Designing With AI

How I came to love design and used AI to alleviate the most frustrating parts of the process.

By Lydia Chilton

DOI: 10.1145/3596926

Right now, it seems like AI is eating the world as it now generates media: text, images, music, code, videos, and who knows what else. The hype is in full tilt: Will it take our jobs? Will it raise our children? Will it cure cancer? Will it answer simple questions correctly? Although there are some very exciting and impressive things happening right now, AI by itself is not enough. AI isn't going to produce great works of art. People will. AI can help, but we have to know how and where to use it. We need a framework that can incorporate AI into human workflows.

The single most important thing I learned in grad school was the science of design. It changed my life. I used to think good ideas were just magic. They came to you in the shower, or not at all. And that people like Mozart, Picasso, Einstein, and Tesla just had a lot of them, and other people had less of them. Or maybe none at all. I was afraid to think I might be in the "none at all" category.But one day, I was stuck in a cab with a professional designer. It was during rush hour in Beijing, so we were going to be there for a long time. I'd always been confused by his role, so I asked him directly: "So you're a designer... what do you do?" And he explained the design process to me.

You don't start with a solution or even a problem. You start by learning about a context. For human-centered design, that's usually people in an environment, their history, their activities, their needs and challenges, and current practices. Then you isolate a problem or opportunity—hopefully one you have some insight into addressing. But that's just the beginning.

Once you isolate a problem, you don't just pull a solution out of thin air. First, you generate many ideas: whole ideas, half ideas, safe ideas, crazy ideas, and everything in between. You can also build on ideas or bring ideas together to make even more ideas. Overall, the best way to have a good idea is to have a lot of ideas. But ideas are only a starting point. Ideas need to be instantiated into the world. The best way to do that is to create prototypes—quick and dirty implementations of the idea that lets you test whether it is feasible, and what works and what fails. Then you can ideate and prototype again.

At the very heart of all design is the process of flare and focus. First explore a space, then pick where you want to drill in deeper in that space. Explore the context before you pick a problem; have many ideas before you pick a few; and build multiple prototypes before you pick which ones make the cut. That's where good ideas come from—many rounds of flare and focus. That is design.

I became obsessed with the design process. It wasn't magic; it was something anyone could do. Even me! For a while, it felt like an amazing secret that I had discovered. And I applied it

Image by Andrij Borys Associates with the assistance of Shutterstock AI

to everything: school, life, home decorating, event planning, everything. But after I got past the initial excitement, I started to see some of the problems with the design process.

The first problem is that whereas flare and focus is a great model, it's not a perfectly accurate one. To my great frustration, I had been in many brainstorming meetings with smart people and hundreds of sticky notes; we all felt the creative juices flowing, but I couldn't honestly say that any promising ideas came out of the session. I began to doubt the process. Maybe you still do need magic to get good ideas. I needed to know: Does brainstorming really lead to good ideas? If so, how?

I decided to develop a more concrete model of the design process. I focused on only one problem—generating visual metaphors. Visual metaphors are an advanced graphic design technique to draw attention to a message. They combine two objects in a novel and a useful way to convey a message symbolically. They often strike people as highly creative. Figure 1 shows several examples of visual metaphors for messages: a new drink is made in New York City (like a skyscraper is), a new drink is cool and lemony (like a drink cozy made from a lemon peel), and an article about digital medicine (like getting a health check-up over the phone). These are examples made by people using a system we developed called VisiBlends [1].

In VisiBlends, the design process is broken into such small steps that it is clear there is no magic involved. It was just brainstorming many visual symbols of two concepts (like Starbucks and summer), and finding two that share the same shape, so they can be blended together (see Figure 2). From this model, it was clear that brainstorming is essential because if you don't have enough options of images to symbolize each concept, the likelihood that two symbols will be similar enough to blend is low. You can get lucky, but you don't need to depend on luck if you have more symbols. That's how brainstorming works.

Obviously, this VisiBlends design workflow is tailored to one specific problem and would have to be generalized to other problems, but to me, it was a kind of existence proof that the design process really did work without magic.

