The Journal of Media Art Study and Theory
Volume 3, Issue 1, 2022
Automating Visuality
www.mast-journal.org
MAST

On the Truth Claims of Deepfakes: Indexing Images and Semantic Forensics

Rebecca Uliasz
Duke University

Abstract

When news media shared a video of outgoing president Donald Trump acknowledging the victory of president-elect Joe Biden, some social media users conspired that it was a deepfake, a synthetic image made with machine learning (ML) algorithms, despite evidence to the contrary. Employing this example in the following, I focus on how images generate veracity through the interrelated actions of human and machine learning (ML) algorithms. I argue that ML presents an opportunity to revisit the semiotic infrastructures of images as an approach towards asking how photorealistic images produce truth claims in ways that exceed the purely visual. Drawing from photographic theories of the image index and diagrammatic understandings of ML, I argue that *meaning*, described here as what images do in the world, is a product of negotiation between multiple technological processes and social registers, spanning data sets, engineering decisions, and human biases. Focusing on Generative Adversarial Networks (GANs), I analyze sociopolitical and scientific discourses around deepfakes to understand the ways in which ML affords hegemonic ways of seeing. I conclude that ML *operationalizes* the evidentiary power of images, generating new thresholds of visibility to manage uncertainty. My aim is to critically challenge post-truth paranoias by analyzing how ML algorithms come to have ethicopolitical agency in visual culture, with implications for how images are made to *matter* in post-truth media ecologies.

Keywords
deepfake; machine learning; generative adversarial network; truth claims; indexicality; diagram; post-truth

Introduction

> Whatever the truth of images in media might be, we will have to reckon with their radically contradictory reputation as "everything" and "nothing," the most valuable and powerful elements of the messages transmitted by media, or the most trivial, degraded, and worthless. (Mitchell 35)

When news media shared a video on January 7, 2021, of outgoing president Donald Trump denouncing Capitol protestors and acknowledging the victory of then president-elect Joe Biden, some social media users conspired that it was faked. Users across Facebook, Twitter, and Parler claimed Trump's "head doesn't seem to move . . . properly," citing "his cadence" and "the obvious face movements they're attempting to pass as real" as evidence that the video was a deepfake, despite that it was released by the White House and posted to Trump's own social media account (@ddale8; "Fact Check").

Deepfakes belong to the category of "synthetic media" which includes digital video, image, text, and audio created autonomously by or with the assistance of machine learning (ML) (Babu; Paris and Donovan).[1] Because they are believable in a visual sense, they contribute to public paranoia concerning our ability to perceive and understand digital images culturally (Chesney and Citron; Mirsky and Lee). As one Reddit user pointed out, in contemporary media

[1] They are typically traced to November 2017, when a Reddit user called "deepfakes" made a post containing machine-generated pornography of actresses Gal Gadot and Scarlett Johansson (Paris and Donovan).

ecologies “video evidence can no longer be taken as gospel; anything can be discounted as fake” (@Eric_Fapton). That this claim was made of the video of Trump’s address highlights existing tensions within media theories of the image, paradoxes that earn them the contradictory reputations as “everything” and “nothing.” Images appear to our minds and eyes as resemblances of things while at the same time only ever appearing in mediation, circulating through networks that shape their productions, distributions, and receptions. With the increasing role that ML plays in influencing human perception and knowledge, the visual accuracy of an image, understood as its resemblance to a thing in the world, may not testify to its connection to its referent in the traditional sense that a photorealistic image might. At stake with this example are questions of how ML participates in the production of the ambiguities that typify post-truth media ecologies, and how its participation factors into the shifting onto-epistemological status of images in visual culture today.

While the relationship of visual media to truth takes many discursive, abstract, and historical forms, this article narrows in on digital images that aim towards photorealism, mobilizing deepfakes to track the shifting function of photorealism in networked culture. One part of this inquiry is asking how imagistic indexicality is negotiated by ML. The pre-digital photographic index was understood by some, following Charles Sanders Peirce’s definition developed through his second trichotomy of sign types, to function by way of an existential connection between a sign and its object, a trace of the real that confirms its referentiality. Some have claimed that digitization severs the photograph’s indexical links because image information is converted into digital code, rendering the image “programmable” (Manovich). Others point out that the rhetoric of crisis is often evoked in regards to the digital, where the threat of the “death” of the index would usher in an era in which any photorealistic image can be forged, and thus doubted (Paulsen). It seems that further algorithmic developments in the field of image synthesis have only intensified such anxieties around imagistic veracity.

