Fake Attention

MICHAEL JEFFEERSON
Cornell University

Keywords: Artificial Intelligence, Machine Learning, Error, Methodology

Rather than taking a techno-positivist position on the digital, computation might be approached critically as a medium with the capacity to affect, manage, and disrupt. Within this context, architectural applications of machine learning might productively agitate the stability of our defined disciplinary conventions (particularly in relationship to methods of production motivated by typology). These moments of friction where conventions come into contact with external systems of AI technology are explored. This paper proposes to fold these tendencies back into the way we think about making architecture; back into our processes and pedagogies to develop a reciprocal and discursive relationship with technology. By leaving space to foster attention, the paper's mission is to develop skills that allow us to see the world anew and to become aware of the coded ways in which both technologies and the physical stuff of the world are motivated by hidden systems and to open up new frameworks for reconstructing our existing practices and conventions.

Increasingly, artificial intelligence surrounds us and mediates our experiences. Machine learning, a subset of AI, relies on complex algorithms to manage vast amounts of data often oriented toward optimization and efficiency: superimposing new systems onto our digital and built environments that affect our patterns of life. Often blackboxed, the processes of machine learning when applied to image-making frequently produce uncanny visuals that are familiar but rendered with distorted effects. These disruptions offer a glimpse into the patterns of machine learning and an opportunity for intervention. In one sense the errors that occur are merely awkward and seemingly harmless. Yet, these errors also bear relationships with malignant realities of the world. Because machines learn from the world around it, it is not surprising that systems that influence the world are then drawn into AI's machinations resulting in outputs that are inherently biased, sometimes in nefarious ways.

As opposed to focusing on the potential of machine learning as a tool for optimization, one may consider what is learned by slowing down our attention, instead focusing on the seeming errors of machine-generated "fakes" as sites of investigation. These moments of slippage reveal the means by which machine learning algorithms operate while simultaneously causing us to look at routine objects in a new light (for example, machine-generated images of faces, cats, shoes, etc.), thereby calling attention to the substrate of codes that define our cultural practices and built environment. By drawing these traces into the foreground we might better understand both the systems we take for granted and the protocols of the machine. This paper proposes to fold these tendencies back into the way we think about making architecture; back into our processes and pedagogies to develop a reciprocal and discursive relationship with technology.

MACHINES LEARNING FROM HUMANS

Microsoft's CaptionBot, launched in 2016 as an experimental AI, encouraged users to submit an image of which it would promptly attempt to describe in a caption. The AI fashioned itself (supposedly) in first person as "us[ing] Computer Vision and Natural Language to describe contents of images," noting "I am still learning. So sometimes I get things wrong."[1] And indeed more than sometimes, CaptionBot did get things wrong. In a collection titled "Not Really Confident", Ben Sisto created a compendium of images of abstract art that he inputted into the CaptionBot and documented its absurd results: an Ellsworth Kelly painting given the caption "I am not really confident, but I think it's a close up of a toothbrush" or a Raoul Ubac painting captioned "I am not really confident, but I think it's a cake made to look like a zebra."[2] While humorous, the erroneous results betray some fundamentals about AI and machine learning; namely that the datasets used for training AI bear a large impact on their evaluative capacity. Having learned well to classify toothbrushes and cakes, but without sufficient training on abstract painters, CaptionBot provided results that were both precise (one could imagine how the paintings might be interpreted as a toothbrush or a cake) and wrong (they are clearly paintings).

What CaptionBot reveals in an admittedly flippant gesture is that despite our (meaning humans') view of AI as an alien repository of applied knowledge, it is firmly rooted in the real world; and, therefore, subject to inheriting issues of systemic bias. Like many AI tools, from CaptionBot to text-to-image generators like DALL-E and Midjourney, their application is seemingly fun and playful while belying a much more serious set of nefarious potentials. To underscore this through example, Microsoft's precursor to CaptionBot, a chatbot named "Tay", only lasted a few hours before human users "taught" it how to be racist. Microsoft's claim that "the more you chat with Tay the smarter she gets," suggests that describing AI as "smart" should be qualified: AI is not a neutral intelligence but a conditioned progression of learned experiences upon which we bestow the term "intelligence" even as we assume machines to be impartial.[3] And, aside from the notion that "Tay" was anthropomorphized (gendered even), describing machines as learning has implicitly suggested that AI is sentient when it is not. This stems from linguistic confusion as machine learning (ML) has moved from the terrain of Computer Science into the mainstream. To learn is inherently suggestive of sentient beings like humans. But, as described by Meredith Broussard, "computer scientists know that machine 'learning' is more akin to a metaphor in this case: it means that the machine can improve at its programmed, routine, automated tasks. It doesn't mean that the machine acquires knowledge or wisdom or agency, despite what the term learning might imply."[4] Rather, the training of ML algorithms depends on robust and repetitive analyses of datasets (comprised of large quantities of content such as images of cats or social media profiles) aimed toward specific metrics of success (such as successfully identifying cats in an image or curating content for social media users).

