at the earliest stages of design, this technology seems very well equipped to help with brainstorming and divergent thinking, rather than most of the digital design tools of the past that need a well-defined problem, resolve

ambiguity, or add precision to the design output. AI diffusion models aren't precise, and that's the point. This leads to a large shift in control.

Is the human engaged in image-making or is it more like image-finding?

One navigates through the model's latent space and discovers surprise results and interpretations of the original prompts. The model is directly making images, and the human is only directly giving instructions to the model. My own experiences and experiments have led me to map some workflows that take advantage of this hybrid human/machine dynamic to augment my own creative design process.

The Process: Reprioritizing the Language of Design

How does a hybrid process work in reality? Generative AI image-making generally starts with language. One describes what to generate with plain language and then the diffusion model draws from its latent space and denoises a novel set of images from static. All of that said, getting images of what you asked for in the prompt and getting images of what you wanted aren't necessarily the same thing (Vermillion 2023). This tension between the human's written description and the machine's interpretation and drawing ability is interesting in the way that it is mediated by language.

One now has to ask: What words, names, places, etc. do I and the machine have a shared understanding of?

This "skill" of writing prompts in order to gain more control over the outputs, or to simply explore the diffusion model's latent space, requires writing very descriptively in a search for words that generate the desired object, space, or effect. Having explored this topic extensively, I've been asked to run workshops for various companies and universities on this very topic: Writing descriptive prompts to generate specific spaces and effects. Without getting into too many details, it really does seem like certain frameworks for approaching prompt writing really do help in order to have productive conversations with these large-language-models.

Counterintuitively, the less time spent drawing gives one more time to better develop the idea—and to develop a vivid description that effectively communicates the idea—for use as prompts. Even if the generated images are disappointing or limited in their immediate value, the human creative is still forced to pause and think and then write about the goals, concepts, and desires of a project. And the immediacy of the image outputs from the model facilitates an iterative investigation of words and images in a rapid cycle as demonstrated in Figure 3.



Figure 3: A series of investigations showing the process of starting with a simple prompt and then building complexity while comparing outputs. ³

Hybrid Workflows

What do hybrid workflows that take advantage of AI's non-determinate generative outputs look like? AI's affordance of speed greatly enhances the breadth and depth of an iterative design inquiry as one can explore many more ideas and much more territory in the design space within the same amount of time as traditional design exploration tools. This luxury of speed allows one to improvise methods of creating new and strange juxtapositions generated from the latent space, while letting us see things in new ways or from a different direction and challenging us with alternatives to our first ideas. AI is very good at creating situational juxtapositions in order to challenge design preconceptions, to blend media to tell a story, or to perform simple brainstorming by combining disparate things together.

Often, I will start a session with Midjourney by trying a prompt that combines two different things together in new ways such as the examples in Figure 4.

³ The following prompts - starting with a simple prompt and building complexity - were used in this iterative investigation: "a romantic alleyway in the old city"; "a set of diners eating in a romantic alleyway in the old city"; "a set of diners eating under an immersive spatial installation hanging over a romantic alleyway in the old city"; "a set of diners eating under an immersive spatial installation composed of hundreds of parametric intricate glass sculptures hanging over a romantic alleyway in the old city"; "a set of diners eating under an immersive spatial installation composed of hundreds of purple parametric intricate glass sculptures with undulating and turbulent patterns hanging over a romantic alleyway in the old city".



Figure 4: A series of images generated in Midjourney.⁴

Not only do the images that are generated by AI often surprise me, but the expressive language that I start to insert into the prompts also subverts my very understanding of the design problem that I'm trying to address and leads me to reprioritize the set of ideas in my head. Trying to communicate to a computer in plain English, as it turns out, can be a very long but satisfying exercise of trial and error, all the while developing more and more descriptive language and generating uncanny results.

