

El futuro no ocurrirá, sucederá como una imagen: La distancia entre los objetos y sus representaciones artificiales

CHRIS HAMAMOTO & FEDERICO PÉREZ VILLORO

En una escena de la película de ciencia ficción de Stanley Kubrick *2001: Una odisea del espacio*, el “computador algorítmico programado heurísticamente”, mejor conocido como HAL, observa cómo el protagonista de la película, Dave Bowman, y su copiloto intentan tener una conversación secreta. HAL es omnipresente en la nave espacial y controla todas las funciones del Discovery One en su camino de la Tierra a Júpiter. Encerrados en una cápsula de “actividad extravehicular” más pequeña, los astronautas apagan todos los micrófonos dentro de la cámara y comprueban que no puedan ser escuchados. Tras determinar que HAL ya no podía oírles, comparten su preocupación de que la máquina esté funcionando mal, llegando a la conclusión de que, tal vez, es necesario apagarla. Sin que la tripulación lo sepa, HAL los *observa* mientras discuten su plan, analizando los movimientos de sus bocas para determinar lo que dicen. Al darse cuenta del peligro que corre, HAL sabotea a la tripulación desatando un conflicto en el que el protagonista logra desconectar la computadora para salvar su propia vida.

Al ser la primera vez que se retrató en una película el punto de vista de una máquina en primera *persona*, la imagen de HAL, amenazante pero sin forma más allá del *ojo rojo* de una cámara, estableció una plantilla para la manera en la que imaginamos la *visión* de las máquinas y nuestra relación con la inteligencia artificial en la cultura popular. Desde el pistolero de la película *Westworld* de 1973 que, en una representación sorprendentemente premonitoria, *veía* el mundo como fotogramas pixelados reminiscentes a los métodos modernos de segmentación de imágenes, hasta la película de 1991, *Terminator 2*, en donde el personaje T-800 de Arnold Schwarzenegger visualiza datos aumentados sobre su mirada. La mayoría de las impresiones en pantalla de la visión artificial son oclocéntricas, pero la forma en que las máquinas procesan información visual no coincide realmente con



En 2019, durante una residencia de investigación en OCAT en Shenzhen, Chris Hamamoto y Federico Pérez Villoro, en colaboración con el ingeniero Nic Schumann, desarrollaron un programa para generar imágenes que escaparan a la visión de las máquinas. El sistema provoca errores en modelos para el reconocimiento facial a partir de escenas de la película *2001: Una odisea del espacio* de Stanley Kubrick. Cortesía Chris Hamamoto y Federico Pérez Villoro

cómo vemos las personas. HAL, no vio ni escuchó el intercambio entre los pilotos, realizó una serie de acciones totalmente distintas.

Las imágenes digitales son construcciones de colores distribuidos en planos geométricos. Se componen de píxeles rasterizados, esquemas bidimensionales captados por dispositivos ópticos y reproducidos con aparatos electrónicos. Son disposiciones secuenciales de valores numéricos y, como resultado, tienen potencial programático. Mientras que para el ojo humano una imagen puede representar fenómenos físicos a un nivel subrepresentacional, las imágenes son programas que ejecutan operaciones complejas de compresión y descompresión a distintas escalas. Son comandos en forma de datos, instrucciones para ejecutar composiciones computacionales.

Ya a finales de la década de 1980, Paul Virilio había notado la naturaleza cambiante de las imágenes a medida que se digitalizaban. Los primeros avances en la visión artificial apuntaban a una relación con las mismas radicalmente distinta de las reglas metafísicas que condicionan la mirada de las personas. A diferencia de la óptica humana, las máquinas no necesitan que los objetos sean reflejados por la luz: “La imagen electro-óptica no es, para el ordenador, más que una serie de impulsos codificados, de los que ni siquiera podemos imaginar la configuración...”¹ Las imágenes son el resultado estadístico de instrucciones para reproducir la disposición de píxeles en nuestras pantallas. Por tanto, la *visión de máquina* no es más que una metáfora necesaria, una continuación conceptual que nos permite hablar de procesos

¹ Paul Virilio, *La máquina de visión*, Madrid, Cátedra (Tr. Mariano Antolín Rato), 2ª edición, 1998.