My aim is to critically challenge such anxieties through analyzing the ways in which ML *operationalizes* the generation of visual truth claims. The operational notion of machinic images is defined in non-representational discourses of the image (Farocki). The meaning of an operational image would be derived from the action it prompts in a technical system rather than a result of its "reality effects" (Barthes). Today, our visual paradigm would seem linked to the "algorithmic turn" (Uricchio), a condition in which images are produced through processes that perform statistical calculations on the vast quantities of image information contained in online data sets. Our opening example suggests that the algorithmic turn is linked to a qualitative shift in the way machinic operations intervene in very human parameters of visibility, with consequences for visual cultures and their epistemologies (Ananny and Crawford; Pasquinelli and Joler). The "truth" of the image, if it is to be found, is in this instance a product of negotiation between various technological infrastructures, social relations, and epistemic practices.

To locate the "epistemic agency" (Werner) of deepfakes, the *how* of how deepfakes make meaning, requires we move between the multiple semiotic registers through which they operate, which span technical and social scales. Deepfakes are complex epistemic things, linked to infrastructures composed of data, models, and algorithms that pattern machine behaviors and human knowledge practices. The issue at stake is not just one of deciphering what a deepfake represents but also what it does, which is related to how ML is used to claim meaning and the types of meaning it makes possible. The following inquiry into how deepfakes come to matter in the context of post-truth discourses is one that asks, in N. Katherine Hayles's words, how human-computer interactions "form semiotic relationships that exceed the limits of biological cognition alone" and to what end (Hayles 41). Its aims are twofold: first, to describe the ontology of synthetic images through the non-sensuous semiotic processes of ML, and second, to understand the significance of these imaging operations for the politics and epistemologies of "truthful" visual media today.

Indexing Images and Truth Claims

Deepfakes fill a certain cultural role as harbingers of post-truth politics. Their advent, argues Yves Citton, "strikes a fatal blow to the trust one could (in most cases) put in what looked like (and was presented as) indexical representations of reality" (47). In this case, the mere existence of deepfakes, and synthetically generated media more broadly, forces one into a state of trustlessness, where one suspects that behind every imagistic truth claim there lurks the possibility of a lie, and therefore the possibility of deceit.

Concerning photorealism, some have argued that the concept of the "death" of the index stems from a misdiagnosis of the image index as primarily a "trace" that would mark its material connection to the real. Such accounts point out that images are by virtue subject to processes of manipulation (Doane; Paulsen). Both digital images and traditional chemical photographs alike are "subject to elaborate procedures before a picture will result" (Gunning 40). Photographic images, rather than containing a direct "imprint" of a real object, are realized through lenses, film stock, and other material constraints. Instead of the material trace, some have argued that we understand digital images with regard to the *deixis,* the "pointing" function of the index (Paulsen). This particular type of indexicality focuses on the way a sign is mediated, how it "reaches out" to point to something that may not yet be known but is called to attention, such as how a pointing finger indexes something yet to be seen. If material traces are what gave optical photographs a trustworthy existential link, the pointing function of the index in general is to communicate something across time, "to reference a real without realism" (Doane 4). The index can be understood as the very *possibility* of communication through mediation.

As such, the stratum of the index as deixis is always somewhat ambiguous, open to interpretation and doubt. Still, others remark that despite this indeterminacy inherent to the deixis, images still somehow possess the power to "tell the truth." Rather than presenting contradiction, Tom Gunning notes that the ambiguity of the index is precisely what allows for things to be said about an image. For Gunning, digitization doesn't undermine the image's indexical qualities but shows us how truth claims are technically and socially produced by way of establishing iconic resemblances. For instance, Photoshop tools might give users unprecedented freedom to manipulate an image of a face. But for a portrait to function as an accurate image of its subject, it can only be transformed in a way that maintains this iconic relationship (Gunning 41). This quality of "visual accuracy" compels our belief in the image's referent while also giving it a mutable characteristic. In this sense, a fake image must maintain visual accuracy, a relation to a notion of "likeness" that is configured through technical or artistic manipulations. Put differently, a preformatted and agreed-upon sense of iconic similarity is required to produce accurate (described as "believable" in the case of deepfakes) fake images. As Gunning argues, "indexicality intertwines with iconicity" in our assessment of photorealistic images—the former establishes the possibility of interpretation while the latter grounds meaning through qualitative resemblance to a known object (Gunning 41).