AI's capacity to perform well on a particular task but not necessarily to behave wisely (or ethically) produces profound consequences. Broussard provides such an example when discussing the likelihood of survival for passengers aboard the Titanic based on how much they had paid for their tickets. According to and learning from the dataset of passengers and their demographic information, the trained machine learning model predicted a greater likelihood of death for lower-paying passengers. Not necessarily wrong, but extrapolated to the real world, one might imagine how various insurance rates could then be applied to passengers paying higher or lower ticket prices. If you paid less for a ticket, the data indicates you are more likely to die, and consequently your insurance rates go up. This theoretical example supports real world price optimization efforts that are not abstract or theoretical but are actually practiced in the real world. [5] Again Broussard explains:

"Price optimization is used in industries from insurance to travel—and it often results in price discrimination. A 2017 analysis by ProPublica and Consumer Reports found that in California, Illinois, Texas, and Missouri, some major insurers charged people who lived in minority neighborhoods as much as 30 percent more than people who lived in other areas with similar accident costs.... In an unequal world, if we make pricing algorithms based on what the world looks like, women and poor and minority customers inevitably get charged more. Math people are often surprised by this; women and poor and minority people are not surprised by this."

> —Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World.*[6]

This is to say that AI is not inherently biased, racist, or even mischievous. But, because it learns from the real world, it upholds extant consequential and abusive systems so long as its metrics of success are undergirded by them. As problematic as AI might be in reinforcing our own biases, it nevertheless reveals them to us in new ways, calling on us to pay attention to both AI's protocols and our own ethical use of these tools.

HUMANS LEARNING FROM MACHINES

As AI and machine learning evolves and learns from the world, it leaves traces on our patterns of thinking and working. It situates itself into our architectural lives through software interfaces, the tools with which we fabricate, and the protocols by which we organize, share, and disseminate work. Through these mechanisms, it mediates our engagement with architecture in every facet despite that our field is consumed with the production of physical stuff (artifacts and buildings). That computation has in many ways become habitualized does not diminish the ways in which emerging technologies including AI and ML leave imprints on how we think about and make architecture. Rather, the introduction of AI technologies into our traditional modes of production implant new ways of working and thinking that retool established norms and methods we have inherited in the discipline of architecture. These shifts are fundamental in ways that are both radical and routine.

In deceptively simple ways the effects of computation on our patterns of thinking and working serve as a profound epistemological shift. Take Mario Carpo's discussion of Google's Search Don't Sort campaign for Gmail in the mid-2000's in which he suggests how technology and big data fundamentally dismantle traditional taxonomic systems. The introduction of Gmail initiated an approach to the organization of information that radically redeployed our engagement with the computer.[7] The sorting of information into a computer's folders has roots in a Linnaean taxonomic approach by which information is grouped and sorted hierarchically and carried over into the

