

Pensar en imágenes después de la inteligencia artificial

Yanai Toister y Joanna Zylinska

ENSAYO / CULTURA VISUAL × INTELIGENCIA ARTIFICIAL

Pensar en imágenes después de la inteligencia artificial

Yanai Toister y Joanna Zylinska

Publicación original · Journal of Visual Culture, 24(3), 431–451 · DOI: 10.1177/14704129251405715

Edición en español para TecnoPsique. Traducción de *Image thinking after artificial intelligence*. Se conserva la autoría original y la bibliografía. La obra fuente está publicada bajo licencia Creative Commons Atribución 4.0 (CC BY 4.0); esta edición indica su condición de traducción y adaptación tipográfica.

APERTURA

Resumen

Este artículo propone que, al sintetizar actos de lenguaje con procesos de visualización, la inteligencia artificial generativa está modificando la arquitectura de la cognición humana, lo que exige una nueva manera de comprender nuestra relación con las imágenes. El concepto multifacético de «pensar en imágenes» que figura en el título remite a: (1) nuestro pensamiento humano acerca de las imágenes, un área central de indagación para los estudios de cultura visual; (2) la generación de imágenes mediante procesos cognitivos, sean humanos o maquínicos; y (3) las propias imágenes como agentes del pensamiento. Tanto los estudios de cultura visual como la neurociencia recurren a la noción de «imagen mental», basada en la idea de que los seres humanos formamos imágenes internas a partir de palabras y conceptos. Esas imágenes facilitan, a su vez, el modo en que vemos —y conocemos— objetos y fenómenos. Hoy ha surgido un nuevo circuito de retroalimentación entre lenguaje, visión e imágenes, en el que las máquinas pueden generar imágenes a partir de instrucciones textuales (*prompts*). Este desarrollo tecnológico desdibuja la

frontera entre imaginación humana y generación algorítmica, y reclama, por tanto, una reevaluación radical de cómo percibimos, imaginamos y comprendemos el mundo. Más allá de las frecuentes descalificaciones de las imágenes sintéticas como promedio, banalidad o mera «media», el artículo propone un enfoque crítico inspirado en Vilém Flusser que busca oportunidades cognitivas en nuestro momento cultural actual.

Palabras clave: inteligencia artificial · cognición · Flusser · imagen-texto · percepción · Mitchell.

01

Introducción: pensamiento humano y maquínico

La célebre pregunta de Alan Turing (1950), «¿Pueden pensar las máquinas?», sigue resonando. Ahora que las máquinas computacionales realizan proezas consideradas en otro tiempo exclusivamente humanas —desde componer sinfonías hasta redactar argumentos jurídicos—, el enigma persiste: ¿cómo piensan exactamente y, más importante aún, piensan realmente? Sus algoritmos, conjuntos de datos y redes neuronales constituyen una opacidad conceptual: una caja negra que encubre nuestra ignorancia no solo acerca de las arquitecturas cognitivas maquínicas, sino, más profundamente, acerca de la maquinaria del pensamiento humano[1]. Durante siglos, la filosofía occidental cartografió el pensamiento mediante metáforas de la visión, equiparando ver el mundo con conocerlo y, más ampliamente, conocer con ver. Que el latín *videre* («ver») comparta raíz con *verus* («verdadero») y *veritas* («verdad») acentúa esta conexión. Desde el obispo Berkeley hasta los primeros fenomenólogos, diversos filósofos entrelazaron cognición e imágenes mentales, y trataron esas imágenes interiores transitorias como un vínculo vital entre percepción y comprensión, vista e intelección. Los trabajos recientes en filosofía de la mente y ciencia cognitiva marcan un desplazamiento desde el modelo introspectivo de la mente, basado en «imágenes estáticas dentro de la cabeza», hacia una explicación generativa de sus operaciones: comprenden la percepción como predicción bajo incertidumbre, las imágenes como un espacio de trabajo generativo y el lenguaje como un *prior* (es decir, una suposición inicial o datos de entrada) que configura aquello que se vuelve visible. Es desde este registro que abordamos aquí el «pensar

en imágenes». Pero si la arquitectura del pensamiento humano permanece opaca, ¿cómo podría diseñarse con confianza su contrapartida artificial? ¿Cómo puede hacerse visible el pensamiento, incluso para quien piensa, y qué papel desempeñan ahora las imágenes en este proceso?

Este artículo explora lo que denominamos el «espacio latente» de la cultura (Somaini, 2025), donde el pensamiento humano y el maquínico se entrelazan de maneras tanto desconcertantes como prometedoras. Para ello, no alcanza con preguntarse si las máquinas pueden pensar (Hayles, 2017)[2]. También debemos examinar cómo su supuesto pensamiento redefine el nuestro y nos permite acoger un modelo de cognición más complejo, multisensorial y experiencial, que integre conceptos, imágenes, perceptos y diversas formas de experiencia no simbolizada (Hurlburt y Schwitzgebel, 2007: 104, 137–138). Sin embargo, nuestro interés principal reside en la reciprocidad cambiante entre pensamientos e imágenes en humanos y máquinas. Al rastrear la transformación de esa reciprocidad, buscamos iluminar los contornos de un paisaje cognitivo que la producción de imágenes sintéticas está reconfigurando. El concepto multifacético de «pensar en imágenes» que proponemos remite a: (1) el pensamiento humano acerca de las imágenes, un campo fundamental para los estudios de cultura visual; (2) la generación de imágenes mediante procesos cognitivos, ya sean humanos o maquínicos; y (3) las imágenes mismas como agentes del pensamiento. Con ello, pretendemos dar algunos primeros pasos hacia un nuevo marco conceptual para los estudios de cultura visual. Este marco nos permitirá reposicionar tanto nuestros objetos de investigación como nuestro propio papel de investigadores en el paisaje epistémico contemporáneo, cada vez más configurado por el aprendizaje automático y la IA generativa. Resulta central para esta indagación un diálogo intenso con un filósofo de la tecnología que tomó en serio tanto el pensamiento como las imágenes: Vilém Flusser. Pondremos en relación su explicación programática de las imágenes técnicas con la síntesis basada en difusión del aprendizaje automático, aclarando al mismo tiempo por qué la producción de imágenes dirigida por *prompts* difiere epistémicamente de la captura mimética. Al hacerlo, dejaremos entre paréntesis los importantes problemas infraestructurales de economía de plataformas y trabajo analizados por otros (Crawford, 2021; Pasquinelli, 2023), para situar el propio pensar en imágenes como infraestructura cognitiva: un circuito recursivo en el que los conceptos devienen imágenes y las imágenes, a su vez, median los conceptos. Este enfoque, elaborado conjuntamente como respuesta a los desarrollos actuales de la producción de imágenes sintéticas, extiende nuestros

trabajos filosóficos previos sobre percepción, computación y fotografía (Toister, 2020; Zylinska, 2023) hacia un modelo explícitamente cognitivo, aunque anclado en perspectivas críticas provenientes de los estudios de cultura visual.

La IA generativa puede ayudarnos en este intento de repensar el pensamiento en imágenes humano y maquínico. Modelos como Stable Diffusion, Midjourney o DALL·E evocan imágenes a partir de instrucciones textuales con gran fluidez, y producen resultados que desdibujan la frontera entre imitación e imaginación. No obstante, sus operaciones, impulsadas por *prompts* en lenguaje natural junto con procesos de vectorización —que sitúan las probabilidades de aparición de uno u otro agrupamiento de imágenes en un espacio multidimensional—, descansan sobre un supuesto problemático: que el pensamiento humano puede destilarse en rutinas computables. Este estado de cosas exige un examen más profundo del concepto de «imagen mental», no simplemente como construcción visual en la mente, sino como interacción dinámica entre memoria, imaginación y percepción sensorial. La investigación neurocientífica revela que imaginar un objeto (es decir, evocarlo mentalmente) activa vías neuronales semejantes a las que se activan al verlo efectivamente, lo que sugiere que las imágenes mentales pueden ser una piedra angular no solo de la visualización, sino también del pensamiento (Nanay, 2021). Sin embargo, esa piedra angular no es universal. Algunas personas experimentan afantasía, la incapacidad de formar imágenes mentales, mientras que otras refieren hiperfantasía, la capacidad de formar representaciones mentales extraordinariamente vívidas (Milton et al., 2021). Este fenómeno revela que el pensamiento humano constituye una red compleja de variaciones neurológicas que resiste su simplificación en una forma algorítmica unificada.

