Rebecca Uliasz
PhD Student, Computational Media, Arts & Cultures, Duke University
**work in progress: please do not circulate without authors permission**

**Seeing Like a Computer: From Probability to Pattern (talk version)**

In dialogue with our organizing themes, I want to tentatively think about agency as the capacity to make decision, or the ability to decide in the face of something incommensurable or incalculable. We will come back to that but what is at stake in my critique is foregrounding a *different temporality* of decision making, unmoored from human semiosis, operating preemptively to collapse speculated futures into the present. Thus, what is at stake in computational decision-making structures, aka algorithms, is the total subsumption of agency through the automation of decision, and thus the automation of politics altogether.

In honing in on ImageNet in its specificity, I will show, by way of example, the relationship between automation and agency given through algorithmic governance as one that is better understood through its mechanization of computational time and space as an infrastructural, rather than visual, object of analysis.

**Part 1: ImageNet**

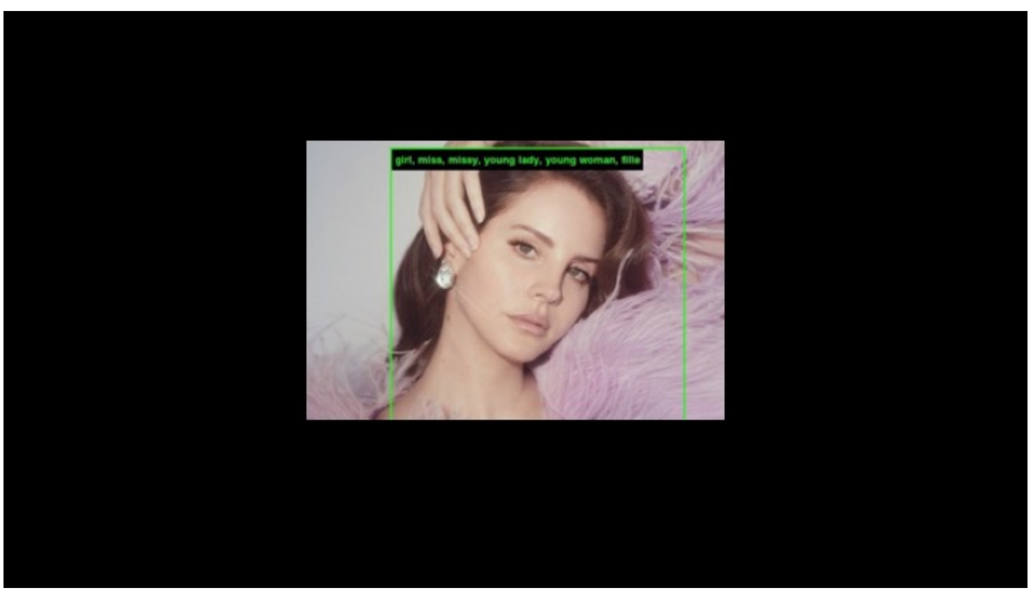
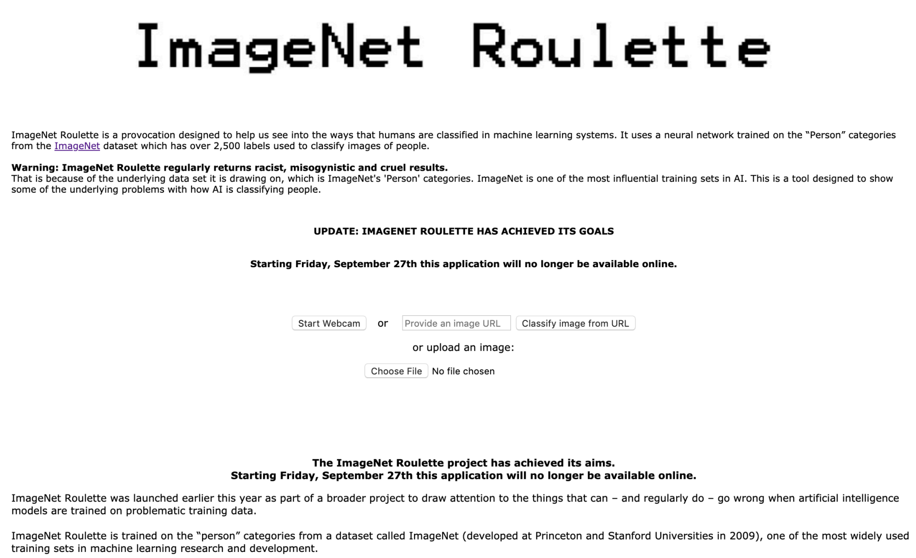


Fig 1. Image classified through Paglan and Crawfords *ImageNet Roulette*

In late September, we all opened our twitter feeds to find an abundance of selfies, a rather unnotable occurrence apart from the fact that our friends faces were bound by a thin green box, stamped with a single noun in the top left corner. A friend's pouty mouth selfie is overlaid with the term "creep"—is this meant to describe her?