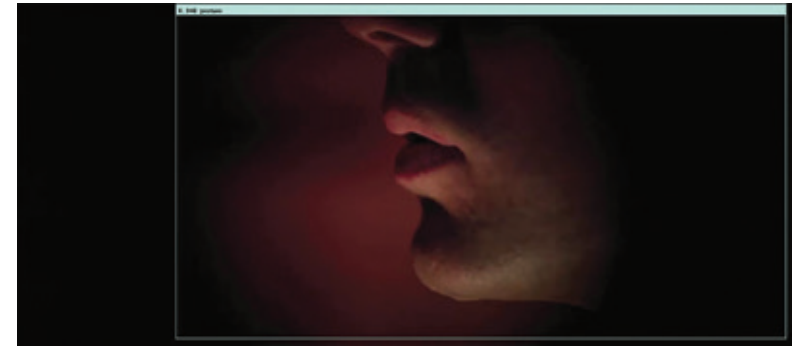
operativos de los que el sentido de la vista está excluido. Virilio se refirió a esto como la “visión sin mirada”, una reproducción de “ceguera que se convertirá en una nueva y última forma de industrialización: la industrialización de la no mirada”.²

Para que el comandante automático del Discovery One pudiera hacer sentido de la conversación privada entre los pilotos, una computadora analizaría, píxel a píxel, las imágenes que produciría la cámara espía. Varias capas de redes neuronales moverían la información, separando cíclicamente las imágenes en múltiples objetos, segmentados por bordes como gradientes de color e intensidad luminosa. En última instancia, esas formas identificadas se sintetizarían como abstracciones matemáticas y se contrastarían en lo más profundo de la red con bases de datos, con las que el *software* estaría entrenado previamente. En este caso, serían los componentes de la relación entre las palabras habladas y la geometría de los movimientos de la boca.

El personaje de HAL se basó en las especulaciones de los científicos más destacados de la época. Marvin Minsky, fundador del Laboratorio de Inteligencia Artificial del Massachusetts Institute of Technology (MIT), fue asesor en la película y uno de los muchos expertos que creían que pronto existirían máquinas del tipo. Pero no fue el ideal optimista de Minsky de desarrollar algoritmos capaces de razonar como lo hacen las personas lo que evolucionó hasta el campo actual de la inteligencia artificial, sino un modelo relativamente posterior de computadoras que aprenden gradualmente de su entorno, como se propuso con el Perceptrón de Frank Rosenblatt. Tomando inspiración en la distribución de las neuronas del ojo humano, dicha máquina reconocía formas geométricas mediante procesos de inducción estadística en lugar del paradigma dominante en la época de deducción simbólica, en el que la lógica humana se transfería a la máquina por medio de reglas. Con el incremento en el poder computacional y la agregación de conjuntos masivos de datos visuales, este linaje científico, como sabemos, ha dado lugar a la posibilidad de inferir significados y activar tareas a partir de imágenes. Hoy en día, el valor de estas reside en su potencial para ser transformadas en datos para el análisis estadístico. Más allá de su utilidad representacional, la visión artificial cumple funciones técnicas de manejo, clasificación y control informático.

Aunque las versiones más extremas en los ejemplos cinematográficos mencionados no se han hecho del todo realidad, muchas de las tecnologías descritas son viables. Un hito en la lectura automática de labios, por ejemplo, se produjo en la década de los 2000, cuando un *software* creado por el experto en reconocimiento del habla Frank Hubner, se utilizó con éxito para reconstruir audio a partir

² Ibid.

