



# Algorithmic anxieties: Trevor Paglen in conversation with Anthony Downey

Trevor Paglen<sup>1</sup> · Anthony Downey<sup>1</sup>

© Springer Nature Limited 2020

In September 2019, I met with Trevor Paglen as he prepared to present a new project at the Barbican Curve in London. “From ‘Apple’ to ‘Anomaly’” explored the images—so-called datasets—that are used to train algorithms. In a subsequent far-reaching conversation, recorded at the Barbican Centre on 26 September 2019, Paglen presented an extended overview of the ideas behind this work, observing how artificial intelligence and “machine learning” utilise datasets to recognise different objects and, more problematically, produce classificatory systems for “recognising” individuals.

Collating and categorising over 30,000 images from ImageNet, the largest dataset in use today for developing algorithms, “From ‘Apple’ to ‘Anomaly’”, involved presenting a relatively small selection of the dataset’s overall images—which number 14 million in total organised over 20,000 different categories—and classificatory labels. Ranging from relatively innocuous terms such as “apple” or “strawberry” to more offensive phrases such as “Failure”, “Loser”, “Non-starter”, “Unsuccessful person”, “Jezebel”, “Drug-addict”, and “Junkie”, the labels and classification systems used to organise the dataset deployed by ImageNet—which human operators input and define—disclose the extent to which the project operates on an in-built bias based on societal mores and prejudices. Through exhibiting the images that created the original datasets alongside the categories used to cluster such images, Paglen’s project pointedly displays the often-hidden aspects of machine learning and the political economy of meaning that is produced through such apparatuses. “When you look at these training sets and the categories”, Paglen observes in the following conversation, “they’ve been collected into, you are looking at embedded world views and the cultural and political forms of seeing and categorising that are being hard-wired into computer vision systems”.

In addition to this project, and working with AI researcher Kate Crawford, Paglen also discussed ImageNet Roulette, a website where users could upload their own photographs to see how the database, using datasets and categories such as those mentioned above, might categorise them. I uploaded three images of myself to ImageNet Roulette before it went offline and the results that came back were varied—ranging from “swot, dweeb, learner, assimilator”, to “performer and psycholinguist”—but relatively inoffensive. I understand from Paglen, however, that others had far more problematic results relayed to them that expressed racial prejudices and criminal overtones. Further categories, ranging from “wrongdoer” to “offender”, were applied to images uploaded by an African-American man, while “stunner, looker, man-trap”, were used to describe a white woman. In each case the disturbing potential of such datasets to perpetuate racial and misogynistic stereotypes is amply demonstrated.

ImageNet Roulette is no longer live, but Crawford and Paglen have written up their findings in an article entitled “Excavating AI”<sup>1</sup>. As a result of the project, ImageNet removed 600,000 images of people stored on its database but, significantly, have not substantially revised their categorical systems of classification. On an operative level, both ImageNet Roulette and “From ‘Apple’ to ‘Anomaly’” highlight how algorithms—as coded sets of instructions—give rise to biased forms of artificial intelligence (AI), the latter being an algorithmic grouping or cluster that can modify and create new algorithms in response to further data inputs and, thereafter, at least theoretically, develop “intelligence”. Presenting audiences and researchers alike with an opportunity to understand the algorithmic anxieties surrounding, for example, the construction of race and gender through AI systems, both projects demonstrate the degree to which evolving determinations of subjectivity are being predefined and established through the operations of often occluded datasets and the opaque operations of machine learning.

✉ Anthony Downey  
Anthony.Downey@bcu.ac.uk

<sup>1</sup> Birmingham City University, Birmingham, UK

<sup>1</sup> See: Kate Crawford and Trevor Paglen, “Excavating AI: The Politics of Training Sets for Machine Learning”, 19 September 2019, <https://excavating.ai>.



**Trevor Paglen**, “From ‘Apple’ to ‘Anomaly’”, Installation view. Barbican Curve, London.  
© Max Colson, 2019



**Anthony Downey** I understand that we are going to start with a short presentation by Trevor, who will give us an overview of the research that went into datasets that went into his new project, namely, “From ‘Apple’ to ‘Anomaly’”.

**Trevor Paglen** Thank you. I think something very dramatic is happening in the world of visual culture today and the world of images that surrounds us; images that are part of the societies that we live in. We’re increasingly embedded within sensing systems, whether that’s cameras that are embedded within urban infrastructures that take pictures of car licence plates, or facial recognition software installed at airports and borders and commercial places. Commercial imaging systems in shopping malls that record your movements through department stores, monitor your facial expressions, and trying to figure out what products you might like, or systems that try to read your lips and decipher what you’re talking about when you’re talking to other people. These kinds of sensing systems are in places that are maybe a little bit less obvious too; for example, if you put a picture on Facebook, a few of your friends might see it and your experience of that resembles sharing a photograph album with your friends and relatives, but in the background of that those images are being scrutinised in great detail by a host of artificial intelligence algorithms that have been trained to try and recognise things and elements in them.