However, after running this process over and over with people, I noticed another big problem with the design process: It's exhausting. Even with a room full of people, thinking of 30 or 40 different symbols for a concept can be tedious and draining. People can easily do five or 10 but getting beyond those initial ideas is cognitively taxing. It is easy to get fixated on a single idea (like finding a dozen different suns for summer), when you need diverse ideas (sun, ice cream, beachball, beach umbrella, swimming pool, air conditioner, lawnmower, barbeque, etc.). What's worse is once you have 30 symbols of both concepts, you have 900 combinations to consider. That's nobody's idea of a good time.

Seeing how much work the design process is, I started to understand why people might just wait for a brilliant idea to magically come to them. For experienced designers, heuristics and a repertoire of previous good ideas to lean on can speed up the process, but it's still work. And for beginning designers, the process can be daunting. The design process would be much more impactful if we could make it less cognitively demanding

Figure 1. Examples of visual metaphors from the VisiBlends system. The blend idea and prototype come from the system, and final results are Photoshopped.

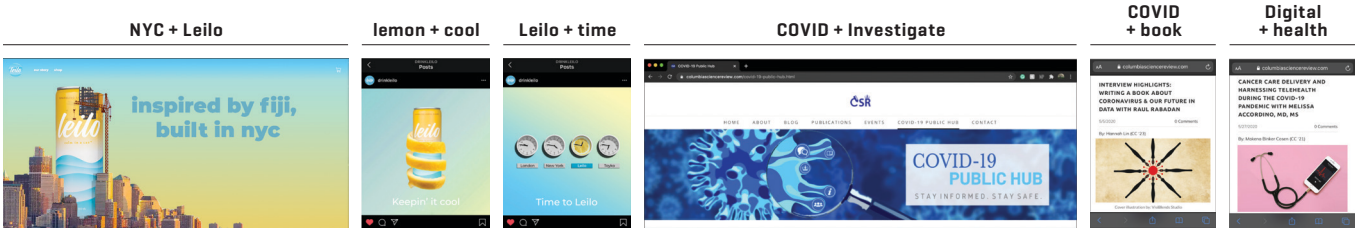
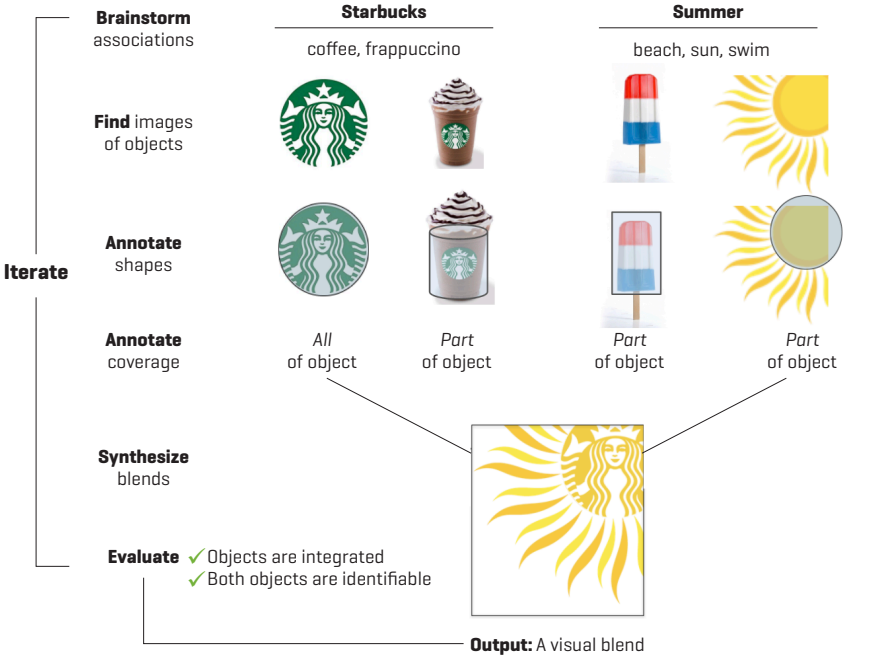


Figure 2. The design process workflow as broken down by the VisiBlends system.



I then turned to finding ways AI could help with the design process. I started to build tools with traditional AI techniques, like building knowledge networks to help find multiple diverse ideas easily [2] and using computer vision tools to detect shapes, extract objects, synthesize blends, and disqualify hundreds of combinations that would not blend well [3]. This worked and it proved AI could do these tasks. But the AI was brittle, and you'd have to retrain it for every new problem. It was a lot of work to set up…practically speaking, it was not worth the effort unless you were making thousands of blends.

