

Ideating with Text and Image Models by Treating Data as Material*

JEFFREY V. NICKERSON, Stevens Institute of Technology, USA

SITONG WANG, Columbia University, USA

LYDIA B. CHILTON, Columbia University, USA



Fig. 1. A diffusion imodel output from the prompt "data as material"

1 INTRODUCTION

Data as material, the focus of this workshop, is a simile: qualities of data and materials are put in correspondence. This paper explores this simile with the assistance of technology, based on recent work demonstrating how generative AI may be useful in ideation [3, 5, 7, 8]. We do this because this case may add more empirical results to the still nascent study of generative AI as an ideation tool. We also do this because more empirical results may help with the development of theories about how humans and AI can work together. During ideation sessions that use AI, humans and machines working together form a larger system, a metahuman system [4]. These systems are novel because humans and machines have different mental models and learn at different speeds [6]. One reason they may be useful in ideation is because machines and humans operate so differently that they can shake each other out of fixation[9]; the pairing can result in a process that is both fun and productive.

Large language models and image diffusion models are built from data. The data is for the most part generated by humans, so the models represent a kind of collective creativity. Because of this, it is reasonable to use these models to ideate on the implications of the simile *Data as material*. This paper does so by using the language model GPT-3[1] to generate text that is selected and modified before being fed to the image model midjourney[2] to produce images. The process is shown in Figure 2. The human sends a prompt to the language model, which returns text that can be used as a script to send to the image model. In this case, it is not a script as in the dialog of a play, but a script that suggests an

*This paper presented at the 2023 CHI workshop entitled: Data as a Material for Design: Alternative Narratives, Divergent Pathways, and Future Directions.

instruction that will produce some kind of visual output, as in the instructions an art director would give to a concept artist, graphic designer, or illustrator. The image model takes the script as input and produces images as output.

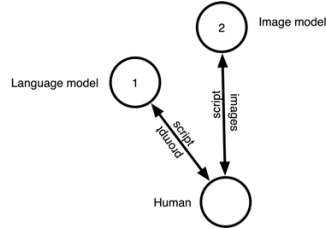


Fig. 2. A human delegates to language models to generate text to send to image models

Specifically, the language model is interrogated to generate ideas or concepts, and then those ideas or concepts are fed as textual prompts to the image model. Sometimes the textual prompts are relayed without editing. Other times edits are made. So the human, in this case one of the authors, remains in the loop, delegating first to text, then selecting and sometimes editing the textual responses, then sending to the image model. The image model generates four low resolution images. The human then selected some of these images to be rendered at higher resolution. There is also an option in the interface to the image model that allows for a request to generate more alternatives starting from one of the preferred images. The regeneration by default uses the same textual prompt, but there is an additional option to also modify the prompt. As in any kind of ideation, many ideas are left on cutting room floor, while the most promising are modified further.

That is the entirety of the process involved. The model is really a human in the loop model, with a human serving as a director and an integrator across AI systems. Shown in Figures 1, 3, 4 are 51 outputs from this process.

2 THE EXPLORATION

The exploration started by asking the language model: "Suggest ways data can be used as a material". The model responded:

- (1) Data can be used to create 3D models and objects. For example, 3D printing technology can use data to create physical objects.
- (2) Data can be used to create interactive visualizations and simulations. For example, data can be used to create virtual reality experiences or augmented reality applications.
- (3) Data can be used to create art. For example, data can be used to create digital paintings or sculptures.
- (4) Data can be used to create music. For example, data can be used to generate musical compositions or soundscapes
- (5) Data can be used to create physical products. For example, data can be used to create custom-made products such as clothing or furniture.