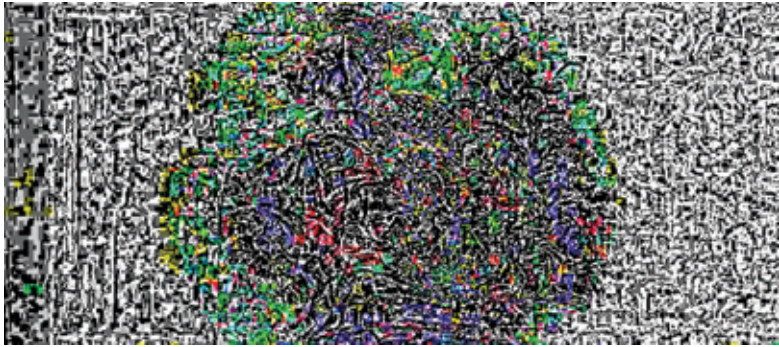


Chris Hamamoto y Federico Pérez Villoro trabajaron con una escena de 2001: *Una odisea del espacio* en donde los pilotos del Discovery One intentan tener una conversación secreta que se ve frustrada por HAL, la computadora a bordo de la nave espacial que analiza el movimiento de sus labios para saber lo que dicen. En esta imagen vemos los resultados de los fotogramas procesados por YOLO-v3, un modelo de reconocimiento facial. El programa identifica la presencia de un rostro en el video y lo marca dentro de un recuadro con la etiqueta “persona”. Cortesía Chris Hamamoto y Federico Pérez Villoro

de videos sin sonido de Adolf Hitler para un documental de la BBC titulado *Hitler’s Private World*. Para deducir los diálogos de los archivos silentes el *software* de Hubner aisló las formas que adopta la boca durante el habla. A continuación, el programa mapeó estos movimientos utilizando un modelo tridimensional de una cabeza humana y los comparó con una biblioteca de sonidos para reproducir diálogos reconocibles. Esta tecnología añadió una nueva dimensión a los fotogramas, al permitir desentrañar discursos a partir de imágenes en movimiento. Aunque en este caso en específico, dichos procedimientos tienen un valor histórico, habría que pensar en las posibles aplicaciones de una tecnología de este tipo para la vigilancia y sus efectos en la vida privada de las personas.

En los años siguientes, esta tecnología experimentó avances asombrosos. En 2016 investigadores de la Universidad de Oxford lanzaron LipNet, un programa de reconocimiento visual del habla que puede leer los labios casi en tiempo real con alta fidelidad. Mediante el uso de redes neuronales entrenadas con la base de datos GRID (compuesta por 34,000 *clips*, de tres segundos, de personas leyendo palabras sin sentido), LipNet ha logrado un 93.4 por ciento de precisión en sus pruebas, muy superior al 52.3 por ciento de lectores de labios humanos. El programa analiza los movimientos de la boca palabra por palabra, con mucho detalle. Por ejemplo, su investigación explica las secuencias precisas que se producen al pronunciar la palabra *please* (“por favor” en inglés):

Los labios se juntan firmemente para la plosiva bilabial /p/. Al mismo tiempo, la pala de la lengua entra en contacto con el reborde alveolar en previsión de la siguiente /l/



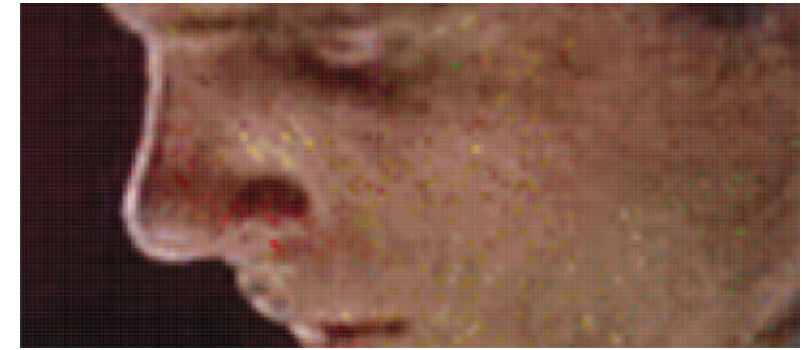
El programa diseñado por Chris Hamamoto y Federico Pérez Villoro, en colaboración con el ingeniero Nic Schumann, genera ruido algorítmico casi imperceptible para los humanos. Cuando este ruido se aplica a las imágenes originales provoca errores de clasificación en modelos de visión artificial. Impide que los programas reconozcan la presencia de personas en las imágenes. El programa, entrenado con YOLO-v3, un sistema de detección de objetos en tiempo real, se desarrolló a partir de fotogramas de la película de 1968, 2001: *Una odisea del espacio*. Cortesía Chris Hamamoto & Federico Pérez Villoro