What we’re seeing is the advent of a new relationship to images, with computer vision and artificial intelligence taking the foreground in processes of seeing. In the past, images needed somebody to look at them in order for them to come into existence. An image that nobody ever saw basically didn’t exist. But that’s not true anymore. There’s a vast world of images that are now machine readable that don’t

need humans to look at them to make sense of them. For the last number of years, I’ve been trying to learn how machines look at images. I want to know what forms of seeing are embedded within technical systems.

In my studio, we’ve been developing a lot of software that allows us, as humans, to try to see through the eyes of machines. For example, we created a whole programming language that allows us to take a picture of a string quartet playing music and run it through software that you would use for a guided missile, or for a self-driving car, or an AI algorithm that’s doing object detection—or something akin to that—and it will draw pictures that represent what that algorithm is seeing when it’s looking at an image.

To do object recognition, and to do what is the cutting-edge computer vision work, you use things called neural networks, which is basically another word for artificial intelligence or machine learning systems. In order to build a neural network that can recognise different objects, the first thing you have to do is start with a taxonomy. In other words, you have to create a giant list of all the objects you want to be able to recognise. For example, let’s pretend we’re going to make a neural network to recognise things in our kitchen—apples, oranges, spoons—you make a list of all the things you want it to be able to recognise. Then what you do is you start to build what’s called a training library or training set. You need to give the neural network hundreds, if not thousands, of examples of each of those objects you want it to learn how to see. You have to feed it thousands of pictures of oranges, or thousands of pictures of spoons, thousands of pictures of plates, forks, and so on. The system will then do a statistical analysis of all of those images and break them down into what I think of as primitive components, or

primitive shapes. So they could be horizontal lines, diagonal lines, vertical lines, and it will invent a series of primitive shapes that it can then assemble in various ways to make sense of more complex objects.

When the system has been trained in this way, you can show it a picture of something it's never seen before and it will analyse that image and it will define the formal components that make up that image. So, a spoon might be some vaguely parallel-ish lines, kind of chrome silver colour, an ellipse on one side of it. You break it down into these shapes and if you find all of these primitive shapes, it's probably more likely to be a spoon than something else. A fork, in this logic, is going to be very similar to a spoon except for the end of the fork; instead of an ellipse, you're going to have a little spiky thing. Bananas will be very different: it'll be made of arcs, and more yellowish colours, and when it finds all of those primitive shapes put together it will say "this is a banana". Once you've trained your network to recognise these different objects you can start doing things like showing it an apple and it says "this is an apple".

The images that you're feeding into the network—images of oranges and apples and the forks, etc.—are called training images or training data. The data are the images that you use to train the neural network how to see. Looking at these collections of training images has been something that I've spent a lot of time working on for the last few years. When I crack these things open, I want to understand what kinds of logic are built into the training data that are used to teach computer vision systems how to see. When we're thinking about training images, we have some prehistories of them: you can think of fingerprints, for example, being a prehistory of training images; or you can think about mugshots being prehistory of training images. We find figures associated with these forms of historical photography in people like Francis Galton or Alphonse Bertillon. In the early 1990s, you start seeing research laboratories and military laboratories creating their own collections of training images in the service of early computer vision.

When you're creating a training set, you have a couple of things you have to make—you have to create an overall taxonomy, which is true of all kinds of training sets. And every time you're creating a taxonomy, there's always a politics to that, because when you're creating a taxonomy you're saying this is a range of categories that are intelligible, and it's always going to be a limited range. And in doing so you're always creating a negative space to define the things that are outside of that, the things and elements that are not intelligible within this given taxonomy.

This brings me to affective computing, or how do you teach a computer your emotional state by looking at your face. When you're creating these kinds of datasets for something like emotions, we can start looking at what kind of assumptions are built into a training set. First of

all, we have an assumption that emotion itself is a sensible taxonomy and that these emotions are expressed on people's faces. So, where do these assumptions come from? In the case of affective computing, these assumptions come from a psychologist called Paul Ekman, who asserted that emotions can be categorised into six basic categories and can be ascertained by looking at someone's face with the eyes providing the proverbial window to the soul. Since publication, Ekman's work has been very profoundly critiqued from the fields of anthropology and psychology, but his theory of emotions lent itself particularly well to computer vision. It posited that there was a discrete number of emotions and those emotions were measurable by looking at someone's face. This has become a paradigm for the underlying epistemology of affective computing, no matter how much it's been criticised in the social sciences. It is a perfect theory of AI and how artificial intelligence operates. It also helps us better understand the politics of taxonomic classifications that we see becoming more widespread today in facial recognition systems operated and driven by algorithms.