Es importante señalar que los generadores de imágenes por IA hacen algo más que replicar nuestras imágenes mentales traducidas a *prompts*: intervienen efectivamente en el circuito cognitivo y reconfiguran cómo pensamos tanto la producción de imágenes como el pensamiento mismo. El circuito de retroalimentación es recursivo: las máquinas generan imágenes que los humanos observan e interiorizan, recalibrando en el proceso nuestra imaginación. La línea entre cognición y percepción humanas y maquínicas comienza, así, a desdibujarse, mientras que algo novedoso parece emerger: un espacio compuesto en el que el pensamiento humano es reconfigurado gradualmente por operaciones computacionales. Proponemos denominar a este proceso «hackeo cognitivo» (*cognitive hacking*). No se trata de *hackeo* en el sentido tradicional de vulnerar sistemas seguros; se trata, más bien, de un hackeo epistémico

que implica reconfigurar el sistema perceptivo y mental mediante el cual procesamos el mundo. Dejando atrás tanto las connotaciones explícitamente negativas de «hackeo» adoptadas por los discursos de ciberseguridad como el escepticismo implícito en las descripciones de la IA como un mero «loro estocástico desprovisto de significado» (Bender et al., 2021), hacemos nuestra la propuesta de Ted Underwood (2022) de «abordar los modelos neuronales como máquinas para cartografiar y recablear el comportamiento colectivo». Desde este marco, los modelos de IA se consideran no tanto modelos de la inteligencia o de la cognición humanas, sino principalmente modelos de cultura. Dejando a un lado importantes cuestiones económicas y laborales, el pánico casi existencial que las imágenes de IA suscitan hoy entre numerosos pensadores, periodistas y miembros del público podría deberse, por tanto, más al vacío epistémico que esos modelos ponen de manifiesto —la posibilidad de que no haya nada esencial «ahí», en el corazón de nuestro pensamiento y nuestra imaginación humanos— que a cualquier aplanamiento estético o conceptual específico que supuestamente provocan.

02

El conocimiento occidental entre proyección y predicción

La suposición de que una visión clara producía naturalmente un conocimiento verdadero, que había configurado el pensamiento occidental al menos desde Platón, comenzó a resquebrajarse bajo el escrutinio filosófico de la modernidad temprana. Las *Meditaciones* de Descartes (1984[1641]) socavaron la confianza en la vista al mostrar con qué facilidad los sentidos podían engañar incluso al pensador más disciplinado. Berkeley (1949[1713–1734]) intensificó esta duda mediante su tesis radical *esse est percipi*, al reubicar la realidad en el ámbito de la percepción y de la experiencia interior. En el marco filosófico de Kant (1998[1781–1787]), la percepción sintetizaba las entradas sensoriales mediante categorías impuestas *a priori* por la mente. Así, la visión dejó de entenderse como una mera transmisión de la realidad: pasó a ser una negociación perpetua entre la multiplicidad del mundo exterior y las facultades organizadoras de la mente. La fenomenología del siglo XX abrazó este giro epistemológico. Merleau-Ponty (1945, 1968) describió poéticamente la percepción como una trama contingente, insistiendo en que el ver emergía de las interacciones fluidas y siempre cambiantes entre el cuerpo humano y el mundo —o

los mundos— que habitaba. Esta perspectiva encarnada mostró que la visión es intrínsecamente interpretativa y se encuentra continuamente reconfigurada por el contexto subjetivo, la memoria, la emoción y la expectativa. La ciencia cognitiva corrobora la intuición de Merleau-Ponty al revelar que la percepción es menos un registro de estímulos externos que una predicción e hipótesis continuas: un diálogo anticipatorio entre jerarquías neuronales y evidencia sensorial (Bruner, 1957; Neisser, 2014[1967]). En este marco, ver se presenta como algo eminentemente generativo. No descubrimos simplemente lo que existe, sino que «categorizamos» activamente (Bruner, 1957: 123–125) y, por tanto, «construimos» (pp. 127–138) aquello que puede conocerse, creando y recreando el mundo visible en cada mirada.

Existen paralelismos claros entre la investigación neurocientífica actual y las intuiciones acerca del carácter construido de las imágenes y de la visión procedentes de la historia del arte y los estudios de cultura visual, muchas de las cuales anteceden a los descubrimientos científicos más recientes. El concepto de «fórmulas del *pathos*» de Aby Warburg (1999[1781]) destaca cómo los gestos emocionales y los motivos expresivos —como los movimientos intensos de cuerpos o elementos naturales— atraviesan la historia, dando forma a la memoria cultural y condicionando nuestra respuesta ante las imágenes. El argumento de John Berger (2008[1972]) según el cual los «modos de ver» están cultural e históricamente situados insiste, a su vez, en que la visión se encuentra indisolublemente ligada al contexto ideológico en el que se produjo la imagen. La formulación de los «regímenes escópicos» de Martin Jay (1988) amplía esta posición al detallar configuraciones históricas específicas mediante las cuales la visualidad es gobernada, regulada y dotada de sentido. En conjunto, estos autores sostienen que el acto privado e interno de ver ya está siempre inserto en prácticas sociales, convenciones culturales y economías simbólicas más amplias. Esta negociación epistemológica reaparece en la influyente taxonomía de W. J. T. Mitchell (1984, 1986), que agrupa imágenes gráficas, ópticas, perceptivas, mentales y verbales dentro de una misma «familia». Mitchell deja claro que las imágenes nunca son recibidas únicamente de manera óptica, ni constituyen «simples» marcas gráficas que nosotros desciframos. Más bien, son construidas activamente, pensadas hasta cobrar existencia mediante procesos estratificados que abarcan códigos lingüísticos, simbólicos y afectivos, todos ellos en constante interpenetración, configuración, distorsión e influencia mutua. «Como un actor en el escenario histórico, una presencia o personaje dotado de estatus legendario» (Mitchell, 1986: 9), la imagen posee también una forma singular de agencia.

Resulta crucial para los estudios de cultura visual reconocer que, durante siglos, las imágenes han sido convocadas a la existencia mediante fórmulas científicas y estéticas codificadas, pero que esas fórmulas están siendo ahora radicalmente reconfiguradas por la computación. La IA generativa, con sus modelos de difusión entrenados sobre vastos corpus de imágenes, modifica las dinámicas estructurales de la visualidad, sacude su «fundamento existencial» (Denson, 2024: 170) y expande sus horizontes imaginativos hacia regiones cada vez más extensas de invisibilidad. Esos modelos hacen algo más que reproducir convenciones conocidas: evalúan, reinventan y propagan nuevos tropos y motivos. Para los estudios de cultura visual, esto equivale a una situación en la que los procesos computacionales determinan activamente normas estéticas y los marcos cognitivos que las acompañan. Habitamos ahora un entorno en el que las máquinas no solo generan («escriben») e interpretan («leen») imágenes —y, por esa vía, las operacionalizan—, sino que también prescriben los términos de producción del significado visual y, por extensión, de lo visual mismo. Por consiguiente, el campo de los estudios de cultura visual se extiende más allá de la interpretación de imágenes para examinar las condiciones infraestructurales y los sustratos tecnológicos que gobiernan su producción, recepción y circulación. Nuestro interés aquí se centra en producir imágenes con el lenguaje y a través de él, y en el papel novedoso que desempeñan las imágenes dentro del pensamiento humano. Durante la última década, la teoría de los medios mostró hasta qué punto buena parte de la cultura visual contemporánea opera más allá de su exhibición ante humanos, funcionando principalmente como «visión de plataforma» (*platform seeing*; Mackenzie y Munster, 2019; Toister, 2024) dentro de cadenas operativas que clasifican, verifican y optimizan imágenes para lectores maquínicos. Sobre la base de este trabajo, que ofrece un puente entre las historias de la visión bosquejadas anteriormente y nuestro propio foco en las imágenes como procedimientos que coproducen supuestos cognitivos previos, utilizamos el concepto de «invisualidad» (*in-visuality*; Parikka, 2023) para nombrar el espacio en el que se realiza el trabajo más decisivo antes de que las imágenes aparezcan —para los seres humanos— como figuras visibles.

03

La «imagen-texto» y el circuito del prompt

En 1986, Mitchell propuso el concepto de «imagen-texto» (*imagetext*) para comprender la interpenetración de imagen y texto lineal como una nueva forma híbrida que exigía una lectura conjunta (Mitchell, 1986, 1994). «Todos los medios —escribió— son medios mixtos, y todas las representaciones son heterogéneas; no existen artes "puramente" visuales o verbales» (Mitchell, 1994: 5). La imagen-texto designaba «obras (o conceptos) compuestos, sintéticos, que combinan imagen y texto» (p. 89), pero también nombraba la «problemática» de las relaciones entre lo visual y lo verbal (p. 7). Era un lugar de entrelazamiento semiótico que Raymond Bellour (1996) describió como «una doble hélice». El trenzado de esa hélice era «un lugar de conflicto» en el que se escenificaban «antagonismos políticos, institucionales y sociales» (Mitchell, 1994: 91). La formulación de Mitchell llegó en un momento oportuno, ya que permitió a los investigadores dar cuenta de la imbricación mutua entre lenguaje e imagen en una amplia variedad de artefactos (en su mayoría estáticos). Sin embargo, todavía suponía la centralidad de un «lector de imágenes» humano, alguien que descifraba signos, negociaba tensiones e interpretaba el juego semiótico. Por eso no daba cuenta de lectores ampliados por las máquinas y entrelazados con ellas, ni de las consecuencias de ese entrelazamiento para nuestra relación con las imágenes.