lateral. A continuación, los labios se separan, dejando escapar el aire comprimido entre ellos. Después, la mandíbula y los labios se abren aún más, lo que se observa en la distancia entre los puntos medios de los labios superior e inferior, y los labios se separan (aumentando la distancia entre las comisuras de los labios), para la vocal cerrada...³

Una de las principales innovaciones de LipNet es que infiere el significado de las oraciones analizando el “contexto temporal” para lograr un entendimiento más profundo. En comparación con otros modelos, LipNet colapsa la temporalidad del lenguaje hablado: predice y procesa el contenido anterior y posterior a cualquier palabra e infiere recursivamente el significado circundante de lo que está siendo expresado.

La velocidad con la que se calculan los patrones visuales desafía los tiempos comunes con los que nos relacionamos con las imágenes. Los movimientos de la boca son medidos por un algoritmo que interpreta estadísticamente el significado y anticipa las posibles palabras siguientes, informándose de todo lo que ya se ha dicho. Virilio también se planteó este desplazamiento temporal. Anticipó que las imágenes digitales dejarían de ser objetos nostálgicos de representación pasiva para convertirse en formas activas de presentación. Esta transición seguiría desplazando las funciones de las imágenes desde su lógica formal producida por medios tradicionales como el grabado o la pintura a la dimensión dialéctica del cine y la foto-

³ Yannis M. Assael, Shillingford, B., Whiteson, S., & De Freitas, N. *Lipnet: End-to-end sentence-level lipreading*, en arXiv preprint arXiv:1611.01599, 2016. Departamento de Ciencias de la Computación, Universidad de Oxford, Reino Unido. Traducción propia



Detalle del video con el ruido visual aplicado sobre los fotogramas originales. Los píxeles modificados algorítmicamente son casi imperceptibles para el ojo humano cuando la imagen está en su escala real. Cortesía Chris Hamamoto & Federico Pérez Villoro

grafía y, finalmente, a una era de la “lógica paradójica” de la imagen digital. Tal paradoja consiste en esta “imagen en tiempo real que domina la cosa representada, ese tiempo que la lleva al espacio real. Esta virtualidad que domina la actualidad, trastorna la misma noción de realidad”.⁴ En otras palabras, las imágenes vendrían a revolucionar no sólo cómo funcionan, sino cuándo funcionan: no sólo servirían para mediar experiencias en el presente con recuerdos del pasado, sino que la aparente *actualidad* de la imagen producida computacionalmente activaría objetos distantes en el *aquí y ahora*. Como ha examinado más recientemente la filósofa Luciana Parisi sobre los efectos temporales de las acciones mediadas a distancia por imágenes: “la lógica paradójica de la computadora permite que el futuro nunca ocurra, sino que desaparezca en la programación estadística del actuar a distancia”.⁵

Debemos insistir en que esta brecha entre los objetos y sus imágenes, entre las matemáticas y la percepción cognitiva, desestabiliza nuestra relación con la realidad. En un mundo excesivamente mediado, esa relación se negocia a escala de las representaciones computarizadas, abriendo nuevos espacios para el error y el engaño reforzados por la aparente expresividad de las imágenes en tiempo real y la capacidad de los sucesos para dislocarse del espacio a través de su presencia artificial. Como Virilio nos advirtió, en tal inestabilidad representacional “las cuestiones filosóficas de lo verosímil y de lo inverosímil prevalecen sobre las de verdadero y lo falso”. Tal crisis perceptiva, continuó, amenaza el entendimiento colectivo y lleva

