laFuga

El Nooscopio de manifiesto

La inteligencia artificial como instrumento de extractivismo del conocimiento

Por Matteo Pasquinelli, Vladan Joler

Tags | **Capitalismo cognitivo | Extractivismo | Filosofía de los medios**

Matteo Pasquinelli (PhD) es profesor de Filosofía de los Medios en la Universidad de Artes y Diseño de Karlsruhe, donde coordina el grupo de investigación sobre Inteligencia Artificial y Filosofía de los Medios KIM. Editó la antología Alleys of Your Mind: Augmented Intelligence and Its Traumas (Meson Press) y, con Vladan Joler, el ensayo visual "The Nooscope Manifested: AI as Instrument of Knowledge Extractivism" (nooscope.ai). Su investigación se centra en la intersección de las ciencias cognitivas, la economía digital y la inteligencia de las máquinas. Para Verso Books está preparando una monografía sobre la historia de la IA titulada provisionalmente El ojo del maestro.

1. Algunos esclarecimientos respecto del proyecto de mecanizar la razón

El Nooscopio es una cartografía de los límites de la inteligencia artificial, concebida como una provocación tanto a la ciencia de la computación como a las humanidades. Todo mapa presenta una perspectiva parcial, una manera de provocar debate. Del mismo modo, este mapa es un manifiesto –de disidentes de la IA. Su principal objetivo es desafiar las mistificaciones de la inteligencia artificial. Primero, en tanto definición técnica de la **inteligencia**, y segundo, como forma política supuestamente **autónoma** de la sociedad y de lo humano. [1] En la expresión "inteligencia artificial" el adjetivo "artificial" remite al mito de la autonomía tecnológica: insinúa las caricaturescas "mentes alienígenas" que se auto-reproducen en silicio, pero sobre todo, en realidad, mistifica dos procesos reales de alienación: el crecimiento de la autonomía geopolítica de las compañías *hi-tech* y la invisibilización de la autonomía de los trabajadores a escala global. El proyecto moderno de mecanización de la razón humana claramente ha mutado, en el siglo XXI, en un régimen corporativo de extracción del conocimiento y colonialismo epistémico. [2] Esto no sorprende, ya que los algoritmos de aprendizaje maquínico (*machine learning*) son los más poderosos algoritmos de compresión de información.

El propósito del Nooscopio es secularizar la IA, llevarla del status de "máquina inteligente" al de instrumento de conocimiento. En lugar de evocar leyendas de cognición alienígena, es más razonable considerar al aprendizaje maquínico como un **instrumento de magnificación del conocimiento**, que ayuda a percibir características, patrones y correlaciones a través de vastos espacios de datos inaccesibles al alcance humano. En la historia de la ciencia y la tecnología, esto no es nuevo: ya ha sido realizado por instrumentos ópticos durante las historias de la astronomía y la medicina. [3] En la tradición de la ciencia, el aprendizaje maquínico es solo un **Nooscopio**, un instrumento para ver y navegar en el espacio del conocimiento (del griego *skopein*, "examinar, mirar" y *noos* "conocimiento").

Tomando la idea de Gottfried Wilhelm Leibniz, el diagrama del Nooscopio aplica la analogía de los medios ópticos a la estructura de los aparatos de aprendizaje maquínico. En la discusión acerca del poder de su *calculus ratiocinator* y "números característicos" (la idea de diseñar un lenguaje numérico universal para codificar y resolver todos los problemas del razonamiento humano), Leibniz propone

una analogía con los instrumentos de magnificación visual como el microscopio y el telescopio. Escribe: "Una vez que los números característicos hayan sido establecidos para la mayoría de los conceptos, la humanidad poseerá un nuevo instrumento que mejorará las capacidades de la mente en mucha mayor medida que los instrumentos ópticos fortalecen nuestros ojos, y reemplazará al microscopio y al telescopio en la misma medida en que la razón es superior a la vista" (Leibniz, 1677/1951, p. 23). Aunque el objetivo de este texto no es reiterar la oposición entre culturas cuantitativas y cualitativas, el credo de Leibniz no necesita ser seguido. Las controversias no pueden ser computadas conclusivamente. El aprendizaje maquínico no es la mejor forma de inteligencia.

Los instrumentos de medición y percepción siempre tienen sus aberraciones intrínsecas. Del mismo modo en que las lentes de microscopios y telescopios nunca son perfectamente curvilíneas y suaves, las lentes lógicas del aprendizaje maquínico encarnan fallas y sesgos. Poder comprender el aprendizaje maquínico y registrar su impacto en la sociedad implica estudiar en qué grado los datos sociales son difractados y distorsionados por estas lentes. Esto se conoce generalmente como el debate sobre sesgos en IA, pero las implicaciones políticas de la forma lógica del aprendizaje maquínico son más profundas. El aprendizaje maquínico no trae una nueva época oscura sino una de racionalidad difractada, en la cual, como se mostrará, una episteme de causalidad es reemplazada por una de correlaciones automatizadas. De manera más general, la IA es un nuevo régimen de verdad, prueba científica, normatividad social y racionalidad, que frecuentemente toma la forma de *alucinación estadística.* Este diagrama manifiesto es otro modo de decir que la IA, el rey de la computación (fantasía patriarcal de conocimiento mecanizado, "algoritmo amo" y *máquina alfa*) está desnudo. Aquí estaremos espiando dentro de esta caja negra.

2. La cadena de montaje del aprendizaje maquínico: Datos, Algoritmo, Modelo

La historia de la IA es una historia de experimentos, fallas maquínicas, controversias académicas y rivalidades épicas por el financiamiento militar, conocidas popularmente como "inviernos de la IA". [4] Aunque la IA corporativa de hoy describe sus poderes en el lenguaje de la "magia negra" y la "cognición suprahumana", las técnicas actuales están aún en una fase experimental (Campolo & Crawford, 2020). La IA está en la misma etapa en que estaba la máquina de vapor cuando fue inventada, antes de que fueran descubiertas las leyes de la termodinámica necesarias para explicar y controlar su funcionamiento interno. De manera similar, hoy hay redes neuronales eficientes para reconocimiento de imágenes, pero no hay una *teoría del aprendizaje* que explique por qué funcionan tan bien y cómo es que fallan tan mal. Como toda invención, el paradigma del aprendizaje maquínico se consolidó lentamente, en este caso durante el último medio siglo. El algoritmo maestro no apareció de la noche a la mañana. Por el contrario, ha sido una construcción gradual de un método de computación que tiene que encontrar aún un lenguaje común. ¿Cómo puede esbozarse, entonces, una gramática crítica del aprendizaje maquínico que sea concisa y accesible, sin entrar en el juego paranoide de definir la Inteligencia General?

Como instrumento de conocimiento, el aprendizaje maquínico se compone de un objeto observable (**datos de entrenamiento**), un instrumento de observación (**algoritmo de aprendizaje**) y una representación final (**modelo estadístico**). El ensamblaje de estos tres elementos se propone aquí como un diagrama espurio y barroco del aprendizaje maquínico, denominado extravagantemente Nooscopio [5] Siguiendo con la analogía de los medios ópticos, el flujo de información del aprendizaje maquínico es como un haz de luz proyectado por los datos de entrenamiento, comprimido por el algoritmo y difractado hacia el mundo por la lente del modelo estadístico.

El diagrama del Nooscopio apunta a ilustrar al mismo tiempo dos lados del aprendizaje maquínico: **cómo funciona y cómo falla**, enumerando sus componentes principales, así como el amplio espectro de errores, limitaciones, aproximaciones, sesgos, fallas, falacias y vulnerabilidades que son propias de su paradigma. [6] Esta doble operación enfatiza que la IA no es un paradigma monolítico de racionalidad sino una arquitectura espuria hecha de técnicas de adaptación y trucos. Además, los límites de la IA no son simplemente técnicos sino que están imbricados con sesgos humanos. En el diagrama del Nooscopio, los componentes esenciales del aprendizaje maquínico se representan en el centro, las intervenciones y los **sesgos humanos** a la izquierda, las limitaciones y los **sesgos técnicos** a la derecha. Las lentes ópticas simbolizan sesgos y aproximaciones que representan la compresión y

distorsión del flujo de información. El sesgo total del aprendizaje maquínico está representado por la lente central del modelo estadístico a través del cual el mundo se difracta.

Las limitaciones actuales de la IA son generalmente percibidas gracias al discurso sobre el sesgo –la amplificación algorítmica de la discriminación por género, raza, capacidades y clase. En el aprendizaje maquínico, es necesario distinguir entre sesgos históricos, sesgos de datos y sesgos algorítmicos, los cuales ocurren en diferentes etapas del flujo de la información. [7] El **sesgo histórico** (o sesgo del mundo) ya es evidente en las sociedades antes de la intervención tecnológica. En cualquier caso, la naturalización de este sesgo, o sea, la integración de la desigualdad en una tecnología aparentemente neutral, es en sí misma dañina (Eubanks, 2018) (véase también (Crawford, 2017)). Parafraseando a Michelle Alexander, Ruha Benjamin lo ha llamado "el nuevo código Jim": "el empleo de nuevas tecnologías que reflejan y reproducen desigualdades existentes pero que son promovidas y percibidas como más objetivas o progresistas que los sistemas discriminatorios de épocas previas" (Benjamin, 2019, p. 5). El **sesgo de datos** se introduce a través de la preparación de los datos de entrenamiento por parte de operadores humanos. La parte más delicada del proceso es el etiquetado de los datos, en el cual las taxonomías viejas y conservadoras pueden causar una mirada distorsionada del mundo, tergiversando las diversidades sociales y exacerbando las jerarquías sociales (ver más abajo el caso de ImageNet).

