INTERNATIONAL JOURNAL OF DIGITAL LAW – IJDL
ano 01 · n. 01 · janeiro/abril 2020 - Publicação quadrimestral
DOI: 10.47975/digital.law.vol.1.n.1
01
ISSN 2675-7087

IJDL

International Journal of DIGITAL LAW

NUPED
Núcleo de Pesquisas em Políticas Públicas e Desenvolvimento Humano da PUCPR

FORUM

IJDL – INTERNATIONAL JOURNAL OF DIGITAL LAW

Luís Cláudio Rodrigues Ferreira
Presidente e Editor

Av. Afonso Pena, 2770 – 15º andar – Savassi – CEP 30130-012 – Belo Horizonte/MG – Brasil – Tel.: 0800 704 3737
www.editoraforum.com.br / E-mail: editoraforum@editoraforum.com.br

Impressa no Brasil / Printed in Brazil / Distribuída em todo o Território Nacional

IN61 International Journal of Digital Law – IJDL. – ano 1, n. 1 (abr. 2020) –.– Belo Horizonte: Fórum, 2020-.

Quadrimestral; Publicação eletrônica
ISSN: 2675-7087

1. Direito. 2. Direito Digital. 3.Teoria do Direito. I. Fórum.

CDD: 340.0285
CDU: 34.004

Coordenação editorial: Leonardo Eustáquio Siqueira Araújo
Aline Sobreira
Capa: Igor Jamur
Projeto gráfico: Walter Santos

DOI: 10.47975/IJDL/1corvalan

Inteligencia Artificial GPT-3, PretorIA y oráculos algorítmicos en el Derecho

GPT-3 Artificial Intelligence, PretorIA, and Algorithmic Oracles in Law

Juan Gustavo Corvalán[*] [**]
Universidad de Buenos Aires (Ciudad de Buenos Aires, Buenos Aires, Argentina)
corvalanjuang@gmail.com
https://orcid.org/0000-0001-9565-2818

Recebido/Received: 07.02.2020/ February 7th, 2020
Aprovado/Approved: 18.03.2020/ March 18th, 2020

Resumen: El artículo trata sobre el tema de la inteligencia artificial y el aprendizaje automático. Aborda el tema de la superinteligencia y el aprendizaje automático como género. Describe el tema del aprendizaje profundo y los oráculos artificiales. Relaciona la causalidad y la capacidad predictiva de la inteligencia artificial. Indica los primeros resultados de GPT-3, así como su impacto en la ley. Concluye el texto informando el caso PretorIA, con un enfoque en tres dimensiones. Al final, afirma que, a tarea de entrenamiento, a fin de cuentas, modula y condiciona el ejercicio de competencias humanas complementadas con oráculos algorítmicos y asistencia digital.

Palabras-clave: Inteligencia artificial. GPT-3. Algoritmos de Oracle. Aprendizaje profundo. *Big data*.

Abstract: The article deals with the topic of artificial intelligence and machine learning. Addresses the topic of superintelligence and machine learning as a genre. Describe the topic of deep learning and

Como citar este artigo/*How to cite this article*: CORVALÁN, Juan G. Inteligencia Artificial GPT-3, PretorIA y oráculos algorítmicos en el Derecho. *International Journal of Digital Law*, Belo Horizonte, ano 1, n. 1, p. 11-52, jan./abr. 2020.

[*] Director del Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la Universidad de Buenos Aires (Ciudad de Buenos Aires, Buenos Aires, Argentina). Doctor en Ciencias Jurídicas, Universidad del Salvador. Director del Posgrado en Inteligencia Artificial y Derecho de la UBA. Director de la Diplomatura en Derecho 4.0 de la Universidad Austral. Cocreador de Prometea, la primera Inteligencia Artificial predictiva al servicio de la Justicia. Co-creador de PretorIA, y Director académico de la implementación de ese sistema en la Corte Constitucional de Colombia. Director General de Proyecto en el marco del Convenio de Implementación de Prometea en la Corte Interamericana de Derechos Humanos. Actualmente se desempeña como Fiscal General Adjunto en lo Contencioso Administrativo y Tributario ante el Tribunal Superior de Justicia de la CABA.

[**] Mi agradecimiento especial a Gerardo Simari, uno de los referentes en materia de inteligencia artificial en Argentina. Muchas gracias por las sugerencias, comentarios y la revisión de los aspectos vinculados a la descripción de las diferentes técnicas de IA. Por otra parte, como siempre, a mi querida amiga María Elena Lumiento, por la revisión y sugerencias vinculadas a las cuestiones estrictamente vinculadas a la dogmática penal. Muchas gracias a Victoria Carro por la colaboración en las cuestiones vinculadas a causalidad.

artificial oracles. It relates the causality and the predictive capacity of artificial intelligence. It indicates the first results of GPT-3, as well as its impact on the law. The text concludes by informing the PretorIA case, with a three-dimensional approach. In the end, he affirms that the training task modulates and conditions the exercise of human skills complemented with algorithmic oracles and digital assistance.

Keywords: Artificial intelligence. GPT-3. Oracle algorithms. Deep learning. Big data.

1 Introducción

La frase que hoy se ha popularizado como "humanidad aumentada",[1] en realidad es un proceso histórico que se viene desarrollando a partir de los avances tecnológicos a lo largo de los siglos como la rueda, el papel, la imprenta, el vapor y la electricidad. Lo novedoso que nos trae esta cuarta revolución industrial, es que se masificaron tecnologías que reemplazan, complementan y/o mejoran lo que antes solo podíamos lograr con nuestra capacidad intelectual.

Desde hace varios años trabajamos desde una doble perspectiva. Por un lado, desarrollar sistemas predictivos (oráculos artificiales)[2] en el ámbito del Sector Público y la Justicia, que sean compatibles con los principios de un Estado constitucional y con los derechos humanos.[3] Por otro, entender y describir el fenómeno de la inteligencia artificial (en adelante IA), bajo la lógica de que se trata de un conjunto de tecnologías de propósito general.[4] Esto significa que tenemos por delante a la innovación más disruptiva de toda la historia humana. La IA, como la revolución de las revoluciones, cambiará profundamente a nuestra especie y a todos los sectores de las sociedades del siglo XXI. El corazón de la IA es el llamado *machine learning* o aprendizaje de máquina que presenta muchos matices de intervención humana,

1 Véase: SADIN, Eric. *La humanidad aumentada*...

2 En adelante, nos vamos a referir a los sistemas predictivos basados en inteligencia artificial, bajo el concepto de oráculos artificiales.

