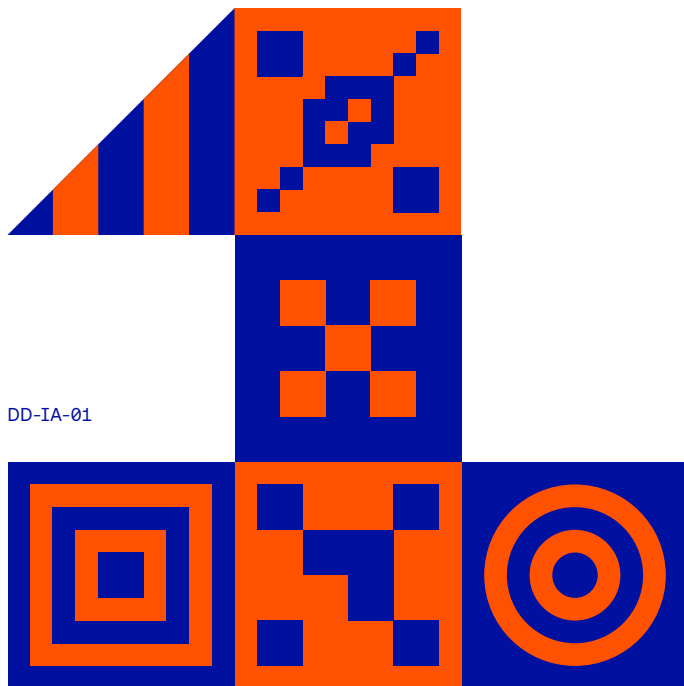
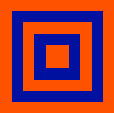


Algunos conceptos básicos





Inteligencia Artificial

Sistema para el análisis y toma de decisiones a través de operaciones computarizadas y automatizadas, basadas en un **conjunto de datos**. El término Inteligencia Artificial se usa para distintas tecnologías, con la característica común de simular la inteligencia humana, aplicada a problemas para ofrecer soluciones basadas en la predicción o el reconocimiento de patrones.

La IA se considera “débil” o “estrecha” cuando ejecuta una tarea específica, y “fuerte” o “general” si contextualiza diferentes problemas específicos y ejecuta tareas independientes a partir de dicha contextualización, como una inteligencia humana.

Algunos sistemas de IA, con distintos niveles de complejidad y autonomía, están presentes en aplicaciones como asistentes virtuales y sistemas de reconocimiento de voz, o en dispositivos especializados como los drones o vehículos autónomos.

Algoritmo

Fórmula o conjunto de reglas para articular instrucciones, procedimientos o procesos para resolver un problema o realizar una tarea. En el ámbito de la **IA**, los algoritmos se incorporan en los códigos de programación de un sistema informático para operar de forma automatizada, y así orientan a las máquinas sobre cómo buscar respuestas a una pregunta o soluciones para un problema.

Los problemas asociados a la toma de decisión suelen ser atribuidos a algoritmos o fórmulas mal diseñadas, que



toman en consideración o ponderan datos con resultados equivocados, discriminatorios o dañinos.

Datos

Son unidades de información, representaciones alfanuméricas de hechos singulares y de estadísticas que pueden ser analizadas por diversos medios. Constituyen el ingrediente esencial para el **aprendizaje automático** y la aplicación de la **IA**; por esta razón, suele pensarse que con mayor cantidad de datos, pueden entrenarse mejor los sistemas automatizados para que tomen mejores decisiones.

Dato personal

Unidad de información que por sí sola o en relación con otra información, permite identificar a una persona física. Son datos personales el nombre, el número de identificación legal, el número de teléfono, la edad y la dirección domiciliaria o laboral, entre otras. La anonimización de los datos personales resulta, muchas veces, un proceso necesario para su adecuado manejo.

Datos abiertos - Open Data

Práctica de mantener un conjunto o base de datos disponible para que cualquier persona pueda utilizarla, reutilizarla y compartirla libremente, sin restricción de confidencialidad, derechos de autor o patentes. Los datos se consideran utilizables cuando se encuentren en un formato común, que pueda ser leído y procesado por máquinas.