El **sesgo algorítmico** (también conocido como sesgo maquínico, sesgo estadístico o sesgo del modelo, al cual el diagrama del Nooscopio le presta particular atención) es la amplificación adicional del sesgo histórico y del sesgo de datos producida por los algoritmos de aprendizaje maquínico. El problema del sesgo se originó principalmente por el hecho de que los algoritmos de aprendizaje maquínico están entre los más eficientes para la **compresión de información**, lo cual da lugar a cuestiones de resolución, difracción y pérdida de información. [8] Desde tiempos antiguos los algoritmos son procedimientos de naturaleza económica, diseñados para obtener un resultado en el menor número de pasos, consumiendo la menor cantidad de recursos: espacio, tiempo, energía y trabajo (Pasquinelli, en prensa). La carrera armamentística de las compañías de IA está, aún hoy, ocupada en encontrar los algoritmos más rápidos y simples con los cuales capitalizar los datos. Si bien la compresión de información produce una tasa máxima de ganancia en la IA corporativa, desde un punto de vista social produce discriminación y pérdida de diversidad cultural.

Mientras que las consecuencias de la IA son popularmente comprendidas bajo la cuestión del sesgo, el entendimiento común de sus limitaciones técnicas se conoce como el problema de la **caja negra**. El efecto de caja negra es una cuestión presente en las redes de aprendizaje profundo [9] (que filtran tanto la información que se vuelve imposible invertir sus cadenas de razonamiento), pero se ha vuelto un pretexto genérico para la opinión de que los sistemas de IA no solo son inescrutables y opacos, sino incluso "alienígenas" y fuera de control. [10] El efecto de caja negra es parte de la naturaleza de cualquier máquina experimental en un estadio temprano de desarrollo (ya se mencionó que la máquina de vapor permaneció como un misterio por algún tiempo, aun cuando ya se la había puesto a prueba exitosamente). El problema real es la retórica de la caja negra, fuertemente relacionada con las teorías conspirativas que predican que la IA es un poder oculto que no puede ser estudiado, conocido o políticamente controlado.

3. Los datos de entrenamiento: el origen social de la inteligencia maquínica

La digitalización masiva, que se expandió con Internet en los noventa y escaló con los centros de datos en los dos mil, ha puesto a disposición vastos recursos de datos que, por primera vez en la historia, son libres y desregulados. Un régimen de extractivismo del conocimiento (también conocido como Grandes Datos -*Big Data*- fue empleando gradualmente algoritmos eficientes para extraer "inteligencia" de esas fuentes abiertas de datos, mayoritariamente con el propósito de predecir comportamientos de consumo y vender publicidad. La economía del conocimiento se transformó en una forma novedosa de capitalismo llamada capitalismo cognitivo y, luego, *capitalismo de vigilancia,* por diferentes autores (Corsani et al., 2004; Zuboff, 2019). Fueron el desbordamiento informacional de Internet, los vastos centros de datos, microprocesadores más veloces y algoritmos para compresión de datos los que sentaron las bases para el surgimiento de los monopolios de la IA en el siglo XXI.

¿Qué tipo de objetos culturales y técnicos son los conjuntos de datos que constituyen la fuente de la IA? La calidad de los **datos de entrenamiento** es el factor más determinante de la así llamada "inteligencia" que extraen los algoritmos de aprendizaje maquínico. Hay una perspectiva importante a tener en cuenta para entender la IA y el Nooscopio. Los datos son la principal fuente de inteligencia. Los algoritmos son la segunda, son las máquinas que computan ese valor e inteligencia en un modelo. Sin embargo, los datos de entrenamiento no aparecen nunca en bruto, independientes e imparciales (ya son ellos mismos "algorítmicos") (Gitelman, 2013). El diseño, formateo y edición de los datos de entrenamiento es una empresa laboriosa y delicada, que es probablemente más significativa para los resultados finales que los parámetros técnicos que controlan el algoritmo de aprendizaje. El acto de elegir una fuente de datos en lugar de otra es la marca profunda que deja la intervención humana en el dominio de las mentes "artificiales".

El conjunto de datos de entrenamiento es una **construcción cultural,** no sólo técnica. Usualmente comprende datos de entrada que se asocian con datos ideales de salida, como imágenes con sus descripciones, llamadas etiquetas o metadatos. [11] Un ejemplo canónico podría ser una colección de un museo y su archivo, en el cual las obras de arte están organizadas por metadatos tales como autor, año, medio, etc. El proceso semiótico de asignar un nombre o una categoría a una imagen nunca es imparcial. Esta acción deja otra profunda huella humana en el resultado final de la cognición maquínica. Un conjunto de datos de entrenamiento para aprendizaje maquínico usualmente se compone siguiendo estos pasos: 1) producción: trabajo o fenómenos que producen información; 2) captura: codificación de la información en un formato de datos producido por un instrumento; 3) formateo: organización de los datos en un conjunto; 4) etiquetado: en aprendizaje supervisado, la clasificación de los datos en categorías (metadatos).

La inteligencia maquínica se entrena con vastos conjuntos de datos que se acumulan en maneras que no son ni técnicamente neutrales ni socialmente imparciales. Los datos en bruto no existen, ya que dependen del trabajo humano, datos personales y comportamientos sociales que se acumulan durante largos períodos, a través de redes extensas y taxonomías controversiales. [12] Los principales conjuntos de datos para aprendizaje maquínico (NMIST, ImageNet, Labelled Faces in the Wild, etc.) se originaron en corporaciones, universidades y agencias militares del Norte Global. Pero mirando con más cuidado, uno descubre una profunda división del trabajo que inerva el Sur Global a través de plataformas de *crowdsourcing* usadas para editar y validar datos. [13] La parábola del conjunto de datos de **ImageNet** ejemplifica los problemas de muchos de los datos para la IA. ImageNet es un conjunto de datos de entrenamiento para aprendizaje profundo que se ha vuelto un *benchmark* de facto para los algoritmos de reconocimiento de imágenes: efectivamente, la revolución del Aprendizaje Profundo se inició en 2012 cuando Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton ganaron el desafío anual de ImageNet con la red neuronal convolucional AlexNet. [14] ImageNet fue iniciada por la científica de la computación Fei-Fei Li en 2006. [15] Fei-Fei Li tuvo tres intuiciones para construir un conjunto de datos confiables para el reconocimiento de imágenes. Primero, descargar millones de imágenes libres de servicios web como Flickr y Google. Segundo, adoptar la taxonomía computacional **WorldNet** para etiquetar imágenes. [16] Tercero, subcontratar el trabajo de etiquetar millones de imágenes a través de la plataforma de *crowdsourcing Amazon Mechanical Turk*. Al final del día (y de la cadena de montaje), trabajadores anónimos de todo el planeta reciben unos centavos por la tarea de etiquetar cientos de imágenes por minuto de acuerdo a la taxonomía WorldNet: su trabajo resultó en la ingeniería de una construcción cultural controversial. La científica de IA Kate Crawford y el artista Trevor Paglen han investigado y revelado la sedimentación de categorías racistas y sexistas en la taxonomía de ImageNet: véase la legitimación de la categoría "persona fracasada, perdedora, quedada, poco exitosa" para cien imágenes arbitrarias de personas (Crawford & Paglen, 2019).

El voraz extractivismo de datos de la IA ha causado una inesperada regresión en la cultura digital: en el inicio de los dos mil Lawrence Lessig no podía predecir que el gran repositorio de imágenes online bajo licencias **Creative Commons** se volvería, una década después, un recurso desregulado para tecnologías de vigilancia por reconocimiento facial. De manera similar, los datos personales son continuamente incorporados, sin ninguna transparencia, a conjuntos privatizados de datos para aprendizaje maquínico. En 2019, el artista e investigador de IA Adam Harvey reveló, por primera vez, el uso no consentido de fotos personales en datos de entrenamiento para reconocimiento facial. La revelación de Harvey causó que la Universidad de Stanford, la Universidad Duke y Microsoft retiraran

sus conjuntos de datos en medio de un gran escándalo por **infringir el derecho a la privacidad** (Harvey & LaPlace, 2019; Murgia & Harlow, 2019). Los conjuntos de datos online provocan problemas de soberanía de datos y derechos civiles que las instituciones tradicionales son muy lentas en contrarrestar (ver el Reglamento General de Protección de Datos Europeo) [17] . Si 2012 fue el año en el que comenzó la revolución del aprendizaje profundo, 2019 fue el año en que sus fuentes se descubrieron como vulnerables y corruptas.

4. La historia de la IA como la automatización de la percepción

La necesidad de desmitificar la IA (al menos desde un punto de vista técnico) es algo que se entiende también en el mundo corporativo. Yann LeCun, director de IA de Facebook y padrino de las redes artificiales convolucionales, insiste en que los sistemas de IA actuales no son versiones sofisticadas de la cognición, sino de la percepción. De la misma manera, el diagrama del Nooscopio deja al descubierto el esqueleto de la caja negra de la IA y muestra que la IA no es un autómata pensante, sino un algoritmo que lleva a cabo **reconocimiento de patrones**. La noción de reconocimiento de patrones contiene cuestiones sobre las que es necesario profundizar. A todo esto, ¿qué es un patrón? ¿Se trata de una entidad meramente visual? ¿Qué significa considerar a los comportamientos sociales como patrones? ¿El reconocimiento de patrones es una definición exhaustiva de inteligencia? Lo más probable es que no lo sea. Para traer algo de luz sobre estas cuestiones sería bueno emprender una breve arqueología de la IA.

