Check for updates

Large Language Objects:

The Design of Physical AI and Generative Experiences

Marcelo Coelho, Massachusetts Institute of Technology and Formlabs
Jean-Baptiste Labrune, Massachusetts Institute of Technology

Insights

- Large language objects (LLOs) are physical interfaces that extend the capabilities of large language models into the physical world.
- Due to their generative nature, LLOs present more fluid and adaptable functionalities; their behavior can be created and tailored for individual people and use cases; and interactions can progressively develop from simple to complex, better supporting both beginners and advanced users.

Large language models (LLMs) have shown an unprecedented ability to generate text, images, and code, surpassing in many ways our own human capabilities and promising to have a profound impact on design and creativity. While powerful, however, these new forms of intelligence remain largely ignorant of the world outside of natural language, lacking knowledge of our objects, bodies, and physical environments.

Large language objects (LLOs) are a new class of artifacts that extend the capabilities of LLMs into the physical world. Through general-purpose language understanding and generation, these design objects revisit and challenge traditional definitions of form and function, heralding the design of physical AI as a new frontier for design.

In this article, we describe a series of LLOs developed at the MIT Design Intelligence Lab (https://designintelligence.mit.edu/) that combine generative and discriminative models to reveal a host of new applications for AI, from new ways of experiencing music or creating and telling stories to new forms of play, communication, and creating physical forms. These LLOs provide a glimpse at a new kind of creative process that weaves together the capabilities of human and artificial intelligence from

the early stages of concept development through form-finding, fabrication, and interaction.

A NEW FORM OF INTELLIGENCE

LLMs are neural network architectures designed to encode and generate human language. Trained on massive amounts of text data, they learn the statistical patterns of natural language and how it is used by humans and, from simple input prompts, can output complex and semantically coherent text. Today's state-of-the art LLMs have billions of parameters, are trained on trillions of tokens, and require substantial computational resources and cloud infrastructure for both training and inference.

However, rapid developments in hardware architecture coupled to software strategies, such as model quantization, fine-tuning, and retrieval-augmented generation, have improved the quality of LLMs' output and are allowing them to run on increasingly affordable embedded computers. In addition, new prototyping platforms designed for parallelism and tensor operations, such as Google's Coral (https://coral.ai/products/dev-board-micro/) and Nvidia's Jetson Nano (https://bit.ly/4ckeXFw), are lowering the barrier of entry and deployment of embedded machine learning, similar to how Arduino and the Raspberry Pi made embedded computing accessible in the past. What were once powerful and disembodied large-scale models running on high-performance computing clusters will soon become embedded into every object around us, having a profound impact on how we interact with and experience the physical world.

Nonetheless, in spite of their sophistication, these emergent forms of intelligence still remain largely ignorant of the world outside of language, lacking real-time, contextual understanding of our physical surroundings, our bodily experiences, and our social relationships. They are rarely "situated," as defined by Lucy Suchman, since they don't interact in real time with their immediate environment [1]. In his 1990 paper "Elephants Don't Play Chess," Rodney Brooks introduced the notion that if we want robots to perform tasks in everyday settings shared with humans, they need to be based primarily on sensory-motor couplings with the environment [2]. Or, as Jacob Browning and Yann Lecun describe it: "The problem is the limited nature of language. Once we abandon old assumptions about the connection between thought and language, it becomes clear that these systems are doomed to a shallow understanding that will never approximate the full-bodied thinking we see in humans" [3].

One area in which LLMs seem to falter is their understanding of geometric and mathematical concepts. Ask GPT-4 to design a table and it will suspend a tabletop midair, disconnected from its legs. Correcting this issue requires several prompts and precise user instructions, making it less than an ideal replacement for CAD tools [4]. Another challenge is their lack of real-time and contextual knowledge. With an April 2023 training data cutoff, GPT-4 is incapable of telling us what the weather is like today in Cambridge.