These were then paraphrased and fed into the image model. For example, the metaverse suggestion resulted in images 5(e)-5(k). The physical products suggestion resulted in images 4(n)-4(o)

A different approach was then taken based on the idea of simile. If data is to be taken as material, then one way to generate images is to start with words related to materials. The prompt "List some verbs that reflect what you can do with material" came back with the list: "Sew, Trim, Iron, Paint, Drill, Glue, Hammer, Bind, Braid, Knit, Weave, Crochet, Assemble, Fabricate, Construct, Mold, Weld, Carve, Burnish, Cut, Sand, Polish, Measure, Forge". Each of these verbs

were turned into gerund phrases and used as prompts. *Assembling data* yielded 4(a), *constructing data* yielded 4(b)-4(g), and *Fabricating data* yielded 4 (t)-4(u). *Drilling Data* yielded 4(p)-(s): these are some of the more evocative images generated, as they are offshore platforms that extract data rather than oil from below the ocean.

Folding data yielded 4(v). The image suggested a building, so an author remixed the image with the edited prompt "Folding data as a building" which transformed the image into 5(w). This is not unusual in the practice of image generation from models: an image is returned that resembles something else. The resemblance can be reinforced by editing the prompt. The same thing happened with weaving data: 5 (s), (x) and (y) were vaguely architectural, and the prompt was changed to *weaving data as a building*, yielding 5(t)-5(w).

Another approach also exploited the simile. The AI model was asked to generate names of different materials and responded with

1. Wood 2. Metal 3. Glass 4. Ceramics 5. Plastic 6. Fabrics 7. Stone 8. Rubber 9. Leather 10. Concrete

The model was then asked: "Explain how data is like wood" to which it replied:

Data is like wood in that it can be cut, molded, and reshaped to create something new or to fit certain needs. Just like wood, data needs to be organized and structured in order to be useful. Data can be chopped up and repurposed to create more complex solutions, just like wood can be used to craft furniture and other items from raw materials. Both data and wood need to be managed carefully and securely for the best results.

The response is thought-provoking because it emphasizes construction as a common theme. The prompt "data as wood" yielded 4(m) and "data as cement" yielded 4(h).

3 CONCLUDING THOUGHTS AND IMAGES

Data as material suggests we can do with data what we can with material: assemble, construct, drill, hammer, fabricate, glue, fold, and weave. Indeed, these all can be done with data, and the results can be easily illustrated. Material like wood, cement and fabric exist in a three-dimensional space. Fabric starts off two-dimensional, but with folding it can encompass three-dimensional structures. Data, though, starts off with far more dimensions. Each column in table or each node on a graph can be thought of as a dimension. The dimensions are rarely orthogonal, but instead have complex dependencies. Understanding this is the usually the provenance of data science and related disciplines. But language and text models together allow for fluid exploration. It is as if we have a materials factory, passing raw material to the first machine, then inspecting and modifying the more refined result before sending to another machine which may transform the material, then inspecting, modifying, or accepting the final product. Data, in the forms of text and image, are the materials that undergo transformation. Figure 5 provides the concluding thoughts in the form of images by enlarging two of the images from Figure 4. Together they illustrate metaphorically the processes of drilling down into and then building from data.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grants 2128906, 2129020, 2202578, and 1909803.