El arquetipo para reconocimiento de patrones es el Perceptron de Frank Rosenblatt. Fue inventado en 1957, en el Cornell Aeronautical Laboratory (New York, EE.UU.) y su nombre es una abreviatura de "Perceiving and Recognizing Automaton" (Rosenblatt, 1957). Dada una matriz de 20x20 fotoreceptores, el Perceptron puede aprender cómo reconocer letras individuales. Un patrón visual se graba como una impresión en una red de neuronas artificiales que disparan de manera orquestada con la aparición de imágenes similares y activan una sola neurona de salida. La neurona de salida dispara 1 = verdadero si una imagen dada fue reconocida, o 0 = falso si cierta imagen no fue reconocida.

La automatización de la percepción, considerada como un montaje de pixeles a lo largo de una cadena de montaje computacional, estaba originalmente implícita en el concepto de redes neuronales artificiales de McCulloch y Pitts (Pitts & McCulloch, 1947). Una vez que el algoritmo para reconocimiento visual de patrones pudo sobrevivir el "invierno de la IA" y hacia finales de los años 2000 logró volverse eficiente, se lo aplicó al conjunto de datos no-visuales, lo que verdaderamente inauguró la edad del Aprendizaje Profundo (Deep Learning) (la aplicación de técnicas de reconocimiento de patrones a toda clase de datos, no solo visuales). Hoy en día, en el caso de los autos autónomos, los patrones que hay que reconocer son objetos que puede haber en una calle. En el caso de la traducción automática, los patrones a reconocer son las secuencias de palabras más comunes que aparecen en textos bilingües. Sin importar su complejidad, desde el punto de vista numérico del aprendizaje maquínico, nociones tales como imagen, movimiento, forma, estilo y decisión ética pueden ser descritas como distribuciones estadísticas de patrones. En este sentido, el reconocimiento de patrones realmente se ha vuelto una nueva **técnica cultural** que puede ser usada en muchas áreas. A efectos explicativos, describimos al Nooscopio como una máquina que opera bajo tres modalidades: **aprendizaje**, **clasificación**, y **predicción**. En términos más intuitivos, podemos llamar a estas modalidades: extracción de patrones, reconocimiento de patrones, y generación de patrones.

El Perceptron de Rosenblatt fue el primer algoritmo que preparó el terreno hacia el aprendizaje maquínico en el sentido contemporáneo. En una época en la que "ciencia de la computación" todavía no se había adoptado como definición, se denominó al campo como "geometría computacional", mientras que Rosenblatt mismo lo llamó "conexionismo". Ahora bien, el objeto de estas redes neuronales era el de calcular una inferencia estadística. Lo que computa una red neuronal no es un patrón específico, sino la **distribución estadística de un patrón**. Apenas debajo de la superficie del marketing antropomórfico de la IA, es posible encontrar otro objeto técnico y cultural que es necesario examinar: el **modelo estadístico**. ¿Qué es el modelo estadístico en el aprendizaje maquínico? ¿Cómo se calcula? ¿Cuál es la relación entre un modelo estadístico y la cognición humana? Estos son temas centrales a esclarecer. En términos del trabajo de desmitificación que hay que llevar a cabo (también el de evaporar algunas preguntas ingenuas), sería bueno reformular la pregunta cliché

"¿Puede pensar una máquina?" en términos de las preguntas teóricamente más sólidas "¿Puede pensar un modelo estadístico?", "¿Puede desarrollar consciencia un modelo estadístico?, etc.

5. El algoritmo de aprendizaje: la compresión del mundo en un modelo estadístico

Es bastante usual que a los algoritmos de IA se los describa como fórmulas alquímicas, capaces de destilar formas "extrañas" de inteligencia. Pero, ¿qué hacen realmente los algoritmos de aprendizaje maquínico? Son pocos, incluso dentro de quienes están en la búsqueda de una AGI (Inteligencia Artificial General), los que se molestan en hacer esta pregunta. Algoritmo es el nombre de un proceso mediante el cual una máquina realiza un cálculo. El producto de tales procesos de máquina es un modelo estadístico (más precisamente denominado "modelo estadístico algorítmico"). En la comunidad de programadores, el término "algoritmo" está siendo, cada vez más, reemplazado por "modelo". Esta confusión terminológica surge del hecho de que el modelo estadístico no existe por separado del algoritmo: de alguna manera, el modelo estadístico existe dentro del algoritmo bajo la forma de una memoria que está distribuida entre sus parámetros. Es por esta misma razón que es esencialmente imposible visualizar un modelo estadístico algorítmico, tal y como se hace con funciones matemáticas simples. Aun así, el desafío vale la pena.

En el aprendizaje maquínico hay muchas arquitecturas de algoritmos: Perceptron simple, redes neuronales profundas, máquina de vectores de soporte, redes bayesianas, cadenas de Markov, *autoencoders*, máquinas de Boltzmann, etc. Cada una de estas arquitecturas tiene una historia diferente (casi siempre arraigada en agencias militares y corporaciones del Norte Global). Las redes neuronales artificiales comenzaron como estructuras computacionales muy simples que fueron evolucionando y ganando complejidad y que ahora podemos regular mediante unos pocos hiperparámetros que expresan millones de parámetros. [18] Por ejemplo, las redes neuronales convolucionales se describen mediante un conjunto limitado de hiperparámetros (número de capas, número de neuronas por capa, tipo de conexión, comportamiento de las neuronas, etc.) que proyectan una topología compleja de miles de neuronas artificiales con millones de parámetros en total. El algoritmo comienza como una página en blanco y, durante el proceso llamado entrenamiento, o "aprendizaje desde los datos", va ajustando sus parámetros hasta que alcanza una buena representación de los datos de entrada. En el reconocimiento de imágenes, como ya vimos, el cálculo de millones de parámetros tiene que resolverse en un salida binaria simple: 1 = verdadero, se reconoce cierta imagen; o 0 = falso, una imagen dada no se reconoce. [19]

Para intentar una explicación accesible de la relación entre el algoritmo y el modelo, demos una mirada al algoritmo complejo Inception v3, una red neuronal convolucional profunda para el reconocimiento de imágenes diseñada por Google y entrenada sobre el conjunto de datos ImageNet. Se dice que Inception v3 tiene una precisión del 78% a la hora de identificar la etiqueta de una imagen, pero el rendimiento de "inteligencia maquínica" en este caso puede ser medido también por la proporción entre el tamaño de los datos de entrenamiento y el algoritmo (o modelo) entrenado. ImageNet contiene 14 millones de imágenes con etiquetas asociadas que ocupan alrededor de 150 gigabytes de memoria. Por otro lado, Inception v3, que está pensado para representar la información contenida en ImageNet, solo tiene 92 megabytes. La relación de compresión entre los datos de entrenamiento y el modelo describe también de manera parcial la tasa de difracción de la información. En una tabla de la documentación de Keras se compara estos valores (número de parámetros, profundidad de las capas, dimensión del archivo y precisión) para los principales modelos de reconocimiento de imágenes. [20] Es una forma un tanto tosca, pero efectiva, de mostrar la relación entre modelo y datos, para mostrar cómo se mide y evalúa la "inteligencia" de los algoritmos en la comunidad de programadores.

Los modelos estadísticos siempre han tenido influencia sobre la cultura y la política. No surgieron únicamente con el aprendizaje maquínico: el aprendizaje maquínico es precisamente una nueva forma de automatizar la técnica del modelado estadístico. Cuando Greta Thunberg advierte "Escuchen a la ciencia", lo que realmente quiere decir, al ser ella una buena estudiante de matemática, es "Escuchen a los modelos estadísticos de la ciencia del clima". Sin modelos estadísticos, no hay ciencia del clima: sin ciencia del clima, no hay activismo climático. La ciencia del clima es, de hecho, un buen ejemplo por dónde empezar, si es que queremos comprender los modelos estadísticos. El calentamiento global se pudo calcular luego de, primero, recopilar un gran conjunto de datos de las temperaturas de la superficie de la Tierra cada día del año y, segundo, aplicar un modelo matemático

que traza la curva de las variaciones de la temperatura en el pasado y proyecta el mismo patrón hacia el futuro (Edwards, 2010). Los modelos climáticos son artefactos históricos que se prueban y debaten dentro de la **comunidad científica** y, hoy por hoy, incluso por fuera de ella [21] que ha sido desarrollado por el Centro Nacional para la Investigación Atmosférica en Boulder, Colorado, desde 1996. El CESM es una simulación numérica totalmente acoplada [22] del sistema de la Tierra que consiste en componentes atmosféricos, oceánicos, de hielo, de superficie terrestre, del ciclo de carbono, y otros. El CESM incluye un modelo de clima que brinda simulaciones de vanguardia del pasado, presente y futuro de la Tierra. http://www.cesm.ucar.edu] Los modelos de aprendizaje maquínico, por el contrario, son opacos e inaccesibles al debate por parte de la comunidad. Dado el grado de mitificación y los prejuicios sociales en torno a sus construcciones matemáticas, la IA ha inaugurado la era de la *ciencia ficción estadística*. El Nooscopio es el proyector de este gran cine estadístico.