Further investigation reveals that this social media fad is part of a larger project on artificial intelligence and their associated data sets by researcher Kate Crawford and artist Trevor Paglan. Paglan and Crawford created a website application, called ImageNet Roulette, that implements computer vision algorithms trained on the enormous dataset ImageNet, a massive taxonomy of labeled images including all variety of things from cats to cupcakes to corvettes to creeps. ImageNet contains over 20000 categories each housing hundreds of tagged images in a nested hierarchy. Since 2010 it has become arguably the most important data set for training object recognition algorithms.


ImageNet Roulette

ImageNet Roulette is a provocation designed to help us see into the ways that humans are classified in machine learning systems. It uses a neural network trained on the "Person" categories from the ImageNet dataset which has over 2,500 labels used to classify images of people.

**Warning: ImageNet Roulette regularly returns racist, misogynistic and cruel results.**
That is because of the underlying data set it is drawing on, which is ImageNet's 'Person' categories. ImageNet is one of the most influential training sets in AI. This is a tool designed to show some of the underlying problems with how AI is classifying people.

**UPDATE: IMAGENET ROULETTE HAS ACHIEVED ITS GOALS**

**Starting Friday, September 27th this application will no longer be available online.**

Start Webcam or Provide an image URL Classify image from URL

or upload an image:

Choose File No file chosen

**The ImageNet Roulette project has achieved its aims.**
**Starting Friday, September 27th this application will no longer be available online.**

ImageNet Roulette was launched earlier this year as part of a broader project to draw attention to the things that can – and regularly do – go wrong when artificial intelligence models are trained on problematic training data.

ImageNet Roulette is trained on the "person" categories from a dataset called ImageNet (developed at Princeton and Stanford Universities in 2009), one of the most widely used training sets in machine learning research and development.

Fig 2. ImageNet Roulette homepage, now deactivated

Paglan and Crawford are interested in the "person" category of ImageNet, not oft used in object recognition and for good reason—this category contains thousands of human classifications from "redneck" to "drug addict" to "spinster" presented in a taxonomic form that suggests the validity of the labels and their relations to the images. Their critique hinges on the assertion that taxonomies themselves contain a politics—the “sets” in a dataset assert a claim to normative principles through the division of types at the exclusion of outlier information that would threaten their integrity. For example, the category “adult body” contains a subset for adult male body and adult female body, asserting an underlying claim that an adult body may be gendered in one of two ways. The issue with this logic, in their words “to name a thing is in turn a means of reifying the existence of this category,” an assertion of an epistemological truth that allows for concretization of racist, sexist, or xenophobic judgements. We might see the mass circulation of the open source ImageNet and its incorporation in machine learning algorithms as a continuation of oppression through algorithmic means.

Paglan and Crawford describe their methods of excavating the big dataset as “*opening up the black box to look at how these ‘engines of seeing’ currently* ***operate***.” It contains what Paglan calls *invisible images*, defined as the means through which machines automate social organization, exclusive of the human subject. In what follows, I want to call into question the assumptions that are made in the “opening up the black box proposition”.

To some, the proposition of a world of “machine to machine” images suggests that machine learners communicate at levels beyond the human sensory threshold, and thus, are inaccessible to us. I seek to trouble the polarizing discourse on the opacity of machine learning. Here, a list of terminology—opacity, concealment, perception—is inexorably tied to the register of humanist reason rooted in the primacy of vision, the ancient associations between vision and knowledge production, between knowledge and power. Computer vision, the algorithmic process of making meaning from digital images or visual information, is a site where the multiple meanings of “vision” are conflated, problematizing critical analysis of the broad social and political operations of these technologies. Clarification of terms is necessary to critique the politics of automated cognition in its ability to authorize and transform techniques of governance.

**Part 2: The Operational Image**

The digital image has come into a new relationship with power with the advance of machine learning. In the following I will argue that these images are not apprehended visually, but rather organize our logics of perception through patterning, structuring what can and cannot be seen. The digital image, ironically, authorizes techniques to make things disappear, or make you see things that quite literally do not exist-- it is often written off for excluding the human altogether. Rather, I will argue that the digital image that feeds artificial intelligence is what film-maker Harun Farocki calls an *operative image*, it does not represent an object foreclosed to human perception, but is part of an operation, it is functional.