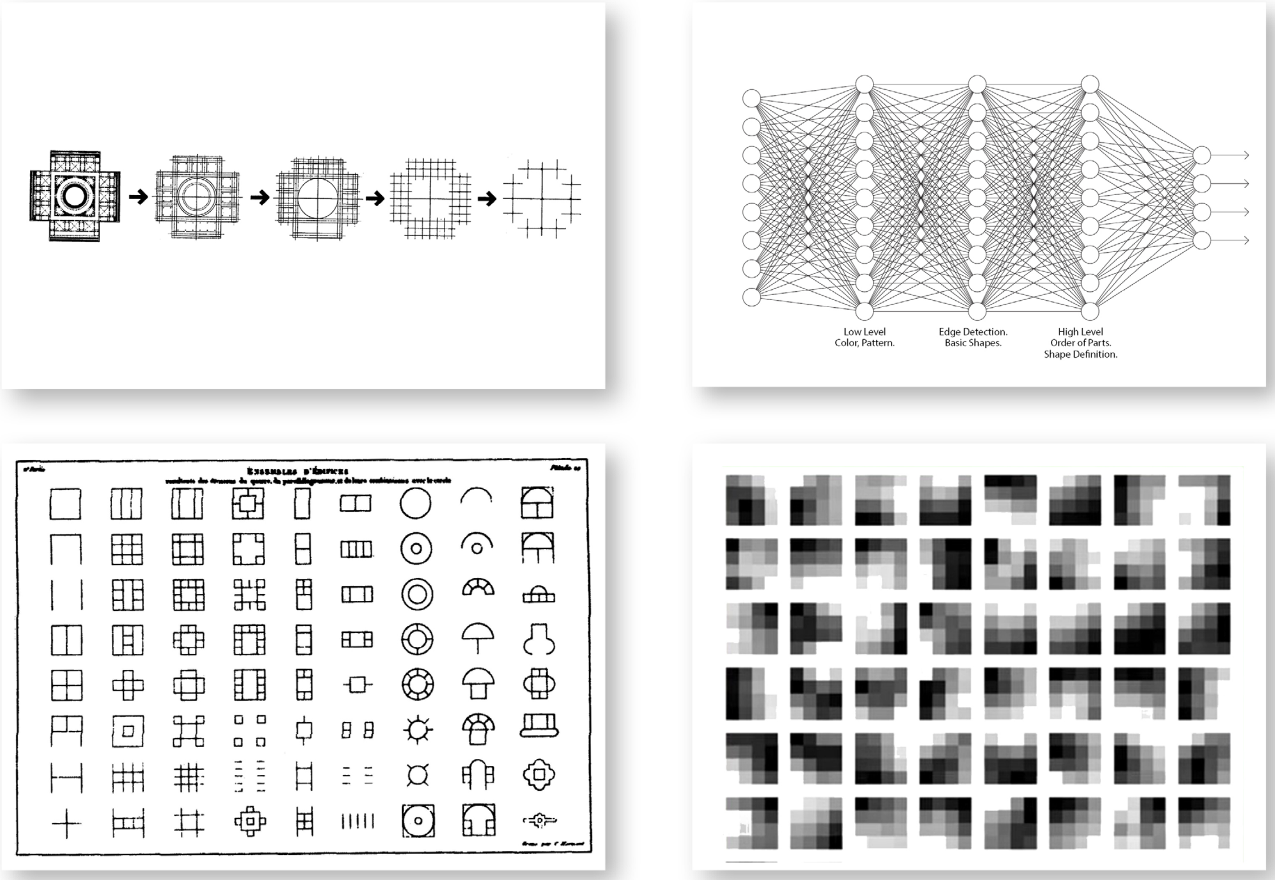

Figure 1. J.N.L. Durand methods of abstraction (left column) and Neural Network processes of classification (right column).

way architects such as Blondel, Buffon, and Le Roy among others introduced typological classification systems into architecture. Yet these protocols are not necessary to contemporary computation, which can process vast data with ease without the patterns of organization that we find necessary. Google's encouragement to "search, don't sort," is a recognition of computation's prowess and instrumentalized a new and habitual pattern to file management in which taxonomic file organizations are no longer necessary. By locating information with ease, our accrued systems for accumulating, organizing and designed methods of retrieval are rendered obsolete. We once took pride in clearing and organizing email inboxes; today these "folders" have been replaced by an endless scroll of searchable data.

In this way the modus operandi of the machine has been imprinted onto how we think and act in the world: often occurring at radically different scales. Philipe Schaerer's Diary, a tableau of over 225,000 data entries, suggests how the intimacy with which we document, recall our experiences, and mark time are reconfigured by technology.[8] The series visualizes the organization of images according to date as the vast array that we are familiar with in our engagement with smartphones; the scaleless pinching and pulling not only of images but images that represent the passage of time. In Schaerer's case, these images comprise and demarcate over a decade of his life.

DEEP FAKES AND DEEP STRUCTURES

AI and ML models have become increasingly convincing and proficient at producing images; at times generating "deep fakes" that are impossible to distinguish from images of real things. Yet, frequently there are traces in their products that give away their computational origin. StyleGAN, a deep learning model prominently known for its capacity to generate deep fake images of people who do not in reality exist is one such example of the moments where the falseness of the image reveals itself.[9] Amidst the seamless blending of faces into one another, intermittent flashes occur where objects appear on faces or in hair, earrings mismatch, or growths with facial qualities appear in the background.[10] These errors are dead giveaways of deep fakes. But, moreover, they are significant reveals of the black-boxed protocols of ML models and an example of how "failure is essential to understanding the nature of norms."[11] In place of applying analytical attention to erroneous images, this process might instead be inverted. As suggested by Andrew Atwood "... one way to test the stability and strength of any convention is to