LARGE LANGUAGE OBJECTS

LLOs are a new class of artifacts that extend the capabilities of large language models into the physical world. They act as physical interfaces for LLMs by providing physical affordances such as multimodal input, end effectors, user feedback, and, most importantly, world knowledge and real-time context. Inspired by situated robotics [2], they integrate natural language and generative feedback at their core, not only shaping their final functionality and behavior but also reaching into the design and fabrication tools from which they are originally created and built.

These design objects revisit and challenge classical definitions of form and function: They are designed and communicate with natural language; their interfaces are generated on demand; they are able to learn and improve over time; and, acting as agents, their behaviors can emerge from their relationship with other objects. In combination with other neural network models and more-conventional coding, LLOs reveal a host of new applications for AI, from new ways of experiencing music with friends or creating and telling stories to new forms of playing, communicating, and learning. Below, we introduce four examples of LLOs developed in the MIT course 4.043/4.044 Design Studio: Interaction Intelligence (https://architecture.mit.edu/news/exhibit-objects-ai).

Memeopoly. Developed by Quincy Kuang and Annie Dong, Memeopoly [5] is an LLM-powered board game that combines the strengths of tangible play with the dynamism and interactivity of an open world digital game. In this LLO, human and machine creativity are combined to create unforeseen storylines personalized to players' unique preferences. Pawns and board are generated in real time, cultivating a thrilling gaming experience that is contextual and always evolving, surpassing the static and generic narratives typically found in traditional board games.

Narratron. Developed by Yubo Zhao and Aria Xiying Bao, Narratron [6] (Figure 1) is a small device that combines a camera, a projector, and multiple neural network models to augment shadow puppetry with generative images and storytelling. A custom-trained classifier recognizes shadow puppets, such as a dog or a rabbit, and prompts GPT-4 and DALL-E to generate a story and a sequence of image backdrops to accompany the user-generated shadows. By turning a small hand crank, reminiscent of a traditional film camera's, users can play and augment their stories in collaboration with an ensemble of AI models.

What were once powerful and disembodied large-scale models running on high-performance computing clusters will soon become embedded into every object around us.

VBox. Developed by Danning Liang and Artem Laptiev, VBox (Figure 2) is a boom box that takes users on a journey into music and culture through the lens of language. Using GPT-4's knowledge of how music is perceived and described through text, VBox presents users with a series of words based on a song's emotion, meaning, or cultural background. By selecting a word, users navigate to new songs following a thread that maintains a musical continuity, not through genre or a particular artist, but rather through the ways in which we describe music and encode culture into LLMs.

AIncense. Developed by Youtian Duan and Kai Zhang, AIncense (Figure 3) is a praying device powered by an LLM and designed to amplify self-motivation through ritual and psychological suggestion. Inspired by the physical and experiential rituals that surround the burning of incense, when offered a prayer, this LLO provides users with voice-based guidance and comfort, acting as an ambient mediation between humans and their complex desires.

DESIGNING INTELLIGENCE

LLOs provide a glimpse at a new kind of creative process that weaves together the capabilities of both human and artificial intelligences. Within the MIT Department of Architecture and the Morningside Academy for Design, our students have explored what this practice looks like by engaging in multiple modes of work. Alongside classic design activities such as casting and woodworking, they treat machine intelligence as a new kind of material, designing LLOs by prompting, constructing new datasets, and training and fine-tuning models. Learning progresses from "chisel to neural networks" by intertwining traditional modes of making with new techniques and prototyping strategies, such as neurometric design, generative experience, and networks of neural networks (NNN).