Considered from the standpoint of photography theories of the image, deepfakes testify to the ongoing socio-technical value we place on visual accuracy which manifests in our continued investment in imagistic realism as truthful. As Gunning argues, the truth claim of photorealistic images can even be said to *depend* on the ways in which we value the visual domain to provide a "continuing sense of the relation between the photograph and pre-existing reality" (45). Nonetheless, there are technical differences between the icon-index function of photographs and the operations of ML. Deepfakes do away with the photographic image's previous relationship with an optical lens. Instead, they are given their form by statistically weighted models trained on data sets of pre-existing images. As we will explore, their indexicality is operationalized by algorithms, with consequences for how ML systems generate their own truth claims, producing images that take on prescriptive rather than representative qualities.

GANs and Diagrams

It is part of the "trick" of ML that we treat their outputs iconically, i.e., as pictures that communicate meaning through their qualitative resemblance to some object. As Leif Weatherby and Brian Justie explain, we fall in the trap of "naïve iconic interpretation" when we construe machine behavior as intelligence because a system expresses behavior we take to be *like* intelligence (383). In such instances, we draw upon metaphysically ungrounded assumptions about *what* intelligent behavior looks like.[2] That we socially and culturally ascribe verisimilitude to certain outputs produced by ML speaks to the bias inherent to iconic interpretation, but also raises questions about how the technical operations specific to ML factor into qualitative interpretations. Framed as a question, when ML intersects with the visual field, how does it change cultural and social assumptions of likeness? How is meaning, understood in the pragmatic sense as the action of a sign, intermediated between cognitive registers which span the technical and social? In Peirce's tripartite, the icon is the sign type "from which information may

[2] With concerns related to my own in this article, Weatherby and Justie attend to the ways in which deep neural networks treat icons (as images) indexically. They argue that the indexical pointing function *can* result in an iconic (imagistic) output, although the operations of the net are "indexical all the way through" (390).

be derived" by a cognitive interpreter (309). Arguing that the icon sign can function non-trivially, or by means other than visual similitude, Frederik Stjernfelt argues that the icon works in Peirce's semiotics to describe non-sensuous semiotic operations.[3] As Stjernfelt describes, a particular type of icon, which Peirce calls a diagram, is at stake whenever the "rational relations" between parts, which mirror the relations between the parts of the object they correspond to, can be manipulated to gain new information about the object. Diagrammatic manipulation involves the operationalization of similarity to yield "implicit possibilities involved in the icon" (72). In this sense, ML can be described as a particular *type* of icon, a diagram, which captures likenesses as the relationships between heterogeneous parts in vector space rather than by visual similitude (Mackenzie).

The interplay between the iconic and indexical elements of a GAN makes verisimilitude possible. A GAN (see fig. 1) contains a diagrammatic relation of nodes, layers, and thresholds in which indexical relationships are algorithmically calibrated to achieve optimal icon outputs, and icons are converted to indices to provide information about how this optimization process should occur. A GAN takes a defined training set of images (icons), recasts them as vectorial representations (indexes), and learns to output images (icons) that look similar enough to this training set to a human observer such that the human cannot decipher between the two—a sort of visual Turing Test (Mirsky and Lee). This means the output images, although synthetically produced, "pass" as real, in a conventional sense.

[3] Stjernfelt gives an *operational* concept of Peirce's account of diagrams as non-trivial icons as "moving pictures of thought." This concept works to separate the icon from psychologism: "The decisive test for its iconicity rests in whether it is possible to manipulate the sign so that new information as to its object appears" (90).

To describe this process in some technical detail, a GAN is comprised of two competing neural networks (NNs).[4] A generator (G) network takes N-dimensional random variables (noise) as input. G captures a probability distribution function P(x) of samples drawn from the data set and aims to output an image with the same distribution. A discriminator (D) takes an input and determines whether this input was drawn from the training set or, rather, from the generator. G is trained to maximize the error of D ("fool" it into taking fake images as part of the training set) while D aims to minimize this error—a zero-sum game. In this game, or training process, indexical pathways between nodes in hidden "deep" layers are optimized over time to find the ideal routes to maximize the possibility of producing the desired types of outputs. The result is that the output image tends toward the "likeness" of the input image set, even while it remains uniquely singular.