3 Estamos convencidos de que el uso de las tecnologías emergentes del paradigma 4.0, pueden generar un salto trascendental en el modo y en el tiempo en que se ejecutan las tareas de en el campo jurídico.

4 Se considera que la IA tiene un doble papel de tecnología de propósito general y herramienta para la innovación, la IA ha logrado protagonismos en los debates en múltiples esferas bajo la promesa de cambiar la forma en que vivimos y nuestra percepción del mundo, ver: GÓMEZ MONT. *Constanza et al. La inteligencia*...

otros sistemas que autoaprenden, automejoran y, en menor medida, otros que son creados por otro sistema de IA.[5] Pero antes de avanzar en todo ellos, es importante considerar tres grandes cuestiones asociadas a todos estos sistemas inteligentes.[6]

Por un lado, las técnicas de IA se basan en detectar y reconocer patrones de información en los datos.[7] Esto se logra a partir de combinar ordenadores, internet, algoritmos y lenguajes de programación para resolver problemas o tomar decisiones que antes solo podían ser realizadas por nuestras capacidades cognitivas.[8] Por otra parte, en todos los sistemas de IA también se pueden presentar aprendizajes supervisados y no supervisados. Incluso, ambos se pueden combinar. Por último, los sistemas inteligentes se basan en algoritmos para funcionar.[9] Y este concepto se asocia a un *conjunto de instrucciones, reglas o una serie metódica de pasos que puede utilizarse para hacer cálculos, resolver problemas y tomar decisiones.*[10] Los algoritmos son a la informática, lo que los códigos procesales, procedimentales y los protocolos son al campo jurídico.[11] En la informática, los algoritmos se usan para "escribir código" en el lenguaje informático, y de esa forma se obtienen resultados en un lenguaje binario de 0s y 1s. En síntesis, los algoritmos son la base de sistemas de IA que ejecutan instrucciones a partir de técnicas de aprendizaje automático. Sobre ellas transforman datos en patrones de información y luego en conocimiento que permite automatizar tareas, elaborar predicciones o previsiones y realizar detecciones inteligentes.[12]

Como afirma Naciones Unidas (ONU), la manera de entendernos y nuestra relación con el mundo tiene lugar desde la perspectiva de los algoritmos. Son una parte fundamental de las sociedades de la información, ya que cada vez más gobiernan las operaciones, decisiones y elecciones que antes quedaban en exclusivas manos de los seres humanos.[13] Ahora bien, aunque por ahora resulta imposible reproducir en máquinas a un órgano tan complejo como el cerebro, hay que considerar que los ingenieros en aviación no copiaron las técnicas de aprendizaje de

---

5 El "paisaje técnico de la inteligencia artificial (IA)", ha evolucionado significativamente desde 1950 cuando Alan Turing se planteó la pregunta de si las máquinas pueden pensar. Des 2011, se han producido avances en el subconjunto de IA llamado "aprendizaje automático", en el que las máquinas aprovechan los enfoques estadísticos para aprender de datos históricos y realizar predicciones en situaciones nuevas. La madurez de las técnicas de aprendizaje automático, junto con los grandes conjuntos de datos y el aumento del poder computacional están detrás de la expansión actual de la IA. Ver OCDE. *Inteligencia artificial*...

6 Véase en: ONU. *La Resolución Nº 73/348*...

7 Ampliar en DOMINGOS, Pedro. *The master*...

8 "En la base de la inteligencia artificial están los algoritmos, códigos informáticos diseñados y escritos por seres humanos que ejecutan instrucciones para traducir datos en conclusiones, información o productos". ONU. La resolución N° 73/348...

9 CORVALÁN, Juan Gustavo. *Administración*..., CORVALÁN, Juan Gustavo. *Digital and Intelligent*...,

10 BENÍTEZ, Raúl et al. *Inteligencia*..., p. 14. Ver, además, ONU. *La Resolución N° 72/540*...

11 CORVALÁN, Juan Gustavo. *Desafíos*..., 9 set. 2019; CORVALÁN, Juan G., GALETTA, Diana U. *Intelligenza*...

12 Véase ONU. *Resolución N° 73/348*..., STRINGHINI, Antonella. *Administración*..., p. 199-215.

13 Véase ONU. *Resolución N° 72/540*...

los pájaros para construir los aviones modernos,[14] ni que los submarinos o barcos "nadan" tal cual lo hacen las personas. Con el avance de la IA sucede un fenómeno similar al que se presenta en las discusiones acerca de nuestra inteligencia. En general, tendemos a considerar "no inteligentes" tareas que se vuelven sencillas y habituales con el paso del tiempo y este fenómeno se presenta, aun en mayor medida, cuando se naturalizan los resultados generados por máquinas inteligentes.

2 IA débil, blanda, restringida o estrecha

Aunque es difícil ponerse de acuerdo en un concepto omnicomprensivo, hay un elemento común en muchas definiciones de la inteligencia humana: la *capacidad de procesar información para resolver problemas en función de alcanzar objetivos.*[15] Y en esta capacidad de procesamiento se ubica un factor crucial: el reconocimiento de patrones. En nuestro cerebro se presentan dos grandes procesos simultáneos. Por un lado, lo que se conoce bajo el nombre de etiquetas emocionales. En estas etiquetas se apoya el cerebro para seleccionar la información más relevante para la toma de decisiones. Son marcas que imprime en los pensamientos y experiencias almacenadas en la memoria, que contienen información afectiva en cada recuerdo. Por ejemplo, peligroso, agradable o molesto. Cuando nos encontramos con una situación o estímulo etiquetado, entonces poseemos información útil para decidir rápidamente qué acción debemos tomar.

Por otra parte, el reconocimiento de patrones y el pensamiento jerárquico o el llamado modelo jerárquico de la estructura de la inteligencia biológica.[16] Esta forma de razonar, pensar y clasificar los objetos se vincula con una estructura compuesta de diferentes elementos ordenados según un patrón. Ambos procesos, están asociados a la definición de inteligencia humana, entendida como la *capacidad de procesar información para resolver problemas en función de alcanzar objetivos (Ray Kurzweil)*. Todo esto, está relacionado con la flexibilidad, velocidad y precisión para adaptarnos a los entornos. La IA se basa en obtener, por otros métodos artificiales, lo que alcanzamos con la inteligencia humana: el reconocimiento de patrones para alcanzar objetivos o resolver problemas. Ésta es una concepción amplísima y macro de lo que hace la IA. Sin embargo, como sucede con algunas categorías del derecho público como la discrecionalidad, podemos hablar de dos sentidos:[17] débil y fuerte.