Take the UTK Face dataset, this was created at the University of Tennessee at Knoxville. This consists of 20,000 images that have annotations for people's age, gender, and ethnicity. Age is an integer 0–116; gender is either 0—male or 1—female; race is an integer denoting 0–4 denoting white, black, Asian, Indian, and others. However, the idea that you can measure an individual's gender by looking at their face, that gender is binary, and that you can tell someone's race by looking at them, and that race is white/black/Asian/Indian or "other" is obviously preposterous. And there are histories here, relating to the racial classifications in this dataset, that further recall classificatory systems used by the South African apartheid regime in the 1970s, where each person in that scheme was classified as either black, white, coloured, or Indian. And these categorisations mattered: they affected what your civil liberties were, and your relationship to legal, social, and economic systems of power.

The most ambitious training set today, and the most widely cited, is ImageNet, and it is images used for its training set that provides the basis for "From Apple to Anomaly", the current installation in The Curve here at the Barbican Centre. The dataset was first published in 2009, having been created by researchers at Stanford and Princeton University. Consisting of over 14 million images, which have been labelled into more than 20,000 categories, it has since become the gold standard for training sets. It was an attempt—in the words of its creators—to "map out the entire world of objects". It's a massive database and it's intricate. There are, for example, 1478 pictures of strawberries, 932 pictures of strawberry ice cream, and 604 pictures of strawberry daiquiris. There are classes for "apples", "apple butter", "apple fritters", "apple dumplings", "apple jelly",



Trevor Paglen, “From ‘Apple’ to ‘Anomaly’”, Installation view. Barbican Curve, London. © Max Colson, 2019



“apple juice”, “apple pie”—and that isn’t even all of the apple-related elements. It goes on and on and on.

When we look at ImageNet, we see a world view coming into view. A world view that is contained within the training set itself. We see this happening at many levels. First of all, we have this overall taxonomy, but we can also look at the individual categories and ask what kinds of concepts become reified through such systems; we can see how things like the category for apple might be relatively uncontroversial, but as we go further into the dataset, it does become more controversial. At some point, for example, the computer vision system does not so much describe people as much as it judges them. When we looked more closely at ImageNet, we found that there are about 2800 categories of people in there. These are categories defining different kinds of people, and certain images have been associated with those categorical definitions. When we look at these categories, we very quickly start finding ones that are not only judgemental, but also classist, ableist, racist, homophobic, misogynistic, and just outright cruel. We have categories such as “bad person”, “debtor”—like you can tell what someone’s bank account is by looking at their face in the weird epistemology of ImageNet? You also have “Swinger”, “Tramp”, “Ball buster”, “Ball breaker”—the latter defined as “a demanding woman who destroys men’s confidence”. You can also find “Failure”, “Loser”, “Non-starter”, “Unsuccessful person”, “Jezebel”, “Drug-addict”, and “Junkie”. These are all pictures that the researchers had collected by “scraping” the Internet, using accounts such as Flickr for example, and then

hiring Amazon Turk workers to label those images and sort them into these 20,000 categories.<sup>2</sup>

What we have here is a layer in the training set of the categories, a layer where human input is involved, and those categories betray a world view. When we go down a layer to the actual images, we need to ask what does it mean to look at an image and label it? We look at an image, but what do we see? A woman on holiday on the beach has been labelled a “kleptomaniac”. A man labelled a “loser”. Another labelled an “anti-semitic”. Sigourney Weaver, for example, is labelled a “hermaphrodite”. When you look at these training sets and the categories they’ve been collected into, you are looking at embedded world views and the cultural and political forms of seeing and categorising that are being hard-wired into computer vision systems. Systems whose creators would often like you to imagine are neutral—that it’s about maths, algorithms, and science—but when you crack open the hood, so to speak, there all sorts of questionable things going on.

**AD** Thank you Trevor. On the subject of lifting the hood, it appears that “From Apple to Anomaly” is very much about the back end of algorithmic reasoning; the parts we do not see but play a significant role—as you note—in defining the epistemological categories that produce knowledge. To this end, the images that make up your project are often about making visible that which is invariably rendered invisible in the very process of machine learning. These are often

<sup>2</sup> Amazon’s Mechanical Turk project, according to its website, is a “crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually.” See: <https://www.mturk.com>.

invisible datasets that produce algorithms that, in turn, train artificial intelligence.

**TP** Exactly.

**AD** As you were speaking, I was thinking that this project is very much about visualising invisibility, so to speak. It has become a cliché that we're somehow drowning in images, that there is a glut of images, whereas the reality is that we're only seeing about 1% of the images that are currently circulating via the neural networks that define artificial intelligence and machine learning systems. The other 99% are manifesting a new world order—and that world order is closed to us, inasmuch as we do not see that system at work.