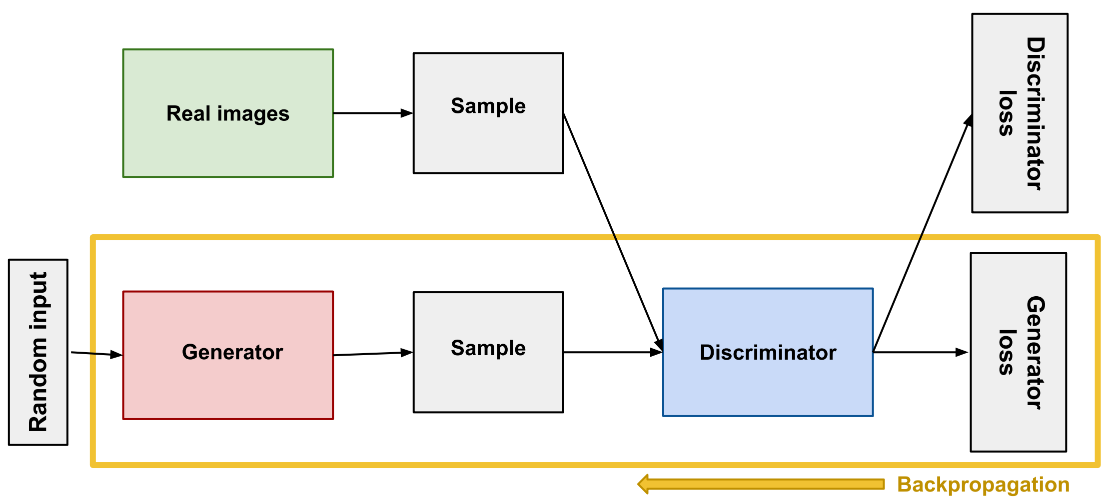

Fig. 1. "Backpropagation in generator training." Google Developers, https://developers.google.com/machine-learning/gan/generator. Accessed 1 Aug. 2021.

[4] This method was designed by MIT Deep Learning textbook author Ian Goodfellow and colleges in 2014. For surveys of further developments in deepfake methods, see Nguyen et al.; Mirsky and Lee.

Although the technological novelty of a GAN is that it can generate completely new categories instead of depending on a priori rules, we should see this as less of a departure from human cognition than an intensification of iconic reasoning. In a certain sense, the possibilities of establishing likenesses are expanded by a GAN, for it incorporates semiotic processes that operate at speeds and granular scales outside of human perceptive registers into its diagrammatical operations, opening the possibilities for deepfake generation beyond the purview of what is humanly sensible. At the same time, the GAN is optimized such that it may automate claims to imagistic verisimilitude following specific prefigured criteria. It is because of this prefiguration or biasing of the GAN in a certain way, that GANs preserve the truth function of images. Like photorealistic images before, deepfakes preserve iconic continuities. This has to do with the way *meaning* is articulated as a problem to be *solved* through the optimization of ML algorithms. Providing an example of how meaning is problematized by ML, Jonathan Roberge and Michael Castelle describe a deep learning object recognition model trained on the ImageNet database, a popular dataset containing over 14 million categorized images. Instead of using the taxonomies of ImageNet's object categories, "ML/DL's conquest of iconicity—its ability to calculate the likeness between a picture of a tiger and an arbitrary value or category denoted to as 'tiger'—is performed through a layered, directional (and thus indexical) flow of linear and nonlinear transformations . . . by producing categorial outputs, ML/DL necessarily morphs into something prescriptive" (9). Problem solving for ML is a matter of discovering the right rules to calculate resemblance to some input rather than classifying an existent object. In the case of a GAN, the optimization of the model is the problem to which the solution has already been decided upon, in the sense that one has already specified what sort of output one would like to get.

It would seem that the indexical operations of a GAN are what gives them their epistemic agency, as they carve out the semiotic paths through which meaning can flow. In doing so, they both produce and constrain the possibilities of their outputs. But the iconic aspects of ML operations should also be examined, not because they perform some sort of ideological trick but because such operations are entwined with epistemology at the level of social reality. The reality imagined by engineers through ML is one where diagrammatic substitutions can be applied methodologically to find the ruleset to achieve the types of outputs one would like to see in the world. Part of the value in reading ML diagrammatically is to understand how it imposes certain regularities on an epistemology based on the type of problem it is trained to solve. For instance, a GAN trained on images selected from the tiger category of the ImageNet database would learn to expand a random noise distribution into an image statistically similar to those tiger images and visually accurate to a human observer. "Tigerness" is the problem space the GAN operates within. Despite variations, "tigerness" will always be maintained because it is the solution given in advance of the formalization of a procedure, which is discovered over time through the optimization of interactions between training algorithms, vectorized images, and human decisions.