---

[14] KURZWEIL, Ray. La *singularidad*..., p. 161.

[15] Ampliar en GARDNER, Howard. *La inteligencia*..., p. 52; ROECKELEIN, Jon E. *Dictionary*...; MANES FACUNDO-NIRO, Mateo. *Usar el*..., p. 114-115; SIEGEL, Daniel J., *Viaje*...; MARINA, José Antonio. *El cerebro*..., p. 37-42.

[16] Sobre todas estas cuestiones, ampliar en: MANES, Facundo; NIRO, Mate. *El cerebro*..., p. 269-270. SIGMAN, Mariano. *La vida*..., p. 133-134.

[17] La clasificación más habitual que se realiza ha sido introducida por SEARLE, J. R. *Minds*..., p. 417-457.

En la comunidad internacional, se llama IA "débil", "restringida", "estrecha" o "blanda" al procesamiento de datos e información para resolver problemas a partir de utilizar algoritmos inteligentes sobre la base de aplicar diferentes técnicas informáticas. La idea básica es obtener resultados específicos en ciertas actividades o ámbitos concretos que antes solo podían obtenerse a partir de nuestros cerebros.[18] Mientras que los humanos transitamos un camino biológico de aprendizaje evolutivo, la IA se basa en algoritmos, datos históricos, computadoras, programación humana y, sobre todo, aprovechando tres características que superan por mucho nuestras capacidades cognitivas: velocidad de procesamiento, posibilidad de conectarse y articular con otros sistemas de forma instantánea y, por último, la capacidad casi infinita de almacenamiento de los datos e información.

Este concepto de IA débil o restringida es el que sustenta el género aprendizaje de máquina o *machine learning* que abarca una serie de técnicas más o menos sofisticadas. La especie más conocida es el aprendizaje profundo (*deep learning*). Algunos autores se refieren a esta clase de técnica, basada en un tipo de redes neuronales artificiales – RNA.[19] El empleo de estas técnicas requiere de grandes cantidades de datos para ser "entrenada" y, por su modo de funcionar, se asemeja a una "caja negra" (*black box*). Esto quiere decir que, por ahora, no es posible determinar –al menos en parte– el paso a paso de la lógica de procesamiento de datos que sucede en el interior del sistema cuando se trata de redes neuronales. En otras palabras, no se puede conocer en un 100% lo que sucede en las "capas ocultas de la red".[20]

3 IA fuerte, dura o general y la llamada "superinteligencia"

La IA fuerte representaría la transformación más importante de este siglo.[21] Representa la fase final de transición de la IA débil, que son todos los sistemas de IA que desarrollamos en este artículo y que son catalogados como IA débil o blanda. Este tipo de IA se vincula con dos grandes fenómenos. En primer lugar, con el hecho de alcanzar algunos aspectos claves de la especie humana: el sentido común, la

[18] En las organizaciones públicas, la IA permite llevar adelante la transición de una burocracia imprenta o digital, hacia una burocracia inteligente. Ampliar en CORVALÁN, Juan Gustavo. *Prometea*..., p. 29; CORVALÁN, Juan Gustavo. *Hacia*...

[19] Una red neuronal artificial puede ser comprendida como una combinación masiva de unidades de procesamiento simple, que aprenden del entorno a través de un proceso de aprendizaje y almacenan el conocimiento en sus conexiones. Véase HAYKIN, Simon. *Neural*...

Véase también: UNESCO. *El Correo de*..., p. 8; ONU. *La Resolución N° 72/540*.

[20] Téngase en cuenta que la referencia a capas es específica de redes neuronales.

"La IA moderna es, básicamente, una caja negra, que logra un desempeño superior al humano sin que las personas comprendan cabalmente cómo se obtiene ese resultado" (Comisión Económica para América Latina y el Caribe, CEPAL. *Datos*..., p. 171.

[21] KURZWEIL, Ray. *La singularidad*..., p. 339.

capacidad de "sentir", de reconocer el entorno y la llamada "autoconsciencia". En segundo lugar, así como se alude a una inteligencia general humana que es producto de abarcar diferentes áreas de contenido, se busca desarrollar una Inteligencia Artificial General (IAG), que se traduce en una capacidad general de aprender. Es decir, se trata de simular el comportamiento o la inteligencia humana en un plano integral.[22]

Como estos sistemas todavía no se han desarrollado, es importante considerar que es meramente conjetural la diferencia entre IA fuerte e IA débil, y está fuertemente correlacionada con los alcances que se le asigna a la inteligencia humana y a dónde se desee poner el acento. Por ejemplo, cuando las máquinas simulan o actúan como si fueran inteligentes en ciertos ámbitos o tareas concretas, se conoce como *débil, blanda o estrecha*. En cambio, cuando se afirma que las máquinas "realmente" demuestran inteligencia y no solo la simulan, entonces estaríamos en presencia de la *IA fuerte.*[23]

Ahora bien, por un lado, es importante tener presente que todavía no se han desarrollado sistemas de IA que posean sentido común y la habilidad de manejar diversos campos de conocimiento a la vez,[24] aunque GPT-3 es un "primer borrador avanzado" de una IA que aspira a lograr eso. Por otra parte, como hemos sostenido en otros trabajos, en vez de poner el foco en disquisiciones conceptuales acerca de lo que es o no es IA, es urgente abordar los beneficios, riesgos, desafíos, daños y, esencialmente, cómo diseñar ecosistemas de regulación que permitan que su desarrollo sea compatible con los derechos humanos.[25] Por eso es tan relevante separar cajas negras de cajas blancas y, a su vez, comprender cómo se puede

---

22 En algunas ocasiones también se habla de IA "general", pero los términos no son exactamente iguales. Toda IA fuerte será general, pero a la inversa no tiene por qué siempre darse. En la comunidad científica, hay un debate intenso entre especialistas acerca de si esta clase de IA llegará y, eventualmente, cuándo hará su aparición. Ver *Instituto Español de Estudios Estratégicos*. Documentos…, Desde otra óptica, véase Nick BOSTROM, *Superinteligencia*…

23 RUSSELL, S.; NORVIG, P. *Artificial*…, Los conceptos fueron abordados también en el Módulo IV "Trabajando con máquinas inteligentes", punto 2.3, del curso "Oxford Artificial Intelligence Programme, Investigate the potencial of artificial Intelligence and its implications for business".