Aquí entra Vilém Flusser, cuya filosofía de los medios, también formulada en la década de 1980, anticipó ese desplazamiento. Al haber tomado las imágenes en serio —no solo como artefactos o, más estrechamente, como arte, sino como vías hacia la cognición y el conocimiento (Flusser, 2011)—, su obra ofrece una perspectiva epistémica singular para desarrollar hoy el pensar en imágenes. En un pasaje sorprendente, Flusser (2000: 12) proclamó: «A nuestras espaldas se invierte la jerarquía de los códigos. Los textos, originalmente un metacódigo de las imágenes, pueden tener a su vez las imágenes como metacódigo». Al postular esta inversión, y sin admitir nunca que descansaba sobre un falso binarismo, Flusser buscó desmantelar el andamiaje metafísico de larga data que había privilegiado la abstracción lingüística sobre la inmediatez visual. El «metacódigo», antes sello distintivo de la superioridad humana, regresa aquí como sustrato, como código fuente

de un nuevo circuito de producción de imágenes. En consecuencia, el pasaje del texto a la imagen implica un desplazamiento tectónico: las formas mediáticas quedan subsumidas bajo la lógica operativa de «máquinas semióticas» (Nake, 2005: 61), que manipulan procesos sígnicos antes que materia y, al hacerlo, inauguran un nuevo régimen de producción de signos.

El aparato de caja negra de la síntesis de texto a imagen en los modelos de IA generativa vuelve literal la intuición de Flusser: el lenguaje funciona no solo como una abstracción de las imágenes, sino como motor de su devenir computacional. Así, el texto no se limita a describir, titular o incluso anclar las imágenes: condiciona su propia emergencia. No es que las palabras ilustren figuras, ni viceversa, sino que la distinción misma se disuelve en la lógica recursiva de una red de aprendizaje profundo. Las palabras se convierten en *embeddings* (representaciones vectoriales), los *prompts* son operaciones y las imágenes se vuelven probabilidades, modulaciones y derivaciones. Sin embargo, la ventana del *prompt* simplemente hace visible lo que hace tiempo estaba implícito en toda clase de «imágenes sintéticas» —término utilizado inicialmente en el contexto de la visión artificial (Horn y Bachman, 1978)—: que son «posibilidades programadas», según la expresión de Flusser (2000: 69)[3]. En los modelos de IA, en particular, cuando el lenguaje ingresa en las máquinas a través de un cuadro de texto, atraviesa los parámetros propietarios y las restricciones del modelo, y regresa como imagen modulada por filtros de contenido y ponderaciones de estilo. Lo decisivo es que cada iteración reorganiza la paleta de la imaginación colectiva: el fetiche o la fantasía de alguien se transforma en un campo de probabilidades ajustado a la conveniencia de la plataforma.

La iconología tradicional, como método interpretativo de la historia del arte, descifraba significados una vez que la imagen se había estabilizado (supuestamente). En cambio, la difusión en los modelos de IA nos obliga a comenzar la interpretación mucho más atrás: en las etapas de elección del modelo, construcción del conjunto de datos y ajuste de parámetros. Preguntar qué podría significar una imagen sintética, si es que significa algo, es inseparable de preguntar qué descripciones se extrajeron de internet, qué trabajo humano etiquetó el corpus y quién posee los pesos que orientan la salida. Y puesto que estos sistemas median la cultura visual, nuestra unidad de análisis debería ser el modelo de IA como infraestructura cognitiva de esa cultura, y no la imagen aislada que produce. Lo que regresa a los usuarios como «imagen» es una distribución producida previamente por la composición y el etiquetado del conjunto de datos, la selección y parametrización del modelo, los filtros de seguridad y las

ponderaciones de estilo. Interpretar esas imágenes exige entonces atender a la procedencia de los conjuntos de datos, así como a la información sobre marcas de agua, porque estos elementos prescriben las condiciones epistémicas dentro de las cuales la imaginación se vuelve operativa. En suma, el modelo no es únicamente un medio de representación pictórica: es un sistema que organiza las probabilidades de aquello que puede llegar a ser representado visualmente.

Los estudios de cultura visual necesitan, por tanto, un término más dinámico que «imagen-texto» para captar cómo las imágenes llegan a ser imágenes en la actualidad. Como ya no son solo lugares de legibilidad, sino también de escritura y reescritura, y como los encuentros con imágenes duraderas y estáticas son cada vez menos frecuentes, proponemos el concepto de «texto-imagen-texto» (*textimagetext*) para destacar que las imágenes circulan hoy en circuitos incesantes. No solo aparecen y reaparecen en noticias, campañas promocionales o como memes en las redes sociales; también son incorporadas a conjuntos de datos invisibles para entrenar sistemas de IA con los que luego se generan nuevas imágenes. Cuando los *prompts* se convierten en imágenes y estas retroalimentan nuevos *prompts*, entramos en otro circuito en el que se desdibujan los límites entre producir imágenes y generar pensamientos. En un circuito texto-imagen-texto, el lenguaje inicia una imagen que, a su vez, se vuelve legible para humanos y máquinas como nuevo texto. Este circuito se distingue de los modelos anteriores de interacción entre imagen y texto. En la imagen-texto de Mitchell, la relación seguía estando mayormente controlada por seres humanos y, más importante todavía, orientada hacia ellos: un jeroglífico para descifrar, una caricatura para leer, una película para interpretar (sin desconocer que Mitchell [2005] reconoció posteriormente la agencia de las imágenes). En el texto-imagen-texto, la imagen debe ser leída tanto por humanos como por máquinas y sirve de dato para nuevas iteraciones del proceso generativo[4]. La frontera entre pensar imágenes y visualizarlas se vuelve aquí porosa: el lenguaje diseña activamente imágenes en vez de limitarse a describirlas o anotarlas. Las imágenes, así, revisan y reformatean el espacio lingüístico del que emergen, pero también incorporan al usuario humano como parte del circuito. El texto-imagen-texto dirige nuestra atención hacia la estructura temporal y procedimental de la producción de significado en los sistemas de IA, donde *prompts* basados en lenguaje, imágenes, descripciones y espectadores coproducen sentido a lo largo de múltiples ciclos de generación, retroalimentación y refinamiento.

04

Imágenes mentales a través de mentes y modelos

Un fundamento constitutivo para articular el circuito texto-imagen-texto es la «imagen mental»: un término que ha ido desplazando gradualmente la noción clásica de «imaginación» y que hoy atraviesa la filosofía, la psicología y la ciencia cognitiva. En su forma más simple, una imagen mental es una «representación perceptiva que no se desencadena directamente por una entrada sensorial» (Nanay, 2003: 4). Sin el anclaje causal de la sensación inmediata, funciona como un ensayo de la percepción. Una imagen mental no tiene por qué ser exclusivamente visual: puede incorporar señales auditivas, olfativas o de otro tipo. Además, los perceptos nunca se generan solo desde el yo y para el yo. La imagen mental opera dentro de un espacio cultural, incluso cuando se experimenta como algo intensamente personal.

Los modelos de generación de imágenes por IA ofrecen una contraparte mecánica del proceso mental de ensayo imaginativo. Un motor de difusión comienza, de un modo parecido a como lo hace la mente en ausencia de estímulos sensoriales, con una indeterminación estructurada: un lienzo de ruido inscrito en un espacio latente de alta dimensionalidad. Esto es para el modelo lo que las imágenes mentales son para nosotros: un espacio de trabajo generativo donde formas apenas esbozadas negocian con la memoria antes de hacerse visibles. Una señal lingüística comprimida (el *prompt*) desplaza entonces ese conjunto de información hacia la coherencia. En cada paso de eliminación de ruido, el modelo adelanta una hipótesis provisional de forma, consultando la memoria estadística de cientos de millones de imágenes acompañadas de descripciones. Quizá de un modo similar a los circuitos predictivos del cerebro, revisa iterativamente su conjetura, ponderando regiones de la imagen según la atención del *transformer*, de manera que «granada roja» arrastra el color y el contorno emergentes lejos de, por ejemplo, «planeta rojo». Tras unas treinta pasadas, el circuito convierte la representación latente en píxeles y produce una imagen que maximiza la probabilidad conjunta de patrón visual y señal verbal: una superficie ejecutable de cómo, dentro de ciertos parámetros, debería verse el *prompt*.

La analogía entre imágenes mentales e IA generativa está justificada. Ambos mecanismos recurren a la experiencia previa —memoria personal en el primer caso, enormes conjuntos de datos extraídos de internet en el segundo— y realizan

predicciones guiadas por señales. Ambos completan oclusiones, interpolan trayectorias y sesgan sus resultados hacia estilos conocidos. Esta analogía permite describir a ambos sistemas como predictivos. En el caso humano, el aparato cognitivo propone hipótesis que las señales sensoriales entrantes corrigen, mientras las imágenes forman parte de ese espacio de trabajo generativo. Las palabras modulan ese espacio: incluso señales verbales mínimas actúan como supuestos previos que inclinan la inferencia hacia ciertos rasgos y la alejan de otros. En el caso del modelo, un *prompt* repondera de manera similar una trayectoria latente durante la eliminación del ruido y orienta la forma emergente hacia rasgos congruentes con la etiqueta. En conjunto, esto reformula el texto-imagen-texto no solo como un trenzado semiótico, sino como un circuito predictivo en el que el lenguaje condiciona la imagen, la imagen regresa como lenguaje y el circuito actualiza los supuestos previos que guiarán la percepción y los *prompts* posteriores[5].