6. Todos los modelos están mal, pero algunos son útiles

"Todos los modelos están mal, pero algunos son útiles": el *dictum* canónico del estadístico británico George Box ha encapsulado durante mucho tiempo las limitaciones lógicas de la estadísticas y del aprendizaje maquínico (Box, 1979). Esta máxima, sin embargo, se suele usar para legitimar los sesgos de la IA corporativa y estatal. Los informáticos argumentan que la cognición humana refleja la capacidad de abstraer y aproximar patrones. Entonces, ¿cuál es el problema con que las máquinas también hagan aproximaciones? Dentro de este argumento, se repite retóricamente que "el mapa no es el territorio". Esto suena razonable. Pero lo que debería ponerse en cuestión es que la IA es un mapa muy comprimido y distorsionado del territorio y que este mapa, como muchas formas de automatización, no está abierto a la negociación comunitaria. La IA es un mapa del territorio sin el acceso ni el consentimiento de la comunidad. [23]

¿Cómo hace el aprendizaje maquínico para trazar un mapa estadístico del mundo? Veamos el caso específico de reconocimiento de imágenes (la forma básica del **trabajo de percepción**, que ha sido codificado y automatizado como reconocimiento de patrones). [24] Dada una imagen para ser clasificada, el algoritmo detecta los bordes de un objeto como la distribución estadística de píxeles oscuros que están rodeados de píxeles claros (un patrón visual típico). El algoritmo no sabe qué es una imagen, no percibe una imagen como lo hace la cognición humana, sólo calcula píxeles, valores numéricos de brillo y proximidad. El algoritmo está programado para grabar solo el borde oscuro de un perfil (esto es, *ajustarse* al patrón deseado) y no todos los píxeles de la imagen (eso resultaría en sobreajustar y repetir todo el campo visual). Se dice que un modelo estadístico está entrenado de manera exitosa cuando puede **adaptarse** con elegancia solo a los patrones importantes de los datos de entrenamiento y aplicar esos patrones también a los nuevos datos "en la naturaleza". Si un modelo aprende el conjunto de datos de entrenamiento demasiado bien, reconocerá sólo coincidencias exactas de los patrones originales y pasará por alto aquellos "en la naturaleza" con similitudes cercanas. En este caso, el modelo está **sobreajustado** porque ha aprendido meticulosamente todo (incluido el ruido) y no puede distinguir un patrón de su fondo. Por otro lado, se dice que el modelo está **subajustado** cuando no puede detectar los patrones significativos de los datos de entrenamiento. Las nociones de sobreajuste de datos, ajuste y subajuste se pueden visualizar en un plano cartesiano.

El desafío de proteger la precisión del aprendizaje maquínico radica en poder encontrar un buen equilibrio entre el ajuste y el sobreajuste de los datos, algo que es difícil de lograr debido a la serie de sesgos de las máquinas. Aprendizaje maquínico es un término que, tanto como "IA", antropomorfiza una pieza de tecnología: el aprendizaje maquínico *no aprende nada* en el sentido propio de la palabra, como lo hace un humano; el aprendizaje maquínico simplemente mapea una distribución estadística de valores numéricos y establece una función matemática que, con suerte, se aproxima a la comprensión humana. Ahora bien, dicho esto, el aprendizaje maquínico puede, por esta razón, arrojar nuevas luces sobre las formas en que los humanos comprendemos.

El modelo estadístico de los algoritmos de aprendizaje maquínico también es una aproximación en el sentido de que adivina las partes faltantes del gráfico de los datos: ya sea a través de una **interpolación**, que es la predicción de una salida *y* dentro del intervalo conocido de la entrada *x* en el conjunto de datos de entrenamiento, o mediante una **extrapolación**, que es la predicción de la salida *y* más allá de los límites de *x*, lo que suele introducir altos riesgos de inexactitud. Esto es lo que significa "inteligencia" hoy en día dentro de la inteligencia artificial: extrapolar una función no lineal más allá de los límites de los datos conocidos. Como dice acertadamente Dan McQuillian: "No hay

ninguna inteligencia en la inteligencia artificial, ni siquiera aprende, aunque su nombre técnico sea aprendizaje maquínico, se trata simplemente de minimización matemática" (McQuillan, 2018a).

Es importante recordar que la "inteligencia" del aprendizaje maquínico no se basa en fórmulas exactas de análisis matemático, sino en algoritmos de **aproximación por fuerza bruta**. La forma de la función de correlación entre la entrada *x* y la salida *y* se calcula algorítmicamente, paso a paso, a través de tediosos procesos mecánicos de ajuste gradual (como el descenso por el gradiente, por ejemplo) que son equivalentes al cálculo diferencial de Leibniz y Newton. Se dice que las redes neuronales están entre los algoritmos más eficientes porque estos métodos diferenciales pueden *aproximar* la forma de cualquier función si se cuenta con las suficientes capas de neuronas y abundantes recursos computacionales. [25] La aproximación gradual por fuerza bruta de una función es la característica principal de la IA de hoy en día, y es solo desde esta perspectiva que se pueden entender sus potencialidades y sus limitaciones –en particular su creciente huella de carbono (el entrenamiento de redes neuronales profundas necesita de cantidades exorbitantes de energía debido a que el descenso por el gradiente y otros algoritmos de entrenamiento similares operan sobre la base de ajustes continuos infinitesimales (Strubell et al., 2019).

7. Del mundo al vector

Las nociones de ajuste de datos, sobreajuste, subajuste, interpolación y extrapolación pueden ser fácilmente visualizadas en dos dimensiones, pero los modelos estadísticos generalmente operan en espacios de datos multidimensionales. Antes de ser analizados, los datos se codifican en un **espacio vectorial multidimensional** que está bastante lejos de ser algo intuitivo. ¿Qué es un espacio vectorial y por qué es multidimensional? Cardon, Cointet y Mazière describen a la vectorización de datos de esta manera:

> Una red neuronal requiere que los valores de entrada para el cálculo tomen la forma de un vector. Por lo tanto, el mundo debe ser codificado de antemano en forma de una representación vectorial puramente digital. Mientras que ciertos objetos como las imágenes se pueden descomponer naturalmente en vectores, otros objetos deben ser "incrustados" [26] dentro de un espacio vectorial antes de que sea posible calcularlos o clasificarlos usando redes neuronales. Este es el caso del texto, que es el ejemplo prototípico. Para ingresar una palabra en una red neuronal, la técnica *Word2vec* lo "incrusta" en un espacio vectorial que permite medir su distancia respecto de las otras palabras en el corpus. Las palabras adquieren, así, una posición dentro de un espacio de varios cientos de dimensiones. La ventaja de tal representación está en las numerosas operaciones que permiten una transformación semejante. Dos términos cuyas posiciones inferidas están cerca en este espacio son semánticamente similares; se dice que estas representaciones están distribuidas: el vector del concepto "departamento" (- 0.2, 0.3, - 4.2, 5.1…) será similar al de "casa" (- 0.2, 0.3, - 4.0, 5.1…). (…) Mientras que el procesamiento del lenguaje natural fue el pionero en "incrustar" palabras en un espacio vectorial, hoy somos testigos de una generalización de procesos de incrustación que se extiende progresivamente a todos los campos de aplicaciones: las redes se están volviendo puntos simples en un espacio vectorial con *graph2vec*, los textos con *paragraph2vec*, las películas con *movie2vec*, los significados de las palabras con *sens2vec*, las estructuras moleculares con *mol2vec*, etc. Según Yann LeCun, el objetivo de los diseñadores de máquinas conexionistas es poner el mundo en un vector (*world2vec*). (Cardon et al., 2018)

El **espacio vectorial multidimensional** es otra razón por la cual la lógica del aprendizaje maquínico es difícil de comprender. El espacio vectorial es otra nueva técnica cultural, con la que vale la pena familiarizarse. El campo de las humanidades digitales, en particular, ha estado cubriendo la técnica de vectorización a través de la cual nuestro conocimiento colectivo es generado [27] y procesado imperceptiblemente. La definición original de ciberespacio elaborada por William Gibson profetizó, muy probablemente, la llegada de un espacio vectorial y no de la realidad virtual: "Una representación gráfica de datos extraídos de los bancos de cada computadora en el sistema humano. Complejidad inconcebible. Las líneas de luz oscilaban en el no espacio de la mente, los grupos y las constelaciones de datos. Como las luces de una ciudad que se aleja" (Gibson, 1984, p. 69).

Sin embargo, debe enfatizarse que el aprendizaje maquínico todavía se parece más a la artesanía que a la matemática exacta. La IA sigue siendo una historia de *hacks* y trucos, en lugar de intuiciones

místicas. Por ejemplo, un truco para la compresión de información es la **reducción de dimensionalidad**, que se utiliza para evitar la maldición de la dimensionalidad, que es el crecimiento exponencial de la variedad de características en el espacio vectorial. Las dimensiones de las categorías que muestran baja varianza en el espacio vectorial (es decir, cuyos valores fluctúan solo un poco) se agrupan para reducir los costos de cálculo. La reducción de dimensionalidad se puede usar para agrupar significados de palabras (como en el modelo *word2vec*) pero también puede conducir a una **reducción de categorías**, que puede tener un impacto en la representación de la diversidad social. La reducción de dimensionalidad puede reducir taxonomías e introducir sesgos, normalizando aún más la diversidad mundial y borrando identidades únicas (Samadi et al., 2018).

8. La sociedad de los robots de clasificación y predicción

La mayoría de las aplicaciones contemporáneas del aprendizaje maquínico se pueden describir en base a las dos modalidades de clasificación y de predicción, que delinean los contornos de una nueva sociedad de control y gobierno estadístico. La clasificación se conoce como *reconocimiento de patrones*, mientras que la predicción se puede definir también como *generación de patrones*. Un nuevo patrón se reconoce o se genera cuando se interroga el núcleo interno del modelo estadístico.

La **clasificación** por aprendizaje maquínico se suele emplear para reconocer un signo, un objeto o la cara de una persona, y para asignarles una categoría (etiqueta) que les corresponde de acuerdo a la taxonomía o a la convención cultural. Se alimenta al modelo con un archivo de entrada (por ejemplo, la foto de un rostro que fue capturada por una cámara de vigilancia) para determinar si se ajusta o no a su distribución estadística. En caso afirmativo, se le asigna como valor de salida la etiqueta correspondiente. Desde los tiempos del Perceptron, la clasificación ha sido la aplicación original de las redes neuronales: con el *Deep Learning*, esta técnica se ha vuelto ubicua en los clasificadores de reconocimiento facial desplegados tanto por las fuerzas policiales como por los fabricantes de teléfonos inteligentes.