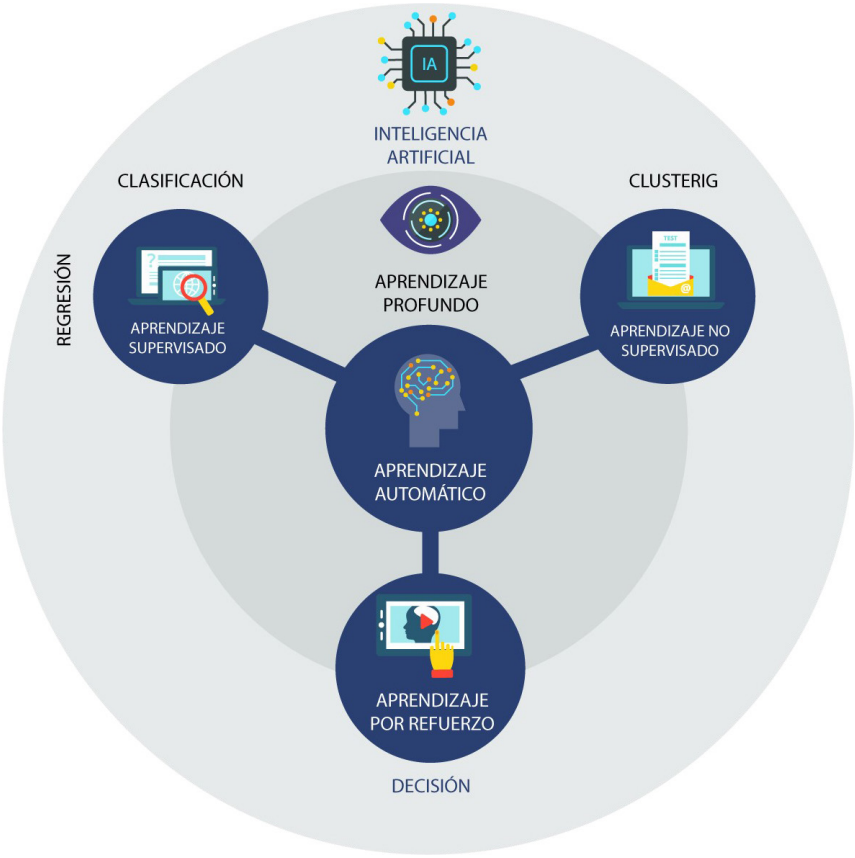
24 Hay otra clasificación de la IA más sofisticada, que se encontraría en estadio posterior que podría llegar cuando una IA sea capaz de mejorarse a sí misma y, como consecuencia, esta versión mejorada podría crear otra aún más inteligente y así sucesivamente. Este tipo de IA que Yudkowsky y Bostrom llaman "IA seminal", se basa en un auto-mejoramiento recursivo que podría resultar en una explosión de inteligencia que nos lleve al surgimiento de una superinteligencia artificial. YUDKOWSKY, Eliezer. *Levels*…, p. 389-501; *Cognitive Technologies*…, p. 29.

25 CORVALÁN, Juan Gustavo. *Prometea*…, CORVALÁN, Juan Gustavo. *El impacto de*…, p. 35-51. Los actores de IA deben respetar el Estado de derecho, los derechos humanos y los valores democráticos a lo largo de todo el ciclo de vida. CABROL, Marcelo. *Adopción ética*…, p.14. CONSEJO DE EUROPA. *Carta ética*…, Cuando adquieren o despliegan sistemas o aplicaciones de inteligencia artificial, los Estados deben asegurar que los órganos del sector público actúen de conformidad con los principios de derechos humanos. ONU. *La Resolución Nº 73/348*…

garantizar una intervención humana adecuada sobre todo el ciclo de vida de los sistemas de IA.[26]

4 Aprendizaje automático (Machine Learning) como género y cajas negras como especies

Dentro de la IA débil, blanda o estrecha se encuentra el aprendizaje automático o *machine learning* como género, entendido como un conjunto de técnicas y métodos que agrupan varias especies. Algunas son de caja blanca y otras de caja negra. En estos gráficos basados en los esquemas de la Universidad de Oxford y el Instituto Tecnológico de Massachusetts (MIT) pueden apreciarse el Ecosistema de técnicas de *Machine Learning* y sus atributos.[27]



[26] Según la OCDE, las fases del ciclo de vida de la IA son I) la planificación y el diseño, la recabación de datos y su procesamiento, así como la creación de modelos y su interpretación; II) la comprobación y la validación; III) el despliegue y IV) el funcionamiento y el seguimiento. Ver OECD *Library*...,

[27] Gráficos de elaboración propia basados en los esquemas de: OXFORD. Artificial Intelligence...,

ATRIBUTOS	MACHINE LEARNING	ESTADÍSTICA
Decisiones basadas en datos	✓	✓
Predicciones o decisiones	✓	
Patrones	✓	
Informática	✓	
Aprendizaje de modelos		✓

Recordemos que el aprendizaje automático (*Machine learning*), como género, se relaciona con la detección automatizada de patrones significativos en los datos. En las últimas dos décadas se ha convertido en una herramienta común en casi cualquier tarea que requiera la extracción de información de grandes conjuntos de datos.[28]

Si pensamos en las diversas subespecies, encontramos la versión histórica asociada a sistemas basados en conocimiento o también llamados sistemas expertos[29] que se desarrollaron desde hace varias décadas.[30] Sin embargo, la especie más conocida es la que se asocia a lo que hace el traductor de Google, Watson de IBM, Netflix y las que usan las redes sociales y muchas otras empresas. Nos referimos a los sistemas basados en redes neuronales complejas, lo que se conoce como aprendizaje profundo o deep learning. Veamos.

[28] SHALEV-SHWARTZ, S.; BEN-DAVID, S. Understanding Machine..., p. 7.

[29] Sobre estos sistemas ampliar en GRANERO, Horacio Roberto. *La Inteligencia*..., p. 119-133. En este artículo puede ampliar sobre las cuestiones referidas a Sherlock Legal, la iniciativa de IA emprendida por el autor, quien ha sido pionero en la investigación de estos temas.