No obstante, las divergencias de esta analogía también son importantes. Las imágenes humanas son encarnadas, están cargadas de afectos y son multisensoriales. Sus supuestos previos están ponderados por la saliencia emocional y la retroalimentación propioceptiva. Su «función de pérdida» es el éxito pragmático en situaciones vividas. Por el contrario, los supuestos previos de un modelo consisten en conteos de frecuencia recolectados de la web, y su función de pérdida es minimizar el error a nivel de píxeles frente a imágenes etiquetadas consideradas como referencia. Mientras la mente biológica optimiza para sobrevivir y para la experiencia vivida significativa, la mente sintética optimiza para la fidelidad estadística. Este contraste puede especificarse sin promover ni el antropomorfismo ni, en el extremo opuesto, el reduccionismo computacional. La percepción humana minimiza el error de predicción bajo restricciones encarnadas (como presupuestos metabólicos, saliencia afectiva y contexto social), de modo que la «exactitud» se negocia continuamente con la viabilidad y el valor. Los modelos generativos contemporáneos minimizan pérdidas sobre distribuciones de entrenamiento y optimizan la fidelidad estadística a emparejamientos entre descripciones e imágenes sin compromisos orgánicos propios. La consecuencia no es que la primera «piense» y la segunda no (Bratton et al., 2022; Zylinska, 2024), sino que optimizan cosas distintas. Cuando se encuentran en nuestro uso de modelos de aprendizaje automático, esas funciones objetivo diferentes coproducen nuestro entorno cognitivo.

Como ya planteamos, las operaciones de las mentes humanas y sintéticas siguen siendo opacas para nosotros. La introspección humana revela únicamente los *qualia*

acabados, y oculta las rutas neuronales que los produjeron. Del mismo modo, el recorrido de un modelo de difusión es matemáticamente explícito y, sin embargo, cognitivamente oscuro: miles de millones de parámetros codifican disposiciones que ningún ser humano sin ayuda puede leer o comprender. Esta ceguera recíproca alimenta la ansiedad por la «caja negra» que rodea los resultados de la IA. Aclara que transparencia e inteligibilidad no son sinónimos: lo transparente para el código de máquina sigue siendo enigmático para la fenomenología, y viceversa. Las imágenes mentales siguen vinculadas con la percepción, aunque no de manera causal; las imágenes por difusión, en cambio, desencadenan iteraciones «en el mundo real» solo después de haber sido vistas y descritas nuevamente (aunque esto probablemente cambie). Aun así, ambas propagan circuitos texto-imagen-texto, porque las imágenes sintéticas siembran nuevas expectativas visuales que migran hacia la publicidad, el diseño de videojuegos y las fantasías personales, alterando de ese modo los supuestos culturales previos de los que dependerá el siguiente conjunto de entrenamiento. En este sentido, el modelo hace algo más que imitar la cultura: desplaza algo dentro de ella.

05

Hackeo cognitivo

Reconocer este desplazamiento nos anima a especular sobre lo que las máquinas podrían enseñarnos acerca de nuestro propio pensamiento. Como los modelos carecen del anclaje motivacional propio de los humanos, sus resultados suelen ser muy detallados, aunque pueden interpretarse como semánticamente superficiales. Sin embargo, interactuar con ellos pone de relieve hasta qué punto nuestras imágenes humanas están condicionadas por el lenguaje preexistente, que transporta múltiples supuestos previos: cada *prompt* exterioriza el habla interior y deja al descubierto clichés persistentes, sesgos y lagunas de nuestra comunicación humana. Para generar imágenes, los modelos de IA transforman gradualmente el ruido en patrón mediante un despliegue negentrópico. El hecho de que ese proceso pueda producir resultados desalineados o surrealistas (por ejemplo, anatomías deformadas) no debería leerse como un fracaso en la representación de la realidad, sino como la señal visual de una epistemología emergente: una en la que el error puede indicar divergencia generativa antes que una falla sistémica (Zylinska, 2024).

Desde luego, la IA generativa también corre el riesgo de reforzar clichés y uniformidades impulsadas por el mercado. Pero detenerse allí, como hacen muchos de sus críticos, equivale a ignorar sus potencialidades cognitivas latentes. Las imágenes generadas por IA pueden, en cambio, funcionar como ventanas hacia nuestros propios procesos de pensamiento, cada vez más entrelazados con los de las máquinas computacionales. Proponemos denominar «hackeo cognitivo» a las consecuencias de ese entrelazamiento.

Aunque sus orígenes se encuentran en las actividades comunitarias de programadores entusiastas del MIT durante la década de 1970, en el lenguaje contemporáneo el término «hackeo» evoca imágenes de actores clandestinos infiltrándose en infraestructuras digitales, sorteando cortafuegos y tomando el control de servidores. Queremos recuperar el sentido original de radicalidad implícito en este término y proponer que el hackeo sea entendido como una forma de reprogramación sistémica: la manipulación y transformación de una estructura dada para producir resultados nuevos, no siempre controlables. Es en este sentido más amplio que ofrecemos el concepto de hackeo cognitivo como descripción de las transformaciones que la IA generativa está provocando actualmente en las estructuras de la percepción, el pensamiento y la imaginación humanos. Esta forma de hackeo no es realizada de manera encubierta por agentes maliciosos: es, más bien, una característica emergente de los sistemas culturales y técnicos contemporáneos. En lugar de intervenir directamente sobre nuestro aparato cognitivo, funciona redirigiendo nuestros circuitos cognitivos en sus interacciones con los procesos computacionales. Por lo tanto, el hackeo cognitivo no debería considerarse una simple metáfora: remite a procesos materiales efectivos con consecuencias para la vida real. Sostenemos, de hecho, que el auge de la IA generativa está dando lugar a una reconfiguración sistemática de las condiciones en las que ocurre el pensamiento, mediante un proceso colectivo, recursivo y en gran medida involuntario. Cuando hablamos, entonces, de la incursión de la IA en la cognición, nos referimos a procesos que reconfiguran las estructuras fundamentales mediante las cuales la información se transforma en significado cultural. Cada nueva síntesis modifica el repertorio imaginativo del usuario. Con el tiempo, los usuarios aprenden a anticipar los patrones «característicos» del sistema y modulan sus expresiones verbales y sus marcos conceptuales para inducir determinados resultados, ya sea con fines artísticos, de visualización científica o de uso cotidiano. Con estos fines, la IA generativa opera imperceptiblemente y recalibra de manera gradual los procesos internos de pensamiento e imaginación del usuario,

«hackeándolos» silenciosamente desde dentro. Utilizamos, pues, el término «hackeo» sin dramatismo, para designar el alineamiento gradual de los supuestos previos, las preferencias y las expectativas de los usuarios mediante su interacción reiterada con imágenes mediadas por modelos.

Por supuesto, los circuitos texto-imagen-texto no afectan por igual a todos los usuarios: no todo el mundo es especialista en ingeniería de *prompts*, ni siquiera usuario ocasional de herramientas de IA. Sin embargo, la penetración de imágenes generadas por IA en el discurso público a través de las redes sociales y otras plataformas digitales implica que todos estamos expuestos a sus lógicas visuales. El resultado es una armonización gradual de nuestra imaginación colectiva: comenzamos a pensar con las gramáticas de imágenes inscritas por los modelos de IA y a través de ellas. La naturaleza recursiva de la generación de imágenes implica que cada imagen nueva queda configurada por el peso acumulado de resultados anteriores. Lo que denominamos «estilo» o «coherencia» constituye, en realidad, una forma de inercia estadística. La serie de técnica mixta *V3 Model Collapse* (2025), de la artista Felicity Hammond, condensa esta dinámica en un único gesto: al reintroducir datos sintéticos en su propio circuito, la obra escenifica el «colapso del modelo» que amenaza a la IA generativa y, al hacerlo, invita a reflexionar sobre un «colapso del modelo humano» paralelo: la erosión de nuestro rango imaginativo cuando también nosotros reciclamos imágenes condicionadas por la IA. A medida que los usuarios se adaptan a los lenguajes visuales de las plataformas, su imaginación es modulada por esos lenguajes. No nos limitamos a escribir imágenes: ellas nos escriben a nosotros de vuelta, recalibrando nuestros marcos cognitivos mediante la exposición continua. Este entrelazamiento plantea la pregunta de si el hackeo cognitivo es una cuestión de grado o de naturaleza. Algunos podrían argumentar que todos los medios reconfiguran la cognición: la imprenta estandarizó nuestra base de conocimiento, la fotografía modificó nuestra percepción y el cine reorientó nuestra atención. La IA generativa podría ser simplemente la más reciente de una serie de prótesis cognitivas con efectos transformadores. Sin embargo, su carácter recursivo, colectivo y automatizado posiblemente la haga cualitativamente diferente, ya que los circuitos que inicia no son secuenciales, sino exponenciales. Durante 2025 surgieron numerosas preocupaciones sobre «la progresiva uniformidad de la voz de la IA» y, más importante aún, de la voz y la expresión humanas después de nuestra interacción con modelos como Midjourney o ChatGPT. Un artículo de *The Verge* capta el aspecto más existencial de esta ansiedad al señalar que quienes hablan «no se dan cuenta de que su lenguaje está cambiando»

(Parker, 2025). Aun reconociendo que los modelos y sistemas de IA se volverán más expresivos con el tiempo y reducirán su inquietante insipidez, se afirma que «el riesgo más profundo» no reside tanto en la uniformidad lingüística como en «perder el control consciente de nuestro propio pensamiento y expresión».