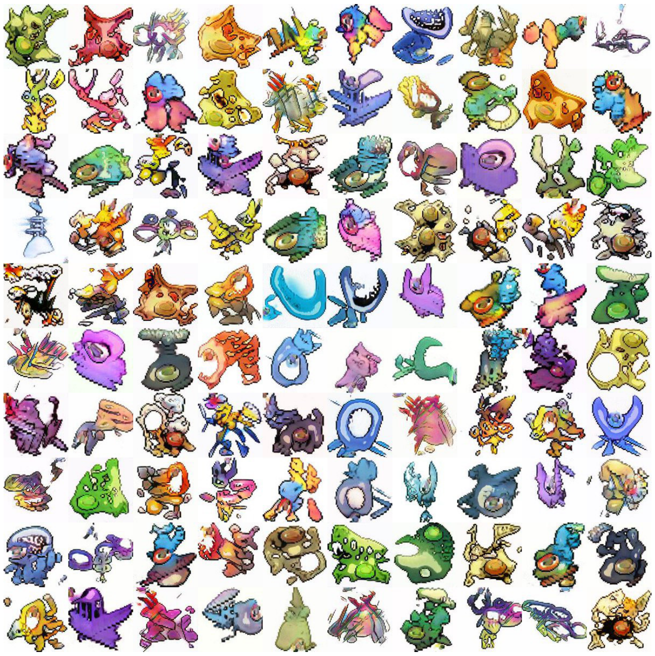

Figure 2. AI-generated Pokemon using StyleGAN. Michael Friesen.

agitate or rearrange its order."[12] Machine learning models offer us such an opportunity. By examining disciplinary methods of classification and typological production through the lens of neural networks, we might locate moments of friction where architectural conventions come into contact with technologies such as machine learning models. How might the processes of generating architecture and deep fake images compare with one another? How might the tendencies of each fold back on to one another and create a discursive relationship?

Traditionally in the discipline of architecture buildings are classified through analytical sequences of abstraction. Among the most didactic methods of classification and production—and thus selected as a foil—is J.N.L. Durand's work presented in his Recueil et parallèle des édifices de tout genre, anciens et modernes and Précis des leçons d'architecture. Durand's approach to classification, stemming from the Enlightenment Era, importantly profited from a systematized process by which like types were superimposed to extract similarities and eliminate differences to arrive at basic root forms, or deep structures; the legible set of geometric diagrams for which Durand is most commonly known today. If architecture's classification efforts are aimed toward producing diagrammatic clarity, the processes of classification in neural networks produce just the opposite. Hidden within complex arrangements of algorithms performing blackboxed tasks, the sequences of abstraction of machine learning models locate patterns in visual data that are indecipherable for humans. If traditional forms of abstraction in architectural analysis yield a singular root form to identify type, machine learning in contrast identifies granular features architects would find irrelevant: for instance, the accumulation of certain line segment features construct the digit "2." (Figure 1) Opaquely, the machine accumulates its own series of abstractions through low level recursive processes (akin to basic Photoshop contrast filters) to analyze and distill these features that together allow an image to be classified. Each process within the neural network results in an output image upon which other image processes are performed. Each sequence of operations then successively lowers the level resolution of an image (e.g. an original input image with a pixel resolution of 28x28 would ultimately distill to a 4x4 group of pixels). Over the course of the network the machine learning model becomes trained to evaluate patterns impossible for a human to detect by breaking the digits down into individual characteristics that comprise their figuration: for instance the individual curves, semi-circles, and straight lines that accumulate to form the digit "2" if it were graphically deconstructed.[13]

On one hand then, Durand's basic diagrams abstract architecture to its most basic geometric organizations, canceling out differences, and eliminating particular traits to foreground a typological organization. And on the other, the machine learning models detect a vast array of complex patterns in data that as an ensemble denote type. Crucially, in both cases, these methods of sorting and classifying content can be inverted as generative diagrams. For Durand, the internal diagrams can be employed to construct infinite variations that share the same abstracted deep structure. A machine learning model, though, does not operate within the context of a basic diagram: instead it replicates and accumulates patterns of traits that are assembled, hybridized, superimposed and genetically linked to one another. Take, for instance, an image of AI-generated Pokemon in which the AI's task is to mimic and a replicate the visual features that, in this case, could be interpreted as a Pokemon.[14] (Figure 2) In its efforts we see that it creates instances that look like the anime characters but are not. Parts are rearranged, some are all tails and legs without heads or torsos, some have no eyes, some are blobs. And yet, they are still comprised of the colors, textures, and knobby limbs by which we can recognize them as Pokemon. As humans we can hold two competing ideas in our minds: that these images are both Pokemon and they are not: they maintain their characteristics while missing the mark. They are all feature and no structure. This quality of AI production has been described aptly by Google AI researcher François Chollet: "Humans are largely unable to reproduce the visual likeness of something. But they know what the parts are (2 wheels + 2 pedals + handlebar + saddle). On the other hand, a [deep learning] model is excellent at reproducing local visual likeness (what it's fitted on), yet it has no understanding of the parts and their organization."[15] This sets up a discursive problem when processes for producing Deep Fakes are intermixed with methods of ascertaining Deep Structures.