AI doesn't have to replace the human touch. In fact, often I will sketch ideas while generating images in Midjourney, and just to come full circle, I often feed my sketches into Midjourney when simple words won't do to communicate with the model such as the example shown in Figure 1 (located at the beginning of this article). By combining an old graphite line drawing and an old color rendering, the AI diffusion model is able to blend the two together to visualize new possible architectures and extending my imagination to generate options I might have never thought of.

Other explorations of mine change the context of something familiar to see what is generated, for instance situating a scaled model or maquette within a new full-scale context such as Figure 5. By photographing a 3D-printed study model and then placing it in a real or fictional place, one can evaluate scale but interestingly, the diffusion model also modified the building and the context in interesting generative ways.

⁴ A series of images generated in Midjourney. (left) "a tall parametric desert pavilion for burning man made from inflated translucent polychromatic latex membranes in the shape of an elaborate branching bifurcating tree blowing in the wind", (center) "an immersive spatial installation made of bulbous glass bladders filled with colorful fluids and hanging from the ceiling of a tall gallery space, high key, bright and airy, photorealistic", (right) "an immersive spatial installation of undulating and hanging yellow extension cables densely covering the ceiling of a large production facility".

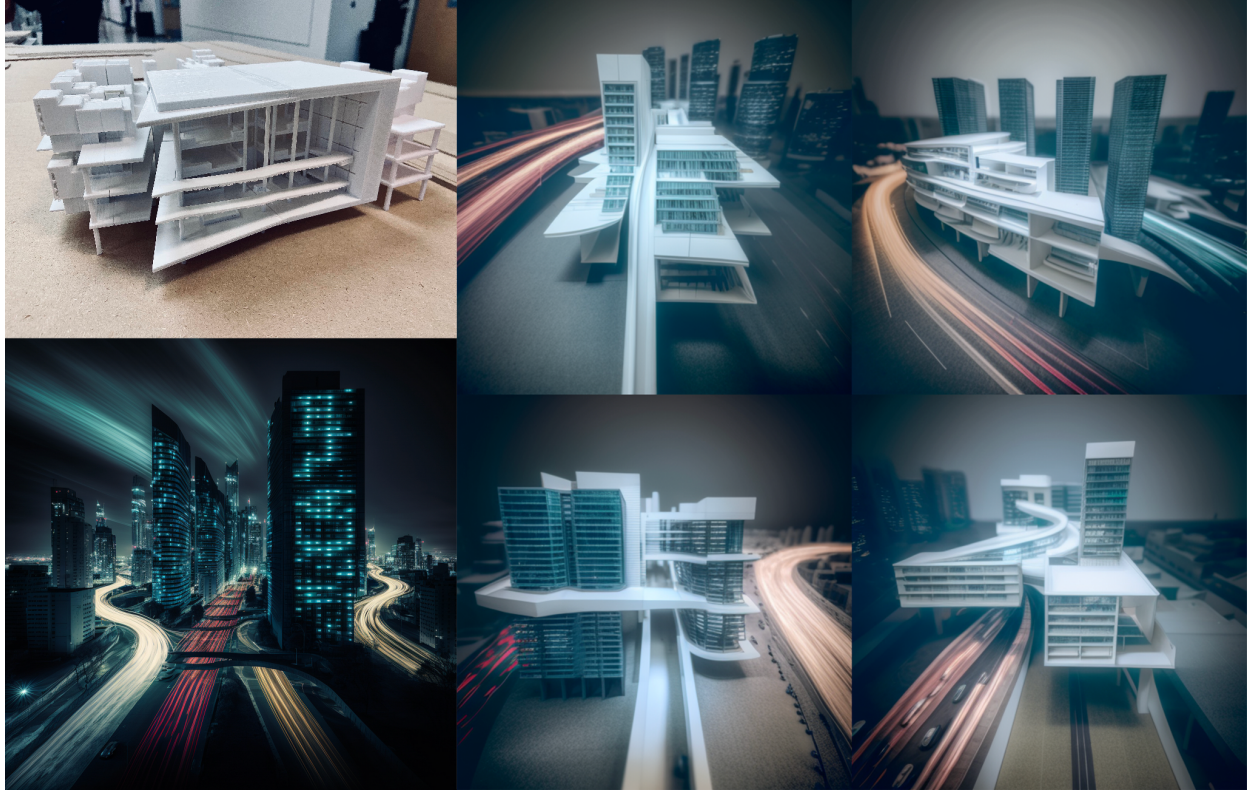


Figure 5: Top left: A photograph of a 3D-printed architecture model; bottom left: an evening rendering of a cityscape; right: a 2 x 2 combinatory sequence of architecture and context generated in Midjourney.⁵