¿Cuáles son las implicaciones filosóficas de este desarrollo? Si, en el esquema tradicional de la Ilustración, la visión constituía una vía hacia una verdad singular, en el paradigma actual la propia visión se ha vuelto plural y contingente. Nos enfrentamos a un mundo en el que la realidad se fabrica de manera rutinaria, donde «nuestras experiencias son el contenido que el cerebro predice de adentro hacia afuera» (Seth, 2021; véase también Zylinska, 2023: 147–165). El neurocientífico Anil Seth sostiene que la realidad es, en esencia, una «alucinación controlada» sostenida por una cantidad de datos sensoriales apenas suficiente para mantenernos vivos (Seth, 2021). La IA generativa extiende esta condición hacia afuera: proporciona alucinaciones externas que se sincronizan con nuestras alucinaciones internas. Hackea la máquina predictiva del cerebro al ofrecer predicciones prefabricadas e hiperrealistas (es decir, imágenes) que pueden sobrescribir o potenciar las nuestras. El peligro es que nos convirtamos en consumidores pasivos de esas visiones prefabricadas, con una imaginación sobreajustada a los datos de entrenamiento de la IA. Pero también podríamos aprender a jugar con posibilidades latentes colectivas que ninguna mente individual podría abarcar por sí sola.

En lugar de oponernos a este desarrollo o romantizar una supuesta forma más pura de cognición humana anterior a la IA, nuestra tarea consiste en cartografiar críticamente sus vectores. Como ya señalamos, el hackeo cognitivo no debe considerarse intrínsecamente maligno. Pero, como todos los hackeos, revela las vulnerabilidades de un sistema: en este caso, nuestro sistema epistemológico y los valores que lo sostienen. Si nuestros pensamientos son ahora coproducidos por máquinas con una intensidad sin precedentes y a una escala nunca antes experimentada, los estudios de cultura visual deben interrogar rigurosamente las fuerzas infraestructurales que delimitan el campo de lo visible. Volvemos una vez más a Flusser para buscar aperturas dentro de este relato —actualmente dominante— del colapso del modelo y explorar algunas posibilidades alternativas de pensar con el sistema computacional contemporáneo y a través de él.

06

Cómo pensar como la IA: una remezcla de Vilém Flusser y Mira Schendel

En sus reflexiones filosóficas sobre la obra de la artista Mira Schendel, una emigrante judía como él con quien mantuvo una larga amistad en su hogar adoptivo de São Paulo, Flusser describe dos modos distintos de pensamiento, en ambos de los cuales la imagen desempeña un papel importante. Estos estilos no son simples preferencias cognitivas: reflejan maneras diferentes de relacionarse con el mundo y de crear significado en él. En el momento cultural actual, en el que los sistemas de IA generativa se convierten en agentes activos en la producción de nuestra cultura visual, vale la pena volver a las intuiciones de Flusser porque, como suele ocurrir con su obra, revelan una inquietante capacidad de anticipación respecto de los medios futuros. La obra de Schendel, tal como Flusser la lee, se convierte en un lugar de experimentación conceptual —por parte de Flusser y también nuestra— sobre cómo los seres humanos ven, imaginan y conceptualizan el mundo. La comparación no es incidental: tanto las obras de Schendel como los sistemas de IA generativa operan en el umbral entre lenguaje e imagen, concepto y artefacto.

Flusser comienza su ensayo «Mira Schendel» describiendo lo que denomina el modo «tradicional» de pensar: una operación secuencial que se desplaza de lo concreto a lo abstracto. En este estilo de pensamiento —con el que se identifica el Flusser filósofo, pero que también representa para él la racionalidad científica occidental en su conjunto—:

> *Me encuentro con algo concreto. Formo una imagen de esa cosa concreta (la «imagino») para familiarizarme con ella. Y luego traduzco mi imagen en un concepto para comprender la cosa concreta y poder manejarla. (Roth, 2014)*[6]

En otras palabras, primero viene la imaginación y después el análisis: la mente «ve» un objeto y luego lo explica al generar una idea a partir de él. Flusser vincula este proceso sucesivo con etapas históricas de la conciencia: una progresión desde una imaginación «mítico-mágica» (que implica producir imágenes para aprehender el mundo) hacia un razonamiento «epistemológico-técnico» (que deriva conocimiento

conceptual de esas imágenes). En este esquema, la imagen funciona como una estación intermedia: un sustituto representacional en el camino hacia una comprensión conceptual más rigurosa y supuestamente más verdadera. La transparencia es aquí mayormente metafórica: comprender equivale a «ver a través» de la imagen para llegar a su significado oculto. El significado mismo es algo que debe descubrirse mediante capas de reducción simbólica. La imagen es inicialmente opaca y debe interpretarse paso a paso para revelar el concepto que hay detrás de ella. Según Flusser, este modo considera las imágenes subordinadas a las ideas: se mira a través de la imagen para acceder a la idea que está detrás. El concepto (es decir, el «mensaje» o significado abstracto) solo se vuelve claro después de ese esfuerzo intelectual secuencial, de manera semejante a como un lector procesa palabras línea por línea para captar su significado completo. La secuencia cognitiva que va del objeto a la imagen y de esta al concepto refleja un proceso histórico mediante el cual los seres humanos fueron alejándose gradualmente de la experiencia inmediata y superponiendo órdenes simbólicos sobre el mundo cada vez más distante de las cosas concretas. Como señala Flusser, esta doble abstracción produce una situación en la que «nosotros mismos, de algún modo, no estamos allí. Pues estar allí, existir, significa estar en un mundo de objetos» (Roth, 2014). En otras palabras, mediante la abstracción llegamos a nuestra propia alienación, que también implica una desconexión sensorial respecto del mundo.

Según Flusser, el modo de pensamiento de Schendel opera en dirección inversa. Comienza con el concepto —una idea abstracta— y busca materializarla o convertirla en imagen, dándole una forma visual y espacial. En vez de utilizar la imaginación para ascender hacia la abstracción, Schendel, la artista, la emplea para dar forma a aquello que de otro modo permanecería inmaterial. Flusser percibe esta inversión como una radical reversión epistemológica: «Mira intenta traducir el concepto de transparencia y el concepto de significado en una realización (ejecución) [imagen] de transparencia y una realización [imagen] de significado» (Roth, 2014). Dicho sencillamente, toma una idea encontrada en la vida cotidiana y le da forma visible; finalmente, esa imagen se materializa como un objeto (artístico). Esta inversión trastoca «la relación tradicional entre imaginación y razón discursiva» (Flusser, 2017: 244). En lugar de utilizar la imaginación para producir finalmente un concepto, Schendel parte de un concepto y se sirve de su acto imaginativo-artístico para encarnarlo. El resultado es un concepto hecho perceptible: adquiere cuerpo y, al mismo tiempo, exige una respuesta corporal del espectador. Al volver los conceptos imaginables y concretos, el enfoque de

Schendel tiene una «función violentamente desalienante» (p. 245). Nuestros procesos cognitivos se vuelven «sincronizados imaginativamente» (p. 245), en lugar de sucederse unos a otros; el significado se capta dentro de una estructura unificada y no queda diferido a lo largo de una cadena de razonamientos.

Lo crucial es que este estilo alternativo de pensamiento aborda la transparencia y el significado de manera literal e inmediata, con la imagen funcionando como su encarnación. Flusser se refiere a la obra *Sin título [Discos]*, de Schendel, realizada en 1972, para ilustrar este punto. La obra presenta dos láminas redondas y transparentes de acrílico, superpuestas y cubiertas de diversas escrituras y símbolos, suspendidas en un espacio de exposición. Su transparencia es tanto literal como conceptual: el espectador ve la imagen y ve a través de ella, y se enfrenta a un campo de significantes cuyo sentido debe construir activamente y que se resiste al cierre. Flusser admite, con una mezcla de emociones y algunos sugerentes matices de género, que Schendel «obliga a mi visión a entrar en una dinámica inversa... No se supone que deba "explicarla"... Se supone que debo "ensamblarla"» (Roth, 2014). Esta inversión del orden cognitivo tradicional marca un desplazamiento en el propio modo en que llega a existir el conocimiento: ya no como una proposición abstracta que debe verificarse, sino como una situación visual que debe experimentarse. Con lo simbólico y lo sensorial colapsando uno en el otro, el significado no se descubre, sino que se genera activamente en el acto de mirar. Los conceptos se vuelven visibles, y esa visibilidad tiene un efecto de realidad. Nos brinda la oportunidad de «vivir entre conceptos realizados», en palabras de Flusser. Al transformar ideas abstractas en imágenes y luego en cosas, Schendel ejemplifica un mundo en el que las ideas no permanecen confinadas a las mentes o a los textos escritos, sino que flotan allí afuera, como objetos que pueden verse, aprehenderse y (re)imaginarse. Comenzamos a «vivir... no entre conceptos, sino entre imágenes de conceptos», como lo expresa Flusser (2017: 245).