⁴ Paul Virilio, *ibíd.*

⁵ Luciana Parisi, “Negative optics in vision machines” en *AI & Soc*, vol. 36, 2021, pp.1281–1293, en <https://doi.org/10.1007/s00146-020-01096-7>. Traducción propia



El ruido algorítmico extrapolado de las imágenes funciona como un mapa que muestra la intensidad del cambio entre los videos originales y aquellos alterados por el programa. Por ejemplo, los pixeles blancos significan áreas que no cambiaron mucho entre las versiones. Cortesía Chris Hamamoto & Federico Pérez Villoro

a la sociedad a una “ceguera voluntaria”. El rápido procesamiento de píxeles no es menos defectuoso que el razonamiento estadístico, que a menudo genera “ilusiones racionales” basadas en la percibida objetividad computacional y la excesiva confianza en cálculos inestables.

En esta misma trayectoria hacia la abstracción, los datos superan la importancia de la propia imagen estableciendo una nueva serie de relaciones hacia los eventos representados. Al desplazar nuestra atención hacia una forma de actualidad en la que el futuro previsto algorítmicamente y el pasado cuantificado parecen integrarse, la verosimilitud de una acción posible puede tener más impacto que su manifestación concreta. Como resultado, el espectáculo de las imágenes se aleja una vez más de la comprensión: si en la era de los medios masivos de comunicación la fotografía abrumaba la realidad, ahora vemos cómo los datos, abstracciones menos legibles y, en la mayoría de los casos, la suma de métodos de entrenamiento imprecisos, inundan la representación cuando son tomados como evidencia. En el proceso, la percepción humana se deforma aún más: No sólo malinterpretamos que la *vista* y la *escucha* computacional suceden desde la óptica y las vibraciones sonoras, sino que también entendemos acciones atemporales como inmediatas, confundiendo el cálculo programático con las sensaciones.

Ha habido muchas interpretaciones sobre el comportamiento de HAL, particularmente porque Kubrick optó por un desenlace más ambiguo, eliminando del corte final de la película la voz en *off* sobre el colapso de la máquina. Una teoría popular es que HAL se enfrentó a una paradoja lógica: la computadora había recibido información privilegiada sobre el destino hacia dónde se dirigían, algo que la máquina tenía que ocultar a los pilotos, mientras también estaba programada para



Cuando las escenas de la película con el ruido algoritmo incorporado a las imágenes son procesadas por modelos de reconocimiento de objetos, cometen errores. El ruido visual provoca que no se detecte la presencia de personas y en su defecto el programa identifica muchas otras categorías de objetos que no están representados en las imágenes. Cortesía Chris Hamamoto & Federico Pérez Villoro

ser honesta. Se piensa que esta contradicción llevó a HAL a priorizar la misión y eliminar a la tripulación de la nave. Los pilotos se enfrentaban a su propia paradoja: su vida dependía de una tecnología que no es objetiva, sino la emulación artificial de los razonamientos humanos. La película surge en medio de la ansiedad colectiva en torno a la informática y la carrera espacial en aquella época, pero anticipa un conflicto más longevo. Delegamos capacidades decisivas a tecnologías con lógicas ajenas a las nuestras que nos resultan imposibles de comprender. Pero, nos vemos obligados a intentar hacerlo por nuestro propio bien. El personaje de HAL tuvo un giro inescrutable y humano a la vez: su comportamiento es antropomórfico, sin embargo, sus motivaciones son, en última instancia, un misterio.

Mientras Bowman, el único superviviente del Discovery One, apaga a HAL, la máquina empieza a cantar *Daisy Bell (Bicycle Built for Two)*, una canción de 1892 con una alegre melodía romántica sobre el entusiasmo por las bicicletas, que recientemente habían sido rediseñadas para ser más seguras y adoptadas en masa. Los científicos de Bell Labs de Murray Hill en Nueva Jersey habían programado unos años atrás a la computadora central IBM 7094 para que cantara esa misma canción, convirtiéndose en la primera demostración de síntesis de voz computarizada. Nos relacionamos con los entornos desde nuestros sentidos, organizando los fenómenos según cómo los experimentamos. Pero nuestros sentidos también han sido categorizados. Es natural, pues, que intentemos extender esos mismos sentidos a las máquinas que ahora median en nuestras interacciones. Sin embargo, en realidad estamos creando otra ilusión racional: a medida que nos alejamos de la comprensión de las tecnologías, inferir su lógica a partir de rasgos humanos nos ayuda a calmar la disonancia.