Another example of this way of work is from my collaboration with Harper's BAZAAR magazine in Serbia to create the first hybrid photography/AI fashion magazine cover and editorial in May 2023. In this case, the goal was to combine conventional photography of a model and wardrobe along with fictional, fanciful settings that I generated in Midjourney (Figure 6). This process points the way for a truly hybrid process of human-crafted prompts, AI generation, human generation and creativity, along with human and software post-production, and all by a multidisciplinary team of creatives.

Conclusion

As we continue to investigate these new tools to discover their abilities and affordances, we must always ask how they might reshape and redefine the ways in which we work. This is a self-reflection opportunity for those of us in creative disciplines to better understand how we design and how these tools will augment or evolve the design process.

⁵ Top left: A photograph of a 3D-printed architecture model; bottom left: an evening rendering of a cityscape; right: a 2 x 2 combinatory sequence of architecture and context generated in Midjourney.



Figure 6: From left to right: Harper's BAZAAR Digital Cover and editorial photographs from 30 May 2023. The model and wardrobe are real, however, none of the animals, landscapes, or architectural settings actually exist, but rather they were generated in Midjourney.⁶

Already, a diffusion model feels like having our very own creative team, at our fingertips, ready to brainstorm any idea we can write down, from the normative to the outlandish. And just like having a talented creative team, we now prepare to be surprised at the results every time we roll the dice with a prompt—the varied interpretations, the surprise juxtapositions. Oftentimes the most interesting results are the ones that subvert our preconceptions, forcing us to look at our own ideas in very different ways, and extending our imaginations.

In all of the use cases described here, the diffusion model is generating the imagery from the human-made prompts and then the human is responsible for critically questioning and curating results. Human thinking and critical discernment are still needed to design and make art that has value, but at the same time, these tools are here, they aren't going away and it's up to creatives to better understand how to use AI creatively.

Perhaps it's time for us to stop defining creativity as only a human endeavor, but rather one of hybrid capabilities that combines a diffusion model's latent space with human critical thinking and imagination.

⁶ From left to right: Harper's BAZAAR Digital Cover and editorial photographs from 30 May 2023. The model and wardrobe are real, however, none of the animals, landscapes, or architectural settings actually exist, but rather they were generated in Midjourney.

Bibliography

Boden, Margaret A. (1990). *The Creative Mind: Myths and Mechanisms*. London: Weidenfeld & Nicholson.

Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. *Advances in Neural Information Processing Systems*, 34, 8780-8794. <https://doi.org/10.48550/arXiv.2105.05233>

Gibson, James (1977). "The Theory of Affordances", in Robert Shaw and John Bransford (eds), *Perceiving, Acting, and Knowing*, London: Wiley.

Kudless, Andrew (2022). "Diffused Narratives". in *Gradient Journal*, Feed #2. Online source: <https://gradient-journal.net/articles/diffused-narratives>

Leach, Neil. (2022). *Architecture in the Age of Artificial Intelligence: An Introduction to AI for Architects*. London: Bloomsbury.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). "Learning Transferable Visual Models from Natural Language Supervision". In *International Conference on Machine Learning* (pp. 8748-8763). PMLR. <https://doi.org/10.48550/arXiv.2103.00020>

Vermillion, Joshua (forthcoming, 2023). "Getting What You Asked For". In Frank Jacobus and Brian M. Kelly (eds). *Artificial. Intelligent. Architecture: New Paradigms in Architectural Practice and Production*. New York: ORO Editions.