Then a miracle happened. Foundation models in AI took off; GPT-3 and DALL-E 2 took my breath away. The models had such general capabilities, but were by no means perfect. They weren't brittle and didn't need to be retrained for every new task. But the downside of these models is they need specific descriptions of the output—you can't just say "make me an ad for a new drink called Leilo," or "make me a news illustration about digital health." That is too big of a creative leap for AI to make, so you need to be more specific. But you can't be too specific because the AI might not be able to create it. For example, when I described visual blends explicitly it does pretty terribly. Figure 3 shows DALL-E images for the prompt, "a Starbucks logo in the shape of the sun." They're all embarrassingly bad.

Visual metaphors are probably too specific for DALL-E to create, but we wanted to explore how DALL-E could create news illustrations. News illustrations don't have to be visual metaphors, they can be anything, so we were open to the AI suggestions. Unfortunately, just using the headline as a prompt doesn't work well. The outputs don't capture the tone, or the meaning of the article, and they're not particularly eye-catching. In Figure 4, there are two DALL-E images to portray the headline "Climate change is speeding up sound in the ocean. That's bad for marine life." The first image was prompted by the headline, and it's not great. The subject is a boat with black smoke (probably polluting). A boat is not an ideal subject, and the colors don't convey the negativity in the tone of the piece. Additionally, the style is somewhat off. It's not photorealistic, but it's also not really a cartoon. The whole thing is confusing. But a better prompt can fix this.

To get more relevant and aesthetic images, it's better if users describe their own subject and style that convey the meaning and tone of the piece. The second image in Figure 4 is an improvement. It shows an underwater graveyard that has more gravitas and more closely matches the tone.

Although specifying a subject and style for a headline definitely gets better results, unfortunately it's hard work. It requires brainstorming many subjects and styles and testing what combinations DALL-E will generate well. To alleviate this burden, we built a system called Opal to help structure and organize the flare and focus process [4].

Opal uses GPT-3 to suggest concrete subjects from the headline (see Figure 5). This includes abstract keywords like "global warming" and "marine life" as well as more concrete subjects like "whale" and "fire." GPT also suggested abstract emotions for the piece, like "panic" and "horror" as well as concrete subjects related to those emotions, like

"graveyard." It also suggests artistic styles like "documentary photography," "glitch art," and "abstract impressionism." It takes a few minutes to generate a lot of options, but it's worth the short wait. Like any brainstorm, not all the outputs are good, but some good ideas and combinations come out of it.

And although the computer is making suggestions and prototyping images, the system works best if a person guides the generation. The best image for the article about climate change harming marine life was a combination of subject and style words that the user put together from different machine suggestions. "Whale dying in a red ocean on fire photo" captured the marine life with a sympathetic subject (the whale), and "red ocean on fire" captures the grave tone and is quite eye-catching and editorial.

Figure 6 shows more examples of people using Opal to make news illustrations. For the headline "Me, TV, and COVID-19," the illustration shows the malaise of being indoors during the COVID pandemic with the prompt "TV in the style of Dali." For the headline, "Hyrule, Heroics, and Healing," the illustration shows the cathartic power of video games with the prompt "a hug in the style of pixel art." For the headline "Why Metternick Matters," the illustration shows a classic European diplomat making the "V" sign with the prompt "diplomat in the style of contemporary art." For each example, the prompt used a subject and a style relevant to the tone or topic of the headline.

Opal and VisiBlends both use a single round of flare and focus, but typically for hard problems, the design process requires multiple rounds of flare and focus. Often, the first round is to formulate the problem before you start to solve it, and after the problem is framed, it can take multiple interactions of flare and focus to hone in on a solution. I wanted to capture that iterative aspect of design, so we picked a problem that was harder than visual blends or news illustrations: pop culture blends.

Pop culture blends are images that blend a brand, product, or organization with a domain of pop culture (see Figure 7). For example, every year on Star Wars Day (May 4th), brands find

Figure 3. DALL-E images for the prompt "a Starbucks logo in the shape of the sun." These examples show the challenge of getting a specific image out of text-to-image generative AI.