The operative image is crucial to computer vision, an integral component of *deep learning,* which requires vast troves of perceptual information to train algorithms to make decisions. It is a precondition for deep learning because it quantifies the human sensorium through enumerating visual information. It acts as a mediator between lived social experience and the algorithmic process of deep learning, which ironically operates on a GPU but in a markedly non-optical mode. Big data make it possible for authority, ie those that own the means of computer vision, to execute power over subjects, not through "knowing" a population, but rather through the operation of *extraction*. To frame computer vision as an extractive operation suggests that an integral dimension of computer vision is its ability to produce future patterns from non-patterned information. In other words, the patterning of information is a means for computer vision to continuously invent a new "outside" for further capital appropriation. This productive capacity is a means for power to literally make meaning, and therefore value, out of any social relation whatever.

What do these images have in common? *Find out!*

Fig 3. ImageNet search engine, ImageNet home page

Explains Nora Khan, machine vision algorithms operate through a logic of "seeing, naming, knowing," whereby *naming*, the ability to "correctly" identify and label an image in an efficient and accurate manner (accuracy in this sense being whatever was deemed so by the programmer) is collapsed with *knowing* or “making meaning from.” But from what position does one ground critique in the face of malleable meaning? If machine vision algorithms in fact *produce* meaning, I want to shift the site of critique to the technicity of the algorithm itself. What sorts of sovereign techniques does computer vision make possible when its in motion?

**Part 3: The Financial Image**

Seeing like a computer requires what Louise Amoore calls a *risk calculus*—a derivative form used by state power that incorporates uncertainty in order to array possible futures. Amoore explains that techniques of governance have shifted post 9/11 to algorithmic means of calculating possible futures states—a sensibility shot through with catastrophe gave authority to new calculation techniques to incorporate unaccountable contingencies. Amoore explains that in the absence of sufficient data, the security algorithm “if a and b, in association with c, then x” is used to ontologically associate unknown values. This abstraction involves a continuous disaggregation and reaggregation of noncausal data, constantly shuffling the relations until a pattern deemed significant emerges.

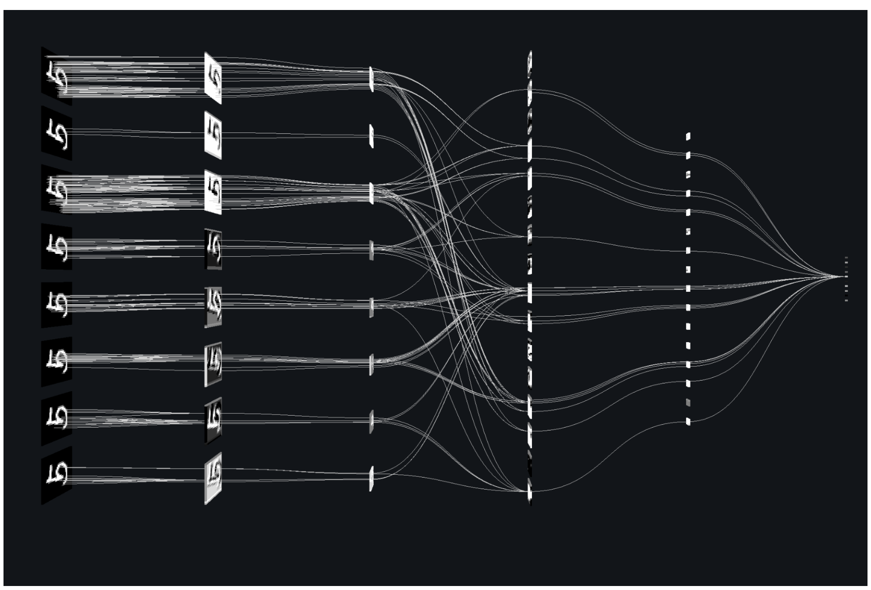


Fig 4.