La **predicción** por aprendizaje maquínico se utiliza para proyectar tendencias y comportamientos futuros de acuerdo a los del pasado, es decir, para completar una pieza de la información conociendo solo una parte de ella. En la modalidad de predicción, se usa una pequeña muestra de datos de entrada (una imprimación [28]) para predecir la parte que falta de la información, siguiendo una vez más la distribución estadística del modelo (esta podría ser la parte de un gráfico numérico orientado hacia el futuro o la parte que falta de una imagen o de archivo de audio). Por cierto, existen otras modalidades de aprendizaje maquínico: la distribución estadística de un modelo se puede visualizar dinámicamente a través de una técnica llamada exploración espacial latente y también, en algunas aplicaciones de diseño recientes, *exploración de patrones*. [29]

La clasificación y la predicción por aprendizaje maquínico se están convirtiendo en técnicas ubicuas que constituyen nuevas formas de vigilancia y de gobierno. Algunos aparatos, como los vehículos autónomos y los robots industriales, pueden ser una integración de ambas modalidades. Un vehículo autónomo está entrenado para reconocer diferentes objetos en una calle (personas, automóviles, obstáculos, señales) y predecir acciones futuras en base a decisiones que un conductor humano ha tomado en circunstancias similares. Incluso si reconocer un obstáculo en la calle parece ser un gesto neutral (no lo es), identificar a un ser humano según las categorías de género, raza y clase (y en la reciente pandemia de COVID-19 como enfermo o inmune), como lo hacen cada vez más las instituciones estatales, es el gesto de un nuevo régimen disciplinario. La arrogancia de la clasificación automatizada ha provocado el resurgimiento de las técnicas reaccionarias lombrosianas que se creían ya confinadas a la histora, técnicas como el reconocimiento automático de género (AGR), "un subcampo de reconocimiento facial que tiene como objetivo identificar algorítmicamente el género de las personas a partir de fotografías o videos" (Keyes, 2018).

Recientemente, la modalidad generativa del aprendizaje maquínico ha tenido un impacto cultural: su uso en la producción de artefactos visuales ha sido recibida por los medios de comunicación como la idea de que la inteligencia artificial es "creativa" y puede hacer arte de forma autónoma. Una obra de arte que se dice que fue creada por IA siempre esconde un operador humano, quien aplicó la modalidad generativa de una red neuronal entrenada en un conjunto de datos. En esta modalidad, la red neuronal se ejecuta *hacia atrás* (pasando de la capa de salida más pequeña hacia la capa de entrada más grande) para generar nuevos patrones después de haber sido entrenada para clasificarlos, un

proceso que generalmente se mueve desde la capa de entrada más grande hacia la capa de salida más pequeña. La modalidad generativa, sin embargo, tiene algunas aplicaciones útiles: puede usarse como una especie de verificación de la realidad para revelar qué es lo que el modelo aprendió, es decir, para mostrar cómo el modelo "ve el mundo". Por ejemplo, se la puede aplicar al modelo de un vehículo autónomo para verificar cómo se proyecta la situación en la calle.

Una forma algo famosa de ilustrar cómo un modelo estadístico "ve el mundo" es Google DeepDream. DeepDream es una red neuronal convolucional basada en Inception (que está entrenada en el conjunto de datos ImageNet mencionado anteriormente) que fue programada por Alexander Mordvintsev para proyectar patrones alucinatorios. Mordvintsev tuvo la idea de "poner la red al revés", es decir, convertir un clasificador en un generador, utilizando algo de ruido aleatorio o imágenes de paisaje genéricas como entradas (Mordvintsev et al., 2015). Descubrió que "las redes neuronales que fueron entrenadas para discriminar entre diferentes tipos de imágenes tienen también mucha de la información que se necesita para generar imágenes". En los primeros experimentos de DeepDream, plumas de pájaros y ojos de perros comenzaron a surgir por todas partes en tanto que las razas de los perros y las especies de pájaros están muy sobrerrepresentadas en ImageNet. También se descubrió que la categoría "mancuerna" se aprendió siempre adherida a un brazo humano surrealista. Prueba de que muchas otras categorías de ImageNet están mal representadas.

Las dos modalidades principales de clasificación y generación pueden ensamblarse en arquitecturas más complejas como en las redes generativas antagónicas (RGA). En la arquitectura RGA (GAN, en inglés), una red neuronal con el rol de *discriminador* (un clasificador tradicional) tiene que reconocer una imagen producida por otra red neuronal que tiene el rol de *generador*, formando un bucle de refuerzo que entrena a los dos modelos estadísticos simultáneamente. Para algunas propiedades convergentes de sus respectivos modelos estadísticos, las RGAs demostraron ser muy buenas a la hora de generar imágenes altamente realistas. Esta habilidad ha provocado su abuso en la fabricación de "falsificaciones profundas" (*deep fakes*). [30] Con respecto a los regímenes de verdad, una controvertida aplicación similar es el uso de RGAs para generar datos sintéticos en la investigación del cáncer, en la que redes neuronales entrenadas en conjuntos desbalanceados de datos de tejidos cancerosos han comenzado a alucinar cánceres en donde no había ninguno (Cohen et al., 2018). En este caso, "en lugar de descubrir cosas, estamos inventando cosas", observa Fabian Offert, "el espacio de descubrimiento es idéntico al espacio de conocimiento que la RGA ya tenía. (...) Mientras pensamos que estamos viendo a través de la RGA –mirando algo con la ayuda de una RGA– en realidad estamos viendo *dentro* de una RGA. La visión por RGA no es realidad aumentada, es realidad virtual. Las RGAs confunden el descubrimiento y la invención" (Offert, 2020). La simulación por RGA de cáncer cerebral es un trágico ejemplo de alucinación científica impulsada por la IA.

9. Fallas de un instrumento estadístico: la no detección de lo nuevo.

El poder normativo de la IA en el siglo XXI debe ser sometido a escrutinio en estos tres términos epistémicos: ¿qué significa enmarcar el conocimiento colectivo como patrones, y qué significa trazar espacios vectoriales y distribuciones estadísticas de los comportamientos sociales? De acuerdo con Foucault, en la Francia de la modernidad temprana, el poder estadístico ya se utilizaba para medir las normas sociales y discriminar entre comportamiento normal y anormal (Foucault, 2004, p. 26). La IA extiende fácilmente el "poder de normalización" de las instituciones modernas –entre otras la burocracia, la medicina y la estadística (originalmente, el conocimiento numérico que poseía el Estado sobre su población)–, que ahora pasa a manos de las corporaciones de IA. La norma institucional se ha vuelto una norma computacional: la clasificación de sujetos, de cuerpos y de comportamientos, ya no parece ser un asunto para los registros públicos, sino, por el contrario, para algoritmos y centro de datos. [31] La "racionalidad data-céntrica", concluye Paula Duarte, "debería entenderse como la expresión de la colonialidad del poder" (Ricaurte, 2019).

Sin embargo, siempre persiste un salto, una fricción, un conflicto, entre los modelos estadísticos de la IA y el sujeto humano supuestamente medido y controlado. Este salto lógico entre los modelos estadísticos y la sociedad usualmente se discute como **sesgo**. Se ha demostrado ampliamente cómo el reconocimiento facial representa de manera errónea a las minorías sociales y cómo los servicios de logística y entregas manejados por IA evitan los barrios de población negra (Ingold & Soper, 2016). Si las discriminaciones de género, raza y clase son amplificadas por los algoritmos de IA, esto es parte

de un problema mayor de discriminación y normalización en el núcleo lógico del aprendizaje maquínico. La limitación lógica y política de la IA radica en la dificultad de la tecnología para **el reconocimiento y la predicción de un evento nuevo**. ¿De qué modo el aprendizaje maquínico lidia con una anomalía realmente única, un comportamiento social poco común, un acto innovador o una disrupción? Las dos modalidades del aprendizaje maquínico muestran una limitación que no es solamente un sesgo.

Un límite lógico de la clasificación por aprendizaje maquínico, o del reconocimiento de patrones, es la incapacidad de reconocer la **anomalía única** que aparece por primera vez, tal como una nueva metáfora en poesía, un nuevo chiste en la conversación cotidiana, o un objeto inusual (¿una peatona, una bolsa plástica?) en el escenario de la calle. La **no detección de lo nuevo** (algo que nunca "ha sido visto" por un modelo y por ello nunca ha sido clasificado antes en una categoría conocida) es un problema particularmente peligroso para los autos con conducción automática y que ya ha causado muertes. La predicción del aprendizaje maquínico, o la generación de patrones, evidencian fallas similares para la suposición de futuras tendencias y comportamientos. Como técnica de compresión de la información, el aprendizaje maquínico automatiza la dictadura del pasado, de taxonomías pasadas y de patrones de comportamiento sobre el presente. Este problema puede denominarse la **regeneración de lo viejo**: la aplicación de una visión homogénea de espacio-tiempo que restringe la posibilidad de un nuevo evento histórico.