Conjunto o base de datos - Dataset / Database

Grupos de elementos que guardan una relación entre sí y que almacenan **datos** como números, fechas o palabras, que pueden ser procesados para producir información.

La calidad de la base de datos será directamente proporcional a las capacidades disponibles para construirla, esto es, para recabar los datos de forma inclusiva y rigurosa, por lo que cualquier incompletitud (por ejemplo, respecto de personas que habitan zonas extremas), inexactitud, desactualización o sesgo en la recolección, serán reproducidas en el conjunto de datos.

Macrodatos - Big Data

Manejo de bases de datos que son demasiado grandes o complejas para los métodos tradicionales de procesamiento o almacenamiento, por lo que requieren una aproximación diferente para su gestión. Los datos con muchos casos (filas) ofrecen mayor poder estadístico, mientras que los datos con mayor complejidad (más atributos o columnas) pueden llevar a una mayor tasa de errores, si no son manejados con las herramientas adecuadas. Para abordar posibles riesgos, se considera relevante abordar dimensiones como: el volumen de los datos, su velocidad de generación, la variedad de su origen, la importancia de la veracidad de los datos y la evaluación de su valor. Este es el modelo conocido como las 5V del Big Data.



Minería de datos - Data mining

Proceso de recopilación o extracción de datos desde distintas fuentes, usualmente a través de sistemas automatizados, con el propósito de construir bases de datos o de identificar patrones en la información. Aunque es una forma de procesar información, puede también servir para hacer inferencias o identificar personas contra su voluntad.

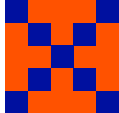
Ciencia de datos - Data Science

Campo científico que integra disciplinas como las matemáticas, la estadística, el estudio de probabilidades, la computación y la visualización de datos, para extraer conocimientos desde un conjunto heterogéneo de datos (imágenes, sonidos, textos, mediciones físicas, datos genómicos, vínculos de redes sociales, entre otros). Los métodos y herramientas derivadas de la aplicación de la **IA** son parte de este campo.

Código abierto

Es un modelo de desarrollo que consiste en garantizar que las personas puedan acceder al código fuente de un programa o aplicación de forma gratuita, facilitando su auditabilidad, su integración a otros sistemas, su reutilización y adaptación en contextos diferentes al original.

Se destaca en oposición al código cerrado, que por justificaciones de propiedad intelectual o secretos industriales, se mantiene bajo reserva o sin posibilidad de modificación por personas no autorizadas expresamente, limitando su auditabilidad y escrutinio.



Computación en la nube - Cloud Computing

Conjunto de tecnologías que permiten el acceso remoto a través de internet a recursos informáticos tales como programas y aplicaciones, almacenamiento de archivos y procesamiento de datos. De este modo, los recursos están alojados en grandes centros de procesamiento de datos, por lo que no se requieren servidores locales y no es necesario instalar programas en los computadores finales, permitiendo la provisión de servicios a distancia de “software como servicio”. Es utilizada también en el desarrollo de Internet de las Cosas (IoT), sistemas de **aprendizaje automático** y **big data**.

Interfaz de programación - Application Programming Interface (API)


Mecanismo que permite programar componentes adicionales en un sistema informático, o hacerlo interactuar con otros programas, a partir de un conjunto de definiciones operacionales y protocolos de comunicación.

Bot / Chatbot

Programa computacional de distinto nivel de complejidad capaz de interactuar con personas a través de una conversación mediante audio o texto, usado normalmente para facilitar un “diálogo” entre una persona y una institución o sistema.

Aprendizaje automático - Machine Learning

Rama de la Inteligencia Artificial dedicada al desarrollo de técnicas y **algoritmos** que, mediante su aplicación reiterada y



la acumulación de información, permiten a un sistema mejorar su desempeño (“aprender”) en la resolución de una pregunta o problema propuesto a partir de inferencias y de forma automatizada.

Aprendizaje no supervisado

Una forma de **aprendizaje automático** en que el sistema no recibe información sobre cómo deben ser los **datos** resultantes, solo se integran en un **conjunto de datos** no estructurados. El **algoritmo** deberá identificar posibles patrones de **datos**, las relaciones entre ellos, y proponer como resultado una estructura, sin requerir validación humana.