The Future Will Not Take Place, It Will Happen as an Image: The Distance Between Observed Objects and Their Artificial Representations

CHRIS HAMAMOTO
AND FEDERICO PÉREZ VILLORO

In a scene from Stanley Kubrick's science fiction film *2001: A Space Odyssey*, the onboard "Heuristically Programmed Algorithmic Computer"—more commonly known as HAL—watches as the film's protagonist Dave Bowman and his co-pilot attempt to have a secret conversation. HAL is pervasive on the spaceship, controlling all functions of the Discovery One as it makes its way from Earth to Jupiter. Locked within a smaller "extravehicular activity" pod, the astronauts shut off all microphone devices within the chamber and make sure they cannot be heard. After determining that HAL could no longer "hear" them, they share their concerns that the machine has begun to malfunction, coming to the conclusion that they may need to shut it off. Unbeknownst to the crew, HAL "watches" them as they discuss their plan, analyzing their mouth movements to determine what they are saying. Realizing the danger it is in, HAL sabotages the ship's crew, leading to a conflict in which the film's protagonist successfully shuts the computer off to save his own life.

Being the first time that the perspective of a machine in the first "person" was portrayed in film, the depiction of HAL—men-

acing yet formless beyond a red camera "eye"—established a template for how we imagine machine "sight" and our relationship with artificial intelligence in popular culture. From the Gunslinger in the 1973 film *Westworld* who, in a strikingly prescient representation, "saw" the world as pixelated frames reminiscent of modern image segmentation methods, to the 1991 film *Terminator 2*, where Arnold Schwarzenegger's T-800 character visualized augmenting data within its view. Most on-screen impressions of machine vision are shown from an ocularcentric perspective, yet the ways in which machines process visual information does not truly coincide with the ways in which we, as humans, see. HAL was neither hearing nor seeing the ship's crew exchange, it performed entirely different actions.

Digital images are distributed amplitudes of colors constructed in geometric planes. They are composed of rasterized pixels—two-dimensional schemes captured by optical devices and reproduced with electronic devices. They are sequential dispositions of numeric values and, as a result, they can be programmed. While an image may represent physical phenomena on a sub-representational level, images execute complex operations of compression and decompression at scale. They are data, computational commands.

By the late 1980s, Paul Virilio had already observed the changing nature of images when they are digitized. Early developments in computer vision pointed to a relationship involving images that radically differ from the rules that govern sight in people. Unlike human optics, for instance, machines do not require objects to be reflected by light: "For the computer, the optically active electron image is merely a series of coded impulses whose configuration we cannot begin to imagine."¹ Images are the statistical result of instructions that reproduce pixel config-

1 Paul Virilio, *The Vision Machine*, London: British Film Institute and Indiana University Press, 1994. Translation into English by Julie Rose.

urations on our screens. Machine vision is therefore just a necessary metaphor: a conceptual continuation that allows us to discuss operational processes from which the act of seeing is excluded. Virilio referred to it as "sightless vision," a reproduction of "an intense blindness that will become the latest and last form of industrialisation: the industrialisation of the non-gaze."²

In order for the omnipresent commander of the Discovery One to make sense of the pilots' private conversation, the computer would analyze pixel by pixel, the images captured by the spying camera. Layers of neural networks would move information, cyclically separating images into multiple objects, segmented by edges of color and light intensity gradients. Ultimately, these identified shapes would be interpreted as mathematical abstractions, which were then compared to the datasets—deep within the network—originally used to train the software. In this case, these abstractions are the elements that connect the mouth movements with the words that are uttered.