Resulta interesante que, en el aprendizaje maquínico, la definición lógica de un asunto de seguridad también describe el límite lógico de su potencial creativo. Los problemas característicos de la **predicción de lo nuevo** están relacionados lógicamente con aquellos que caracterizan la **generación de lo nuevo**, porque el modo en el que el algoritmo de aprendizaje maquínico predice una tendencia en una serie temporal [32] es idéntico al modo en el que genera una nueva obra de arte a partir de patrones aprendidos. La trillada pregunta "¿la IA puede ser creativa?" debería reformularse en términos técnicos: ¿El aprendizaje maquínico puede crear obras que no sean imitaciones del pasado? ¿El aprendizaje maquínico puede extrapolarse más allá de los límites estilísticos de sus datos de entrenamiento? La "creatividad" del aprendizaje maquínico se limita a la detección de estilos a partir de sus datos de entrenamiento y luego a la improvisación aleatoria dentro de esos estilos. En otras palabras, el aprendizaje maquínico puede explorar e improvisar sólo dentro de los límites lógicos que están establecidos por los datos de entrenamiento. Para todos estos asuntos, y su grado de compresión de información, sería más preciso denominar el arte de aprendizaje maquínico como *arte estadístico*.

Otro error del aprendizaje maquínico, del que no se habla, es que se suele adoptar la correlación estadística entre dos fenómenos para explicar la causación de uno por el otro. En estadística, normalmente se entiende que la **correlación no implica causación**, en el sentido de que una sola coincidencia estadística no es suficiente para demostrar causalidad. Un ejemplo trágico es el trabajo del estadístico Frederick Hoffman, que en 1896 publicó un reporte de 330 páginas, destinado a agencias aseguradoras, para demostrar una *correlación racial* entre ser afroamericano y tener una baja expectativa de vida (O'Neil, 2017, Capítulo 9). Haciendo minería de datos superficial, el aprendizaje maquínico puede construir correlaciones arbitrarias que luego son percibidas como reales. En 2008 esta falacia lógica fue abrazada con orgullo por el director de *Wired*, Chris Anderson, quien declaró el "final de la teoría" porque "la avalancha de datos vuelve obsoleto al método científico" (Anderson, 2008). [33] De acuerdo a Anderson, quien no es ningún experto en el método científico ni en la inferencia lógica, si la correlación estadística es suficiente para que Google lleve adelante su negocio de avisos, por ello también debe ser lo suficientemente buena para descubrir paradigmas científicos automáticamente. Incluso Judea Pearl, pionero de las redes bayesianas, cree que el aprendizaje maquínico está obsesionado con el "ajuste de curvas", registrando correlaciones sin ofrecer explicaciones (Pearl & Mackenzie, 2018). Semejante falacia lógica ya se ha vuelto una falacia política, si uno tiene en cuenta que las fuerzas policiales en todo el mundo han adoptado algoritmos predictivos de vigilancia. [34] Según Dan McQuillan, cuando se aplica de este modo el aprendizaje maquínico a la sociedad, se vuelve un aparato biopolítico de **acción preventiva**, que produce subjetividades que pueden ser subsecuentemente criminalizadas (McQuillan, 2018b). Básicamente, el aprendizaje maquínico obsesionado con el "ajuste de curvas" impone una *cultura estadística* y reemplaza la episteme tradicional de causación (y de responsabilidad política) con una de correlaciones guiada ciegamente por la automatización de la toma de decisiones.

10. Inteligencia antagónica vs. Inteligencia artificial

Hasta aquí hemos mostrado paso a paso las difracciones y alucinaciones estadísticas del aprendizaje maquínico a través de las múltiples lentes del Nooscopio. En este punto, la orientación del instrumento debe ser revertida: las teorías científicas tanto como los dispositivos computacionales están inclinados a consolidar una perspectiva abstracta: la "visión desde ningún lugar" científica, que a menudo es sólo el punto de vista del poder. El estudio obsesivo de la IA puede arrastrar al académico a un abismo de computación y la ilusión de que la forma técnica ilumina la social. Como observa Paola Ricaurte: "El extractivismo de los datos asume que todo es fuente de datos" (Ricaurte, 2019). ¿Cómo emanciparnos de una mirada dato-céntrica del mundo? Es tiempo de darse cuenta de que no es el modelo estadístico el que construye al sujeto, sino más bien el sujeto el que estructura el modelo estadístico. Los estudios internalistas y externalistas sobre IA deben difuminarse: las subjetividades hacen a las matemáticas del control desde dentro, no desde fuera. Apoyando lo que una vez dijo Guattari de las máquinas en general, la inteligencia maquínica también está constituida de "formas híper-desarrolladas e híper-concentradas de ciertos aspectos de la subjetividad humana" (Guattari, 1989/2013, p. 2).

En lugar de estudiar exclusivamente cómo funciona la tecnología, la investigación crítica también estudia cómo se rompe, cómo los sujetos se rebelan contra su control normativo y los trabajadores sabotean sus engranajes. En este sentido, una forma de hacer resonar los límites de la IA es mirar hacia ciertas prácticas de hackeo. Hackear es un método importante de producción de conocimiento, una sonda epistémica crucial en la obscuridad de la IA. [35] Los sistemas de aprendizaje profundo para reconocimiento facial han disparado, por ejemplo, formas de activismo de contra-vigilancia. Mediante técnicas de ofuscación facial, los humanos han decidido volverse ininteligibles ante la inteligencia artificial: esto es, volverse, en sí mismos, **cajas negras**. Las técnicas tradicionales de **ofuscación** contra la vigilancia inmediatamente adquieren una dimensión matemática en la era del aprendizaje maquínico. Por ejemplo, el artista e investigador de IA Adam Harvey ha inventado un camuflaje textil llamado HyperFace que engaña los algoritmos de visión computacional [36] para que vean múltiples rostros donde no hay ninguno (Harvey, 2016). El trabajo de Harvey presenta la pregunta: ¿qué constituye un rostro para un ojo humano, por un lado, y para el algoritmo de visión computacional, por el otro? Los fallos (*glitches)* neuronales de HyperFace explotan esa brecha cognitiva (*cognitive gap)* y revelan cómo se ve un rostro humano para una máquina. Esta brecha entre percepción humana y maquínica ayuda a introducir el campo creciente de ataques antagónicos.

Los **ataques antagónicos** explotan puntos ciegos y regiones débiles en el modelo estadístico de una red neuronal, generalmente para engañar a un clasificador y hacerlo percibir algo que no está allí. En el reconocimiento de objetos, un ejemplo antagónico puede ser una imagen alterada de una tortuga, que parece inocua para el ojo humano pero que resulta mal clasificada como un rifle por una red neural (Athalye et al., 2018). Pueden llevarse a cabo ejemplos antagónicos como objetos 3D e incluso calcomanías para señales de tráfico que pueden desorientar a los autos autónomos (que pueden leer una velocidad máxima de 120 km/h donde en realidad es de 50 km/h) (Morgulis et al., 2019). Los ejemplos se diseñan sabiendo lo que una máquina nunca ha visto antes. Este efecto también se alcanza haciendo ingeniería inversa del modelo estadístico o contaminando el conjunto de datos de entrenamiento. En este último sentido, la técnica de **contaminación de los datos** apunta al conjunto de datos de entrenamiento e introduce datos adulterados (*doctored*). Al hacer eso, altera la precisión del modelo estadístico y crea una puerta trasera que eventualmente puede ser explotada por un ataque antagónico. [37] El ataque antagónico parece apuntar a una vulnerabilidad matemática que es común a todos los modelos de aprendizaje maquínico: "Un aspecto interesante de los ejemplos antagónicos es que un ejemplo generado por un modelo a menudo es mal clasificado por otros modelos, incluso cuando tienen arquitecturas diferentes o han sido entrenados con conjuntos de entrenamiento disjuntos [38] " (Goodfellow et al., 2015). Los ataques antagónicos nos recuerdan la discrepancia entre la percepción humana y la maquínica y el hecho de que el límite lógico del aprendizaje maquínico es también un límite político. La frontera lógica y ontológica del aprendizaje maquínico es el sujeto rebelde o el evento anómalo que escapa a la clasificación y al control. El sujeto del control algorítmico contraataca. Los ataques antagónicos son un modo de sabotear la cadena de montaje del aprendizaje maquínico mediante la invención de un obstáculo virtual que pueda sacar de quicio al aparato de control. El ejemplo antagónico es el *sabot* en la era de la IA.

11. Trabajo en la era de la IA

Es necesario aclarar la naturaleza de la entrada (*input)* y la salida (*output)* del aprendizaje maquínico. La IA presenta problemas no solo respecto del sesgo de la información, sino también respecto del trabajo. La IA no es sólo un aparato de control, sino también un aparato productivo. Como recién mencionamos, hay una fuerza de trabajo invisible implicada en cada paso de su cadena de montaje (composición del conjunto de datos, supervisión del algoritmo, evaluación del modelo, etc.). Oleoductos de tareas sin fin inervan del Norte Global hacia el Sur Global; plataformas colaborativas de trabajadores de Venezuela, Brasil e Italia, por ejemplo, son cruciales para enseñarles a los autos autónomos alemanes "cómo ver" (Schmidt, 2019). Contra la idea de la operación de una red de inteligencia alienígena, debe enfatizarse que, en la totalidad del proceso computacional de la IA, el trabajador humano nunca ha salido del bucle, o de modo más preciso, nunca ha salido de la cadena de montaje. Mary Gray y Siddharth Suri acuñaron el término "**trabajo fantasma**" para referirse al trabajo invisible que hace que la IA parezca artificialmente autónoma.

Más allá de algunas decisiones básicas, la inteligencia artificial contemporánea no puede funcionar sin humanos en el bucle (*loop*). Ya sea que esté ofreciendo una selección relevante de noticias o llevando a cabo un complicado pedido de pizza mediante mensaje de texto, cuando la inteligencia artificial (IA) se obtura o no puede terminar el trabajo, miles de negocios convocan a gente para completar silenciosamente el proyecto. Esta nueva cadena de montaje digital suma el *input* de trabajadores distribuidos, envía partes de proyectos más que productos, y opera a través de un servidor de sectores económicos durante todas las horas del día y de la noche. (Gray & Suri, 2019)

La automatización es un mito; debido a que las máquinas, incluyendo a la IA, solicitan constantemente la ayuda humana, algunos autores han sugerido reemplazar "automatización" por el término más preciso **heteromatización** (Ekbia & Nardi, 2017). La heteromatización significa que la narrativa usual de la IA como *perpetuum mobile* sólo es posible gracias a un ejército de reserva de trabajadores.