**TP** You can think about them as infrastructure. You're looking at these masses of images in the installation "From Apple to Anomaly", but they aren't actually there for us to look at as such—they are hardwired into the technical systems that produce the artificial intelligence and machine learning systems that are learning to look at us.

**AD** The show is called "From Apple to Anomaly", which also raises some interesting semantic questions: apple is a normative noun, and anomaly is a relational noun, inasmuch as what you define as an apple will have a general (normative) consensus, but what you define as an anomaly will be inevitably relative. And this is the problem: there are cultural, political, and historical ways of thinking about an anomaly. It is therefore a cultural, political, and historic construct as well as a word. Could you talk about that? I think that gets to the core of what you're talking about, this new algorithmic order, and how it comes into being.

**TP** There are so many different ways to approach that point. I struggle with it a lot, and when writing the "Excavating AI" paper with Kate Crawford, we discussed this a lot.<sup>3</sup> To the extent that the word anomaly is problematic, we can't even say that an apple *is* an apple. What's an apple? Is it allegorical—a form of knowledge related to sin, as in the bible? The point here is that it has all kinds of cultural associations and depending on the context we refer to it in, these can be invoked. "Anomaly", however, is not a thing as such, it's a relational concept, just something we think is weird or out place! There are examples in the exhibition too: 'pest' for example. Pest is like the lifeform that you don't like, for whatever reason, but the concept is inherently relational, inherently historical, and it is specific to whatever context it's being used in, and who is invoking it in what context. So, what does it mean to universalise that phrase "pest" and to apply it to an image? And by universalise, I mean hardwire it into a technological system that is about definitions.

As you start going backwards towards nouns that you might think are more "nouny", so to speak, the politics becomes more urgent: who is a "worker", who is a "leader", who is a "man", who is a "woman", who is a "loser"? You start to see the imposition of a very particular and historical point of view, where a specific politics is being hardwired into those algorithmic systems—based on datasets—that are being let out into the world so as to judge people.

**AD** One example that struck me is "porker" which refers to a young pig fattened or slaughtered, but it is of course an idiomatic phrase. And one of the most difficult things to learn in a language is an idiom. A "porker" in British English could also refer to an obese individual or—through the derivative "porkie"—could reference a lie or a falsehood, as in "you're telling a porkie", which I think comes from the Cockney Rhyming slang "pork pie"—telling a lie! Where do you begin to explain that to a machine through datasets?

**TP** This is precisely what I'm playing with throughout the installation of the project.

**AD** Yes, I like the fact that you included SPAM. Which has become associated with spam mail, that is unsolicited email or phishing, but for this particular dataset it was a tin of SPAM, which has a very British resonance. Which brings me to the first image in the exhibition. The first image in the exhibition is not from a dataset, it's a very specific image of Rene Magritte's oil painting *Ceci n'est pas une pomme* (*This is not an apple*), which he painted in 1964. Like his earlier painting in 1929 of a pipe, *The Treachery of Images* (*This is Not a Pipe*), it poses the question of reality and painterly reality; or the real and its representation. I was taken by the fact that you begin this entire show with a historical reference to aesthetics. Despite the fact that we are for a large part talking about "operational images", to use the phrase Harun Farocki employed to define the ways in which, in our age, images are produced by machines to be seen by other machines, rather than the corporeal, embodied eye, this is also largely a question of aesthetics and the politics of representation. What is a computer seeing that we are not seeing?

**TP** For me, the print of *Ceci n'est pas une pomme* that frames the show is basically posing a question: what is an apple? What does a representation of an apple mean? And, more significantly, who gets to decide what an image of an apple means? In the case of the Magritte painting, it's a picture of an apple that says "this is not an apple". That, for me, is an allegory of self-representation—it is an acknowledgement that representations are always relational and they are based on consensus. And these consensus can change. We could think here about queer or feminist theory and how it tries to change the meaning of images and how we interpret them. To be able to define the meaning of an image and what our own image is involves a significant amount of agency and power over self-representation. I think Magritte's painting is pointing in that direction, towards the

<sup>3</sup> See, "Excavating AI: The Politics of Training Sets for Machine Learning", op cit.







Trevor Paglen, “From ‘Apple’ to ‘Anomaly’”, Installation view. Barbican Curve, London. © Max Colson, 2019

politics of representation. To a system of computer vision imposing its will on that image and saying “no, I don’t care what you think you’re doing here Magritte, this is an apple”: this reveals the underlying politics about who gets to decide what the meaning of images are. It also reveals who gets to create those classifications.