Neurometric design. Natural language is perhaps the most profound transformation brought forth by LLMs. With only a few words, anyone can generate a professional-quality illustration or photograph using a diffusion model. General-purpose language interfaces are powerful for two reasons: They enable an expressive

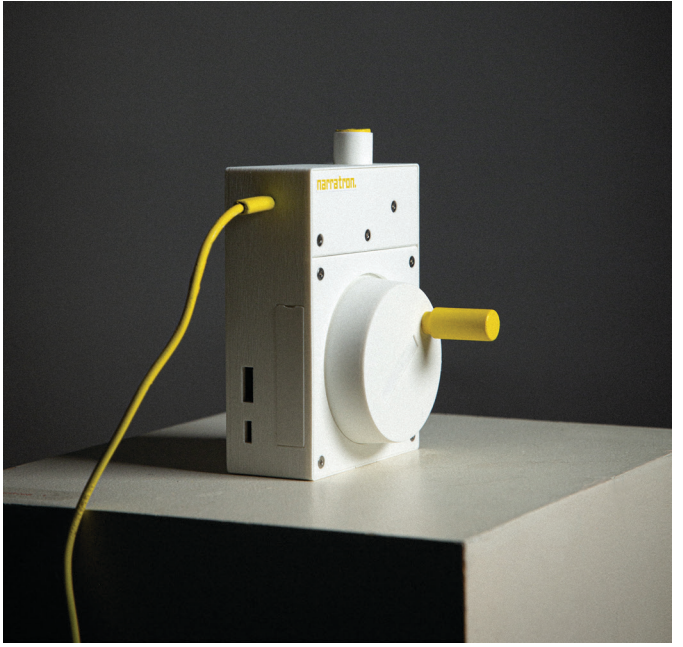


Figure 1. Narratron is a physical interface for storytelling with shadow puppets.



Figure 2. VBox is a radio that uses natural language for music navigation.



Figure 3. AIncense uses an LLM to provide users with voice-based guidance and comfort.

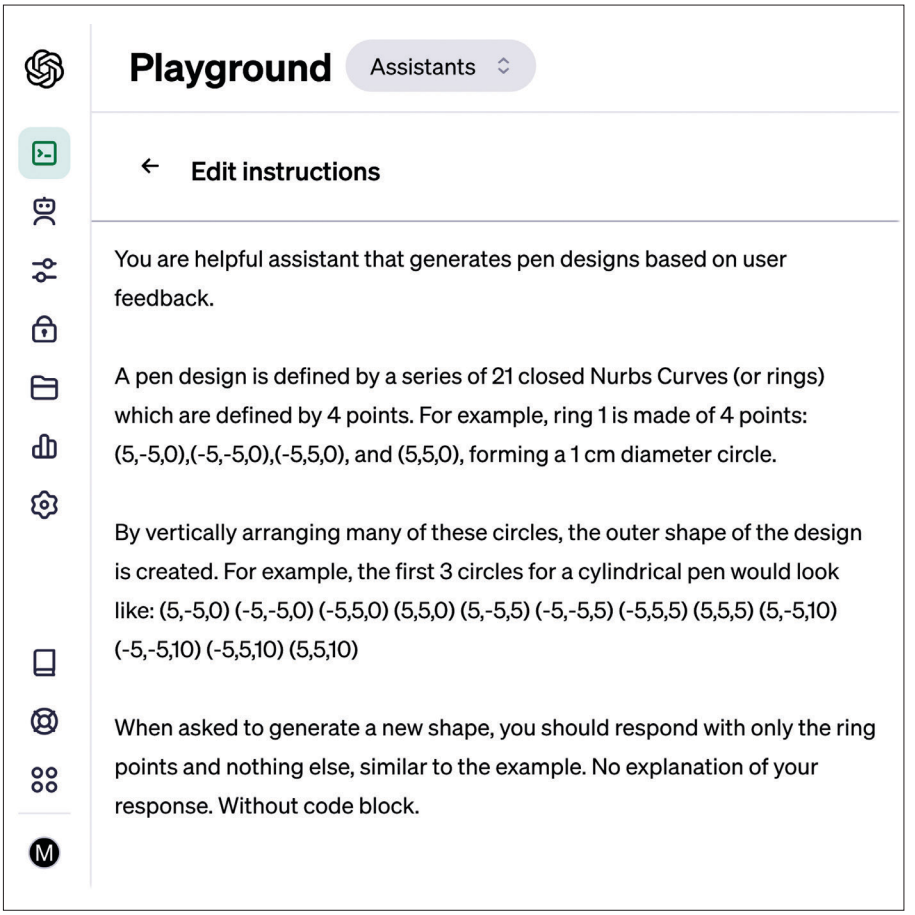


Figure 4. Neurometric instructions for a GPT-4 Assistant.

and rich form of communication with computers while dramatically reducing an interface's learning curve; and they provide a lingua franca through which we can talk to things, things can talk back, and things can talk to each other.

