laFuga

Imágenes invisibles

(tus fotografías te miran)

Por Trevor Paglen

Tags | **cultura digital | Cultura visual- visualidad | Técnica | artes visuales | Estudios visuales | Estados Unidos**

Trevor Paglen es un artista, geógrafo y teórico estaunidense. Su trabajo abarca la creación de imágenes, la escultura, el periodismo de investigación, la escritura, la ingeniería y muchas otras disciplinas. Entre sus principales preocupaciones están aprender cómo ver el momento histórico en el que vivimos y desarrollar los medios para imaginar futuros alternativos. Paglen ha expuesto su creación en el festival transmediale.08,Berlín; The Andy Warhol Museum, Pittsburg; Institute of Contemporary Art, Filadelfi a; Yerba Buena Center for the Arts, San Francisco; Kunstraum München; y el Massachusetts Museum of Contemporary Art, North Adams. Entre otros libros ha publicado Torture Taxi: On the Trail of the CIA's Rendition Flights (escrito junto a AC Thompson; Melville House, 2006), que describe el "extraordinario programa de interpretación" de la CIA; I Could Tell You But Then You Would Have to be Destroyed by Me (Melville House, 2007), en donde examina la cultura visual de los llamados programas militares "negros" o secretos; y Blank Spots on a Map (Dutton/NAL/Penguin, 2009).

I.

Nuestros ojos son objetos corpóreos. Asimismo, la historia de nuestra cultura visual se ha construido de objetos corpóreos. La historia de las imágenes es una historia de pigmentos y tinturas, óleos, acrílicos, nitrato de plata y gelatina – materiales que uno puede utilizar para pintar una caverna, una iglesia o una tela. O para producir una fotografía o imprimir imágenes en las páginas de una revista. La emergencia de la pantalla en la segunda mitad del siglo XX no es la excepción: tubos de rayos catódicos y monitores de cristal líquido emiten luz en una frecuencia que nuestros ojos perciben como colores, y con una densidad que percibimos como formas.

Hemos desarrollado una buena comprensión de las complejidades de la visión humana; de las formas sinuosas a través de las cuales las imágenes se infiltran y determinan la cultura, su tenue relación con la vida cotidiana y con la verdad, y del modo en que son utilizadas para servir (y resistir) al poder. Los conceptos teóricos que utilizamos para analizar la cultura visual clásica son robustos: representación, sentido, espectáculo, significación, mimesis, etc. Por siglos, estos conceptos nos han ayudado a comprender el funcionamiento de la cultura visual.

Pero en la última década un cambio radical está aconteciendo. La cultura visual está cambiando de forma. Se está distanciando del ojo humano y tornando invisible. La cultura visual humana deviene así solo una forma particular de visión, una excepción a la regla. La mayor parte de las imágenes está siendo producida por máquinas y para máquinas, donde el ojo humano rara vez forma parte del proceso. La emergencia de la visión maquínica [1] ha sido no sólo poco examinada sino también mal comprendida por aquellos que hemos comenzado a notar los cambios que están aconteciendo ante nuestros ojos.

Las imágenes invisibles y la visión maquínica se están tornando elementos cada vez más presentes. Su rápida expansión comienza a tener un profundo impacto en la vida humana, eclipsando incluso los cambios traídos por la cultura de masas del siglo XX. Las imágenes empiezan a intervenir en la vida cotidiana ya no a través de la representación y la mediación, sino a través de operaciones, activaciones y ejecuciones. Las imágenes invisibles nos están constantemente observando, examinando e incitando, guiando nuestros movimientos, infligiendo dolor e induciendo placer. Pero todo esto es difícil de ver.

Los críticos culturales han sospechado por largo tiempo que existe una diferencia radical entre las imágenes digitales y la cultura visual del ayer. Sin embargo, han tenido dificultades para definir exactamente esta diferencia. En los 1990, por ejemplo, se habló mucho de la falta de un "original" como característica central de la imagen digital. Más recientemente, la proliferación de imágenes en las redes sociales y sus consecuencias para las relaciones inter-subjetivas se han convertido en un tema de extensa discusión entre los teóricos de la cultura y sus críticos. Pero estos análisis son incapaces de definir de manera precisa los cambios que están ocurriendo.