Synthetic Images and SemaFor

I have argued that the diagrammatic logic of ML facilitates and intensifies the possibilities of imagistic truth claims but what remains is the question of how ML techniques are used to maintain *specific* ways of seeing and knowing the world. Here, I use the term "regime of recognition" following Louise Amoore, who describes the way that ML algorithms don't just recognize already existing categories in the world, in the sense that statistical classification algorithms might make use of a priori classifications in a dataset such as faces, threats, vehicles and so on (67). Instead "they actively generate recognizability as such" by deciding "what or who is recognizable as a target of interest in an occluded landscape" (69). Regimes of recognition describe a field of power in which the politics of ML techniques affect the parameters of what matters and what is brought to attention (Jacobsen).

Before recent developments in ML, digital forensics research was premised on the idea that a target of interest bears some accountable indexical link to its technical infrastructures. For instance, if an image were tampered with or manipulated there would be some form of recognizable digital fingerprint. Computer-generated artifacts added to a digital photograph could be discerned through their use of "idealized models of geometry, lighting, camera optics and sensors," which could be classified using statistics-based analysis (Farid 72). Three key technological developments in the early 2000s posed novel challenges to digital forensics— the invention of the high-powered NVIDIA GPU, the creation of large data sets for computer vision research, and the 2012 creation of AlexNet (DARPAtv, "Artificial Intelligence Colloquium"). AlexNet is used today as the architecture for most self-supervised learning. It is a deep convolutional neural network (CNN) that was designed with the notion that objects in the "real world" are imperfectly represented in images—they may be occluded or ambiguous in some way (Krizhevsky et al.). To overcome the problem of indeterminacy, the creators of AlexNet trained it on millions of images labeled by Amazon Mechanical Turk workers, capturing similarities as statistical distributions of weightings and probabilities rather than fixed categories. In turn, AlexNet would be able to recognize all future instances of an object of interest, even instances that varied in ways that were not previously recognizable.

Whereas digital forensics before these developments depended on access to "white box" statistical models which could link fake images to their manipulation techniques, ML is by virtue black-boxed (Pasquinelli and Joler), which means its learning models evade previous detection techniques. As the director of DARPA Semantic Forensics (SemaFor) Program explains, “detecting GAN generated media requires detailed training data from those specific GANs” and detailed knowledge of those generators (DARPAtv, "Semafor"). Where this knowledge is not available, SemaFor aims to develop “semantic algorithms that automatically detect, attribute, and characterize falsified multi-media to defend against large scale automated disinformation attacks” (DARPAtv, "Semafor").

These three specified goals—detection, attribution, and characterization—each lay a specific claim to meaning. They are each given problems to be solved in regulating maliciously manipulated images. Take SemaFor’s example of the application of this multi-modal approach toward assessing an article describing a protest in front of the US Capitol. *Detection* would aim to solve the problem of classifying semantic inconsistencies and manipulation errors at the level of a single output image. For instance, do the person in this image’s earrings match? *Attribution* would solve the problem of source or authorship e.g., is there formatting constancy in the images in the source in which they appear? After screening the protest images to determine if they are fake, ML might be used to perform a textual analysis of a body of journalism by one purported author to recognize inconsistencies within the vocabulary and style. Once their inauthenticity is established, it could screen these articles for terms associated with a high likelihood of malintent. Finally, *characterization* would attempt to solve the problem of moral categorization, deciding on the intent behind media manipulation and prioritizing the fake content for human review. At the level of characterization, the protest article, for instance, could be flagged for posing a threat to influence voter decisions (DARPAtv, "Semafor"). Semantic forensics moves us beyond the scientific detection of data patterns that already exist within the image. It is instead figured as a site from which one can speculate on the possible harmful social effects of images. In each case, semantic meaning "moves beyond low-level statistical fingerprints" to include high-level information such as "words, phrases, signs, visual elements, audio elements; relationships between these elements; and relationships between such elements and the real world" (SemaFor).