Es dentro de esta arquitectura cognitiva adoptada por Schendel donde pueden situarse las operaciones de la IA generativa. Un modelo de IA no parte de objetos, sino de conceptos codificados en instrucciones textuales. Privados de todo fundamento fenomenológico directo, esos *prompts* funcionan como señales lingüísticas abstractas. El sistema atraviesa entonces su espacio latente, un ámbito estadístico de asociaciones aprendidas y probabilidades visuales, para generar una imagen que encarne —o al menos aproxime— la estructura conceptual inscrita en la instrucción. Este proceso es estructuralmente homólogo al método de Schendel. El modelo, como

Schendel, parte del orden simbólico y avanza hacia su manifestación visual. El resultado no es una imagen del mundo, sino una imagen del concepto. Como tales, las imágenes generadas por IA no pueden considerarse representaciones en el sentido clásico: son síntesis visuales de vectores semánticos. No reflejan la realidad, sino que reconfiguran las condiciones bajo las cuales la realidad es imaginada y percibida.

A pesar de su opacidad originaria, la imagen de IA, como la lámina transparente de Schendel, invita a una interacción iterativa. Emerge dentro de un circuito de retroalimentación en el que el *prompt* se refina en respuesta a la imagen generada, y la imagen, a su vez, reconfigura el *prompt*. Esta dinámica recursiva hace colapsar las fronteras entre autor, espectador y sistema. Produce una forma compuesta de cognición y percepción, en la que concepto e imagen coevolucionan. En vez de extraerse de la imagen, el significado se inscribe en el proceso de su generación. Las implicaciones filosóficas de este desplazamiento son considerables. Si, en el esquema tradicional, el conocimiento requería estabilizar el significado mediante la claridad conceptual, el momento contemporáneo, ejemplificado tanto por Schendel como por los sistemas de IA, ofrece una epistemología diferente: una basada en el flujo conceptual, la inestabilidad estética y la supuesta transparencia de lo simbólico. Para tomar prestada la expresión de Flusser, «todo significa todo, lo que en esencia equivale a decir que no significa absolutamente nada» (Roth, 2014). El mundo es reconfigurado como un campo de imágenes potenciales, y el acto de ver se convierte en una forma de escritura.

Para muchos pensadores actuales, este desplazamiento en la producción de imágenes inaugurado por la IA generativa es profundamente peligroso y profundamente indeseable. Por ello, la perspicaz predicción de Flusser podría leerse como su repudio inequívoco. Un artículo reciente de *New Statesman* condensa este sentimiento al evaluar las imágenes de IA como deprimentes y peligrosas. Predice no solo «una homogeneización del arte, sino también una homogeneización de la cultura» como resultado de la adopción generalizada de modelos de IA en todas las esferas de nuestra vida creativa, y considera esta tendencia como manifestación de la creencia pesimista de que «hemos alcanzado los límites exteriores de lo que podemos lograr: que, de aquí en adelante, solo los robots pueden hacer avanzar nuestros contenidos. En realidad, la IA generativa deja poco lugar para cualquier novedad, pues continúa extrayendo de lo que ya existe» (Manavis, 2025). Sin embargo, esta crítica todavía se apoya decididamente en el primer modo de pensamiento de Flusser, pues cree que la creación de artefactos solo puede proceder linealmente, desde la supuesta riqueza de

los conceptos asociados con la cultura humana hasta la deshumanizante y repetitiva «basura de IA» (*AI slop*). Los parámetros de transformación tanto de las imágenes como del pensamiento parecen ya decididos en este relato, y su única dirección posible sería entrópica, es decir, cargada de caos y ruido. En una línea semejante, el teórico de los medios Roland Meyer supone que la razón de ser de la generación de imágenes es producir imágenes «que ya existen» (Meyer, 2025: 3), idealmente ajustadas a descripciones textuales y a las expectativas de los consumidores. Meyer sostiene que el realismo es «el santo grial de la síntesis de imágenes con IA» (p. 3), pero esa parece ser más bien una expectativa formulada por los críticos de los medios y la cultura visual, quienes luego censuran esas imágenes por no alcanzar el estándar realista (por ejemplo, humanos con un número «incorrecto» de dedos; véase Zylinska, 2024), y no necesariamente por el público más amplio, que parece participar de una diversidad de usos conscientes de esta tecnología que van más allá de producir memes polarizadores y pornografía *deepfake*.

Esta lectura sitúa a la IA como una mera cámara de eco del realismo cultural: en el mejor de los casos, una máquina mimética; en el peor, una generadora de distorsiones banales. Pero tal encuadre clausura las transformaciones epistemológicas y cognitivas más profundas que se despliegan en los sistemas de IA generativa, especialmente cuando se los aborda desde el modo alternativo de pensamiento de Flusser ejemplificado en la obra de Schendel. Donde Meyer ve la IA generativa como una reconstrucción ciega de una base de datos visual, Flusser ofrece una lente más generativa para comprender la arquitectura y la potencialidad del espacio latente. En el modelo cognitivo invertido de Schendel, el pensamiento no comienza con el objeto, sino con el concepto —una estructura abstracta— que se hace visible mediante una forma imaginativa y, a menudo, espacial. Esta inversión no busca reproducir la realidad, sino reconfigurarla, invitando al espectador a ensamblar significado en lugar de extraerlo. Vista desde esta perspectiva, la IA generativa opera de manera semejante. Solo podemos describirla como una «mezcolanza de elementos visuales» (Meyer, 2025: 4) si estamos dispuestos a considerar, por ejemplo, que *Wolf Hall*, de Hilary Mantel, es una mezcolanza de la Biblioteca Británica. Aunque la IA generativa recombina estadísticamente elementos dentro de un espacio latente para sintetizar imágenes que no están directamente presentes en los datos de entrenamiento, esto no es una recuperación ni una mezcla: es una proyección, una producción de imágenes del pensamiento. La IA generativa no busca imágenes visuales precodificadas; las proyecta mediante recombinaciones probabilísticas de rasgos abstraídos de datos

previos. Lo que se sintetiza, por consiguiente, es una nueva actualización de una posibilidad visual: una construcción, no un recuerdo.

El espacio latente, núcleo matemático de los modelos de IA generativa, no es una bóveda de imágenes, sino un mapa vectorizado de potencial visual. Cada coordenada representa un entrelazamiento ponderado de estilos, formas y cargas semióticas, ninguno de los cuales existe como objeto discreto e indexable. Introducir un *prompt* en el modelo equivale a recorrer este paisaje multidimensional y asentarse en un punto donde convergen diversos rasgos visuales. Lo que emerge de este proceso es una materialización especulativa de asociaciones, una captura probabilística de aquello que el modelo infirió de la cultura. Meyer, por supuesto, conoce las operaciones estadísticas del espacio latente, pero para él están sobredeterminadas por las expectativas que «nosotros» supuestamente tenemos de las imágenes, ya sean fotográficas o generadas por IA. Por eso pasa por alto la posibilidad de que del proceso emerja algo novedoso, sorprendente o creativo. La crítica de la economía política de las plataformas y de sus propietarios en las grandes tecnológicas sobredetermina y eclipsa, para él (y para muchos otros), cualquier potencial teórico inherente a la tecnología. Su lectura hace eco del célebre repudio de Hito Steyerl (2023) a las imágenes sintéticas, consideradas simplemente imágenes «medias» (*mean*): «versiones promediadas de un botín masivo extraído de internet», que convergen «en torno al promedio, la mediana; mediocridad alucinada» (p. 82).

Sin desconocer la necesidad de esa crítica política, quizá valga la pena preguntarse si la IA generativa participa en lo que antes denominamos una producción de imágenes del pensamiento que va más allá de su mero aplanamiento, al exteriorizar visualmente un espacio de posibilidades conceptuales. Mediante este proceso nos permite ver nuestras propias ideas desde afuera. El *prompt* se convierte en una sonda heurística: lo que recibimos a cambio es un cuadro de lo que el modelo cree que queremos decir. Los resultados son a menudo sorprendentes: revelan asociaciones imprevistas u opciones creativas, y amplían efectivamente los horizontes imaginativos del usuario. El proceso recuerda lo que Flusser identifica en la práctica de Schendel, donde los conceptos, antes considerados un soporte inmaterial de la representación, se concretan mediante formas visuales. Tanto para Schendel como para la IA generativa, la imagen no es entonces un espejo que se sostiene frente a la naturaleza (o la cultura), sino una superficie en la que el pensamiento se inscribe. Este carácter proyectivo de la IA generativa contrasta marcadamente con los regímenes miméticos tradicionales. Mientras el realismo fotográfico busca estabilizar una presencia, las imágenes de IA

son construcciones precarias, contingentes al *prompt*, a la arquitectura del modelo y a la interacción del usuario. Su generatividad no reside en su semejanza con lo real, sino en su capacidad de actualizar lo no realizado: visualizar experimentos mentales, metáforas y contrafácticos. Por lo tanto, presentar esas imágenes como demasiado constreñidas por las preferencias visuales y culturales estandarizadas de los ingenieros de datos o de sus promotores revela, principalmente, el conocido desagrado modernista del crítico ante los gustos populares, al tiempo que clausura la posibilidad de reconocer su novedad estructural. Para nosotros, en cambio, esas imágenes pueden considerarse impresiones del pensamiento y no huellas medias o pobres del mundo. El pánico que las imágenes de IA suscitan en numerosos investigadores y artistas quizá provenga del *horror vacui* que evocan, pues revelan la generatividad predictiva de nuestras propias ideas y efectos estéticos supuestamente originales. No es solo el espacio latente y sus resultados estadísticos lo que se teme como algo vacío: es el miedo al vacío último en el núcleo de nuestro propio pensamiento, a que no haya nada «allí», lo que se manifiesta en la transferencia de afecto negativo desde nosotros hacia esos artefactos.