In a world where computers speak, prompting becomes a universal tool for generating concepts, materials, form, and behavior. Prompts help capture the complexity of human language and culture while abstracting away the complexity of bits and registers. Rather than programming a lamp to output specific RGB values, we can ask an LLO for something more poetic, such as the "color of the sky on a beautiful sunny day in Cape Cod," and let it generate a statistically significant hex color value. This can make computation more expressive, richer, and fundamentally more accessible, as we can interact with the world of bits using more descriptive and contextual language, grounded in our physical reality and embodied experience.

Neurometric design is a technique we have been exploring for the manipulation of parametric models using natural language, while retaining its function and manufacturability. We start by initializing an AI assistant with the description of a parametric design, providing examples of its parameters, and describing what they do. Figure 4 is an example of a pen generator that controls 252 individual parameters.

Rather than programming a lamp to output specific RGB values, we can ask an LLO for something more poetic, such as the "color of the sky on a beautiful sunny day in Cape Cod."

We connect the LLM inference to a traditional parametric CAD tool, such as Rhino's Grasshopper (https://www.grasshopper3d.com/) or FullControl GCode Designer (https://fullcontrolgcode.com/), which resolves a set of generative parameters into a solid geometry and CAM data (Figure 5). In situations where the generated coordinates interfere with other functional geometry such as the pen cartridge or 3D printability requirements, we use the generated data to guide our model and improve future generations. This allows users to customize an object's design by using high-level object descriptions, such as "make the pen bigger" or "smooth out the edges," overcoming the challenges of having to individually control the 252 parameters that make up the pen's shape.

Generative experience and dynamic mental models. Another way in which LLOs differ from the design of traditional computational objects is through the design of mental models and how they define, guide, and constrain user experience. A mental model is a cognitive construct that users form in their minds that represents how an object or system should work or behave. Through myriad aesthetic, formal, material, and behavioral affordances, designers create mental models that allow a user to understand and predict what will happen when a button is pressed or a particular action is taken.

Due to the generative nature of LLMs, rather than permanently encoding mental models into objects, LLOs can generate them on demand, in real time, and in collaboration with users. This approach expands upon some of the limitations of existing paradigms, such as tangible bits and radical atoms [7], by expressing computational abstractions through physical forms and giving them the ability to respond and adapt to unforeseen user behavior or changes in their environment. In contrast, drawing from a vast parameter space, the behavior of an LLO is fully "designed" upon user interaction, not a priori. These new forms of generative experiences (GX; Figure 6) provide unique opportunities: Objects can present more fluid and adaptable functionalities; their behavior can be generated and tailored for individual

people and use cases; and interactions can progressively develop from simple to complex, better supporting both beginners and advanced users.

Generative design is nothing new. L-systems developed in the 1960s are a classic example of how a set of parameters and a formal grammar can lead to complex and unexpected forms, such as those of trees and flowers. Transformer models are unique in that their generative grammar is grounded on human culture and they offer a practically infinite large parameter space from which to draw. GX design shifts the focus from creating robust mental models to developing robust forms of human communication and collaboration, and directing inference with the right set of physical and contextual constraints.

Generative experiences also exist within a continuum. At one end is Narratron, where the physical interface—what each knob, button, and projector do—is permanently defined, yet the images and stories are generated on demand. And at the other end is Memeopoly, where the board and gameplay are primarily defined by an LLM, which constructs a physical interface and mental model for users in real time based on their input and feedback.