Uno de los problemas es que estos análisis siguen asumiendo que las imágenes son algo que los humanos perciben y, por ello, que la relación entre el espectador humano y la imagen es el momento crucial por analizar – pero es precisamente esta presuposición de un espectador humano lo que quisiera cuestionar aquí.

Lo que es realmente revolucionario de la emergencia de las imágenes digitales es el hecho de que ellas son antes que nada legibles por una máquina: ellas pueden ser percibidas por un sujeto humano solo bajo circunstancias particulares y por breves periodos de tiempo. Una fotografía tomada con un teléfono celular genera una imagen legible maquínicamente [2] que no refleja la luz de modo tal que pueda ser percibida por un ojo humano. Solo secundariamente, un software particular conectado a una pantalla de cristal líquido podrá generar una imagen perceptible para el ojo humano. Pero esta imagen solo se le aparece al ojo humano de manera temporal antes de volver a su forma maquínica inmaterial una vez que la pantalla del teléfono celular se apaga. Sin embargo, esta imagen no requiere ser legible por un ojo humano para que la máquina pueda procesarla. Esto define una diferencia radical respecto a las imágenes contenidas en un rollo de película sin revelar. Si bien el rollo de película debe también ser procesado para tornarse visible al ojo humano, el rollo sin revelar no es legible ni por un humano ni por una máquina.

El hecho de que las imágenes digitales sean antes que nada legibles maquínicamente con independencia de un sujeto humano conlleva profundas consecuencias. Permite la automatización en serie de la mirada y, con ello, el ejercicio del poder a escalas mucho mayores (y mucho menores) de lo que era posible hasta ahora.

II.

Nuestro entorno está lleno de ejemplos de aparatos de visión maquínica: lectores automáticos de patentes de automóviles (ALPR) montados en autos policiales, puentes, edificios, y autopistas toman fotografías de cada automóvil que entra en su campo de visión. Empresas como *Vigilant Solutions* que ofrecen esta tecnología utilizan las cámaras para recolectar información sobre la ubicación de cada automóvil, utilizan software de reconocimiento de texto (OCR) para archivar los números de las patentes, y generan bases de datos que son luego utilizadas por la policía y las compañías aseguradoras, entre otros. [3] En la esfera del consumo, equipamiento como el ofrecido por *Euclid Analytics* y *Real Eyes*, entre muchos otros, permiten instalar cámaras en los centros comerciales para rastrear el movimiento de los consumidores, utilizando software específico para analizar quién mira qué producto y por cuánto tiempo. Además, utilizan software de reconocimiento facial para analizar las expresiones de los consumidores y con ello inferir los estados de ánimo de los sujetos observados. La publicidad también ha comenzado a observar y registrar a sus espectadores. Y en el sector industrial, compañías como *Microscan* ofrecen sistemas integrales de imágenes diseñados para detectar defectos en el proceso productivo o en los materiales, y para supervisar el empaque, transporte, y logística de las industrias de automóviles, farmacéuticos, electrónica, etc. Todos estos sistemas son posibles sólo gracias al hecho de que las imágenes digitales son legibles maquínicamente y no requieren de un ojo humano como parte del proceso.

La cultura visual invisible no se limita a la esfera industrial, legal, o de las ciudades "inteligentes", sino que se extiende con fuerza hacia el campo que de otro modo – e ingenuamente – llamaríamos el campo de la cultura visual humana. Me refiero a los trillones de imágenes que los humanos comparten en las plataformas digitales – plataformas que en primera instancia pueden parecer como plataformas hechas por humanos para otros humanos.

En la superficie, una plataforma como *Facebook* parece análoga a los álbumes de fotos tan típicos de la cultura del siglo XX. "Compartimos" fotografías en internet y vemos cómo la gente les pone "me

gusta" y las "comparte". Antiguamente, la gente llevaba consigo una fotografía de sus hijos en su billetera, y mostraba esta fotografía a sus amigos y conocidos, u organizaba una muestra de diapositivas de las vacaciones familiares. ¿Qué puede ser más humano que querer mostrar a los hijos de uno? Las plataformas diseñadas para compartir imágenes digitales en parte imitan estas formas arcaicas, creando "álbumes" para nuestras "selfies", fotos de bebés, gatos, y viajes.