Sin embargo, hay un modo más profundo en que el trabajo constituye a la IA. La fuente de información del aprendizaje maquínico (cualquiera sea su nombre: datos de entrada, datos de entrenamiento, o sólo datos) siempre es una representación de habilidades, actividades y comportamientos humanos; una producción social, en suma. Todos los conjuntos de datos de entrenamiento son, implícitamente, un diagrama de la división del trabajo humano que la IA debe analizar y automatizar. Los conjuntos de datos para el reconocimiento de imágenes, por ejemplo, registran el trabajo visual que conductores, guardias y supervisores usualmente realizan durante sus tareas. Incluso los conjuntos de datos científicos dependen del trabajo científico, la planificación de experimentos, la organización de los laboratorios, y la observación analítica. El flujo de información de la IA debe comprenderse como un aparato diseñado para extraer "inteligencia analítica" de las formas de trabajo más diversas, y para transferir tal inteligencia a una máquina (lo que incluye, obviamente, dentro de la definición de trabajo, a formas extendidas de producción social, cultural y científica) [39] . Resumiendo, el origen de la inteligencia maquínica es la **división del trabajo** y su principal propósito es la **automatización del trabajo**.

Los historiadores de la computación ya han enfatizado que los primeros pasos de la inteligencia maquínica tuvieron lugar durante el proyecto decimonónico de mecanizar la división del trabajo mental, específicamente la tarea del cálculo manual (Daston, 1994; Jones, 2016; Schaffer, 1994). Desde entonces, el proyecto de la computación ha sido una combinación de vigilancia y disciplinamiento del trabajo, de cálculo óptimo de la plusvalía, y de planificación de los comportamientos colectivos (Pasquinelli, 2019a). La computación fue establecida por, y aún impone, un régimen de visibilidad e inteligibilidad, no sólo de razonamiento lógico. La genealogía de la IA como un aparato de poder se confirma hoy por su empleo extendido en tecnologías de identificación y predicción, y sin embargo la anomalía central que aún continúa esperando ser computada es la *desorganización del trabajo*.

Como tecnología de automatización, la IA tendrá un impacto tremendo en el mercado laboral. Si el aprendizaje profundo tiene una tasa de error del 1% en el reconocimiento de imágenes, por ejemplo, significa que aproximadamente el 99% del trabajo rutinario basado en tareas visuales (por ejemplo, la seguridad aeroportuaria) puede ser potencialmente reemplazado (cuando las restricciones legales y la

oposición sindical lo permitan). El impacto de la IA en el trabajo está bien descrito (desde la perspectiva de los trabajadores, finalmente) en un informe del Instituto de Sindicatos Europeos, que resalta "siete dimensiones esenciales que la futura regulación debería atender para proteger a los trabajadores: 1) resguardar la privacidad y protección de datos de los trabajadores; 2) abordar la vigilancia, el seguimiento, y el monitoreo; 3) transparentar los propósitos de los algoritmos de IA; 4) garantizar el ejercicio del 'derecho a la explicación' en relación a las decisiones tomadas por algoritmos o modelos de aprendizaje maquínico; 5) preservar la seguridad e integridad de los trabajadores en las interacciones humano-máquina; 6) promover la autonomía de los trabajadores en las interacciones humano-máquina; 7) permitir a los trabajadores alfabetizarse en IA" (Ponce, 2020).

Finalmente, el Nooscopio manifiesta una nueva Pregunta por la Maquinaria en la era de la IA. La Pregunta por la Maquinaria (the Machinery Question) fue un debate que se desató en Inglaterra durante la revolución industrial, cuando la respuesta al uso de máquinas y el subsecuente desempleo tecnológico de los trabajadores constituyó una campaña social para una mayor educación sobre las máquinas, que tomó la forma del Movimiento del Instituto de Mecánicos (Berg, 1980). Hoy requerimos una Pregunta por la Maquinaria para desarrollar más inteligencia colectiva sobre la "inteligencia maquínica", más educación pública en lugar de "máquinas que aprenden" tanto como de su régimen de extractivismo cognitivo (que reafirma las viejas rutas coloniales, basta sólo con mirar hoy el mapa de las plataformas colaborativas). También en el Norte Global, esta relación colonial entre la IA corporativa y la producción de conocimiento como un bien común debe pasar a primer plano. El propósito del Nooscopio es exponer el cuarto secreto del Turco Mecánico corporativo e iluminar el trabajo de conocimiento invisible que hace a la inteligencia maquínica aparecer como una fuerza ideológica autónoma.

Referencias

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. Wired magazine, 16(7), 16–07.

Athalye, A., Engstrom, L., Ilyas, A., & Kwok, K. (2018). Synthesizing Robust Adversarial Examples. arXiv:1707.07397 (cs). http://arxiv.org/abs/1707.07397

Beller, J. (2006). The cinematic mode of production: Attention economy and the society of the spectacle. UPNE.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity.

Berg, M. (1980). The Machinery Question and the Making of Political Economy 1815–1848. Cambridge University Press. https://doi.org/10.1017/CBO9780511560330

Box, G. E.(1979). Robustness in the strategy of scientific model building. En Robustness in statistics (pp. 201–236). Elsevier.

Campolo, A., & Crawford, K. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. Engaging Science, Technology, and Society, 6(0), 1-19. https://doi.org/10.17351/ests2020.277

Cardon, D., Cointet, J.-P., & Mazières, A. (2018). Neurons spike back: The invention of inductive machines and the artificial intelligence controversy. Réseaux, n° 211(5), 173-220. https://doi.org/10.3917/res.211.0173

Cohen, J. P., Luck, M., & Honari, S. (2018). Distribution matching losses can hallucinate features in medical image translation. En A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 (pp. 529-536). Springer International Publishing. https://doi.org/10.1007/978-3-030-00928-1_60

Corsani, A., Paulré, B., Dieuaide, P., Lazzarato, M., Monnier, J.-M., Moulier Boutang, Y., & Vercellone, C. (2004). Le capitalisme cognitif comme sortie de la crise du capitalisme industriel. Un programme

de recherche. Laboratoire Isys Matisse, Paris.

Crawford, K. (2017, diciembre 5). The trouble with bias (Keynote lecture). Thirty-first Conference on Neural Information Processing Systems (NeurIPS | 2017), Long Beach Convention Center, Long Beach. https://nips.cc/Conferences/2017/Schedule?showEvent=8742

Crawford, K., & Paglen, T. (2019, septiembre 19). Excavating AI: The politics of training sets for machine learning. -. https://www.excavating.ai

Daston, L. (1994). Enlightenment Calculations. Critical Inquiry, 21(1), 182-202. JSTOR.

Daston, L. (2018). Calculation and the division of labor, 1750-1950. Bulletin of the German Historical Institute, 62(Spring), 9-30.

Edwards, P. N.(2010). A vast machine: Computer models, climate data, and the politics of global warming. MIT Press.

Ekbia, H. R., & Nardi, B. A.(2017). Heteromation, and Other Stories of Computing and Capitalism (Edición: Illustrated). The MIT Press.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Publishing Group.

Foucault, M. (2004). Abnormal: Lectures at the Collège de France, 1974-1975 (G. Burchell, Trad.; First edition). Picador.

Foucault, M. (2010). Las palabras y las cosas. Una arqueologia de las ciencias humanas,. Siglo XXI. (Original work published 1966)

Gibson, W. (1984). Neuromancer. Ace Books.

Gitelman, L. (Ed.). (2013). «Raw data» is an oxymoron. The MIT Press.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. arXiv:1412.6572 (cs, stat). http://arxiv.org/abs/1412.6572

Gray, M. L., & Suri, S. (2019). Ghost work: How to stop Silicon Valley from building a new global underclass (Edición: Illustrated). Houghton Mifflin Harcourt.

Guattari, F. (2013). Schizoanalytic Cartographies (A. Goffey, Trad.; 1 edition). Bloomsbury Academic. (Original work published 1989)

Harvey, A. (2016). HyperFace project. https://ahprojects.com/hyperface/

Harvey, A., & LaPlace, J. (2019). MegaPixel Project. MegaPixels. https://megapixels.cc/about/

Ingold, D., & Soper, S. (2016, abril 21). Amazon Doesn't Consider the Race of Its Customers. Should It? Bloomberg.Com. http://www.bloomberg.com/graphics/2016-amazon-same-day/

Jones, M. L.(2016). Reckoning with Matter: Calculating Machines, Innovation, and Thinking about Thinking from Pascal to Babbage. University of Chicago Press.

Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-22. https://doi.org/10.1145/3274357

Krizhevsky, A., Sutskever, I., & Hinton, G. E.(2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. https://doi.org/10.1145/3065386

Leibniz, G. W.(1951). Preface to the general science. En P. Wiener (Ed.), Lebiniz selections. Scribner. (Original work published 1677)

Lipton, Z. C.(2017). The mythos of model interpretability. arXiv:1606.03490 (cs, stat). http://arxiv.org/abs/1606.03490

Malik, M. M.(2020). A hierarchy of limitations in machine learning. arXiv:2002.05193 (cs, econ, math, stat). http://arxiv.org/abs/2002.05193

Mazzocchi, F. (2015). Could Big Data be the end of theory in science? EMBO reports, 16(10), 1250-1255. https://doi.org/10.15252/embr.201541001

McQuillan, D. (2018a, febrero 16). Manifesto on algorithmic humanitarianism. 'Reimagining Digital Humanitarianism, University of London. https://doi.org/10.31235/osf.io/ypd2s

McQuillan, D. (2018b). People's Councils for Ethical Machine Learning. Social Media + Society, 4(2), 2056305118768303. https://doi.org/10.1177/2056305118768303

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A survey on bias and fairness in machine learning. arXiv:1908.09635 (cs). http://arxiv.org/abs/1908.09635

Mezzadra, S., & Neilson, B. (2019). The Politics of Operations: Excavating Contemporary Capitalism. Duke University Press.