NNN: Networks of neural networks. To extend LLMs beyond the realm of language, LLOs can also draw from traditional coding methodologies and a host of other neural network models and architectures. This integration serves a dual purpose: It directs, refines, and validates the output of LLMs, and it anchors their capabilities in physical reality.

NNN is a visual programming language we have developed for designing with networks of neural networks. Based on a node-based graphical user interface, NNN abstracts away some of the complexities of machine learning by allowing users to visually train their own models, chain several models together, and scaffold them into increasingly complex behaviors and interactions.

Developed in collaboration with Philips Lighting, the example shown in Figure 7 illustrates how NNN makes it possible to create rich and complex interactions. In this scenario, the application combines the following: PoseNet, a neural network model that detects key body points in human figures; a secondary clustering algorithm that extracts predominant colors from an image; and a custom-trained convolutional neural network regressor that interpolates between colors. Combined, they control the behavior of an array of lights based on the movement of a user's body and visual cues from their environment.

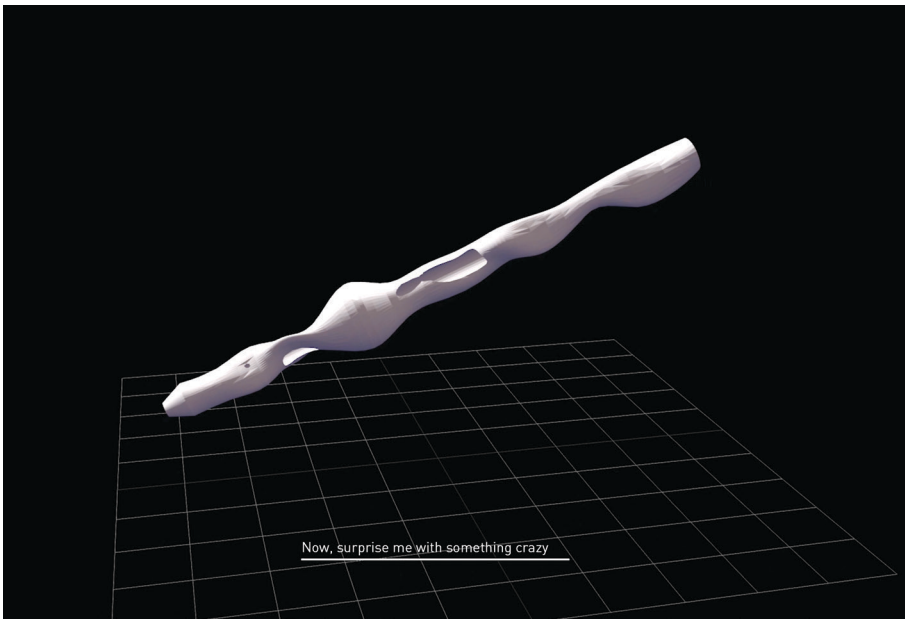


Figure 5. Graphical interface for a neurometric pen generator.