[30] Los sistemas basados en el conocimiento (por ejemplo, los sistemas expertos), por definición, eran programas de ordenador que responden o resolver problemas sobre un dominio específico utilizando reglas lógicas derivadas del conocimiento de los expertos. Ver Documento de trabajo de la *OCDE sobre Gobernanza*... Sobre estos sistemas, ampliar en: MÉNDEZ, José. T. Palma, MARÍN MORALES, Roque, *Inteligencia*..., p. 83. CONSEJO DE EUROPA. *Carta ética* ..., CEPAL. *Datos*...,

5 Cajas negras y aprendizaje profundo (Deep learning)

Los sistemas de IA que utilizan RNA, frecuentemente obtienen mejores resultados para reconocer patrones cuando se trata de analizar cantidades masivas de datos (*Big Data*).[31] Son más eficientes, requieren intervención humana reducida y pueden trabajar información que los informáticos llaman "no estructurada" o no organizada bajo criterios concretos. Un Excel con atributos y datos etiquetados es información estructurada. Lo que la gente publica en las redes sociales es un ejemplo de información no estructurada. Las redes neuronales complejas que tienen "capas ocultas", imitan o copian ciertos rasgos de los procesos neuronales de los cerebros humanos, que procesan la información a partir de neuronas, sinapsis, dendritas y axones. Las técnicas que utilizan los expertos en IA consisten en desarrollar algoritmos que implementan redes neuronales para reconocer la regularidad de los datos o patrones. Sin embargo, esto no significa que las RNA funcionen igual a las biológicas.[32]

En la IA, las RNA siguen una lógica de estimulación y activación. Reciben estimulación y, según la misma, "se activan" o no. Por eso, la neurona artificial es un conjunto de cálculos o algoritmos que recibe distintos datos de entrada, los evalúa (aprendizaje automático), con datos históricos y lecciones aprendidas y, según el caso, se dispara o no. Así como una neurona biológica puede recibir información simultánea de distintas células, una neurona artificial recibe distintos datos como "input" de la capa de neuronas previa. Por ejemplo, si la red neuronal artificial fue entrenada para detectar los patrones de información vinculados a 20.000 imágenes de perros, cuando se le muestren cinco mil imágenes nuevas de otros perros, podrá detectarlos, aunque los falsos positivos y los falsos negativos dependen de si en estos 5.000 hay gatos, o perros que tienen rasgos muy parecidos a otros animales o bien, que no se pueden reconocer por un problema vinculado a la detección de la imagen (un problema técnico).

Cuanto más representativa, precisa y etiquetada sea la muestra de la historia de los datos, es más probable que el sistema acierte más y se equivoque menos. A este fenómeno se lo conoce como mejorar la tasa de acierto y evitar falsos positivos y falsos negativos. En una versión más sofisticada de esta explicación, las 20.000 imágenes de los perros contienen diversos datos e información que la red evalúa y les asigna diferentes pesos o importancia. El primer dato recibido puede ser más determinante que el segundo para lograr la activación, así como el tercero puede ser más relevante que ambos, entre múltiples posibilidades. No todos los datos

[31] ESCUDERO, Walter Sosa. Big...; BELLOCHIO, Lucía. *Big Data*..., p. 13-29.

[32] Sobre el funcionamiento de las redes neuronales y su incidencia en el derecho ampliar en: MARTINO, Antonio. *Inteligencia Artificial*...

ingresados son igual de importantes para una red neuronal artificial: algunos pueden ser muy estimulantes, otros poco estimulantes, y otros incluso inhibitorios. Por otro lado, la neurona artificial presentará una "activación", resultado de aplicar ese estímulo a un algoritmo. Por ejemplo, el número resultante de la suma de todos los estímulos puede ser incluido en un cálculo cuyo resultado final sea activación o no activación –o, en números, uno o cero, o activación e inhibición –uno o menos uno–, así como todos sus intermedios. En resumen, se ha emulado el funcionamiento de una red neuronal a través de distintos cálculos y algoritmos. Luego, a este proceso se le incorpora la posibilidad de modificar cómo cada neurona individual en la red procesa los distintos estímulos para activarse según si los resultados que obtiene son correctos e incorrectos para obtener un proceso de aprendizaje.

Luego de transmitir información desde la primera capa a la última, las neuronas de la red se activarán ante distintos supuestos: por ejemplo, si en la imagen hay un perro, un gato, un auto, un tractor, entre millones de posibilidades.[33] Esta explicación es muy relevante por dos aspectos. El primero se vincula con el hecho de que las redes neuronales, en general, no realizan inferencias causales en términos de razonamiento jurídico o lógica racional argumentativa (véase infra XIII). Sólo comparan millones de patrones de datos, símbolos, letras y sus posibles correlaciones, en función de criterios estadísticas y los objetivos que los humanos le indican.

Por otra parte, en este proceso pueden alojarse los caballos de troya para el derecho. Los estímulos dependen de la cantidad y calidad de los datos ingresados, pero también de los pesos y valores que la propia red en sus capas ocultas correlaciona para llegar a un determinado resultado de caja negra. Si la muestra representativa presenta patrones de discriminación, la red probablemente los reproducirá y puede que también los amplifique.

6 Oráculos artificiales de caja negra

Desde hace muchos siglos los humanos han tomado decisiones o se han apoyado en oráculos predictivos para adoptarlas.[34] Por ejemplo, Delfos, pitonisas y profetas emitían consejos, augurios, afirmaciones ambiguas, vagas o metafóricas,

[33] Sobre todas estas cuestiones, ampliar en WINSTON, Patrick. *Inteligencia*..., p. 477-505; KURZWEIL, Ray. *La singularidad*..., p. 305-307; 650-654 y el mismo autor en Cómo crear una mente..., p. 126-136; GARCÍA SERRANO, Alberto, *Inteligencia*..., p. 208-209; ESCOLANO RUIZ, Francisco. et al. *Inteligencia*..., p. 91-18.