Pero la analogía es engañosa, ya que cuando compartimos una imagen en *Facebook* sucede algo totalmente diferente que cuando aburrimos a nuestros vecinos con una serie de diapositivas. Cuando compartimos una imagen en *Facebook* o en otra red social, estamos alimentando un poderoso sistema de inteligencia artificial con información sobre cómo identificar gente y cómo reconocer lugares y objetos, hábitos y preferencias, aspectos de raza, clase y género, estatus económico, y mucho más.

Independientemente si un sujeto humano vea o no los 2 billones de imágenes que se suben diariamente a *Facebook*, las fotografías en las redes sociales son examinadas por redes neuronales que harían sonrojar al más experto historiador del arte. El algoritmo "Deep Face" de *Facebook*, desarrollado entre 2014 y 2015, genera imágenes abstractas tridimensionales de los rostros de los individuos y utiliza redes neuronales que llegan al 97% de precisión a la hora de identificar individuos – un porcentaje comparable con la precisión del ojo humano, sin considerar que ningún ojo humano podría llegar a reconocer los rostros de billones de personas.

Existen muchos otros: "Deep Mask" de *Facebook* y "Tensor Flow" de *Google* identifican personas, lugares, objetos, ubicaciones, emociones, gestos, rostros, géneros, estatus económicos, relaciones, y muchos otros aspectos.

Adicionalmente, los sistemas de inteligencia artificial se han apropiado de la cultura visual humana y la han transformado en una herramienta gigantesca de entrenamiento. Mientras más imágenes los sistemas de *Google* y *Facebook* ingieran, mayor será la precisión de sus algoritmos, y mayor influencia tendrán en nuestra vida cotidiana. Los trillones de imágenes que nos hemos acostumbrado a considerar como parte de nuestra cultura humana son el fundamento de una automatización de la mirada que poco tiene que ver con la cultura visual del pasado.

III.

Si echamos un vistazo al funcionamiento interno de los sistemas de visión maquínica encontraremos una colección de abstracciones complemente ilegibles por el ojo humano. El territorio de la visión maquínica no es un territorio de representaciones sino un territorio de operaciones y ejecuciones. Está constituido por relaciones performativas más que por las clásicas relaciones representacionales. Pero esto no significa que no exista un fundamento formal al modo cómo funciona la visión maquínica. Todos los sistemas de visión maquínica producen una abstracción matemática de las imágenes, y las características de estas abstracciones son determinadas por el tipo de metadatos que el algoritmo trata de leer. El reconocimiento facial, por ejemplo, implica una serie de procesos técnicos dependiendo de su aplicación, la efectividad deseada, y la disponibilidad de las series de datos utilizadas para su "entrenamiento". La técnica "Eigenface", para dar otro ejemplo, analiza un rostro a través de la substracción de todas las características que son comunes con otros rostros, dejando solamente una "huella digital" de aquello que es único en un determinado rostro. Para reconocer a una determinada persona, el algoritmo buscará la "huella digital" de su rostro.

Las redes neuronales convolucionales (CNN), más conocidas como redes de "deep learning", están constituidas por docenas o incluso cientos de capas internas de software que comparten información unas con otras. Las primeras capas del software desmenuzan una imagen separando formas, gradientes, luminosidades, y ángulos. Estos elementos individuales son luego articulados como formas sintéticas. En las capas más profundas de la red, la imagen sintética producida por las primeras capas es comparada con otras imágenes sintéticas con las cuales la red ha sido entrenada, activando las "neuronas" cuando el software encuentra similitudes, y con ello posibilitando su "reconocimiento".