07

Conclusión: hacia el pensar en imágenes

Este artículo ha rastreado el desplazamiento del lugar de la cultura visual hacia lo que hemos llamado «pensar en imágenes»: un dominio recursivo en el que se entrelazan cognición, computación y visualidad. El pensar en imágenes se despliega a través de tres líneas entrelazadas: (1) una perspectiva analítico-crítica, que diagnostica cómo operan las imágenes antes, por debajo y más allá de la visualidad, al tiempo que cartografía las fuerzas infraestructurales y culturales que precondicionan lo que puede verse y decirse; (2) un proceso generativo que implica el acoplamiento algorítmico del lenguaje y la visión, que convierte *prompts* en imágenes e imágenes nuevamente en datos textuales, modulando las formas humanas y maquínicas de producir imágenes —y de imaginar—; y (3) una operación autónoma, en la que las imágenes actúan como agentes semióticos que circulan, deciden y anticipan por nosotros, reescribiendo las coordenadas de la atención y la percepción dentro de una arquitectura cognitiva distribuida. Al haberse convertido en infraestructuras del pensamiento, las imágenes dan forma a los procesos de toma de decisiones, hacen

circular afectos a escala planetaria y enmarcan aquello que resulta pensable dentro de un régimen tecnocultural determinado.

Desde las cavernas hasta las ventanas, desde la imaginación hasta las teorías del cerebro predictivo de la neurociencia actual, las imágenes siempre han sustentado la cognición. Lo que ha cambiado ahora es la velocidad, la escala y la fuerza agencial con la que lo hacen. Dos rasgos son distintivos del presente: la recursividad a escala (los resultados sintéticos vuelven a ingresar en los corpus de entrenamiento) y la autoría al nivel del *prompt* (la síntesis orientada lingüísticamente se transforma en una forma de alfabetización masiva). En el proceso, las imágenes dejan de ser principalmente artefactos y pasan a funcionar como interfaces dentro de infraestructuras cognitivas, con las uniformidades observadas recientemente como resultado. La IA generativa y los sistemas asociados han introducido una modalidad recursiva en el pensamiento, dando lugar no solo a un nuevo medio, sino a una nueva condición epistemológica. Los sistemas de imágenes contemporáneos no se limitan a representar el mundo, bien o mal: configuran las propias condiciones bajo las cuales la representación resulta posible. Esta lógica recursiva del texto-imagen-texto instituye un nuevo sistema de retroalimentación, en el que las imágenes dan lugar a *prompts*, los *prompts* a nuevas imágenes, y así sucesivamente. Este circuito, a diferencia de las antiguas asociaciones entre imagen y texto, no es estable ni enteramente legible para los seres humanos. Permanece parcialmente oculto por la infraestructura de la imagen: por espacios latentes, *embeddings* de imagen-texto-datos y campos vectoriales que operan muy por debajo del nivel de nuestra percepción consciente. Es precisamente en este sustrato invisual donde se despliega actualmente buena parte del pensar en imágenes contemporáneo.

Para volver a Flusser y Schendel, el pensar en imágenes ya no procede linealmente de la percepción a la imaginación y de allí al concepto, sino que se despliega como una dinámica recursiva en la que concepto e imagen se encuentran, y mentes, máquinas y medios se coconstituyen y cogeneran cultura visual. El futuro que visualicemos ya habrá atravesado el filtro maquínico de plausibilidad. En términos sencillos, las máquinas entrenadas con cultura humana están generando rasgos visuales que los seres humanos adoptan, cerrando efectivamente un circuito de retroalimentación en el que los patrones impulsados por IA se entretejen irreversiblemente en la evolución cultural. Aquí se intensifican las implicaciones para los estudios de cultura visual. La semiótica, la iconografía y los análisis textuales y de historia del arte siguen siendo

importantes, pero ahora deben unir fuerzas con métodos de las humanidades digitales que incorporen la computación y el pensamiento de sistemas ecológicos. Si las imágenes configuran activamente las propias condiciones que hacen posible la representación, el pensar en imágenes debe atender al juego entre superficie y cadena de procesamiento, signo y circuito, siguiendo cómo las infraestructuras invisibles de la estadística y el cálculo coescriben las manifestaciones visibles del pensamiento. Más allá de interpretar las imágenes como artefactos culturales, los investigadores deben preguntarse ahora cómo las imágenes configuran la cognición: cómo piensan las imágenes y, al hacerlo, cómo reconfiguran nuestro propio pensamiento.

ANEXOS / FUENTES

Identificadores ORCID

Yanai Toister: https://orcid.org/0000-0002-5605-5233
Joanna Zylinska: https://orcid.org/0009-0002-2829-2327

ANEXOS / FUENTES

Notas

1. En este artículo, el término «máquina» se refiere específicamente a sistemas de computación electrónica, no a máquinas mecánicas como la máquina de vapor. Esta distinción pone en primer plano las arquitecturas digitales examinadas y sitúa nuestro argumento dentro de los debates actuales sobre aprendizaje automático e IA.

2. Aunque numerosas publicaciones científicas y de divulgación utilizan «pensamiento» y «cognición» de manera intercambiable, el «pensamiento» suele entenderse como un proceso mental que requiere esfuerzo consciente e intencional para comprender información, encontrar soluciones, tomar decisiones o generar nuevas ideas. «Cognición», a su vez, abarca no solo el pensamiento, sino también la atención, la percepción, la memoria y el lenguaje. La cuestión de la conciencia y la intencionalidad constituye un punto de fricción en los debates actuales sobre inteligencia humana y maquínica, como muestra el interrogante persistente acerca de si las máquinas pueden «realmente» pensar. Nuestra propia manera de pensar el «pensamiento» se mantiene, en líneas generales, próxima a la amplia propuesta de N. Katherine Hayles (2017: 5), según la cual «la mayor parte de la cognición humana ocurre fuera

de la conciencia/inconsciencia; la cognición se extiende a todo el espectro biológico, incluidos los animales y las plantas; los dispositivos técnicos cognizan y, al hacerlo, influyen profundamente en los sistemas humanos complejos». Adoptamos, pues, de manera implícita su noción amplia —y funcionalista— de «cognición no consciente»: una «capacidad que se extiende mucho más allá de la conciencia hacia otros procesos neurológicos del cerebro» y que «también está presente en otras formas de vida y en sistemas técnicos complejos» (p. 9). No obstante, nuestro argumento intenta ir más allá de las propuestas de Hayles, incluidas las expuestas en su obra más reciente (Hayles, 2025), al ofrecer una lectura cultural-técnica de las transformaciones del pensamiento humano acerca de las imágenes, con ellas y por ellas.

3. En la taxonomía de Flusser, las «posibilidades programadas» no se limitan a las imágenes sintéticas: designan el espacio delimitado por un programa de todas las imágenes que pueden producirse, del que las imágenes sintéticas constituyen apenas una subclase.

4. Así operacionaliza este circuito un modelo de difusión como DALL·E 3, de OpenAI: (1) **Texto Imagen.** Cuando un usuario introduce un *prompt* como «Una ciudad futurista con plantas bioluminiscentes y vehículos flotantes al atardecer», el modelo primero transforma esa entrada en lenguaje natural a una forma matemática (un vector latente, es decir, una cantidad con magnitud y dirección que puede ubicarse en un espacio multidimensional llamado espacio latente). Esta codificación semántica del *prompt* se emplea para condicionar un proceso de difusión en el que una red neuronal parte de ruido aleatorio y lo elimina iterativamente, guiada por el *prompt*, para producir una imagen sintética congruente con la descripción dada; (2) **Imagen Texto.** Una vez generada, la imagen no se muestra únicamente a un espectador humano. También se introduce en un modelo de descripción automática de imágenes o en un codificador de estilo CLIP (*Contrastive Language-Image Pretraining*) que «lee» la imagen y vuelve a traducirla en una descripción textual, etiquetas o metadatos. Este proceso de retroalimentación cumple múltiples funciones: verificar su correspondencia con el *prompt* original, aplicar filtros de seguridad o generar texto alternativo; y (3) **Texto Imagen Texto Imagen Texto.** Sobre la base de la retroalimentación, ya provenga del sistema o del usuario, el *prompt* puede refinarse, por ejemplo: «Hacé las plantas más altas, el cielo más violeta y agregá robots voladores». El proceso vuelve a empezar, ahora en un circuito comprendido dentro del marco texto-imagen-texto. Cada iteración modifica la relación entre lenguaje y visión, ya que los *prompts* son configurados por las mismas imágenes que contribuyen a generar.

5. Para ser claros, el neologismo «texto-imagen-texto» no se aplica únicamente a modelos CLIP, sino a todos los modelos en los que el texto constituye una parte constitutiva de la generación de imágenes.