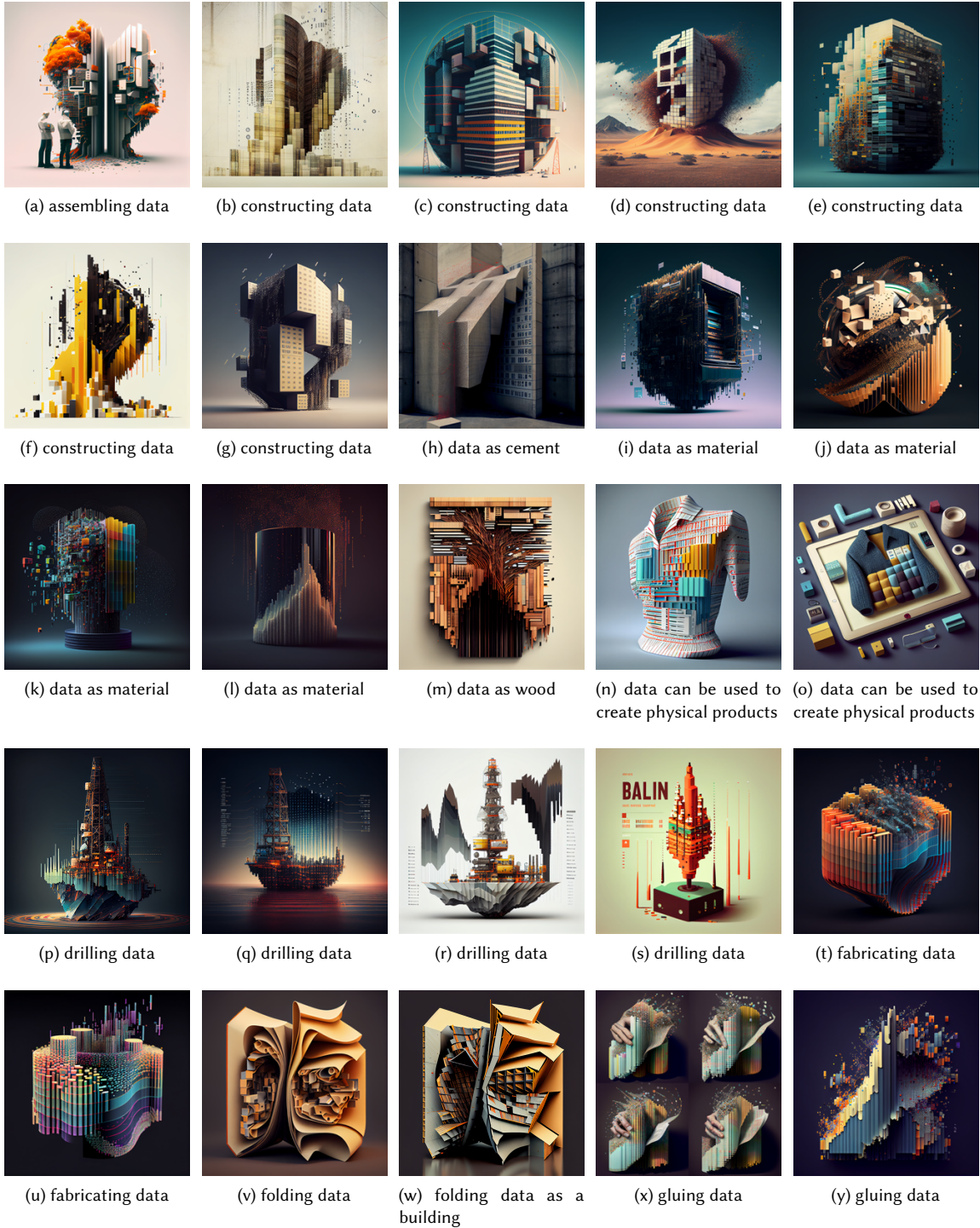


Fig. 3. Images generated from caption text