Podríamos pensar en estos procesos sintéticos y otras estructuras "alucinatorias" al interior de las redes neuronales como una forma de arquetipo Jungiano, una forma de inconsciente colectivo de la

inteligencia artificial – una metáfora atractiva, pero engañosa. Las redes neuronales no pueden inventar sus propias categorías; solo pueden relacionar las imágenes que procesan con las otras imágenes con las que han sido entrenadas. Y este proceso de entrenamiento revela el carácter histórico, geográfico, racial y socioeconómico de sus entrenadores. Si damos una imagen de la "Olympia" de Manet a una red neuronal entrenada con el set "Imagenet", el software reconocerá la imagen como un "burrito". No hace falta mencionar que el clasificador "burrito" es específico a una persona joven viviendo en la zona de San Francisco, donde el burrito estilo "misión" [4] fue inventado. Basta observar brevemente el funcionamiento de las redes neuronales para darse cuenta de que una imagen de alguien sosteniendo algo en su mano será probablemente identificada como un teléfono celular o un controlador de *Nintendo Wii*. Desde una perspectiva más delicada, los ingenieros de *Google* han decidido desactivar el clasificador "gorila" luego de que se hizo evidente que los algoritmos entrenados principalmente con rostros de gente blanca tendían a identificar a los afroamericanos como simios.

El punto es que si queremos entender el mundo invisible de la cultura visual maquínica debemos desnaturalizar nuestra visión humana. Debemos aprender a ver un mundo paralelo, compuesto de activaciones, puntos clave, "eigenfaces", clasificadores, series de entrenamiento, etc. Pero no se trata simplemente de aprender un nuevo vocabulario. Los conceptos formales cargan con presuposiciones epistemológicas, lo que a su vez conlleva consecuencias éticas. Los conceptos teóricos utilizados para analizar la cultura visual humana son insuficientes a la hora de analizar el nuevo territorio maquínico. Esto produce distorsiones, puntos ciegos, e interpretaciones erróneas.

IV.

Existe una tendencia a criticar la lectura algorítmica de imágenes a partir de su falta de precisión – que la "Olympia" sea identificada con un "burrito", y que un afroamericano sea clasificado como un gorila. Estas críticas son fáciles, pero equivocadas. Sugieren simplemente que el problema es un problema de precisión, el cual se solucionaría con más y mejor entrenamiento de las redes neuronales. Basta eliminar los prejuicios de las series de entrenamiento, nos dicen, y el reconocimiento algorítmico de imágenes será inmediatamente menos racista que las interacciones entre humanos. Basta que programemos el algoritmo para que nos vea a todos por igual y los humanos observados no podrán sino ser iguales. Pero este razonamiento no me convence.

La astucia final de la ideología consiste en presentarse a sí misma como verdad objetiva, presentar las condiciones históricas como eternas, y presentar la organización política como un fenómeno natural. Dado que las operaciones algorítmicas operan en un plano invisible y no dependen de un observador humano (y por ello no son explícitamente ideológicas como un cuadro de Napoleón) es más difícil verlas por lo que son: mecanismos de regulación social extremadamente poderosos que sirven intereses de clase y de raza determinados pero que se presentan como una tecnología objetiva.

El mundo invisible de las imágenes no es simplemente un nuevo sistema de taxonomía visual. Es un ejercicio concreto, sagaz, de poder. Un ejercicio ajustado perfectamente a la policía molecular y al movimiento de los mercados – y diseñado para introducir sus tentáculos en aspectos cada vez más ínfimos de nuestra vida cotidiana.

Tomemos el caso de *Vigilant Solutions*. En enero de 2016, *Vigilant Solutions*, la compañía que se jacta de contar con una base de datos de billones de locaciones de automóviles gracias a su sistema de ALPR, firmó un contrato con un puñado de gobiernos regionales en Texas. Como demuestran ciertos documentos obtenidos por la fundación *Electronic Frontier*, el acuerdo fue el siguiente: *Vigilant Solutions* le ofreció sistemas de lectura de patentes de automóviles (ALPR) a la policía local y acceso a su gigantesca base de datos. A cambio, los gobiernos regionales compartieron con *Vigilant Solutions* los archivos de multas impagas y órdenes de arresto. Una lista de patentes "marcadas" asociadas a multas impagas es introducida en el sistema automático de reconocimiento de patentes. Cuando uno de estos sistemas montado en un vehículo policial identifica una patente "marcada", el policía detiene el automóvil y le da dos opciones al conductor: puede pagar in-situ la multa impaga con su tarjeta de crédito (con una comisión del 25% para *Vigilant Solutions*), o puede ser arrestado. Además de este 25%, *Vigilant Solutions* mantiene un registro de cada patente identificada por su sistema

automático montado en los vehículos policiales, incrementando su base de datos la cual es luego monetizada por otros canales. La operación política aquí queda clara. Las municipalidades son incentivadas a incrementar su presupuesto a costa de sus ciudadanos más vulnerables, transformando a los oficiales de policía en recolectores de impuestos, y vendiendo la información policial a compañías privadas. A pesar de la "objetividad" del sistema, no hay duda de que sirve intereses corporativos y gubernamentales a expensas de la población más vulnerable y de la comunidad civil.