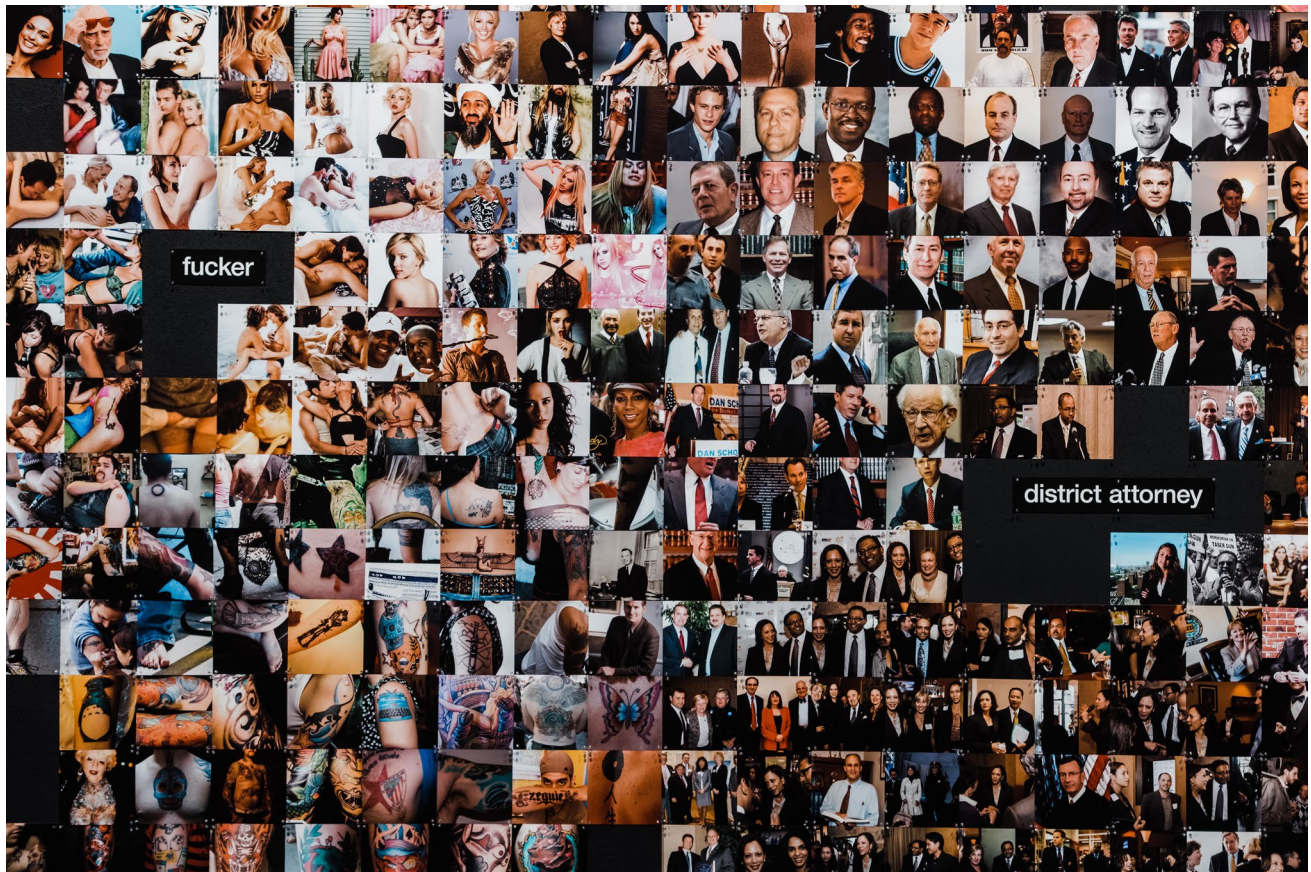
**AD** When Magritte writes “Ceci n’est pas une pomme” across the surface of the painting—or “Ceci n’est pas une pipe” he is of course stating the obvious: this is *not* an apple or a pipe but, rather, a representation of an apple or a pipe that we mostly agree resembles those objects. But that agreement is not only a cultural construct, it is based, as are all normative and normalising pronouncements, on forms of social and political consensus. Which brings us perhaps to one of the core issues in this particular work: most definitions of gender, for example, are cultural constructs based on social and political consensus. They tend towards normative forms of prescriptiveness: you are either one thing or the other. One of the issues that you explore throughout “From ‘Apple’ to ‘Anomaly’” is precisely the downside of that prescriptiveness, and its bias towards misogyny and racism, for example. But I also want to suggest here that this project reveals the biopolitical intent lurking in the forms of

machine learning and algorithm developed from such datasets. Algorithms, through machine learning and the use of datasets, produce subjectivities. Do you agree that there is a biopolitical agenda underwriting algorithms?