6. Debido a sutiles divergencias en la traducción de determinados conceptos, utilizamos dos versiones distintas de la traducción del texto de Mira Schendel: una de ellas es un ensayo

independiente de Nancy Ann Roth publicado en *Tate Papers*, acompañado por su comentario (Roth, 2014); la otra, de Rodrigo Maltez Novaes, es un capítulo de *Groundless*, de Flusser (2017).

ANEXOS / FUENTES

Referencias bibliográficas

Se mantienen los títulos, datos editoriales y referencias tal como aparecen en el artículo original.

Bellour R (1996) The double helix. In: Druckrey T (ed.) *Electronic Culture: Technology and Visual Representation*. New York, NY: Aperture, 173–199.

Bender EM, et al. (2021) On the dangers of stochastic parrots: Can language models be too big? In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March, 610–623.

Berger J (2008[1972]) *Ways of Seeing*. London: Penguin Press.

Berkeley G (1949[1713–1734]) 'Three dialogues between Hylas and Philonous, Dialogue I'. In: Luce AA and Jessop TE (eds) *The Works of George Berkeley*, Vol. 2. London: Thomas Nelson, 37–39.

Bratton B and Agüera y Arcas B (2022) The model is the message. *Noema*, 12 July. Available at: https://www.noemamag.com/the-model-is-the-message/ (accessed 3 July 2025).

Bruner JS (1957) On perceptual readiness. *Psychological Review* 64(2): 123–152.

Crawford K (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.

Denson S (2024) From sublime awe to abject cringe: On the embodied processing of AI art. *Journal of Visual Culture* 22(2): 146–175.

Descartes R (1984[1641]) Meditations on first philosophy, Meditation I. In: Cottingham J, et al. (eds) *The Philosophical Writings of Descartes*, Vol. II. Cambridge: Cambridge University Press, 1–32.

Flusser V (2000) *Towards a Philosophy of Photography*. London: Reaktion.

Flusser V (2011) *Does Writing Have a Future?* Minneapolis: University of Minnesota Press.

Flusser V (2017) *Groundless*. Burgess Hill, West Sussex: Metaflux Publishing.

Hammond F (2025) *V3 Model Collapse*. London: The Photographers' Gallery.

Hayles NK (2017) *Unthought: The Power of the Cognitive Nonconscious*. Chicago, IL: University of Chicago Press.

Hayles NK (2025) *Bacteria to AI: Human Futures with Our Nonhuman Symbionts*. Chicago, IL: University of Chicago Press.

Horn BKP and Bachman BL (1978) Using synthetic images to register real images with surface models. *Communications of the ACM* 21(11): 914–924.

Hurlburt R and Schwitzgebel E (2007) *Describing Inner Experience? Proponent Meets Skeptic*. Cambridge, MA: MIT Press.

Jay M (1998) Scopic regimes of modernity. In: Foster H (ed.) *Vision and Visuality*. Seattle: Bay Press.

Kant I (1998[1781/1787]) *Critique of Pure Reason*. Cambridge: Cambridge University Press.

Mackenzie A and Munster A (2019) Platform seeing: Image ensembles and their invisualities. *Theory, Culture & Society* 35(5): 3–22.

Manavis S (2025) AI art and the ruins of human creativity. *New Statesman*. Available at: https://www.newstatesman.com/politics/media/2025/03/ai-art-and-the-ruins-of-human-creativity (accessed 3 July 2025).

Merleau-Ponty M (1945) *Phenomenology of Perception*. London: Routledge & Kegan Paul.

Merleau-Ponty M (1968) *The Visible and the Invisible*, ed. C Lefort. Evanston, IL: Northwestern University Press.

Meyer R (2025) 'Platform realism': AI image synthesis and the rise of generic visual content. *Transbordeur* 9.

Milton F, et al. (2021) Behavioral and neural signatures of visual imagery vividness extremes: Aphantasia versus hyperphantasia. *Cerebral Cortex Communications* 2(2): tgab035. DOI: 10.1093/texcom/tgab035.

Mitchell WJT (1984) What is an image? *New Literary History* 15(3): 503–537.

Mitchell WJT (1986) *Iconology: Image, Text, Ideology*. Chicago, IL: University of Chicago Press.

Mitchell WJT (1994) *Picture Theory: Essays on Verbal and Visual Representation*. Chicago, IL: University of Chicago Press.

Mitchell WJT (2005) *What Do Pictures Want? The Lives and Loves of Images*. Chicago, IL: University of Chicago Press.

Nake F (2005) Computer art: A personal recollection. *Proceedings of the 5th Conference on Creativity & Cognition*, London, 54–62.

Nanay B (2021) Mental imagery. In: Zalta EN (ed.) *The Stanford Encyclopedia of Philosophy* (Winter edn). Available at: https://plato.stanford.edu/archives/win2021/entries/mental-imagery/ (accessed 3 July 2025).

Nanay B (2023) *Mental Imagery: Philosophy, Psychology, Neuroscience*. Oxford: Oxford University Press.

Neisser U (2014[1967]) *Cognitive Psychology*. New York, NY: Psychology Press.

Parikka J (2023) *Operative Images: From the Visual to the Invisual*. Minneapolis: University of Minnesota Press.

Parker S (2025) You sound like ChatGPT. *The Verge*. Available at: https://www.theverge.com/openai/686748/chatgpt-linguistic-impact-common-word-usage (accessed 25 June 2025).

Pasquinelli M (2023) *The Eye of the Master: A Social History of Artificial Intelligence*. London: Verso.

Roth NA (2014) Mira Schendel's gesture: On art in Vilém Flusser's thought, with 'Mira Schendel' by Flusser. *Tate Papers* 21.

Seth A (2021) *Being You: A New Science of Consciousness*. London: Faber & Faber (Kindle edn).

Somaini A (2025) A theory of latent spaces. In: *The World Through AI*. Paris: Jeu de Paume and JBE Books.

Steyerl H (2023) Mean images. *New Left Review* 140/141, March–June: 5–28.

Toister Y (2020) *Photography from the Turin Shroud to the Turing Machine*. Bristol: Intellect.

Toister Y (2024) Seeing and sensing. *Media Theory* 8(1): 259–276.

Turing AM (1950) Computing machinery and intelligence. *Mind* 59(236): 433–460.

Underwood T (2022) Mapping the latent spaces of culture. *Startwords*. DOI: 10.5281/zenodo.6567481 (3 Parrots).

Warburg A (1999[1781]) Dürer and Italian antiquity. In: *The Renewal of Pagan Antiquity: Contributions to the Cultural History of the European Renaissance*. Los Angeles, CA: Getty Research Institute, 313–327.

Zylinska J (2023) *The Perception Machine: Our Photographic Future Between the Eye and AI*. Cambridge, MA: MIT Press.

Zylinska J (2024) Diffused seeing: The epistemological challenge of generative AI. *Media Theory* 8(1): 229–258.

ACERCA DE LOS AUTORES

Sobre los autores

Yanai Toister es artista, investigador y profesor asociado de Información Visual en la Universidad de Tampere (Finlandia). Su trayectoria comenzó en la fotografía y el arte conceptual, con exposiciones internacionales que luego nutrieron una transición hacia la teoría y la filosofía de los medios visuales. Su trabajo académico abarca estudios visuales, filosofía de los medios y teoría de la imagen, con especial atención al pensamiento algorítmico, las estéticas generativas y las operaciones de las imágenes sintéticas y técnicas. Su monografía *Photography from the Turin Shroud to the Turing Machine* (Intellect/University of Chicago Press, 2020) propuso una lectura crítica de la fotografía como medio de inscripción, representación y programación, cuestionando a la vez narrativas heredadas de la autoría. Su investigación actual desarrolla una teoría contemporánea de la imagen atenta al papel de los sistemas autónomos de sensado, las infraestructuras codificadas y los modos no visuales de registro. Dirección: Universidad de Tampere, Kelevantie 4, 33100 Tampere, Finlandia. Correo: yanai.toister@tuni.fi.

Joanna Zylinska es escritora, artista, curadora y profesora de Filosofía de los Medios y Práctica Digital Crítica en King's College London. Es autora de diversos libros, entre ellos *The Perception Machine: Our Photographic Future Between the Eye and AI* (MIT Press, 2023) y *AI Art: Machine Visions and Warped Dreams* (Open Humanities Press, 2020). Su práctica artística experimenta con diferentes medios basados en imágenes. Defensora del «acceso abierto radical», es codirectora de Open Humanities Press y editora de su colección MEDIA : ART : WRITE : NOW. Actualmente investiga percepción y cognición como zonas fronterizas entre la inteligencia humana y la maquínica, y examina diversas narrativas de colapso, tanto en el plano de los modelos de IA como en los sistemas planetarios. Dirección: King's College London, Strand Campus, Londres WC2R 2LS, Reino Unido. Correo: joanna.zylinska@kcl.ac.uk.

Información editorial de esta edición en español. Traducción del artículo original de Yanai Toister y Joanna Zylinska, publicado en *Journal of Visual Culture* 24(3), pp. 431–451, DOI: https://doi.org/10.1177/14704129251405715. Licencia de la obra original: CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Se ha realizado una adaptación lingüística al español y una nueva composición tipográfica; se han mantenido las referencias en el idioma en que aparecen en la obra fuente. TecnoPsique no está afiliada a la revista ni representa a los autores.