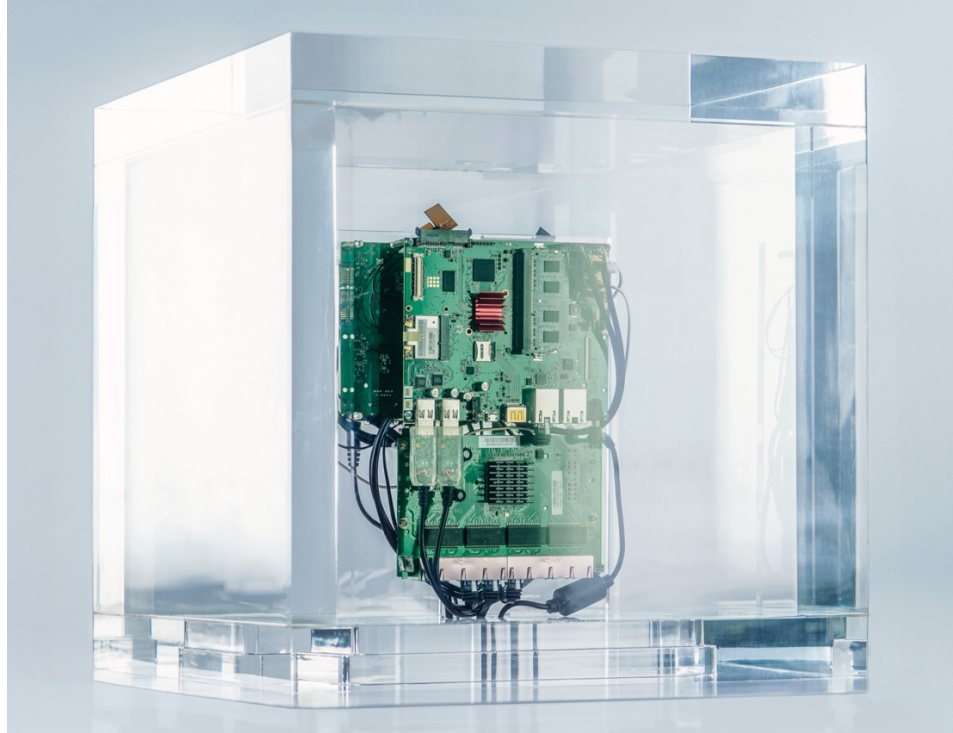


Forbes

Trevor Paglen: "To Expose The Bigotry Of AI, Artist Trevor Paglen Is Putting Computers On Trial"

By Jonathon Keats

January 12, 2021



Trevor Paglen. *Autonomy Cube*, 2015. plexiglass box, computer components. Courtesy of the artist and Metro Pictures, New York, HPI.2020.01.
Trevor Paglen

Back in 2007, a young computer scientist named Fei-Fei Li developed a new curriculum for artificial intelligence. Her curriculum wasn't intended to teach humans about AI, but vice versa. Li conceived a way for AI to learn about us.

Her idea was to create the machine-readable equivalent of an abecedarium. Engaging the nascent gig economy through Amazon Mechanical Turk, she hired hundreds of people to label a vast repository of 3.2 million images, identifying their content so that an AI would learn to recognize other instances. Within a couple years, ImageNet became the ultimate dataset, the 21st century equivalent of *The New-England Primer*.

A decade later, the artist Trevor Paglen brought Li's project full circle by transforming ImageNet into an educational tool for humans. Working with the computer scientist Kate Crawford, he created *ImageNet Roulette*, an app that engaged artificial intelligence to label uploaded selfies. People's response ranged from puzzlement to fury. An Asian woman was identified as a Jihadist, and an African-American man was labeled a wrongdoer. Through *ImageNet Roulette*, people learned what AI had been learning about people, which turned out often to be offensively inaccurate.

**JESSICA
SILVERMAN**

621 Grant Avenue, San Francisco, CA 94108
jessicasilvermangallery.com +1 415 255 9508

Trevor Paglen is to artificial intelligence what Upton Sinclair was to meatpacking. An important exhibition at the Carnegie Museum of Art shows some of what he's uncovered. Equally important, the exhibit provides an introduction to his distinctive form of muckraking.

Paglen would do well as a conventional investigative journalist. Using public records and high-powered optics – and occasionally even scuba gear – he has exposed reconnaissance satellites and the secret “chokepoints” where the National Security Agency taps into the global telecommunications infrastructure.

Paglen made his reputation spying on the spies. However his approach to AI is simultaneously more subtle and more incisive. Instead of investigating the people and companies and government agencies that have made artificial intelligence what it is today, Paglen interrogates AI itself. He does so by putting artificial intelligence to work on problems that reveal ways in which AI is problematic through the results the software generates: a sort of automated self-incrimination.

Instructing a system trained on ImageNet to label people, and showing people what the system sees, is one example of Paglen's cyber-muckraking. Another case, more ambiguous on first viewing, involves image-processing of natural phenomena by computer vision algorithms. For instance, *CLOUD #902 Scale Invariant Feature Transform; Watershed*, shows a storm cloud overlaid with what appears to be a random assortment of geometric shapes. In fact the circles and lines are marks of the AI struggling to process atmospheric conditions using the training it received in an industrial setting.



Trevor Paglen. *CLOUD #902 Scale Invariant Feature Transform; Watershed*, 2019. dye sublimation on wallpaper. Courtesy of the artist and Metro Pictures, New York, HPI.2020.16

Trevor Paglen

**JESSICA
SILVERMAN**

621 Grant Avenue, San Francisco, CA 94108
jessicasilvermangallery.com +1 415 255 9508

The intimations are related to the implications of *ImageNet Roulette*, a permutation on the old programming adage, "garbage in, garbage out". However the circumstances have changed significantly since the maxim became popular in the '60s. Although mid-century computers already had significant power, controlling industrial processes and military weaponry, the systems were self-contained and the code was explicit. Today a dataset such as ImageNet is ubiquitous and deeply entrenched in the decisions made by algorithms far and wide, which may be labeling Asians as Jihadists and African-Americans wrongdoers – and acting accordingly – without anyone knowing it. Other algorithms, inappropriately trained, may catastrophically seek to 'improve' ecosystems by viewing them as factories.

In computer science today, there is a tendency to express social awareness by referring to AI as a black box, a system that makes decisions accountable to nobody because the internal reasoning of the circuitry is unknowable. This critique is typically a prelude to surrender – perhaps with the suggestion that black boxes be permeable in the future – a dangerous and morally dubious position given all the damage AI can do today.

Paglen shows that the black box can be effectively accosted if not fully illuminated, much as humans can be questioned or even tried in court without having their brains dissected and neural networks mapped. He also shows that the errors of algorithms typically reflect the prejudices and poor judgment of the people who educate them. In artificial intelligence, humans are the chokepoint.

Spying on spies requires guile, but when it comes to understanding AI, interrogation can begin without special training in computer science. It simply demands that we more closely examine ourselves.

JESSICA
SILVERMAN

621 Grant Avenue, San Francisco, CA 94108
jessicasilvermangallery.com +1 415 255 9508