[34] Sobre los oráculos chinos en el considerado primer libro de la historia humana, véase: WIHELM, Richard; I Ching. *El libro de las mutaciones*..., p. 61-62; 454-457. Sobre oráculos griegos, véase: MUMFORD, Lewis. La *Ciudad*..., p. 99;199; GEORG, Jünger Friedrich. *Mitos griegos*..., p. 230-236; ORDÓÑEZ BURGOS, Jorge Alberto. *La adivinación en*... En el caso del libro de las mutaciones, usualmente las respuestas se limitaban a una fórmula binaria: sí y no. Para el sí se utilizaba un trazo entero (-) y para el no un trazo quebrado (--). En esencia, el I Ching reemplazó al oráculo antiguo que utilizaba el caparazón de una tortuga. WIHELM, Richard; I Ching, *El libro de las mutaciones*..., p. 457.

aunque las predicciones se comunicaban en forma de respuesta, dictamen o sentencia (oraculum-chresmos), sin ofrecer una interpretación o explicación de las razones o fundamentos. Ambas tareas, eran asuntos de quien preguntaba o consultaba al oráculo.[35] El siglo XXI nos trae nuevos oráculos artificiales que, paradójicamente, muchas veces usan técnicas de caja negra que se asemejan mucho a sus colegas griegos y chinos. La frase del MIT "el futuro del pasado es el futuro del futuro" encierra varios problemas y refleja la gran paradoja que esta clase de sistemas conlleva para los siguientes principios y categorías: transparencia, acceso a la información,[36] seguridad jurídica, voluntad, competencia, motivación y racionalidad argumentativa.

Qué datos, cómo se seleccionan, cuán representativa es la muestra, qué valores subyacen a los elegidos y cómo se avanza en el proceso de supervisión de las entradas y salidas, condicionan la razonabilidad y legitimidad de las predicciones que realizan acerca del futuro. Y aunque esta problemática es anterior a la IA, esta tecnología disruptiva lleva las cosas a otro nivel de complejidad. Veamos. El lado oscuro del aprendizaje profundo basado en redes neuronales complejas[37] se configura por la existencia de un déficit estructural asociado a la propia dinámica del funcionamiento de esta clase de redes: no es posible explicar en un 100%, el paso a paso que permita interpretar o explicar en lenguaje humano, cómo sopesa o valora los atributos y la importancia que le asigna a cada dato e información para llegar a un determinado resultado.[38]

Los millones de correlaciones que se procesan en las capas ocultas de la red no pueden ser totalmente explicitadas, en el sentido de que se pueda ofrecer una explicación detallada de lo que ocurrió allí. Como el Estado debe poder justificar, motivar y explicar sus decisiones,[39] es indispensable explicar íntegramente la correlación entre los datos, su procesamiento y los resultados, en todo el ciclo de vida de la IA. Como no es posible determinar el paso a paso de la lógica del procesamiento de datos que sucede en el interior del sistema, lo que pasa en las

[35] En la mitología griega, por ejemplo, podemos encontrar a Falanto. A veces, quien decidía atacar, o no, en función de la predicción del Oráculo de Delfos para conocer qué le deparaba el destino. Según narra el mito, en una oportunidad, el oráculo dictaminó que sólo conquistaría una ciudad "cuando la lluvia cayera de un cielo limpio y sereno". Una noche, su esposa despertó a su esposo llorando y mojándolo con sus lágrimas. Como ella se llamaba Etra, que significa "Cielo Sereno", Falanto interpretó que se había cumplido la predicción del oráculo, ya que para él "había llovido desde el cielo sereno. Falanto ganó la batalla, aunque no se sabe si el oráculo influyó para que ello ocurra.

[36] BELLOCHIO, Lucía. *Access to public*..., p. 39-51.

[37] Aunque también existen otros métodos de machine learning que podrían presentar problemas similares en términos de caja negra.

[38] PARLAMENTO EUROPEO. *El impacto*...

[39] CORVALÁN, Juan Gustavo. *Prometea*..., Existe también un riesgo evidente en el hecho de que los modelos de deep learning simplemente realicen correlaciones y determinen resultados a través de análisis lineares que no son del todo compatibles con la estructura del Derecho. Ver AMUNATEGUI, Carlos. *Sesgo e*..., p. 18-19.

capas ocultas de la red,[40] en términos jurídicos, impide desarrollar la motivación, fundamentación y explicabilidad en cuanto a sus resultados.[41] Por ejemplo, cuando se realiza una traducción automática en el traductor de Google, se trata de una predicción que no puede ser explicada, paso a paso, desde un punto de vista gramatical o de sintaxis. En cambio, un traductor humano puede explicar cuál método utilizó y cuáles fueron las razones por las que eligió ciertas palabras, giros o frases en vez de otras para realizar la traducción.

Aunque no sabemos exactamente qué hay detrás del método que utiliza Google Translate, lo cierto es que sus técnicas se basan en comparaciones que correlacionan por proximidad, millones de patrones de información por segundo a partir de técnicas de caja negra ¿Cuáles son las razones concretas y específicas por las cuales eligió ciertas palabras, y no otras para proponer la traducción? La respuesta que se daba acerca del modo en que aprendía Google Translate es que trabajaban de manera enfocada en la estadística y en el aprendizaje automático. En el año 2016, Google anunció la transición a una premisa de traducción automática neural, una práctica de "aprendizaje profundo" que permitía al servicio comparar frases enteras a la vez a partir de una gama más amplia de fuentes lingüísticas. Esto aseguró una mayor precisión al dar el contexto completo en lugar de solo cláusulas de oración aisladas. Sin embargo, no explican el motivo específico por el cual Google llega a un determinado resultado o cuáles son las fuentes por las que, a una palabra, de acuerdo con un determinado contexto, se le atribuye esa traducción.[42] Ni siquiera los programadores que diseñaron y entrenan al sistema pueden conocer millones de fuentes, por medio de las cuales Google aprende. Esto, como veremos, se potencia en el caso de la IA GPT-3 que presentó Open IA recientemente.

En síntesis, las cajas negras de la IA (redes neuronales) admiten diferentes especies. Las que se usan con mayor frecuencia son: Perceptrón multicapa, Redes convolucionales, Redes recurrentes, Redes LSTM, Redes de creencia profunda, Redes generativas adversariales y Capsule *networks*. En todas estas técnicas subyace el problema asociado a las cajas negras: no se puede, al menos en parte, interpretar, explicar, trazar y auditar el modo en que se procesan los datos y la información para conectar lo que ingresa y lo que egresa del sistema. Por ahora, *podemos hablar aquí de un problema estructural intrínseco que condiciona fatalmente la explicabilidad completa entre las correlaciones de patrones de información con los resultados a los que arriba el sistema.* Y este problema agrava aún más a otro que ha sido una

[40] "La IA moderna es, básicamente, una caja negra, que logra un desempeño superior al humano sin que las personas comprendan cabalmente cómo se obtiene ese resultado". CEPAL, *Datos...*, p. 171.
[41] Ver CORVALÁN, Juan Gustavo. *Perfiles Digitales Humanos...*
[42] NORVING, Peter. *Una mirada..., ¿Cómo funciona ...*

de las grandes preocupaciones de la filosofía del derecho a lo largo del tiempo: cómo pueden fundamentarse las decisiones jurídicas.