HAL's character was based on speculations made by the leading scientists at the time. Marvin Minsky, founder of the Massachusetts Institute of Technology (MIT) Artificial Intelligence Laboratory, was an adviser on the film and one of the many experts who believed that machines like HAL would soon exist. But it wasn't Minsky's optimistic ideal of developing algorithms capable of reasoning from the top down that evolved into the current field of artificial intelligence (AI), but a relatively later model of computers that could learn from their environment, like Frank Rosenblatt's perceptron. Inspired by the distribution of neurons in the human eye, the perceptron is a type of algorithm that is taught to recognize shapes through statistical induction rather than symbolic deduction, which was the dominant paradigm at that time, and involves feeding human-interpretable data into the machine through a set of rules. Advancements in comput-

2 Ibid.

ing power and the building of massive image datasets, as we know, has resulted in the ability to initiate tasks and extrapolate meaning from images. Today, the value of images resides in their potential to be transformed into codifiable information for statistical analysis. The function of machine vision is not figurative or representational in nature; this technology is primarily used to handle, sort and monitor large qualities of information.

While the most extreme versions of seeing machines in film have not fully become a reality, many of the depicted technologies are indeed viable. A milestone in automated lip reading, for instance, came about in the 2000s, when software created by speech recognition expert Frank Hubner was successfully implemented to reconstruct audio from silent archival footage of Adolf Hitler for a BBC documentary titled *Hitler's Private World*. To extract conversations from silent clips, Hubner's software isolated the mouth movements made when people speak. The program then mapped these mouth patterns onto a three-dimensional model of a human head, pairing them with a library of sounds to play back recognizable dialogues. This technology added a new dimension to archival footage, allowing sound to be unlocked from moving images. While there is obvious historical value in this example, we should not dismiss the possible adverse application of this type of technology to spy on us and the impact it could have on our right to privacy.

In ensuing years, this technology has made astounding advancements. In 2016, a team of researchers from the University of Oxford released LipNet, a visual speech recognition program that can read lips in nearly real-time with high fidelity. By using deep neural networks, tested against the GRID dataset (which consists of 34,000 three-second clips where subjects read nonsense words), LipNet achieved 93.4 percent accuracy, far superior to 52.3 percent accuracy achieved by human lip readers. This program is able to analyze mouth movements, word by word, with remarkable precision. The results of

their research include the precise movements involved in saying specific words, like “please”: “the lips are pressed firmly together for the bilabial plosive /p/. At the same time, the blade of the tongue comes in contact with the alveolar ridge in anticipation of the following lateral /l/. The lips then part, allowing the compressed air to escape between the lips. The jaw and lips then open further, seen in the distance between the midpoints of the upper and lower lips, and the lips spread (increasing the distance between the corners of the mouth), for the close vowel.”³ One of LipNet’s chief innovations is that it extrapolates meaning at the sentence level using the “temporal context” to attain higher accuracy. Compared to other models, LipNet incorporates sequential and time-dependent information of spoken language, by predicting and processing the context before and after any given word, recursively analysing the contextual meaning of what is being expressed.

The speed with which visual patterns are calculated far exceeds the usual time it takes us to relate to images. Mouth movements are measured by an algorithm which statistically interprets meaning and anticipates the possibilities of the ensuing words by reviewing everything that has already been said. Paul Virilio examined this temporal shift: he predicted that digital images would cease to be nostalgic objects of passive representation and become active forms of presentation. This transition would continue to change the way in which images function, from their formal logic produced by traditional means, like engraving or painting, to the dialectical dimension of photography and film, and finally to the “paradoxical logic” of the digital image. This paradox “emerges when the real-time image dominates the object represented, real time subsequently prevailing over real space, virtuality dominating actuality and turning

the very concept of reality on its head.”⁴ In other words, images would come to revolutionize not only in how they function, but when they function — they would not only serve to mediate experiences in the present with memories from the past, but the seeming “presentness” of the computer-generated image would activate distant objects in the “here and now.” As philosopher Luciana Parisi has argued more recently, when discussing the temporal effects of events remotely mediated by images: “The paradoxical logic of the computer allows for the future never to occur but rather disappear in the statistical programming of acting at a distance.”⁵