To unpack this entwinement, we might consider the creative and original potential of machine learning models and methods of architectural production by charting them considering two

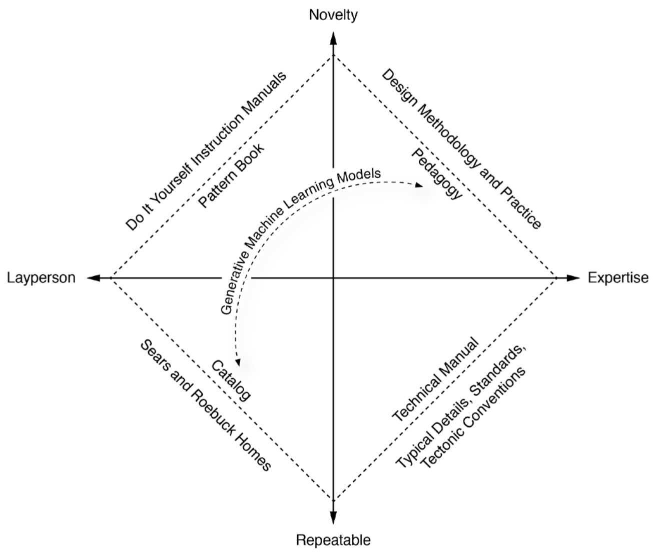
Novelty
Do It Yourself Instruction Manuals
Pattern Book
Design Methodology and Practice
Pedagogy
Generative Machine Learning Models
Layperson
Expertise
Sears and Roebuck Homes
Catalog
Technical Manual
Typical Details, Standards, Tectonic Conventions
Repeatable

Figure 3. Field of Originality. Image credit redacted.

spectrums: 1) one denoting originality that oscillates between repeatability and novelty (down to up respectively), and 2) and another denoting architectural prowess that fluctuates between the layperson and the expert (left to right respectively). (Figure 3) Durand's Precis and other pedagogical approaches are considered to be in the upper right quadrant, as they propose methodologies for generating architectural knowledge (achieving forms of conceptual expertise) and novelty (the methods also give license to innovate). Technical Manuals require expert knowledge but produce standards in the form of repeatable details and codes. Pattern Books and DIY manuals perform as mix and matched associations that produce degrees of novelty while being disseminated to the layperson. And catalogs such as the Sears and Roebuck homes are both repeatable and for the non-expert, a pick and play format of architecture assemblage. Introducing generative machine learning models to the mix, however, collapses the diagram as the AI combines the far edges of these extremes occupying a conflated space between catalog homes and design pedagogies. In training a machine learning model on a robust catalog of images, the algorithm is primed for the seemingly impossible: the repeated production of an infinitely novel set of outputs.

FALSE WALLS

False Walls is an exhibition of fake walls built with real materials aimed at testing an interlocutor relationship between deep fakes and deep structures. Tied up between the computational logics of artificial intelligence and the conventional tectonic arrangements of stud framing, the protocols of the machine are mapped on to traditional methods of construction. One part exhaustive and one part translative, the project featured the production of one machine learning model, over two thousand (digital) walls constructed with typical details, infinite machine-generated walls, and four physically built walls. In place of Durand's system for producing architecture, a conventional tectonic system has been employed (wood stud framed Walls, 12' in length by 8' tall). In place of Durand's strict rules by which a Deep Structure diagram informs the arrangements of elements and parts of building, the typical arrangement of its parts and materials in this assemblage system are embraced. They include: 16" on center spacing of studs, windows with sill plates and headers, openings by supported by king studs, trimmer studs, cripple studs, double top plates and sole plates; all built with dimensional lumber and 4' x 8' sheets of gypsum board.

Within these typical arrangements of materials, a set of additional rules are employed that articulate the particular ingredients for each wall type. Each wall is comprised of openings (windows are 18" x 18", doors 32" wide x 84" tall). Various relationships of the parts denote the color configurations of the walls such as numbers of window and door openings or number and placement of gypsum board panels. This serves as a secondary conceptual and representational apparatus that dictates typological variation as represented through color notation. At once there exist tectonic, conceptual, and representational underpinnings to these images with which the AI is asked to interact.