We can no longer say, in the case of semantic forensics, that verifying likeness is a matter of common-sense knowledge. Rather, the scope of semantic knowledge sets the regime of visibility by determining what types of information are of interest for establishing visual authenticity or, alternatively, falsifying it. Put another way, solving the problems of detection, attribution, and characterization necessitates that an ideal solution is given in advance for each—a formal assumption of what malicious relationships or patterns of artifacts would look like. In the example DARPA provides of the fake protest article, building an algorithm to interpret semantic meaning at each scale involves inbuilt decisions about what a consistent author style would look like, decisions that correlate terms or image features to social harm, and finally, presuppositions that these decisions can be modeled, schematized, and used to screen future media objects. To automate these solutions at scale requires that interpretation sneaks in the backdoor of data analytics to offer preformed answers to the question of what makes a pattern meaningful and to what this pattern should be determined to mean. Pattern finding thus requires decisions, in the political sense, over what types of information count as meaningful and which are spatially proximate or alike enough to offer a form of correlative proof (Apprich et al.).

For DARPA, which serves as a research and development branch of the US Department of Defense, deepfakes pose novel threats to national security because they undermine consensus reality, which could lead audiences to be duped by malign foreign actors. But it would seem that the problems of detection, attribution, and characterization don't deal with establishing an understanding of existent sociotechnical phenomena but instead deal with countering “expected threats” (SemaFor). Taking stock of the synthetic imaging technologies of the present, SemaFor imagines vectors of future development, such as Identity Attack as a Service (IAaaS), new forms of synthetic generation that would have ripple effects up organizational and political scales. For DARPA, threat modeling becomes a form of threat anticipation, a means to “make all actions imaginable in advance” through arraying possible future threats (Amoore 79). SemaFor creates the conditions for acting in the present through operationalizing image interpretation, pre-emptively developing ML models to counter threats that always remain on a horizon of knowability.

Politics of Precomputation

To analyze how algorithms accrue agential power, Amoore adapts the concept of “precomputation,” used in ML to describe how an algorithm is predisposed to be able to recognize the attributes of something in advance of its run time (78). As she explains, to “precompute” is also to imply what a proper procedure should look like and thus comes hardwired with normative social or political assumptions. This point is significant because as ML increasingly becomes part of everyday media ecologies, our ways of being, knowing, and acting in the world are delimited and afforded accordingly. There is a particular logic inherent to ML that is not just relegated to the world of tech but has become part of social life more broadly. This logic operationalizes the information at hand and adjusts the indexical relationships between features to discover the pattern one would need to confirm the precomputed knowledge one has provided as input. Finding the desired output is a matter, as Amoore explains, of establishing the “level of detection that is useful to you,” which establishes a threshold of recognizability (68).

This action calls for decisions on questions such as: what types of information matter in a particular problem space? What ought to be screened out or discarded?

By deciding on precomputational aspects in advance (what type of algorithm or model is being deployed to solve for an unknown x?), one is deciding on "the register of what kinds of political claims can be made in the world, who or what can appear on the horizon, and who or what can count ethicopolitically" (Amoore 69). Precomputing is, plainly put, a way to take action in a condition of unknowability. It permits one to compress reality through the construction of a fragmented and limited simulation where any inputs can be mapped to the desired outputs. In a technical and social reality where data is in surplus, meaning isn't derived from causality but rather can be constructed through the infinitely possible patterning of data points (Halpern et al.). Precomputing establishes a register of perceptibility to describe a complex and unknowable reality through data. The conspiratorial claims around Trump's concession video can be understood correspondingly as a form of paranoid positivism bound to the fragmented (anti-)social realm of a filter bubble, bent on finding the desired outputs through reading patterns into the past to make actionable claims in the present. In this case, visual truth is subject to the rhetorical force of collective belief. For post-truth conspiracy-ridden epistemologies, a truth claim is made on the image to generate, rather than represent, some version of reality in a grasp for control of an unknowable future.

As we saw with GANs, their operations transform icons through diagraming vectorial relations and indexical substitutions. The diagram provides the precomputed elements that make manipulation possible, yielding image icons that are always propositionally accurate within the semantic boundaries determined by the GAN. Despite what post-truth discourses would suggest, fake images make truth claims. What ML shows us in this instance is that the thresholds of regimes of recognition are porous and mutable, positioning photorealism as a site from which one can make claims to reality as a function of power or to its contestation.

Conclusion

When truth claims are made from various networked states, produced through collectives inclusive of technical and social fragments that do not map onto a consensual public sphere, it is important to ask what beliefs and desires govern our networks and how this in turn sets limitations on visibility and knowability. In security discourses, such as DARPA's, forensics isn't about discovering the causal truth behind an image, as if there were a repository of state knowledge through which all effects could be traced back to their origins. Instead, forensics is mobilized to create evidence of futures not-yet-come to prevent technological surprise and thus manage risk as a matter of national security. We could argue that paranoias around control and the loss of control are the animus of our networked reality, where power is maintained through the rejuvenation of its verifying logics. This is today's circular conspiratorial logic—meaning can be found precisely where it is sought out because the logics of verification are subject to any number of technical operations with more or less authority.