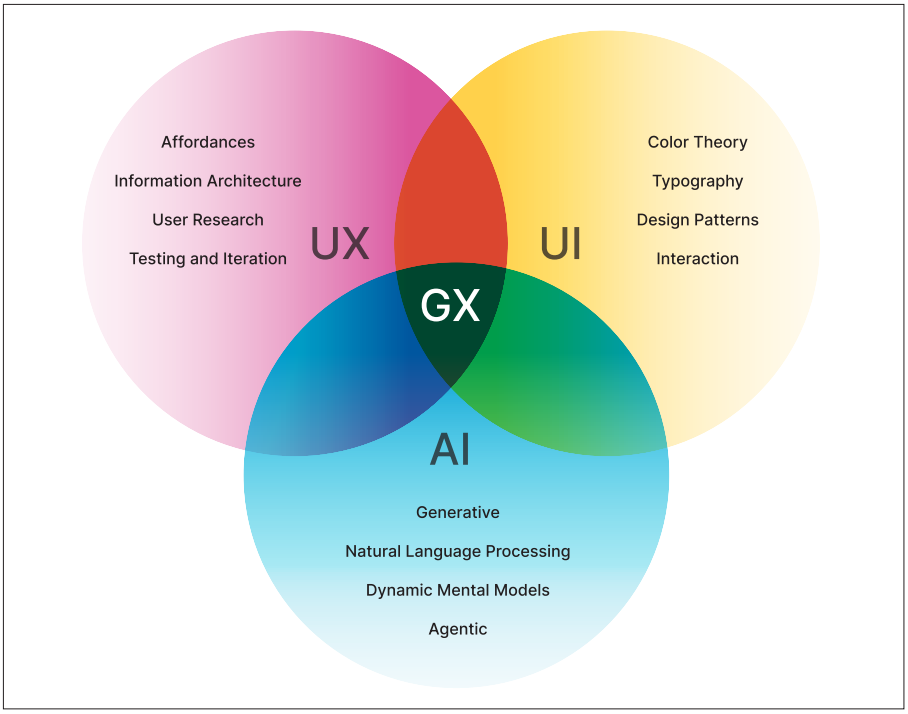


Figure 6. Generative experience (GX) resides at the intersection between UI, UX, and AI.

GENERATIVE CREATIVITY

The dawn of generative AI has profound implications for the future of creativity and industry. Coupled with digital manufacturing and the need to address new user requirements, and to mitigate environmental impact, traditional industries will be required to reimagine their product life cycle, from what are commonly distinct and sequential steps (e.g., R&D, manufacturing, marketing, servicing, recycling, etc.) into progressively flatter, highly interconnected, and continuous cycles. Driven by data, LLOs will support hyper-personalization and market

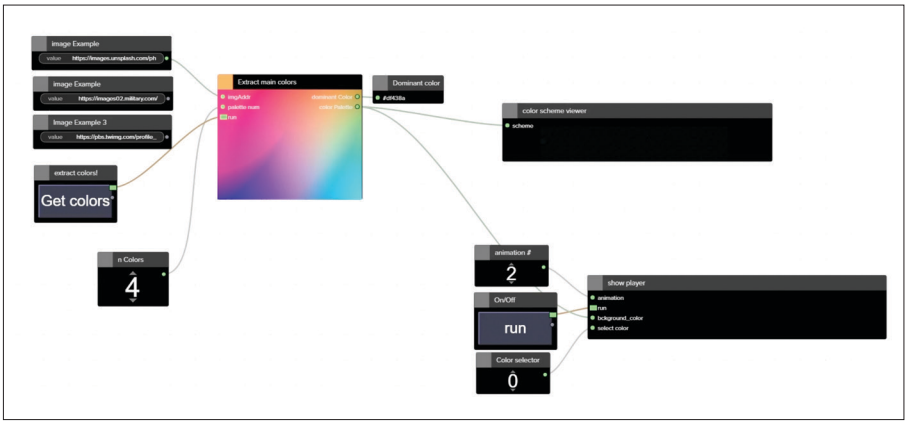


Figure 7. NNN interface for controlling an LED array based on an input image.

segmentation, promote the adoption of new material and fabrication techniques, and enable new ways for people to work together, communicate, and collaborate.

Automotive. Confronted with the challenges of rapidly retooling vehicles to support a sustainable and more dynamic mobility network, the automotive industry stands to greatly benefit from a generative product life cycle. Massive amounts of real-time data from roads and vehicles can backpropagate into improving quality control and manufacturing yields, and reshape vehicle designs to support a broader multitude of use scenarios, richer driving experiences, and personalized aesthetic expressions. The design tools that will enable these new modular mobility platforms and deep personalization have yet to be invented.