The machine learning model was fed the 2,000 images of color-coded digital walls and trained, yielding the capacity of the model to generate an infinite variety of new, fake walls. The interface[16] serves as a mediating environment that spatializes the results in a choose-your-own adventure typological experience. On one hand the infinite grid works within traditional methods of comparison in which the walls are infinitely registered situated and compared against one another. Difference occurs across and through the experience of scrolling. A corresponding digital model interprets machine learning software as a developed surface drawing and literalizes it as a virtual environment. The endless production of images sponsors an endless offsetting of rooms within rooms within rooms; the corresponding spatial experience to the infinite scroll. (Figure 4)

In the final set of built walls, the rules gleaned from the machine are folded back into the material assemblages of stud walls. The scale of attention shifts to a single room, the first room of the infinitely offsetting software space, and to the behaviors inherent in the machine protocols. Awareness of how and why the behaviors occur is important; it reveals the ways machines work that are different than ours. But the power of observation takes us only so far and might be irreconcilable. Rather, we might consider the ambiguous and frustrating territory that these images occupy and the qualities they engender that frustrate our norms (whether in reference to material tectonics or disciplinary methods or beyond). Michael Young's reflections on the nature of deep fakes underscores this:

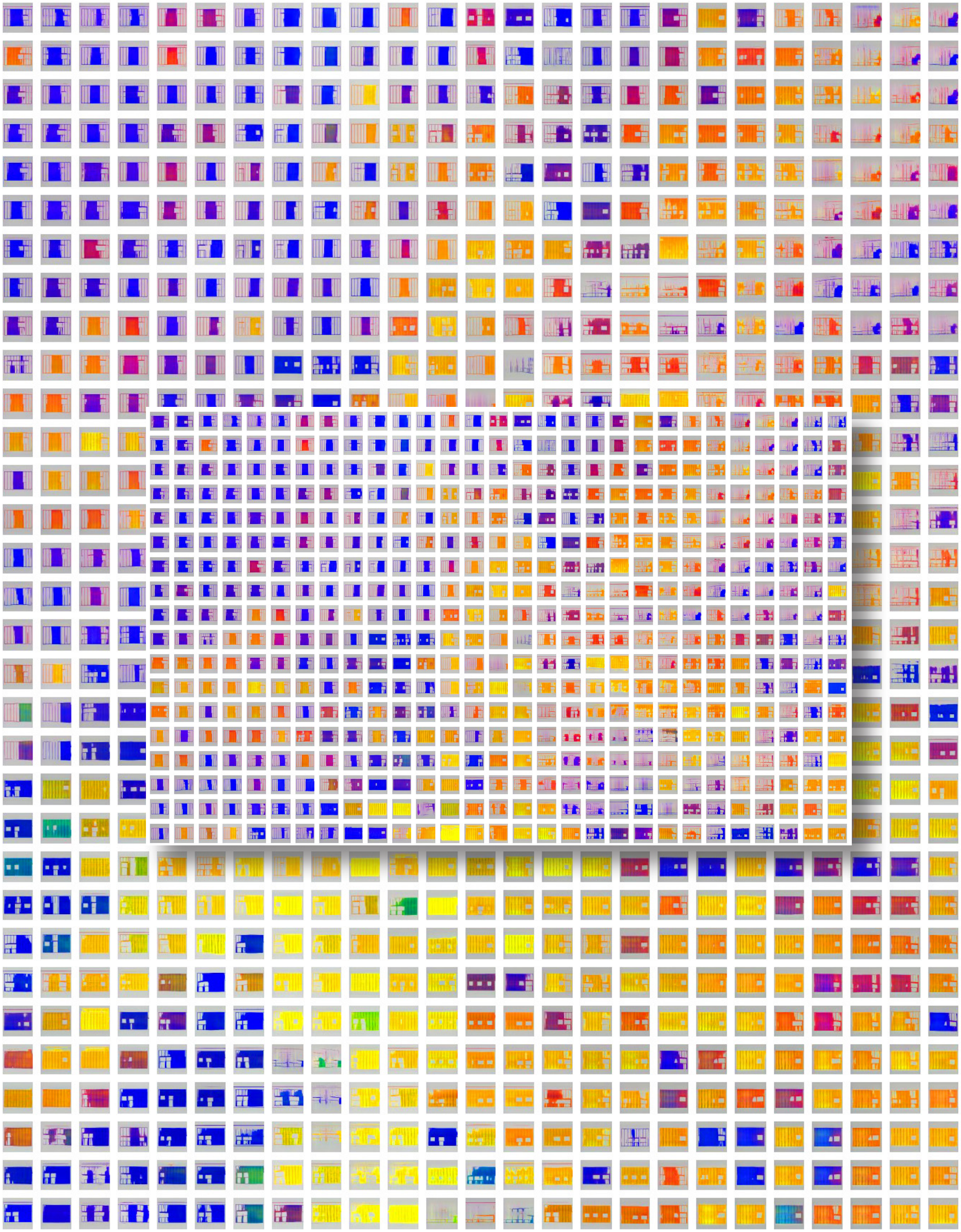

Figure 4. Experiential fields of machine-generated images: endless scroll (below image) and interpretation of ML software-scape as a spatial environment. Image credit redacted.

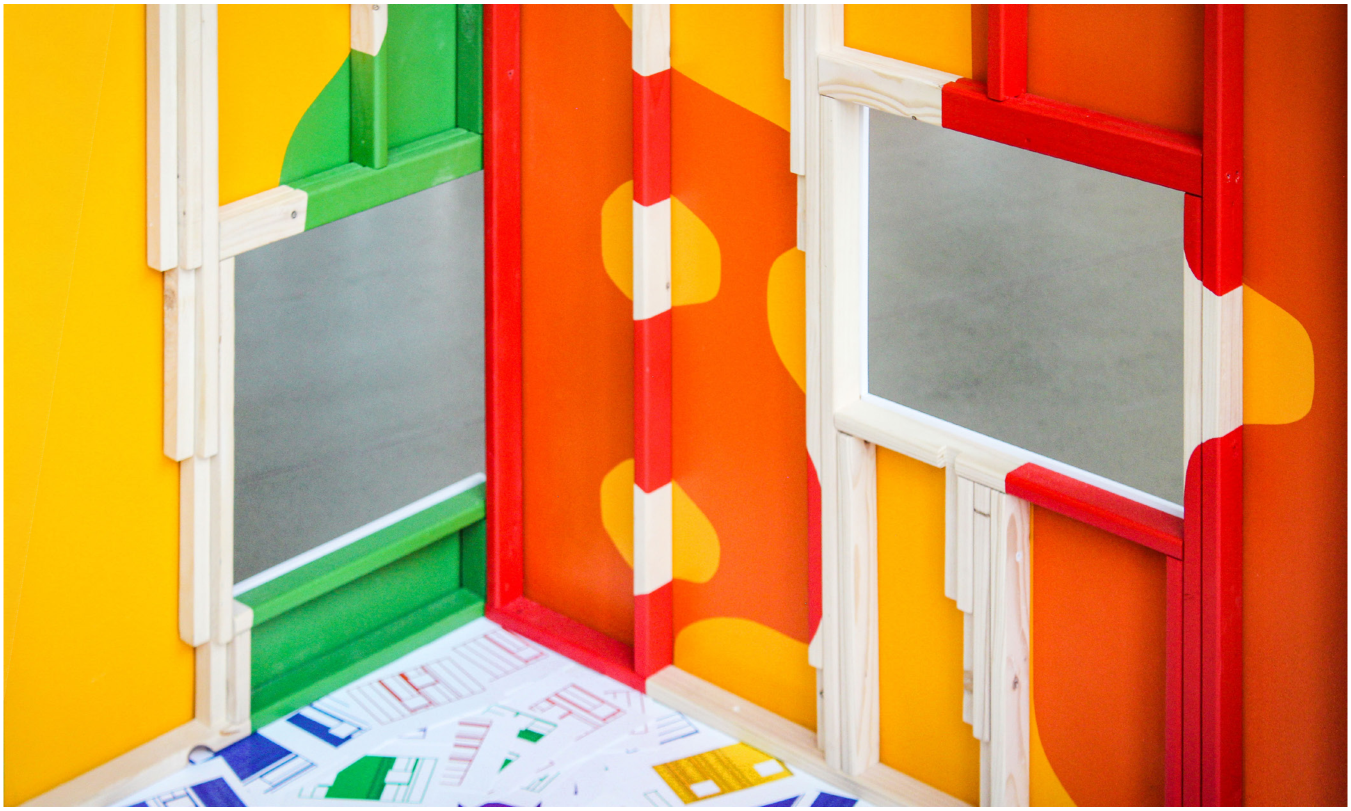

Figure 5. False Walls installation of fake walls built with real materials. Image credit redacted.