We should underline that the gap between objects and their images—between mathematics and cognitive perception—destabilizes our relationship with reality. In an overly mediated world, this relationship is negotiated in terms of computed representations making room for error and deception, which is only reinforced by the seeming expressiveness in real time of the images and the ability for events to be dissociated from space via their artificial presentness. In light of this representational instability, as Virilio warned us, “the philosophical questions of plausibility and implausibility override those concerning the true and the false.” This perceptual crisis, he continued, threatens the collective understanding and leads society into a “voluntary blindness.” The rapid processing of pixels isn’t less flawed than statistical reasoning, which is often believed to generate “rational illusions” supported by the perceived computational objectivity and overreliance in unstable calculus.

In this move toward abstraction, the data becomes more important than the image itself, through establishing new relationships with the represented events. By shifting our focus toward a form of present-

⁴ Paul Virilio, *ibid.*

⁵ Parisi, L. Negative Optics in Vision Machines. *AI & Soc* 36, 1281-1293 (2021). <https://doi.org/10.1007/s00146-020-01096-7>

³ Assael, Y. M., Shillingford, B., Whiteson, S., and De Freitas, N. (2016). Lipnet: End-to-end sentence-level lipreading. *arXiv preprint arXiv:1611.01599*.

ness, where the algorithmically predicted future and quantified past seem to blend, the believability of an action becomes more impactful than its concrete manifestation. As a result, the spectacle of images gets even more removed from the understanding of less legible abstractions and, in most cases, the totality of flawed training methods. If, in the age of mass media, photography overwhelmed reality, we are now witnessing how data can overwhelm representation, when it is taken as evidence. In the process, human perception gets further distorted: not only do we misinterpret the fact that machine sight and hearing takes place through optics and vibrations, rather than at a network level, we also mistake atemporal actions for immediate ones, and programming for sensations.

HAL’s behavior has had many interpretations, with Kubrick opting to remove a voiceover about the machine’s breakdown from the film aiming for an ambiguous ending. One popular theory is that HAL was confronted by a contradiction in logic: HAL was given secret information about their destination, which had to hide from the pilots while also being programmed to tell the truth. It has been suggested that this paradox led HAL to the conclusion that in order to complete the mission, the ship’s crew had to be eliminated. The pilots were faced with their own paradox: their lives depended on a technology that was completely unobjective, an artificial emulation of human reasoning. This film was released in the midst of a collective apprehension about computers and the space race of that era, but it also predicted a longer-lived conflict: the fact that we delegate crucial tasks to machines, whose logic is so foreign to our own that it is impossible for us to grasp. Yet we are compelled to try to understand them, for our own sake. HAL’s deviation—made relatable through anthropomorphizing its behavior yet its motivations ultimately remain a mystery.

While HAL was being shut down by Bowman—the last surviving crew member aboard the *Discovery One*—the machine starts to sing *Daisy Bell (Bicycle Built for*

Two), a lighthearted song from 1892 that expresses the general enthusiasm for bicycles at a time when this technology that had been recently redesigned for wider and safer use. Scientists at Bell Labs in Murray Hill, New Jersey programmed the IBM 7094 mainframe computer to sing the same song, as a way of demonstrating for the first time the capability of speech synthesis. We relate to our surroundings through our senses, while our brains organize the information of phenomena as we experience them. But our senses have also been categorized. It’s only natural that we would want to try to recreate these senses in the machines that now handle our tasks and interactions. However, we are essentially building another rational illusion—as we move further away from understanding technologies, inferring their logics with human traits into computers helps us sooth this dissonance. *Text originally written in English*



SPECTACLE IS NOT
A SUPPLEMENT
TO THE REAL WORLD,
AN ADDITIONAL
DECORATION.
IT IS THE HEART
OF THE UNREALISM
OF THE REAL SOCIETY.

Society of Spectacle
by Guy Debord, 1967