Aprendizaje supervisado

Una forma de **aprendizaje automático** en que las personas proporcionan los **datos** de entrada, salida, y cómo deben ser utilizados, para que la máquina entienda cómo relacionarlos y proponga mejoras que serán luego validadas por otra persona.

Aprendizaje profundo - Deep learning

Rama del **aprendizaje automático** que busca reproducir el funcionamiento del sistema nervioso humano. Consiste en el análisis de grandes **conjuntos de datos** no estructurados, para encontrar patrones sin la necesidad de un entrenamiento previo. Para eso, funciona con capas de procesamiento que se conectan, por ejemplo, como **redes neuronales**: el resultado de la primera capa alimenta la segunda capa y así sucesivamente, produciendo resultados cada vez más abstractos.



Estos algoritmos tienen numerosas aplicaciones; como el reconocimiento automático de sonido, el desarrollo de **sistemas de visión computarizada**, el **procesamiento de lenguaje natural**, diagnósticos médicos y vehículos autodirigidos, entre otras. Es importante que se consideren todos los riesgos de los ámbitos en que se despliegan esos algoritmos.

Análisis predictivo

Método basado en la observación y estudio de los **datos** disponibles hasta el momento, con el objetivo de predecir la probabilidad de ocurrencia de eventos en el futuro. Se ha implementado en políticas públicas que buscan prevenir la vulneración de derechos o asignar beneficios sociales. Sin embargo, sus resultados son discutibles por los posibles **sesgos** en los datos disponibles, y por la opacidad en el procesamiento de dichos datos, por lo que deben estar sujetos a un amplio escrutinio.

Modelo

Una representación abstracta de lo que un sistema de **aprendizaje automático** ha aprendido en el entrenamiento a partir de un **conjunto de datos**. Por ejemplo, un modelo de análisis de datos climáticos puede ser programado para establecer probabilidades que informen el pronóstico diario.

Árbol de decisión - Decision tree

Un tipo de modelamiento predictivo para realizar análisis decisional donde, a través de **aprendizaje supervisado** los



datos son divididos de manera iterativa, a partir de parámetros o criterios predefinidos, para enseñar a las máquinas a tomar decisiones y resolver problemas de categorización o regresión, para obtener un modelo definitivo.

Regresión logística - Logistics regression

Es un proceso estadístico implementado en el **aprendizaje automático** para predecir el resultado de una variable dependiente con base en el análisis de datos anteriores. A través de un proceso de **aprendizaje supervisado**, se utiliza para la clasificación binaria, es decir un resultado o su contrario. Una aplicación práctica es la detección de spam en un servicio de correo electrónico.

Redes neuronales - Neural networks

Base de funcionamiento del **aprendizaje profundo** que consiste en imitar las funciones neuronales propias del cerebro humano. La red neuronal funciona con capas de procesamiento, donde cada neurona es una ecuación matemática simple que se conecta con otra, y esto se repite en toda la capa. Así, en cada capa se elimina información innecesaria y se conserva una representación cada vez más simple y precisa de los datos. Este tipo de algoritmos requieren grandes **conjuntos de datos** y permiten **modelar** múltiples resultados de manera simultánea. Pueden ser utilizadas en tareas como la detección automatizada de rostros humanos (“reconocimiento facial”) o la lectura de caracteres.

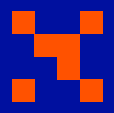


Procesamiento de lenguaje natural - Natural Language Processing (NLP)

Rama de la **inteligencia artificial** aplicada a los sistemas con que interactúan las personas, como los asistentes operados mediante comandos de voz y con respuesta verbal, y está centrada en el desarrollo y optimización de programas que procesan el lenguaje humano con el propósito de deducir su contenido. Dentro del NLP se incluyen el entendimiento de lenguaje (Natural Language Understanding) que consiste transformar el lenguaje humano a un formato legible para las máquinas, y la generación de lenguaje (Natural Language Generation) que consiste en la producción de narrativas habladas o escritas a partir de un **conjunto de datos**.