If the production of visuality has always been tied to questions of knowledge and power, we would do well to take imagistic claims to realism seriously, not just in terms of what images represent but in terms of the conditions of possibility they outline and the semiotic regularities they impose. As we have seen, ML is a technique for abstracting human decisions and biases and mobilizing them to arbitrate decisions outside of the scope of human epistemology. This should not be taken to mean that the claims of ML are somehow more objective, accurate, or truthful than human logic, nor that it tricks us into taking its claims as truths, but rather that it acts in the space of what is unknowable. ML systems allow us to create new indexical relations from within. “The condition of indexicality,” claims Paulsen, is “the openness of the sign to interpretation and doubt” (98). ML expands the field of visibility beyond human sensory ratios, generating new possibilities for interpretation, but it also limits this field through precomputing those possibilities. Its operations are shot through with presumptions about what matters and how it should matter. As a result, we end up foreclosing alternatives, manifesting new worlds that are just superficial variations of the old one. Looking at the links between ML and photorealism lets us recognize how truth claims are syntactically conditioned, semantically bounded, and collectively constructed, and can therefore provide an opportunity to rethink the ontology of images not as representational objects but as ethicopolitical agents enmeshed in active sociotechnical negotiations around what matters and why.

Works Cited

@Eric__Fapton. "What Was /R/Deepfakes and How Did It Influence the Recent Reddit Rule Changes?" *Reddit*, 2017. https://www.reddit.com/r/OutOfTheLoop/comments/7vydoi/what_was_rdeepfakes_and_how_did_it_influence_the/. Accessed 15 Sep. 2021.

Amoore, Louise. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke UP, 2020.

Ananny, Mike and Kate Crawford. "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society*, vol. 20, no. 3, 2018, pp. 973–89.

Apprich, Cramer, et al. *Pattern Discrimination*. Meson Press, 2018.

Babu, Sudharshan Chandra. "A 2020 Guide to Synthetic Media." *Paperspace*, blog.paperspace.com/2020-guide-to-synthetic-media/. Accessed Aug. 2021.

Barthes, Roland. "The Reality Effect." *The Novel: An Anthology of Criticism and Theory, 1900–2000*, Wiley, 1968, pp. 229–34.

Chesney, Robert, and Danielle Citron. *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. U of Texas Law 107 California Law Review 1753, 2019, https://scholarship.law.bu.edu/faculty_scholarship/640. Accessed 20 Mar. 2022.

Citton, Yves. "Could Deep Fakes Uncover the Deeper Truth of an Ontology of the Networked Images?" *The Nordic Journal of Aesthetics*, vol. 30, no. 61–2, 2021, pp. 46–64.

Crary, Jonathan. *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. MIT Press, 1992.

Dale, Daniel [@ddale8]. "Meanwhile on Parler, there are dozens and dozens of posts in which Trump supporters say or wonder if his concession video - which he tweeted out - is a “deepfake.”" *Twitter*, 8 Jan. 2021. https://twitter.com/ddale8/status/1347570822437863424?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1347570822437863424%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.reuters.com%2Farticle%2Fuk-factcheck-trump-consession-video-deep-idUSKBN29G2NL. Accessed 15 Sep. 2021.

DARPAtv. "Artificial Intelligence Colloquium: Media Forensics." *YouTube*, 2019, https://www.youtube.com/watch?v=Crfm3vGoBsM&t=2s&ab_channel=DARPAtv. Accessed 20 Mar. 2022.

---. "Semafor Proposers Day." *YouTube*, 2019, https://www.youtube.com/playlist?list=PL6wMum5UsYvZGmhx5NYRXf0OumdP0H8sN. Accessed 1 Jan. 2021.

Doane, Mary Ann. "Indexicality: Trace and Sign: Introduction." *Differences*, vol. 18, 2007, pp. 1–6.

“Fact Check: Donald Trump Concession Video Not a ‘Confirmed Deepfake.’” Reuters, 2021. https://www.reuters.com/article/uk-factcheck-trump-consession-video-deep/fact-check-donald-trump-concession-video-not-a-confirmed-deepfake-idUSKBN29G2NL. Accessed Sep. 2021.

Farid, Hany. "Digital Image Forensics." *Scientific American*, vol. 298, no. 6, 2008, pp. 66–71.