**TP** I agree 100%—that was the argument we were making in the “Excavating AI” article. Such systems are always going to have a world view built into them and the best thing you can do is pick what kind of world view you want, with the understanding that not only is a world view embedded in these systems but it is also re-imposed on the world that it is subsequently intervening into. There are projects being developed now within the field of machine learning around fairness and transparency, and there are a lot of people trying to technically de-bias training data. It is obviously a bad idea if all CEOs are defined as white men and the term “criminal” turns up images of black men, so I hope that the work that I and other people are doing is addressing these concerns. The technical solutions that will be proposed are focusing on gender equality and racial diversity within the different classifications, but how do you know what someone’s gender identity is if you don’t ask them? How can you just take somebody’s picture and label it “Latino” for example, or even “woman” or a “man”. The prelabelling by people







Trevor Paglen, "From 'Apple' to 'Anomaly'", Installation view. Barbican Curve, London. © Max Colson, 2019

creating the taxonomies and classification systems in use here defines a preconceived notion of what gender is. There is an assumption that you can take people's pictures off the Internet and that Amazon Turk workers can, in turn, figure out what someone's gender is by looking at their picture, rather than asking them. I think that points to some of those biopolitical dynamics. We should also note that some of the images being used in these datasets are being collected from Flickr accounts without the account holders being notified.<sup>4</sup>

**AD** You mentioned epistemology earlier in your introduction, and it got me thinking about knowledge more broadly: what do we know and how do we know it? I was thinking specifically on the work of Michel Foucault and his now seminal volume, *The Order of Things*, which was originally

titled *Les Mots est les Choses* upon its publication in France in 1966.<sup>5</sup>

**TP** Which is actually a play on Magritte and his theories concerning the relationship between pictures and words—which is the other reference in the title of this piece.

**AD** Yes, I only recalled that today when researching the origins of the book's title—apparently Magritte also wrote a number of letters to Foucault upon the book's publication, which he later included in his own book on Magritte, titled *This is not a Pipe!* Foucault's theory of discourse is important here inasmuch as—in my understanding at least—it alerts us to the substructures that allow statements to be made in the first place. When Foucault defines discourse, the epistemological structures that allow statements to be made, he is defining how truths (or agreed systems of thought) come to be socially and politically accepted. I was thinking here of Foucault's influence on subsequent theorists such as Edward Said—specifically the former's work on how

<sup>4</sup> In 2019, it was revealed that a facial recognition database, MegaFace, had "scraped" Flickr accounts for images without the account holder's knowledge or consent. Containing images of 700,000 individuals, the MegaFace dataset has been used by companies to train face-identification algorithms for the purpose of identifying protesters, surveil terrorists, and, according to a recent article, spy on the public at large. See: <https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html>.

<sup>5</sup> The full title in the English translation, published in 1970, is *The Order of Things: An Archaeology of the Human Sciences (Les Mots et les Choses: Une Archéologie des Sciences Humaines)*.





Trevor Paglen, “From ‘Apple’ to ‘Anomaly’”, Installation view. Barbican Curve, London. © Max Colson, 2019



discourse produces the “truth” of a subject—who argued that imperialism, under the guise of orientalism, discursively produced the non-western other as an inferior subject through systems of classifications that appeared self-evident or neutral, but of course had their own inherent biases. Are we seeing something similar emerge here in algorithmic processes, a system of classification that can often appear abstract or neutral—the product of machine learning—that is in reality a system for producing the “truth” of a subject, a means to re-colonise subjects, so to speak, along the lines of normative forms of subjectivity?

**TP** Again, I agree 100%: there are strong echoes of the colonial gaze built into machine learning systems, and it’s not just some kind of abstract process. It is literally measuring and mapping people’s faces and saying these people are “bad” people, and these people are “good” people; these people are “sluts” and these people are “leaders”. We can see the recurrence here of those colonial pseudo-sciences such as phrenology and physiognomy re-emerge as instruments of power and as a means to create so-called inferior people.

**AD** I was also thinking here of Giorgio Agamben’s theory of an apparatus—which is effectively a re-reading of Foucault’s discourse theory.<sup>6</sup> An apparatus is a means, for Agamben, to produce normative and non-normative subjectivities. And algorithms are, effectively, an apparatus, but we cannot see its workings, so to speak. With the advent

of anti- and post-colonial critique, we saw an ongoing and engaged form of criticism that focused on how colonial discourse and, in the visual arts, images operated. In algorithms, especially, if they are proprietary, we cannot do that unless we go to the back end of the workings, to the datasets that are used to train AI systems and algorithms. Despite the fact that it is algorithms, based on the input of image-based dataset that power AI, we do not get to see the images—even though, in another irony, those images have been produced by social media users and others who have uploaded pictures to online platforms.