Dado que los gobiernos buscan nuevas formas de ingreso en una era de recortes presupuestarios, y dado que el capital está permanentemente buscando nuevos aspectos de la vida cotidiana a colonizar, la capacidad para utilizar sistemas algorítmicos de reconocimiento de imágenes para extraer plusvalía de porciones cada vez más pequeñas de nuestra vida es simplemente irresistible. Es fácil imaginar, por ejemplo, un algoritmo de *Facebook* capaz de reconocer a una mujer menor de edad tomando cerveza en la fotografía de una fiesta. Esa información será enviada a la compañía de seguros del automóvil de la mujer, quien ha aceptado alguna aplicación de *Facebook* diseñada para entregar este tipo de información a agencias de seguros, agencias publicitarias, servicios gubernamentales, de impuestos, y policiales. El precio de su seguro de automóvil será reajustado con esta nueva información. Un segundo algoritmo podría revisar su historial de fotografías en busca de conductas similares las cuales podrían ser también traducidas en plusvalía. En el mundo de la cultura visual humana, la fotografía responsable de toda esta cadena de consecuencias hubiese sido olvidada en alguna caja de zapatos. En la cultura visual maquínica esa fotografía no desaparecerá más. Se transforma en un participante activo de su vida, con consecuencias a largo plazo.

Fragmentos cada vez más pequeños de nuestra vida humana están siendo capturados por el capital, sea por la capacidad para escanear automáticamente miles de patentes de automóviles en busca de multas impagas, o por un momentáneo descuido capturado por una fotografía subida a internet. Nuestro seguro de salud será determinado por nuestras fotografías infantiles subidas a internet por nuestros padres sin nuestro consentimiento. El grado de vigilancia policial que recibamos será determinado por la huella digital de nuestro estilo de vida.

La relación entre imagen y poder en el nuevo contexto de visión maquínica es diferente a la relación entre imagen y poder en el contexto de la visualidad humana. En el nuevo contexto se despliegan dos operaciones aparentemente opuestas. La primera consiste en la individuación y diferenciación de todas las personas, lugares, y vidas cotidianas a su alcance – utilizando metadatos para crear una "firma digital" de cada individuo basada en su raza, clase social, lugar de residencia, productos que consume, hábitos, intereses, "me gusta", amistades, etc. La segunda operación consiste en la reificación de estas categorías, eliminando cualquier ambigüedad en la interpretación de los datos de tal modo que los perfiles de metadatos puedan ser operativizados para recolectar impuestos, ajustar pólizas de seguros, ofrecer publicidad personalizada, reforzar la vigilancia policial, etc. El efecto general de estas operaciones es una sociedad que amplía la diversidad (al menos una diversidad de perfiles de metadatos) pero lo hace precisamente porque la diferenciación de perfiles de metadatos permite tanto generar plusvalía como reforzar el control policial de la vida cotidiana.

Los sistemas de visión maquínica son instrumentos de poder extremadamente íntimos que operan a través de una estética y una ideología de la objetividad, pero cuyas categorías están diseñadas para reificar determinadas relaciones de poder. En este sentido, el contexto de sistemas de visión algorítmica produce una especie de hiper-ideología que es particularmente peligrosa dada su supuesta objetividad e igualdad.

V.