Sistema de visión computarizada - Computer Vision System

Un área científica interdisciplinaria dedicada a la exploración, reconocimiento y análisis de imágenes y video. Entre las tareas que aborda está el reconocimiento de objetos, la detección de eventos y movimiento, y el seguimiento de objetos, entre otras. Las técnicas de visión computarizada son utilizadas en aplicaciones como la detección de objetos en vehículos autónomos y el reconocimiento facial automatizado, con lo que un mal desarrollo puede tener consecuencias catastróficas, pero un gran perfeccionamiento puede facilitar labores de vigilancia masiva o discriminatoria.

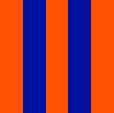


Data brokers

Actores y empresas dedicadas a comercializar datos personales, obtenidos a partir del uso de servicios en internet o desde otras fuentes (incluidos otros data brokers). Están también orientados a la creación de perfiles (por ejemplo, a efectos de orientar campañas de marketing), para lo que recopilan y analizan información de usuarias y usuarios obtenidas de fuentes públicas y privadas. Aunque concentran una gran cantidad de información personal, suelen operar sin ser detectados por las usuarias de tecnologías, que permanecen sin control sobre la calidad de esos datos o sobre cuántas personas o empresas los tienen.

Sesgo algorítmico - Algorithmic Bias

Es un valor de desfase desde el origen de una fórmula, un defecto que lleva a un resultado incorrecto o injusto. En los sistemas de **aprendizaje automático**, dados los efectos sociales que hoy puede tener el sesgo, debe entenderse desde la **ética de la inteligencia artificial**, al menos con dos acepciones: el sesgo inductivo, es decir, los estereotipos o prejuicios de quien está diseñando un sistema de recolección o procesamiento de datos y que se incorporan en el funcionamiento del sistema; y el sesgo de confirmación, es decir un error sistemático introducido por un procedimiento de muestreo o de información disponible, es decir, los sesgos que puedan existir en la conformación de la base de datos. Para reducir estos sesgos es posible impedir, mediante un sistema de **aprendizaje supervisado**, que los resultados sesgados de

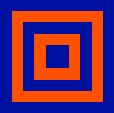


un **algoritmo** se generalicen; sin embargo es necesario auditar regularmente un sistema para prevenir que nuevos sesgos se manifiesten.

Discriminación algorítmica

Es el resultado inapropiado, incorrecto, irregular o injusto que se produce tras la operación de un sistema automatizado de predicción, clasificación o toma de decisión, y que tiende a perjudicar a ciertas personas o grupos. Los **algoritmos** suelen tener **sesgos**, pues tanto en la definición de variables como en el desarrollo del código se imprimen valores de las personas encargadas de su diseño. También puede haber **sesgos** en el **conjunto de datos** con que se entrenan los algoritmos, dependiendo de los criterios con que se han recolectado y procesado dichos **datos**. Esto es especialmente sensible en los sistemas de **aprendizaje automático** para la toma de decisiones y **modelos predictivos**, incluyendo el **procesamiento de lenguaje natural** y **sistemas de visión computarizada**, cuando son implementados en gestión pública o asistencia social, selección y contratación de personal, asignación de créditos o beneficios sociales y aplicación de la ley, entre otras tareas.

Las predicciones algorítmicas pueden reforzar la discriminación hacia grupos tradicionalmente vulnerados, con base en estatus socioeconómico, ingresos, raza, etnia o género, entre otras. Estos procesos se han caracterizado, hasta ahora, por su opacidad. Por eso, recientemente se ha buscado establecer mecanismos como la **IA explicable**, para



la verificación, transparencia y regulación de sistemas de IA, con el fin de reducir los efectos desproporcionados o injustos de estos sistemas.

Transparencia algorítmica

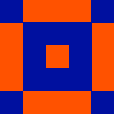
Principio orientado a que los factores que determinan las decisiones tomadas por algoritmos operen de forma transparente y comprensible para quienes utilizan, regulan y son afectadas y afectados por el procesamiento de datos realizado. Se trata de un principio esencial para detectar problemas de funcionamiento y para facilitar la rendición de cuentas en la operación de sistemas automatizados.