**TP** Yes, that is true, there is also something else at work here: there’s another layer beyond the epistemological, and that is to do with commercial systems. It’s not just about assigning categorisations to people because you’re trying to enact some kind of colonial violence, although that has a part to play, it’s also trying to extract value. For example, a company has knowledge of the fact that you drive too fast, and will then modulate your car insurance accordingly to avoid additional risk. Another company might know that you like eating hamburgers because they can—through AI and other means—automatically detect from your Facebook account photographs of you eating hamburgers. They might want to modulate your health insurance based on that. This is how the insurance industry works. You go to any conference about the future of insurance, and this is all they talk about. These forms of classification are about discriminating against or in favour of other people in order to extract value from them. So, when we’re thinking about that epistemological layer, I think we need to add a layer of political economy too. Which is not only about categories, but also about how that apparatus of extractive capital is formed by

<sup>6</sup> Agamben has argued that “apparatuses must always imply a process of subjectification, that is to say, they must produce their subject.” See: Giorgio Agamben, “What Is an Apparatus,” in *What Is an Apparatus?: And Other Essays*, trans. David Kishik and Stefan Pedatella (Stanford, Stanford University Press, 2009), 11.





**Trevor Paglen**, “From ‘Apple’ to ‘Anomaly’”, Installation view. Barbican Curve, London. © Max Colson, 2019



that economic mode of production. I think that's something we need to think about when we're talking about how meaning is produced, and this has reference points in the colonial project. Another interesting element in a data and training set like ImageNet is that you can actually look at it because it was invented at a university. It's meant for people to look at and do this kind of work with it, but we have no idea how many taxonomic systems it has since seeded or played a part in producing.

**AD** We could be presently using apps that have been developed using these data and training sets.

**TP** Yes, exactly. For example, if Getty Images wanted to build a training set that would classify all stock photography, they would start that process with ImageNet, and then reinforce it as they classify their own images. So, you get these geological layers of machine learning that are built on assumptions that would be embedded in datasets such as ImageNet. But, of course, Google has way bigger training data than Stanford University does, and its embedded in their terms of service. The data and training sets that would be used by the big 5 companies, like Amazon, Facebook, and Google, are proprietary; those are closely held secrets and remain invisible to us. And they definitely do not want to show you how it works—this invisible, proprietary politics of classification. The other thing you touched on was the inscrutability of some of the machine learning systems from a technical perspective. In applications like law enforcement and predictive policing, there are massive problems around this in terms of due process.

**AD** I want to return to this notion of value. For so-called big tech, the goal for many of the leading companies would appear to be profit, based on a venture capitalist model

of producing value, but that value is coming from us, not from financial investment per se. It's related to data mining. You've got companies now, such as Bitcoin, Ripple, and Ethereum, whose value is based on the intricacies involved in mining information and making it trustworthy as a medium of exchange in the digital age. Financial value is being defined, and in part replaced, by the value of data. But that data are being mined by us, we're producing that data; we're giving it over to the tech companies, and they are producing value in that which we do not receive.

**TP** I see what you're saying: data are like a fictitious capital.

**AD** Yes, that is a clearer way of putting it—which brings us to the question of labour. There are several ways to talk about labour, one is perceived threat of automation brought about by AI-driven robots—which is perhaps the most common way to talk about labour—but you mentioned the Mechanical Turk project, which is owned by Amazon. This project, which suggests another type of labour, whereby we are uploading data to assist in tasks that computers are currently unable to do—including, but not limited to, identifying and labelling specific content in an image or video. I want to talk about the physical labour going into this and how that is remunerated, which is important to address inasmuch as those who work on these tasks are historically paid very little.<sup>7</sup> And then there's another form of labour going on here and that's to do with how Facebook outsources some

<sup>7</sup> For details of pay rates used by Amazon's Mechanical Turk, see: <https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/>.



of the most horrendous images that come up on their servers to content moderators in so-called developing countries.<sup>8</sup>

**TP** My collaborator Kate Crawford has a fascinating project about this called “Anatomy of AI”, where she looks at all the components and layers needed to produce AI: where do the minerals that power computers come from; where do the power supplies come from to create these systems, and so on.<sup>9</sup> Also, when you’re looking at these training sets, you have massive numbers of images that have to be categorised and put into categories—someone has to do this. So the way they do this is they have online platforms like Mechanical Turk, that are mostly outsourced to central Africa, India, Indonesia, Venezuela, where you basically hire people to look at images and label them. It is similar to CAPTCHA, when you have to click all the images of Stop signs, for example, with people doing that all day long for low pay and without any labour laws to protect them. There is an enormous amount of labour that is underneath the training layer itself. In the installation at the Barbican Curve, that was something that we were trying to mirror. There are 35,000 images in the installation that have been individually printed and individually pinned to the wall. And that was 10 people working for 2.5 weeks, basically around the clock. For me that was a really important part of the project: to create an installation where the amount of labour that went into collating the images was still visible rather than obscured. The content moderation side is another important concern—and I think we need to consider the psychological costs involved in producing AI too and how it is distributed and compensated for across the world.