Figure 4. Two DALL-E generated news illustrations for an article about climate change harming marine life.

The left image was generated using the headline as the prompt, it fails to capture the subject and tone of the article. The right image was generated by a prompt suggested by the OPAL system.

Climate change is speeding up sound in the ocean. That's bad for marine life.

DALL-E: "Climate change is speeding up sound in the ocean. That's bad for marine life."

Climate change is speeding up sound in the ocean. That's bad for marine life.

DALL-E: "Graveyard underwater, photo"

Figure 5. Opal system diagram.

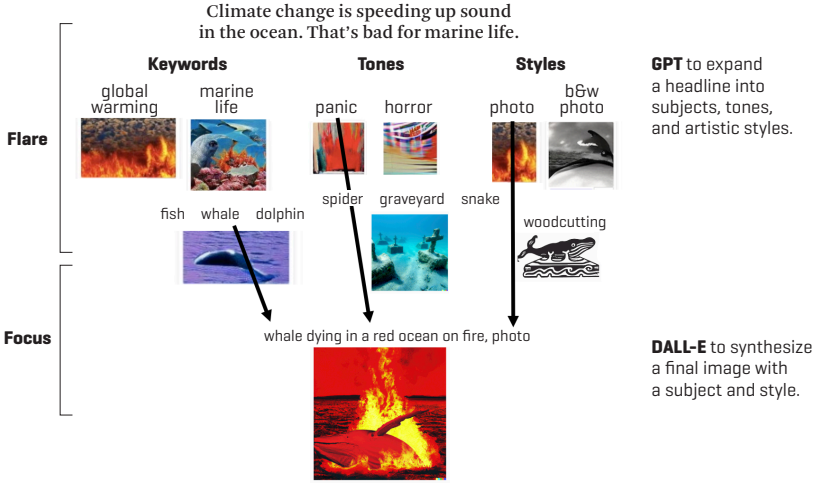


Figure 6. Three examples of news illustrations made using Opal.

Me, TV, and COVID-19

DALL-E: "TV in the style of Dali"

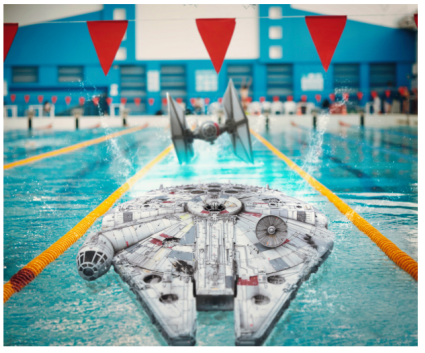
Hyrule, Heroics and Healing

DALL-E: "A hug in the style of pixel art"

Why Metternich Matters

DALL-E: "Diplomat in the style of contemporary art"

Figure 7. Three pop culture blends from the PopBlends system. The idea and prototype were created by the system and the final results were Photoshopped.