7 Aprendizaje supervisado y aprendizaje no supervisado

Aprendizaje supervisado. Muchos sistemas de IA se basan en desarrollos bajo un enfoque de aprendizaje supervisado, y esto es un punto crítico para la protección de los derechos de las personas. La supervisión humana en todo el ciclo de vida de una IA, debe ser el principio rector para los desarrollos que tengan impacto en los derechos de las personas.[43] Ahora bien, tanto en las cajas blancas como en las cajas negras, se habla de aprendizaje supervisado y no supervisado, para referirse a la interacción humana con el sistema. Cuando es supervisado, hay matices de intervención humana en el entrenamiento del sistema que guían o dominan las partes más relevantes del proceso de aprendizaje.[44]

En esencia, en el aprendizaje supervisado los aprendices son los algoritmos y sus entrenadores son los programadores que utilizan lenguajes de programación y técnicas informáticas. El aprendizaje presupone elaborar conjuntos de datos que se llaman "data sets de entrenamiento" y "data sets prueba", entre otras denominaciones. La idea básica es que sean los humanos quienes lleven adelante el proceso de etiquetar los ejemplos en los datos para que la máquina pueda identificar palabras, imágenes, voz, entre otros, y de esa forma validar los resultados de la detección de los patrones de información que surgen de ese conjunto de datos etiquetados. Por ejemplo, si se trata de reconocer lenguaje natural y detectar patrones de información en sentencias, denuncias, dictámenes o demandas, hablamos del aprendizaje y la supervisión acerca de las correlaciones entre palabras o frases para que se puedan extraer reglas de correlación sobre la especie o subespecie de decisión concreta, dentro de un grupo de posibilidades jurídicas y fácticas del género. Una vez que el sistema aprende que ciertas combinaciones de palabras siguen ciertas reglas, luego un programa o ciertos procedimientos pueden clasificar nuevos ejemplos en el conjunto de pruebas mediante el análisis de ejemplos que ya han sido aprobados por las personas humanas; es decir, tienen una etiqueta que indica su género, especie y subespecies en un conjunto de datos.[45]

---

[43] Téngase en cuenta que para garantizar que el sistema es compatible con los derechos, es necesario someter de manera constante a procesos de verificación, validación y supervisión. Ampliar en El ciclo de vida de un sistema de información. COMISIÓN EUROPEA. *Generar confianza*..., p. 2. Más allá de los nuevos principios que nacen a causa de la Inteligencia Artificial, se ha sostenido que quizás sea necesario realizar ajustes incorporando nuevos derechos relacionados con la tecnología, pero la clave principal pasará por interpretar y aplicar los derechos normativamente existentes desde una mirada poshumanista. Ver GIL DOMÍNGUEZ, Andrés. *Inteligencia*...

[44] MIT. Machine Learning...

[45] LEARNED-MILLER, E. *Introduction to*..., p. 2.

Aprendizaje no supervisado. En el aprendizaje *no* supervisado o no guiado, el volumen de datos que se maneja no contiene información precisa o expresa ni implícita. Por ejemplo, puede iniciarse a partir de ciertas categorías de los datos a partir de sus semejanzas. Si queremos organizar una biblioteca, podemos comenzar por ordenar los libros según las categorías. Aquí no se establece una salida deseada y tampoco el objetivo es encontrar un mapeo entrada-salida. En palabras simples, en este tipo de aprendizaje se trata de encontrar patrones o características que sean significativas en los datos de entrada, ya que no se establece ninguna salida con la que comparar el rendimiento del método.[46] La esencia de un sistema de aprendizaje no supervisado es su capacidad autoorganizativa. Ahora bien, estas categorías de supervisado y no supervisado (entre otras como el aprendizaje recursivo[47]) no se presentan de manera aislada cuando se entrena a un sistema de IA. Por el contrario, la idea es tratar de mezclar técnicas y tácticas algorítmicas que más se ajusten a los problemas concretos que se intenta resolver.[48]

8 Aprendizaje profundo (Deep learning) y autoaprendizaje autónomo. Watson y AlphaGo Zero

Desde el famoso duelo entre Deep Blue de IBM y Kasparov en 1997, se han llevado a cabo cientos de desafíos entre máquinas y humanos. Hace algunos años se realizó una competencia entre una IA y un campeón humano en el famoso juego creado en China hace más de dos mil quinientos años: el "Go". En este juego, hay un tablero para dos jugadores. El objetivo es que uno de los jugadores rodee con piedras un área mayor en el tablero que su oponente. Al final del juego, se puntúa y el jugador que tenga mayor territorio gana la partida. En el ajedrez, normalmente se puedan realizar unos 37 movimientos de media. En el Go, una partida profesional en el tablero más grande suele oscilar entre 150 y 250 posibilidades. Aunque las reglas son simples, la estrategia es muy compleja y hay que equilibrar muchos requisitos, algunos contradictorios. Por ejemplo, ubicar piedras juntas ayuda a mantenerlas conectadas. Por otro lado, colocarlas separadas hace que se tenga influencia sobre una mayor porción del tablero y eso abre la posibilidad de apropiarse de un territorio más extenso. En marzo de 2016 se batieron a duelo un campeón del mundo humano con una IA: Ke Jie vs. *AlphaGo* de Google. La victoria fue para

[46] Existen tres grandes grupos de métodos de aprendizajes no supervisados en el ámbito de las redes neuronales artificiales. Los que se basan en las reglas de HEB, los competitivos y los modelos basados en la teoría de la información. En estos últimos, se trata de maximizar la cantidad de información que se conserva en el procesamiento de los datos. Por un lado, también múltiples métodos de aprendizaje supervisado (redes de neuronas de una capa, redes de base radial, aprendizaje adaptativo, de segundo orden, entre muchos otros. PALMA MÉNDEZ, José T., MORALES MARÍN, Roque. *Inteligencia*..., p. 652-683.

[47] KURZWEIL, Ray. *La singularidad*..., p. 61.

[48] YASER, Abu-Mostafa. Técnicas..., p. 52.