**AD** I want to end on a very basic question: what is to be done? I don’t think this is a generic or abstract question. I think something has happened. Something profound has happened. There are algorithms out there that we genuinely don’t know what they are doing, but they are doing something. So what do we do in response? How do we offer something that could potentially disrupt this system, which is creating a new world order, not so much before our eyes, but beyond our view?

**TP** I think that’s a huge question. I don’t know the policy answer to that, but I know how to begin the conversation. I think you have to begin the conversation through reconceptualising how we think about technology. The metaphor that drives me crazy is that “technology is like a hammer, you can use it to build a house or you can hit someone on the head with it”. The most dramatic example of this—nuclear weapons—reveals a vision of the political order inherent in nuclear weapons themselves. If you’re going to have nuclear

weapons, you need to have certain types of infrastructure in place, which means you have to have certain kinds of security measures in place, certain kinds of economies. The existence of nuclear weapons is going to have geopolitical ordering and structure as a consequence of their very presence. In other words, there is a vision of society that is built into the weapon and that the existence of the weapon has to reproduce. I think that that is true for AI systems as well, as it is for all kinds of technology. When we’re talking about machine learning, I can build little models in my studio but to really do this at scale you need to be able to collect all the data of everybody on the planet. There’s five companies in the world that can do that at scale—plus China—and we need to ask: what kind of vision or order of politics is inherent in that fact? We also need to question where the places exist within which we can collectively decide that these sorts of technologies might be a good thing? I think that’s not an impossible thing to imagine—and we should definitely start imagining that it is possible to change them.

**AD** And on that note it is worth noting that there was a time before the Internet and there will be a time after the Internet, but effectively we are living through a moment which will define, at least for a generation, how we interact not with the Internet itself but how it will define our relationship to the world which is exactly what ImageNet Roulette and “From ‘Apple’ to ‘Anomaly’” do with precision and, if I may, human intelligence.

**TP** Thank you!

**Acknowledgements** We would like to thank Alona Pardo, Jon Astbury, Daisy Robinson-Smith, and Elisa Adami for facilitating the above conversation and transcript.

**Trevor Paglen** is an artist whose work spans image making, sculpture, investigative journalism, writing, engineering, and numerous other disciplines. Paglen’s work has had one-person exhibitions at Vienna Secession, Eli & Edythe Broad Art Museum, Van Abbe Museum, Frankfurter Kunstverein, and Protocinema Istanbul, and participated in group exhibitions the Metropolitan Museum of Art, the San Francisco Museum of Modern Art, the Tate Modern, and numerous other venues. He has launched an artwork into distant orbit around Earth in collaboration with Creative Time and MIT, contributed research and cinematography to the Academy Award-winning film *Citizenfour*, and created a radioactive public sculpture for the exclusion zone in Fukushima, Japan. In 2014, he received the Electronic Frontier Foundation’s Pioneer Award for his work as a “groundbreaking investigative artist”. Paglen holds a B.A. from U.C. Berkeley, an MFA from the Art Institute of Chicago, and a Ph.D. in Geography from U.C. Berkeley.

**Anthony Downey** is Professor of Visual Culture in the Middle East and North Africa (Birmingham City University). In 2020, in his capacity as a co-investigator, he was the recipient of an AHRC Global Challenges Research Fund to develop a four-year project on education provision in conflict zones. He sits on the editorial boards of *Third Text* and

<sup>8</sup> See: <https://www.nytimes.com/2018/09/25/technology/facebook-moderator-job-ptsd-lawsuit.html>.

<sup>9</sup> See: <https://anatomyof.ai/img/ai-anatomy-map.pdf>.



*Digital War*, respectively, and is the series editor for *Research/Practice* (Sternberg Press, 2019). Recent and upcoming publications include *Unbearable States: Digital Media, Cultural Activism and Human Rights* (forthcoming, 2021); *Displacement Activities: Contemporary Art and the Refugee Condition* (Sternberg Press, 2020); *Critique in*

*Practice: Renzo Martens' Episode III* (Sternberg Press, 2019); *Don't Shrink Me to the Size of a Bullet: The Works of Hiwa K* (Walther König Books, 2017); *Future Imperfect: Contemporary Art Practices and Cultural Institutions in the Middle East* (Sternberg Press, 2016); and *Art and Politics Now* (Thames and Hudson, 2014).