Computer vision algorithms are ripe with this calculus of abstract association. Deep learning, as described by Ian Goodfellow, is an advanced form of machine learning that uses a layered neural network to perceptualize patterns in order to train a classification algorithm. Writes Goodfellow, deep learning techniques are intended to simulate the intuitive reasoning methods that humans use to make decisions in the real world, where complex variables and incomplete information negate the possibility of relying on a precodified means for arriving at complex decisions. Such algorithms require the "ability to reason in the presence of uncertainty" – where there is unobservable or incomplete information, deep learning algorithms rely on random variables derived from probability theory.(151)

One type of algorithm often used in image recognition applications is called a Convolutional Neural Network, which specializes in processing data in a grid like topology. While explaining the specificity of these algorithms is beyond the scope of this paper, important to their design is the use of convolutional layers called tensors that are used to weight input data according to a specific *kernel*, or array of parameters, referred to as the *bias*, that is used to produce what is called a feature map. Other layers of the neural net perform an operation called pooling, which downsamples information to reduce dimension, retaining only the features deemed most important to the algorithm. In sum, input data is initialized with random weights and processed through a set number of convolutional and pooling layers, and output is compared to the expected result to calculate error. This calculated error is then propagated backward into the hidden layer(s) in order to tune the algorithm, the process repeated a specified number of times or until the desired error rate is achieved. Crucial to this operation is the introduction of randomness—the multiplication of data by any number of possible states—in order to infer patterns within information. In a way, CNN algorithms act as a playground for programmers to tweak the variables of weighted layers until the economically precise outcome is derived.

Algorithms fracture and fragment a dataset, to the extent where it has very little to do with its underlying referent. It is recombined and projected forward to the effect that a new visualization is made to emerge—a diagram of the lines that cut across datasets rather than what is contained within. Computer vision algorithms are programmed to reduce noisy real-world information into patterned forms, overcoming problems of occlusion or missing information—they fill in what cannot be seen by the machine eye. Through the visualization of *something that* *literally does not exist in the collected dataset,* it fills in a gap—the preemptive techniques of computer vision act through the very contingent relations it produces.

**Part 4: I thought I was Seeing Patterns**

Machine learners are used as instruments to make decisions on possibly meaningful correlations. Thus, the sovereignty of the dataset is less about what it contains, and who named it than how it is instrumentalized. In a more worrisome sense, it does not actually matter if your data is even in the set, as risk is not calculated through what is already known.

I want to suggest that the “vision” in computer vision is misunderstood, or at least conceptually misconstrued in the narrow sense, as the faculty of sight. A more nuanced position might even include the ways that the field of vision is often of discursive inheritance, structured by our practices of being and knowing the world. The seeing subject makes a cut in the fabric of the Real. “Vision” in another sense, can also mean “mastery”, an invocation of the all-seeing eye that has the privilege of complete knowledge over subjects surveilled from above. “Vision” however, can also mean discernment, intelligent foresight, or seeing something that is not actually present. vision in another sense is in the future tense, it has a future function, it functions to call from the future, like a prophet or medium. This type of vision, associated with the divine, is divorced from what we might think of as the function of the human visual apparatus—to perceive what is made present to us. Machine learners don’t see, they have visions.

The interplay between big data and deep learning algorithms consists of a number of codified techniques of abstraction used to simplify complexity and account for limited resources so that an economic solution to a problem might be found. They automate a visual economy, surfacing as visualizations that we assign meaning to-- what we come to know as the digital image. This image isn’t one in which we’ve been left out of the loop, but are thoroughly entangled within it, governed through its possible expressions.

Big data sets, like ImageNet, undergird an operational system, comprised of logistical components, governed by extractive dynamics, expressing a preemptive temporal logic that calls from the future. Big data might be thought of as part of a system of operations that at many different moments, performs labor on material to transform it into patterned information, data as commodity form. How might we ground a critical analysis of big data from within a critique of capitalism? Perhaps we need to begin by questioning the presuppositions that are embedded within humanist reason, the givens contained within our methods of deferring to technological solutionism, in order to develop practices for radically inhuman critique.

Works Cited


Amoore, Lousie. *The Politics of Possibility: Risk and Security Beyond Probability*. Durham, NC : Duke University Press , 2013.

Farocki, Harun. "Phantom Images." *PUBLIC* 29 (2004): 12-22.

Goodfellow, Ian, Yoshua Bengio and Aaron Courville. *Deep Learning*. MIT Press, 2016.

Khan, Nora. *Seeing, Naming, Knowing*. New York: School of Visual Arts, 2019.

Mezzadra, Sandro and Brett Neilson. "Extraction, logistics, finance." *The Politics of Operations*. Durham: Duke University Press, 2018.

Paglan, Trevor and Kate Crawford. *Excavating AI: The Politics of Images in Machine Learning Training Sets*. 2019. 2019. <https://www.excavating.ai/>.

Paglan, Trevor. "Invisible Images (Your Pictures Are Looking at You)." 8 December 2016. *The New Inquiry.* May 2019.