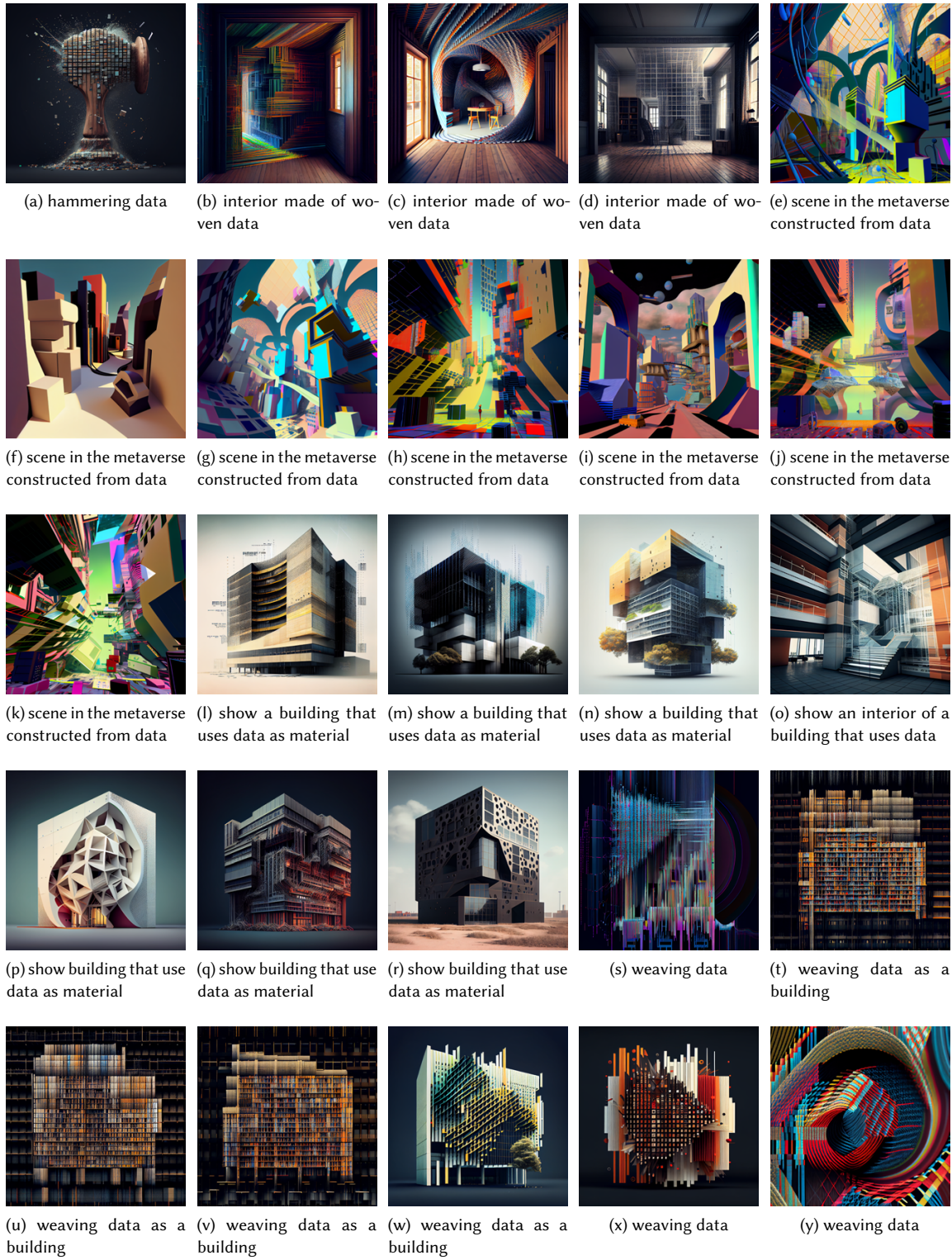
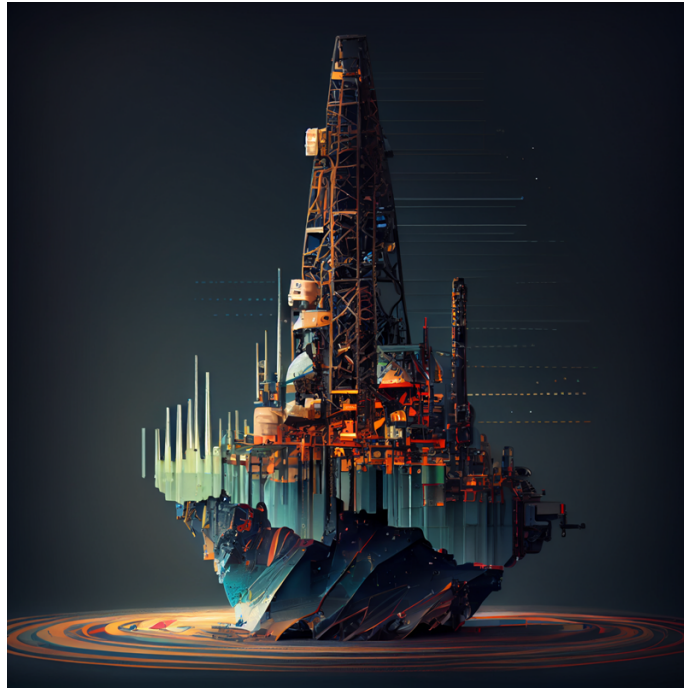
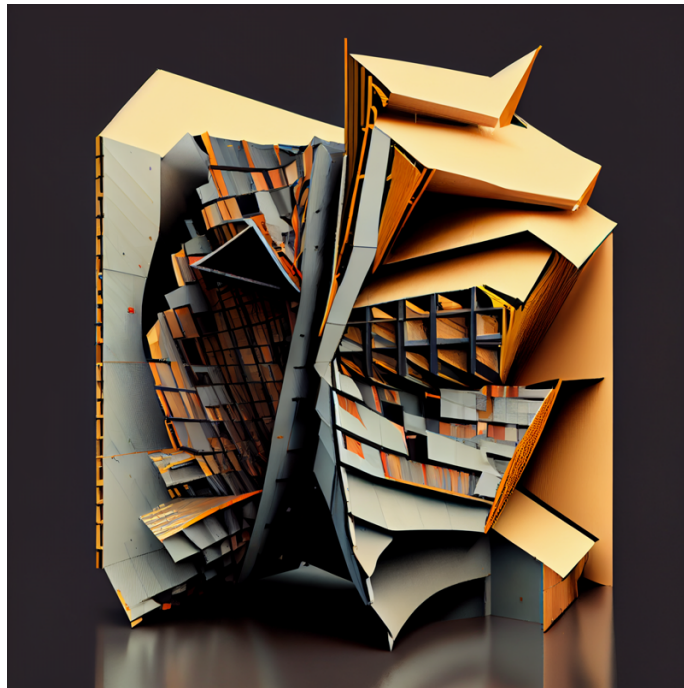