Farocki, Harun. "Phantom Images." *Public*, no. 29, 2004, https://public.journals.yorku.ca/index.php/public/article/view/30354. Accessed 20 Mar. 2022.

Goodfellow, Ian, et al. "Generative Adversarial Nets." *Proceedings of the 27th International Conference on Neural Information Processing Systems*, vol. 2, MIT Press, 2014, pp. 2672–80.

Gunning, Tom. "What's the Point of an Index? Or, Faking Photographs." *Still Moving: Between Cinema and Photography*, edited by Karen Beckman and Jean Ma, Duke UP, 2008. pp. 39–49.

Halpern, Orit, et al. "Surplus Data: An Introduction." *Critical Inquiry*, vol. 48, no. 2, 2022, pp. 197–210.

Hayles, N. Katherine. "Can Computers Create Meanings? A Cyber/Bio/Semiotic Perspective." *Critical Inquiry*, vol. 46, no. 1, 2019, pp. 32–55.

Jacobsen, Benjamin N. "Regimes of Recognition on Algorithmic Media." *New Media & Society*, Oct. 2021. https://doi.org/10.1177/14614448211053555. Accessed 20 March 2022.

Krizhevsky, Alex, et al. "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, vol. 25, 2012. https://dl.acm.org/doi/10.1145/3065386. Accessed 25 Sep. 2021.

Mackenzie, Adrian. *Machine Learners: Archaeology of a Data Practice*. MIT Press, 2017.

Manovich, Lev. *The Language of New Media*. MIT Press, 2002.

Mirsky, Yisroel and Wenke Lee. "The Creation and Detection of Deepfakes: A Survey." *ACM Comput. Surv.*, vol. 54, no. 1, 2021, arXiv:2004.11138. Accessed 20 Sep. 2021.

Mitchell, W.J.T, ed. *Critical Terms for Media Studies*. Co-edited by M.B.N. Hansen, University of Chicago Press, 2010.

Nguyen, Thanh Thi, et al. "Deep Learning for Deepfakes Creation and Detection." *arXiv preprint*, 2019, arXiv:1909.11573. Accessed 20 Sep. 2021.

Paris, Britt, and Joan Donovan. *Deepfakes and Cheap Fakes*. Data & Society Research Institute, 2019.

Pasquinelli, Matteo and Vladan Joler. "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism." *AI & Society*, vol. 36, 2020, pp. 1263–80.

Paulsen, Kris. "The Index and the Interface." *Representations*, vol. 122, no. 1, 2013, pp. 83–109.

Peirce, Charles Sanders. “A Second Trichotomy of Signs.” *Elements of Logic*, edited by Charles Hartshorne and Paul Weiss, Harvard UP, 1931, pp. 247–50.

Roberge, Jonathan, and Michael Castelle. "Toward an End-to-End Sociology of 21st-Century Machine Learning." *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*, Springer, 2020, pp. 1–29.

SemaFor. "Frequently Asked Questions." DARPA, https://www.darpa.mil/attachments/FAQ_SemaFor%202019-09-03_FINALa.pdf. Accessed Dec. 2021.

Stjernfelt, Frederik. "Moving Pictures of Thought." *Diagrammatology: An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics*, edited by Frederik Stjernfelt, Springer Netherlands, 2007, pp. 89–116.

Uricchio, William. "The Algorithmic Turn: Photosynth, Augmented Reality and the Changing Implications of the Image." *Visual Studies*, vol. 26, no. 1, 2011, pp. 25–35.

Weatherby, Leif, and Brian Justie. "Indexical AI." *Critical Inquiry*, vol. 48, no. 2, 2022, pp. 381–415.

Werner, Gabriele. "Discourses About Pictures: Considerations on the Particular Challenges Natural-Scientific Pictures Pose for the Theory of the Picture." *The Technical Image: A History of Styles in Scientific Imagery*, edited by Bredekamp Horst et al., University of Chicago Press, 2021, pp. 8–13.

Rebecca Uliasz is an artist and PhD Candidate in the department of Computational Media, Arts and Cultures at Duke University. Her research interests include emerging issues in digital culture, philosophies of computational media, and new media art. She is one half of the theory-noise performance collective GOVERNANCE, alongside collaborator Quran Karriem. Her work has appeared in *Springer AI & Society: Special Issue, Ways of Machine Seeing*, *Review of Communication, Optimization: Towards a Critical Concept, A Peer Reviewed Journal About*, and transmediale. Email: rebecca.uliasz@duke.edu.