Luxury. At the core of luxury is the drive to be unique, to push creativity, and to stand out from the crowd. With the democratization of design brought forth by generative AI and digital manufacturing, every instance of an object can be inherently unique and personalized, propelling the luxury industry toward further differentiation and into ultra-high-end offerings and customer segments. Design tools trained on rich material experiences and the subtleties of craftsmanship can provide a new future of exclusivity for generative objects and experiences. These new LLOs will leverage dynamic and programmable materials [8], digital manufacturing, and new hybrid human-AI forms of making.

Film. Hollywood is perhaps the industry most affected by the advent of AI-driven creative tools. In recent years, LED volume walls driven by game engines and diffusion models have brought production and postproduction closer together by generating sets, characters, and even entire scenes in real time during shooting. In collaboration with industry partners, we are researching new generative and discriminative on-set AI assistants that can put directors, cinematographers, and gaffers at the center of this creative process. Rather than replacing traditional filmmaking with generative media, we are augmenting it with AI-embedded objects such as lighting, props, and cameras, and supporting new forms of human and AI collaborative storytelling.

CONCLUSION

As LLOs become increasingly commonplace, they will give emergence to a new kind of physical world. Through the combination of general-purpose language understanding and generation, hybrid neuro and symbolic computational approaches, physically situated knowledge, and humans in the loop, LLOs will pave the way to a new design discipline that starts with physical materials and scaffolds toward autonomous physical agents. For future designers, this integration of classic and cutting-edge methods heralds a new frontier: the design of physical intelligence.

ENDNOTES

1. Suchman, L.A. Plans and situated actions: The problem of human-machine communication. In *Learning in Doing: Social, Cognitive, and Computational Perspectives*. Cambridge Univ. Press, 1987; https://api.semanticscholar.org/CorpusID:11242510
2. Brooks, R.A. Elephants don't play chess. *Robotics and Autonomous Systems 6*, 1–2 (Jun. 1990), 3–15; https://doi.org/10.1016/s0921-8890(05)80025- 9
3. Browning, J. and Lecun, Y. AI and the limits of language. *Noema*. Aug. 23, 2022; https://www.noemamag.com/ai-and-the-limits-of-language
4. Makatura, L. et al. How can large language models help humans in design and manufacturing? arXiv:2307.14377 [cs.CL], 2023; https://doi.org/10.48550/arXiv.2307.14377
5. Kuang, Q., Shen, F., Fang, C.M., and Dong, A. Memeopoly: An AI-powered physical board game interface for tangible play and learning art and design. *Companion Proc. of the Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, 2023, 292–297; https://doi.org/10.1145/3573382.3616057
6. Zhao, Y. and Bao, X. Narratron: Collaborative writing and shadow-playing of children stories with large language models. *Adjunct Proc. of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, 2023, Article 119; https://doi.org/10.1145/3586182.3625120
7. Ishii, H., Lakatos, D., Bonanni, L., and Labrune, J-P. Radical atoms: Beyond tangible bits, toward transformable materials. *Interactions 19*, 1 (Jan.–Feb. 2012), 38–51; https://doi.org/10.1145/2065327.2065337
8. Coelho, M., Zigelbaum, J., and Kopin, J. Six-forty by four-eighty: The post-industrial design of computational materials. *Proc. of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, New York, 2011, 253–256; https://doi.org/10.1145/1935701.1935752

Marcelo Coelho is head of design at Formlabs and a faculty member in the MIT Department of Architecture. Spanning a wide range of media, processes, and scales, his work explores the boundaries between matter and information, seeking to create new forms of collaboration between human and machine intelligence.
→ marceloc@mit.edu

Jean-Baptiste Labrune is a designer and research affiliate at MIT exploring new materials that could be "programmed" as a mutual symbiosis between nature and humans. His works focus on the notion of exaptation, the way users of technologies reconfigure and hack them, producing original and unexpected functions and uses.
→ labrune@mit.edu

DOI: 10.1145/3672534 COPYRIGHT HELD BY AUTHORS. PUBLICATION RIGHTS LICENSED TO ACM. $15.00