Fig. 4. More images generated from caption text



(a) drilling data



(b) folding data as a building

Fig. 5. Concluding images

REFERENCES

- [1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>
- [2] David Holz. 2022. Midjourney. <https://midjourney.com/> <https://midjourney.com/>.
- [3] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (UIST '22). Association for Computing Machinery, New York, NY, USA, Article 73, 17 pages. <https://doi.org/10.1145/3526113.3545621>
- [4] Kalle Lyytinen, Jeffrey V Nickerson, and John L King. 2021. Metahuman systems = humans + machines that learn. *Journal of Information Technology* 36, 4 (2021), 427–445.
- [5] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Stan Jastrzebski, Keren Henderson, Jeffrey V. Nickerson, and Lydia B. Chilton. 2023. AngleKindling: Supporting Journalistic Angle Ideation with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.
- [6] Stefan Seidel, Nicholas Berente, Aron Lindberg, Kalle Lyytinen, and Jeffrey V Nickerson. 2018. Autonomous tools and design: a triple-loop approach to human-machine learning. *Commun. ACM* 62, 1 (2018), 50–57.
- [7] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2022. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* (feb 2022). <https://doi.org/10.1145/3511599>
- [8] Sitong Wang, Savvas Petridis, Taeahn Kwon, Xiaojuan Ma, and Lydia B. Chilton. 2023. PopBlends: Strategies for Conceptual Blending with Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.
- [9] Doris Zahner, Jeffrey V Nickerson, Barbara Tversky, James E Corter, and Jing Ma. 2010. A fix for fixation? Rerepresenting and abstracting as creative processes in the design of information systems. *AI EDAM* 24, 2 (2010), 231–244.