Can you swim the 1,500m in less than 12 parsecs? Join the #swimteam #starwarsday

Tame that frizz. Happy #starwarsday from #shampoo

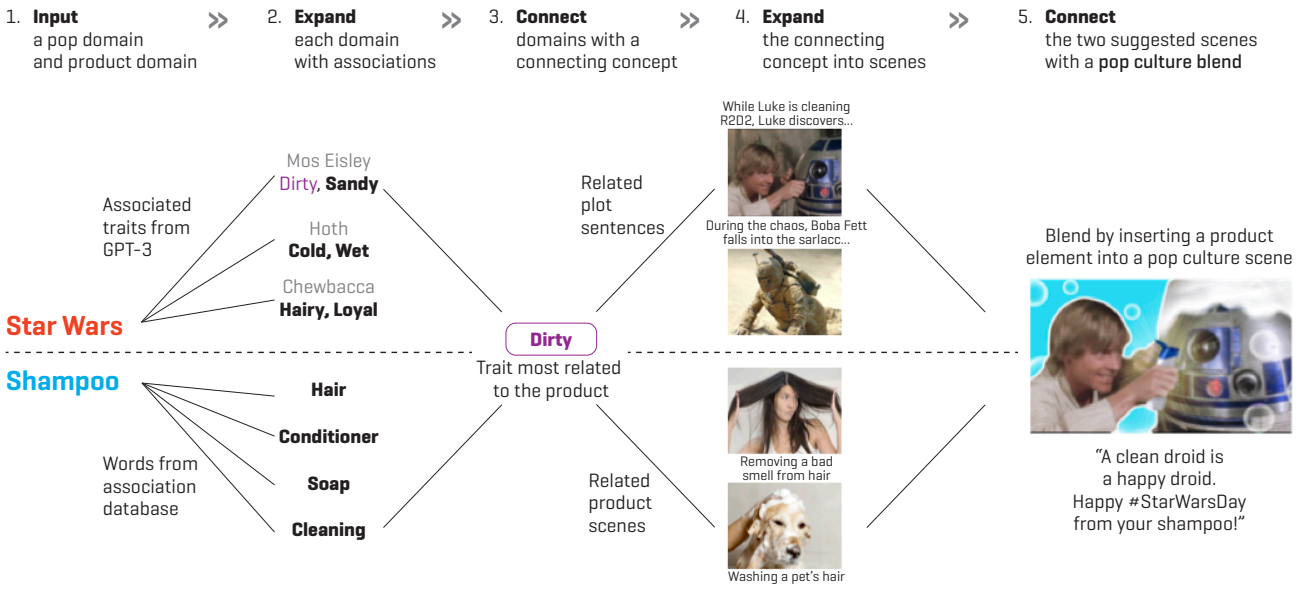
ways to reference memorable "Star Wars" scenes with their products and post it on social media. These images are important for online campaigns because they connect the product to something people already know and like in a way that's already culturally appropriate for social media: memes.

Making pop culture blends is hard because it requires finding a connection between two totally different things. What does shampoo have to do with a space opera? On the surface, nothing. But we can use flare and focus to expand the concepts behind both the brand and the pop culture domain so we can increase the odds of finding a conceptual connection. For example, shampoo is for cleaning, and Tatooine is dirty. The concept of dirty/cleaning could possibly connect the two domains. Once we find a conceptual connection, we do another round of flare and focus to find the images related to the conceptual connection to blend. For example, Luke Skywalker cleaning dirty droids (on Tatooine), and shampoo cleaning hair with lots of suds.

At the heart of the pop culture blends challenge is associative reasoning. What ideas are associated with shampoo? What ideas are associated with "Star Wars"? Which of those associations might be shared? Traditional nat-

Figure 8. The design process workflow for the VisiBlends system.

LLMs are used for brainstorming pop culture associations, product associations, and scenes and activities related to a product. Semantic similarity scores from LLMs are used for finding concepts that connect the pop culture domain to the product.



ural language processing (NLP) methods struggle with associative reasoning and common sense. They are better at extracting facts like the names of characters. However, large language models are quite good at associative reasoning because they process tons of free text on the internet and learn the statistics to associate words and concepts. This opens the possibility to have AI help people with brainstorming and synthesis necessary to find pop culture blends.

To assist the process of finding pop culture blends, we created an AI-infused process in a system called PopBlends [5]. In the first round of flare and focus, users input the pop culture domain (for example, "Star Wars") with the product (for example, shampoo), and the system helps find connecting concepts. To do this the system expands both concepts with a combination of traditional NLP and large language models. It uses traditional NLP to extract all the characters, objects, locations, and organizations from the Wikipedia plot summaries of the "Star Wars" movies. For example, Chewbacca, lightsaber, Tatooine, and Rebel Alliance. Then it uses an LLM to list the activities, adjectives, and catchphrases of each entity. For Chewbacca, the LLM produces activities such as co-piloting the millennium Falcon, adjectives such as hairy, and only one catchphrase, "AAAARRGGG!" Similarly, it expands shampoo into its associations such as hair, conditioner, soap, cleaning, etc. With this wealth of knowledge on both the pop culture domain and the product, it uses semantic similarity metrics to find concepts that apply to both. This is the focus process —searching through all possible pairs to find a connection like dirty/cleaning as seen in Figure 8.