"What may be more important is to understand why we are interested in images that challenge our assumptions about the appearance of reality. There is an innate human desire to understand these thresholds, as Dave Hickey says, "pleasure in art derives less from knowing what we are looking at than from the anxiety of not-knowing just this." Whether this interest is based on fears about the precariousness of our relations to the environment, or from a desire to believe the world can look different from our expectations, is of less importance than identifying the qualities that emerge when relations become unstable."

—Michael Young, *Reality Modeled after Images. Architecture and Aesthetics after the Digital Image.*[17]

The instability that results when attempting to inflict machine behaviors onto material construction challenges the nature of conventions. The act of physical construction exacerbates the struggle with reconciling the inscrutable blurs and hybridizations of fake traits when interfacing with real materials. As with the variety of Pokemon species, the machine has learned particular and repeatable features of the walls, but they are manifested in ways that format indecipherable relations between parts and are tectonically uninterpretable. Learning how these traits are brought together requires new forms of close attention and critical analysis. For instance, the tendencies of curves to be generated (whereas the images used for training included no curvature) represent indices of the machine learning model's training in which 2x4 studs guide smooth transitions between emerging gypsum board panels. The use of color as notation is mixed up, presenting a psychedelic array of colors that are often present in a single wall though no such type exists in the training dataset. These aberrant manifestations suggest a hybridization and transferring of traits between walls and between types that betrays the machine learning model's propensity to collage and fragment like types into irreducable wholes.

The walls produced are both real and not. In building the thing itself, the machine learning model is inverted. Whereas the machine generated images are fake versions of existing walls, the installation doubles down by making real versions of fake images by incorporating the behaviors of the machine into materiality. The conflation of these behaviors and their working back and forth into materiality intertwines the walls so that they are impossibly mediated by both the digital and physical. The result sponsors new effects that could be described as legibly ambiguous. There are no blurs or gradients in the physical set of walls, but the readings of parts are suppressed and encouraged perceptually. The alignments and misalignments between representation and materiality confuse the hierarchies of the project and intertwine deeply rooted methods of producing architecture with the protocols of AI. To do so is to encourage a discursive project that embraces of the instability of the contemporary digital environment; one that reflexively responds to the rules of the machine, of material tectonics, and of architectural disciplinary methods that rely on typology as a framework.

ENDNOTES

1. May, Kyle, Julia van den Hout, Jacob Reidel, Archie Lee Coates, and Jeffrey Franklin. CLOG X Artificial Intelligence, 2018.
2. Ben Sisto. "Not Really Confident," 2016. https://www.bensisto.com/guts-not-really-confident.
3. The Guardian. "Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter," March 24, 2016. https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter.
4. Broussard, Meredith. Artificial Unintelligence. How Computers Misunderstand the World. Cambridge, MA: The MIT Press, 2018.
5. Ibid.
6. Ibid.
7. Carpo, Mario. The Second Digital Turn. Design Beyond Intelligence. Cambridge, MA: MIT Press, 2017.
8. Schaerer, Philipp . "Diary. 2005 - Ongoing." Philipp Schaerer. Accessed October 12, 2022. https://philippschaerer.ch/overview/diary-2005-ongoing/.
9. Hill, Kashmir , and Jeremy White. "Designed to Deceive: Do These People Look Real to You? (Published 2020)." Designed to Deceive: Do These People Look Real to You?, November 21, 2020. https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html.
10. This Person Does Not Exist. "This Person Does Not Exist." Accessed October 12, 2022. https://thispersondoesnotexist.com/.
11. Quote from Akeel Bigrami in May, Kyle, Julia van den Hout, Jacob Reidel, Archie Lee Coates, and Jeffrey Franklin. CLOG X Artificial Intelligence, 2018.
12. Atwood, Andrew. Not Interesting: On the Limits of Criticism in Architecture. ORO Applied Research + Design, 2018.
13. For an expanded explanation see: Pound, Mike . "Computerphile Inside a Neural Network." YouTube, June 30, 2016. https://www.youtube.com/watch?v=BFdMrDOx_CM.
14. Tangermann, Victor . "This AI Draws Horrifying New Pokémon." Futurism, 2019. https://futurism.com/the-byte/neural-network-pokemon.
15. Quote extracted from Kremer, Attay. "Is DALL-E a Genius?" e-flux, October 5, 2022. https://www.e-flux.com/notes/495428/is-dall-e-a-genius.
16. RunwayML. 2020. https://runwayml.com/.
17. Young, Michael. Reality Modeled after Images. Architecture and Aesthetics after the Digital Image. New York, NY: Routledge, 2021.