Los productores culturales han desarrollado estrategias muy eficientes para cuestionar la desigualdad, el racismo y la injusticia en la cultura visual humana. Las estrategias visuales contra-hegemónicas utilizadas por los artistas y productores culturales en la esfera de la visualidad humana por lo general explotan las ambigüedades propias de la cultura humana para producir formas de contracultura – para reivindicar derechos, reclamar, o expandir el campo de representación de ciertos pueblos o posiciones políticas. El influyente trabajo de Martha Rosler "Semiótica de la

cocina", por ejemplo, presentó la imagen patriarcal de la cocina como una prisión; las imágenes de la resistencia y la solidaridad afroamericana de Emory Douglas crearon un nuevo escenario de empoderamiento; las imágenes de Catherine Opie de lo "queer" desarrollaron un vocabulario alternativo sobre el género y el poder. Todas estas estrategias, como muchas otras, se sostienen sobre el hecho de que la relación entre el significado y la representación es flexible. Pero esta idea de ambigüedad, pilar de la teoría semiótica de Saussure a Derrida, simplemente deja de existir en el nuevo contexto de visión maquínica. Esto se traduce en que no haya ninguna estrategia visual proveniente de la cultura visual humana que permita intervenir de manera evidente en el nuevo contexto de visión maquínica.

Enfrentados ante este desafío, algunos artistas y trabajadores culturales están intentando cuestionar los sistemas de visión maquínica a través de formas de ver que son legibles por el ojo humano pero ilegibles por las máquinas. El artista Adam Harvey, por ejemplo, ha desarrollado estilos de maquillaje que impiden el reconocimiento algorítmico del rostro, vestimentas que ocultan las señales infrarrojas de calor corporal, y bolsillos que bloquean la transmisión permanente de nuestra ubicación a través del teléfono celular. Julian Oliver desarrolla la estrategia opuesta, creando máquinas depredadoras de datos que revelan cuan inmersos en información y en sistemas de rastreo nos encontramos. Estos notables proyectos ayudan a los humanos a aprender sobre la existencia de una vigilancia maquínica generalizada. Pero estas estrategias no son universales.

A largo plazo, las estrategias visuales diseñadas para frustrar los sistemas de visión algorítmica están condenadas al fracaso. Ramas completas de la investigación de visión maquínica están dedicadas a crear imágenes "disidentes" que obstaculizan los procesos de reconocimiento visual. Estas imágenes disidentes son luego incorporadas a las series de entrenamiento de los algoritmos, los cuales aprenden cómo superar los distintos obstáculos al reconocimiento visual. Aún más, para poder realmente escondernos de los sistemas de visión maquínica requeriríamos tácticas que resistan no solo a los algoritmos actuales, sino también a los algoritmos que serán creados en el futuro. Para ocultar nuestro rostro de *Facebook*, una debería ser capaz de desarrollar estrategias que no solo impidan el funcionamiento del algoritmo "DeepFace" en su forma actual, sino de cualquier otro sistema de reconocimiento facial producido en el futuro.

Una resistencia real al control policial y a los poderes del mercado que están siendo ejercidos por los sistemas de visión maquínica no provendrá de una tecnología igualmente avanzada. A largo plazo no existe una solución técnica para la intensificación de las desigualdades políticas y económicas que las imágenes invisibles han posibilitado. Para resistir la optimización y depredación propias del nuevo contexto maquínico, uno debe generar ineficiencias deliberadas y aspectos de la vida fuera de la lógica del mercado y del control policial – "refugios" de la esfera digital invisible. Solo la ineficiencia, la experimentación, la autoexpresión, e incluso la violación de algunas leyes, podrán generar la libertad y la representación política necesaria para ello.

Ya no somos nosotros los que miramos a las imágenes – son las imágenes las que nos miran a nosotros. Las imágenes ya no simplemente representan cosas, sino que operan activamente en nuestra vida cotidiana. Debemos comenzar a comprender este cambio si queremos desafiar las nuevas formas de poder que circulan a través de la cultura visual invisible en la cual nos encontramos inmersos.

Notas

1

machine-to-machine seeing

2

machine-readable

3

El artículo "How Britian Exported Next-Generation Surveillance" de James Bridle es una excelente introducción de esta tecnología (N. del A.)

4

misión-style burrito

Como citar: Paglen, T. (2019). Imágenes invisibles, *laFuga*, 22. [Fecha de consulta: 2026-11-04] Disponible en: http://2016.lafuga.cl/imagenes-invisibles/944