Once the system finds a connection, it does another round of flare and focus to find images from each domain related to the connection word. For the pop culture domain, it is important to find not just images of characters and objects, but memorable scenes from the film. For example, a scene from "Star Wars" related to "dirty/cleaning" is when Luke is cleaning R2D2 (on Tatooine) and accidentally finds a message from Princess Leia. The system finds these scenes by searching Wikipedia plot sentences for relevance to the connecting concept, then finding images of high-ranking scenes. For the product domain, we also want "scenes" related to the product, or at least activities and actions they are associated with. For example, a scene/activity associated with shampoo and cleaning is "washing a pet's hair." The system found this using an LLM to suggest scenes related to shampoo and cleaning, then found an image of this scene. Next, users could quickly look through a small set of images to find ones that might be blended. Here the scene of Luke cleaning R2D2 could easily be blended with washing the dog's hair. With a small amount of Photoshop, the user produced a pop culture blend.

But if dirty/cleaning had not produced a good image, that's okay. In the words of Yoda, "there is another." In fact, many others. The benefit of flare and focus is that there are many options, which is important because the best way to have a good idea is to have a lot of ideas. Another connecting concept between shampoo and "Star Wars" is hair. Chewbacca is hairy, and shampoo is for cleaning hair. A lot of hair in Chewbacca's case, and he gets pretty dirty in many "Star Wars" scenes—in the trash compactor and on Tatooine. The system can find these connections as well.

We also experimented with using LLMs to make the connections automatically, and it sometimes did surprisingly well. We asked: "What character in 'Star Wars' would you associate with shampoo? Why?" And it sometimes said: "Chewbacca, because he is hairy." Wow. What a great connective leap. It also said more basic things such as, "Princess Leia because her hair is always perfectly coiffed." Princess Leia is a good connection because she does have iconic hair, but not for the reason the LLM suggested. And it also just outputs nonsense: "I associate Darth Vader with shampoo because his helmet is black." Is that useful? Probably not. So LLMs by themselves are hit or miss, but it's worth giving them a shot. It's also nice to have a process to fall back on if they fail to produce a connection in a single prompt.

After building three systems and experimenting with a number of ways that AI can assist the design process, I'm convinced AI is a great tool for flare and focus. It can suggest multiple diverse ideas that span a design space. It can search through ideas and find ones to combine. It can create prototypes that make an idea more concrete. And sometimes it can make a final image, ready for print.

But AI still isn't like waving a magic wand; people need to drive the process because AI needs guidance to understand the context of a problem, to evaluate outputs, to iterate toward better solutions, and to produce or edit the final versions. There's a lot of human touch that goes into the design process, but for the most cognitively taxing parts that are really just information retrieval and search, AI can alleviate the burden.

Learning about the design process obviously had a big impact on me. One of my greatest joys of teaching is sharing the design process with students to help them have the same epiphany I did: Good ideas aren't magic, they are the fruit of a process. It's a process anyone can follow, but it is a lot of work. And I'm excited about how AI can help make the process faster and easier so people can focus on getting the design right.

References

[1] Chilton, L. B., Petridis, S., and Agrawala, M. VisiBlends: A flexible workflow for visual blends. In *CHI Conference on Human Factors in Computing Systems Proceedings* (CHI 2019). ACM, New York, 2019.

[2] Petridis, S., Shin, H. V., and Chilton, L. B. SymbolFinder: Brainstorming diverse symbols using local semantic networks. In the *34th Annual ACM Symposium on User Interface Software and Technology* (UIST '21). ACM, New York, 2021.

[3] Chilton, L. B., Ozmen, E. J., Ross, S., and Liu, V. VisiFit: Structuring iterative improvement for novice designers. In *CHI Conference on Human Factors in Computing Systems* (CHI '21). ACM, New York, 2021.

[4] Liu, V., Qiao, H., and Chilton, L. B. Opal: Multimodal image generation for news illustration. In the *35th Annual ACM Symposium on User Interface Software and Technology* (UIST '22). ACM, New York, 2022.

[5] Wang, S., Petridis, S., Kwon, T., Ma, X., and Chilton, L. B. PopBlends: Strategies for conceptual blending with large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). ACM, New York, 2023.

Biography

Lydia Chilton is an assistant professor in the Computer Science Department at Columbia University. Her research is in computational design—how computation and AI can help people with design, innovation, and creative problem-solving. Applications include: creating media for journalism, developing technology for public libraries, improving risk communication during hurricanes, helping scientists explain their work, and improving mental health in marginalized communities.

1528-4972/23/06 $15.00