Mitchell, M. (2019). Artificial Intelligence: A Guide for Thinking Humans. Farrar, Straus and Giroux.

Mordvintsev, A., Olah, C., & Tyka, M. (2015, junio 17). Inceptionism: Going deeper into neural networks. Google Research Blog. http://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Moretti, F. (2013). Distant reading. Verso Books.

Morgulis, N., Kreines, A., Mendelowitz, S., & Weisglass, Y. (2019). Fooling a Real Car with Adversarial Traffic Signs. arXiv:1907.00374 (cs, eess). http://arxiv.org/abs/1907.00374

Murgia, M., & Harlow, M. (2019, abril 19). Who's using your face? The ugly truth about facial recognition. https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e

Offert, F. (2020, febrero 1). Neural Network Cultures panel. Transmediale festival y KIM HfG Karlsruhe, Berlin. http://kim.hfg-karlsruhe.de/events/neural-network-cultures?events=neural-network-cultures&post_type=events&name=n

O'Neil, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Reprint edition). Broadway Books.

Pasquinelli, M. (en prensa). The eye of the master: Capital as computation and cognition. Verso Books.

Pasquinelli, M. (2017). Arcana Mathematica Imperii: The evolution of western computational norms. En M. Hlavajova & S. Sheikh (Eds.), Former West: Art and the Contemporary after 1989. The MIT Press.

Pasquinelli, M. (2019a). On the origins of Marx's general intellect. Radical Philosophy, 2(6). https://www.radicalphilosophy.com/article/on-the-origins-of-marxs-general-intellect

Pasquinelli, M. (2019b). Three Thousand Years of Algorithmic Rituals: The Emergence of AI from the Computation of Space. E-Flux, 101. https://www.e-flux.com/journal/101/273221/three-thousand-years-of-algorithmic-rituals-the-emergence-of-ai-from-the-co

Pearl, J., & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect (1 edition). Basic Books.

Pieters, R., & Winiger, S. (2018, enero 23). Creative AI: On the democratisation & escalation of creativity. Medium. https://medium.com/@creativeai/creativeai-9d4b2346faf3

Pitts, W., & McCulloch, W. S.(1947). How we know universals the perception of auditory and visual forms. The Bulletin of mathematical biophysics, 9(3), 127–147.

Ponce, A. (2020). Labour in the Age of AI: Why Regulation is Needed to Protect Workers. European Trade Union Institute (ETUI) Research Paper Series, Foresight Brief(8). https://doi.org/10.2139/ssrn.3541002

Pontin, J. (2007, marzo 25). Artificial Intelligence, With Help From the Humans. The New York Times. https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html

Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. Television & New Media, 20(4), 350-365. https://doi.org/10.1177/1527476419831640

Rosenblatt, F. (1957). The perceptron, a perceiving and recognizing automaton (N.o 85-460-1). Cornell Aeronautical Laboratory Report.

Samadi, S., Tantipongpipat, U., Morgenstern, J. H., Singh, M., & Vempala, S. (2018). The Price of Fair PCA: One Extra dimension. En S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 31 (pp. 10976–10987). Curran Associates, Inc. http://papers.nips.cc/paper/8294-the-price-of-fair-pca-one-extra-dimension.pdf

Schaffer, S. (1994). Babbage's Intelligence: Calculating Engines and the Factory System. Critical Inquiry, 21(1), 203-227. JSTOR.

Schmidt, F. A.(2019). Crowdsourced production of AI Training Data: How human workers teach self-driving cars how to see. En Working Paper Forschungsförderung (N.o 155; Working Paper Forschungsförderung). Hans-Böckler-Stiftung, Düsseldorf. https://ideas.repec.org/p/zbw/hbsfof/155.html

Standange, T. (2016, junio 23). The return of the machinery question. The Economist. https://www.economist.com/special-report/2016/06/23/the-return-of-the-machinery-question

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv:1906.02243 (cs). http://arxiv.org/abs/1906.02243

Suresh, H., & Guttag, J. V.(2020). A framework for understanding unintended consequences of machine learning. arXiv:1901.10002 (cs, stat). http://arxiv.org/abs/1901.10002

Winner, L. (1978). Autonomous technology: Technics-out-of-control as a theme in political thought. MIT Press.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power (1 edition). PublicAffairs.

Notas

1

Sobre la autonomía de la tecnología véase: (Winner, 1978).

2

Sobre la extensión del poder colonial en las operaciones de la logística, los algoritmos y las finanzas véase: (Mezzadra & Neilson, 2019). Sobre el colonialismo epistémico véase: (Pasquinelli, 2019b).

3

Las humanidades digitales denominan a una técnica similar como *lectura distante*, que ha involucrado gradualmente el uso de análisis de datos y aprendizaje maquínico en historia de la literatura y del arte. Véase: (Moretti, 2013).

4

Para una historia concisa de la IA véase: (Cardon et al., 2018).

5

El uso de la analogía visual también apunta a registrar la debilitada distinción entre imagen y lógica, representación e inferencia, en la composición técnica de la IA. Los modelos estadísticos de aprendizaje maquínico son representaciones operativas (en el sentido de las imágenes operativas de Harun Farocki).

6

Para un estudio sistemático de las limitaciones lógicas del aprendizaje maquínico véase: (Malik, 2020).

7

Para una lista más detallada de sesgos de la IA véase: (Suresh & Guttag, 2020). Véase también: (Mehrabi et al., 2019).

8

Los informáticos argumentan que la IA pertenece al subcampo del procesamiento de señales, es decir de la compresión de datos.

9

deep learning

10

Los proyectos tales como la inteligencia artificial explicable, el aprendizaje profundo interpretable y los mapas de calor, entre otros, han demostrado que abrir la "caja negra" del aprendizaje maquínico es posible. Sin embargo, la interpretabilidad y la explicabilidad completas de los modelos estadísticos del aprendizaje maquínico siguen siendo un mito. Véase: (Lipton, 2017).

11

En el aprendizaje supervisado. También el aprendizaje auto-supervisado mantiene formas de intervención humana.

12

Sobre la taxonomía como forma de conocimiento y poder véase: (Foucault, 1966/2010).

13

Tales como el Turco Mecánico de Amazon, cínicamente denominado por Jeff Bezos como "inteligencia artificial artificial". Véase: (Pontin, 2007).

14

Aunque la arquitectura convolucional viene del trabajo de Yann LeCunn a fines de la década de 1980, el aprendizaje profundo comienza con este artículo: (Krizhevsky et al., 2017).

15

Para un recuento accesible (aunque no muy crítico) del desarrollo de ImageNet véase: (Mitchell, 2019).

16

WordNet es "una base de datos léxica de relaciones semánticas entre palabras" que fue comenzada por George Armitage en la Universidad de Princeton en 1985. Ofrece una estructura estrictamente arbórea de definiciones.

17

La regulación de la privacidad de los datos GDPR que aprobó el Parlamento Europeo en mayo de 2018 es, con todo, una mejora, comparada con la falta de regulación en los Estados Unidos.

18

Los parámetros de un modelo que se aprenden de los datos se llaman "parámetros", mientras que los parámetros que no se aprenden de los datos, sino que se fijan manualmente, se llaman "hiperparámetros", que determinan la cantidad y las propiedades de los parámetros.

19

Este valor también puede ser un valor porcentual entre 1 y 0.

20

https://keras.io/applications (Documentación para modelos individuales).

21

Véase el Community Earth System Model (CESM) [Modelo de Sistema Terrestre Comunitario

22

coupled numerical simulation

23

Las escuelas de antropología y etnología post-colonial y post-estructuralista han enfatizado que nunca hay territorio *per se*, sino, siempre, un acto de territorialización.

24

El reconocimiento de patrones es uno entre muchas otras economías de la atención. "Mirar es trabajar", como nos recuerda Jonathan Beller (2006, p. 2).

25

Como lo demuestra el teorema de aproximación universal.

26

embedded

27

rendered

28

primer

29

Para la idea de creación asistida y generativa véase: Roelof Pieters y Samim Winiger (2018)

30

Los *deep fakes* son videos sintéticos similares a los de los medios en los que los rasgos faciales de una persona son reemplazados por los de alguien más, a menudo con el propósito de producir noticias falsas.

31

Sobre normas computacionales, véase: (Pasquinelli, 2017).

32

time chart

33

Para una crítica, véase: (Mazzocchi, 2015).

34

Experimentos realizados por el Departamento de Policía de Nueva York desde finales de la década de 1980. Véase: (Pasquinelli, 2017)

35

La relación entre la IA y hackear no es tan antagónica como puede parecer: a menudo gira en un bucle de aprendizaje, evaluación y refuerzo mutuos.

36

computer vision

37

La contaminación de datos también puede emplearse para proteger la privacidad mediante el ingreso de información anonimizada o aleatoria al conjunto de datos.

38

disjoint

39

Para la idea de inteligencia analítica, véase: (Daston, 2018)

Como citar: Pasquinelli, M., Joler, V. (2021). El Nooscopio de manifiesto, *laFuga*, 25. [Fecha de consulta: 2026-11-05] Disponible en: http://2016.lafuga.cl/el-nooscopio-de-manifiesto/1053