Protección de datos personales

Normativa que regula los derechos de las personas sobre la información que le concierne, en particular, respecto del levantamiento, manejo y procesamiento de esos datos para conocer los detalles de ese uso, corregir errores o solicitar su eliminación. Su especificidad y alcance varía según el marco legal de cada país. Su eficacia es un elemento crucial para el control sobre un elemento fundamental de los sistemas de decisión automatizada: la información relativa a personas.

Legalidad

Principio fundamental del derecho público, que orienta el actuar de la función pública estrictamente bajo el marco de las normas jurídicas determinadas por órganos competentes. Implica que todas las decisiones públicas deben respetar



reglas formales de competencia, como también reglas sustantivas sobre derechos fundamentales y legales.

Necesidad y proporcionalidad

El principio de necesidad es considerado un principio de la función pública que exige que la medida implementada sea requerida para el resguardo de un derecho o el cumplimiento de un objetivo legítimo, tras evaluar la posibilidad de otras decisiones o soluciones igualmente eficaces pero potencialmente dañinas.

Se suma al análisis de necesidad el de proporcionalidad. La proporcionalidad es un principio de la función pública que exige la ponderación equilibrada entre los medios adoptados para conseguir fines legítimos, de forma tal que el logro de esos fines no tenga consecuencias negativas de mayor entidad. Esto implica analizar la idoneidad de las medidas para la consecución de los fines.

En las decisiones públicas relacionadas con IA, la aplicación de estos principios implica una decisión informada sobre la idoneidad de un sistema automatizado para asistir al cumplimiento de funciones públicas, los intereses que se verán afectados por su despliegue, la cantidad de información estrictamente necesaria para su funcionamiento, y la existencia de alternativas, entre otros factores.



Ética de la Inteligencia Artificial - Ethics of Artificial Intelligence

Sistemas de principios que intentan influir en el desarrollo y despliegue de sistemas de IA, para prevenir sus riesgos. Así, para reducir los riesgos de **discriminación algorítmica**, diferentes empresas e instituciones alrededor del mundo cuentan con principios éticos para el diseño, entrenamiento y despliegue de sistemas de inteligencia artificial y de **aprendizaje automático**. Si bien existen decenas de iniciativas sobre ética y principios, en noviembre de 2021 la UNESCO publicó una serie de Recomendaciones Éticas, las cuales han sido adoptadas por 193 Estados miembro, y aunque no son vinculantes, hasta ahora son la normativa con más amplia aceptación a nivel mundial.

IA explicable - Explainable AI (XAI)

Es un conjunto de procesos y métodos que permiten a las personas usuarias comprender los resultados y productos creados por los **algoritmos de aprendizaje automático**. La XAI consiste en superar la idea de una **caja negra** donde se desarrollan los sistemas de **IA** y es crucial para generar confianza en estos sistemas; al describir un **modelo**, su impacto esperado, sus posibles **sesgos** y mecanismos de reparación ante potenciales abusos. Facilita la auditabilidad de los sistemas y también habilita a quienes los desarrollan para ajustar estos sistemas hacia modelos más justos.

Inteligencia Artificial y derechos humanos

Algunos conceptos básicos

Esta publicación fue realizada por Patricio Velasco y Juliana Guerra de Derechos Digitales, con el apoyo del International Development Research Centre (IDRC).



Desde 2019, Derechos Digitales es parte de la red de Cyber Policy Research Centres de IDRC, junto a organizaciones líderes en temas de tecnologías y políticas públicas en el Sur Global. Este reporte corresponde a la línea de trabajo “Inteligencia Artificial e Inclusión”, coordinado por Jamila Venturini, Juan Carlos Lara y Patricio Velasco.

Para más información sobre este proyecto, visita <https://ia.derechosdigitales.org/>

Texto por Patricio Velasco y Juliana Guerra.

Edición y correcciones por Juan Carlos Lara e Ileana Silva.

Diseño y diagramación por Comunas Unidas.



DD-IA-01



Esta obra está disponible bajo licencia Creative Commons Atribución 4.0 Internacional
<https://creativecommons.org/licenses/by/4.0/deed.es>