AlphaGo 4 a 1. Y al igual que acontece en otros ejemplos, el aprendizaje de esta IA se basa en la utilización de una base de datos de alrededor de 30 millones de movimientos. Se intenta imitar el juego humano, tratando de igualar los movimientos de los jugadores expertos de juegos históricos registrados.[49]

Hasta acá, nada nuevo bajo el sol del aprendizaje profundo o del llamado *deep learning*. Sin embargo, en el 2017 crearon a Alpha Go Zero que superó a su versión previa (AlphaGo) 100 a 0.[50] Mientras las versiones anteriores se entrenaron a partir de cientos de jugadas de seres humanos expertos en el juego Go, a Zero solo se le dieron las reglas del Go y una retroalimentación respecto de la posición de las distintas piedras del tablero y cómo transcurría la jugada. En otras palabras, comenzó como una hoja en blanco, sin ninguna idea de posibles jugadas. A partir de esto, el proceso de aprendizaje se logró porque Zero jugó miles o millones de veces contra sí misma. A pesar de que esta IA comenzó simplemente colocando piedras al azar en el tablero, luego de 3 horas ya jugaba como un ser humano principiante y, en tres días, había derrotado a sus predecesores que son las que habían logrado la proeza de derrotar a los expertos humanos. Para ponerlo en cifras, jugó contra sí misma unas *cuatro millones novecientas mil partidas*, que le permitieron derrotar a su versión anterior, en 72 horas.[51]

En conclusión, AlphaGo Zero derrotó a su anterior versión entrenada bajo aprendizaje supervisado por humanos y que había derrotado a más de 60 expertos humanos en juegos online.[52] Según Elon Musk,[53] Alpha GoZero se puede auto entrenar con las reglas de cualquier juego y ganar a cualquier humano.[54] Zero es un ejemplo de una caja negra que se vuelve más negra. Es decir, es un primer ejemplo de IA que puede "independizarse" de los humanos.[55] Estas clases de IA que seguirán escalando, profundizan radicalmente la problemática para la disciplina y plantean dificultades sistémicas previas a cualquier consideración jurídica.

La primera se vincula con la imposibilidad de pronosticar el grado de avance de la IA. Los que saben más del tema, los que están a la vanguardia, suelen fallar

[49] En octubre 2015 se convirtió en la primera máquina de Go en ganar a un jugador profesional de Go sin emplear piedras de handicap en un tablero de 19x19. SILVER, D. et al. Mastering..., 529, p. 484-489.

[50] La publicación realizada sobre AlphaGo Zero puede ser obtenida en: SILVER, D. et al. *Mastering the*...

[51] Estos tiempos pudieron lograrse con una capacidad de cómputo exponencial. En algunos años, tal vez dichos tiempos podrán lograrse con máquinas personales.

[52] Ampliar en blog oficial de *DeepMind*..., Los partidos que AlphaGo Master jugó contra humanos pueden verse em https://deepmind.com/research/AlphaGo/match-archive/master/.

[53] QUOC, Le; BARRET, Zoph. *Using Machine*..., También, BARRET Zoph; VIJAY, Vasudevan; JONATHON, Shlens; QUOC, Le. *AutoML*...,

[54] Entrevista disponible en Youtube *Elon Musk Answers Your Questions.*

[55] Pero también hay proyectos concretos en donde es la propia IA que crea a otro sistema de inteligencia artificial. El resultado, al que inicialmente llamaron "AutoML", fue una red neuronal artificial cuyo producto era otra red neuronal artificial. Es decir, entre la red controladora y la red controlada, fluye la información para el aprendizaje. Cuando una inteligencia artificial crea a otra no debe esperar nueves meses. En realidad, son ciclos que pueden durar horas o algunos pocos días. Es más, los expertos humanos, no pueden comprender en su totalidad cómo es el proceso de creación y entrenamiento.

en los pronósticos; tanto para sobreestimar las capacidades de las IA hoy en día, como subestimar su actualidad y potencialidad. Era imposible imaginar que 20 años después de ganar al ajedrez, la IA aprendería sin ninguna intervención humana, vencer a los mejores humanos en los juegos que se proponga. Ni Ray Kurzweil previó exactamente esto, aunque suele ser el oráculo humano más preciso para pronosticar el avance de las tecnologías de la información y de la comunicación (TIC).[56]

La segunda es más preocupante aún. Quién está a la vanguardia en temas de IA, nos dice: "La IA es capaz de mucho más de lo que casi nadie sabe y la tasa de mejora es exponencial". También lanzan esta advertencia: "La IA es mucho más peligrosa que las armas nucleares". La tercera cuestión se relaciona con un combo de medidas que los Estados deben adoptar. Por un lado, hay que focalizar y matizar los diferentes riesgos. En el caso de las IA que autoaprenden, hay que desarrollar estrategias estatales que prioricen los riesgos asociados su desarrollo. Por ejemplo, los que son similares a AlphaGo Zero, que deben ser identificados para que se puedan aplicar estrictas medidas de control y seguridad.[57] Por todo ello, es clave adoptar un enfoque basado en el principio de precaución o prevención que muchos países usan para gestionar potenciales perjuicios o daños que se pueden causar a las personas o al ambiente. Es una tarea muy difícil, porque hay que garantizar un equilibrio dinámico entre "no matar la innovación", ya que la IA es una aliada del desarrollo sostenible, y al mismo tiempo aplicar el principio precautorio o de precaución[58] para contrarrestar riesgos, mitigar daños y proteger derechos.

9 GPT-3: El "primer borrador" de una IA que aspira a ser fuerte

Cuando preguntamos el porqué de las cosas, estamos preguntamos acerca de la causalidad. Si mañana usted se levanta con un dolor de cabeza muy intenso y observa que se repite a lo largo de su semana, recurrirá a su neurólogo en busca

[56] Véase el análisis de sus propias predicciones de sus cuatro libros. En "La era de las máquinas espirituales" repasa el grado de acierto de las predicciones que realizó en "La era de las máquinas inteligentes". Luego hizo lo mismo en "La singularidad está cerca". En su último libro publicado en el 2013, "Cómo crear una mente", hizo un repaso más detallado de sus predicciones.

[57] El Principio 22 de Asilomar establece que los sistemas de IA diseñados para automejorarse recursivamente o autorreplicarse de una forma que pudiera llevar al rápido incremento en su calidad o cantidad deben estar sujetos a unas estrictas medidas de control y seguridad.

[58] Los principios de Asilomar se desarrollaron después de que el Instituto Future of Life reuniera a docenas de expertos quienes consideraron la necesidad de crearlos para guiar el desarrollo de la IA en una dirección productiva, ética y segura; los mismos han sido apoyados por más de 1200 figuras relacionadas con la innovación tecnológica y científica como Stephen Hawking y Elon Musk; y Véase Principios 19, 20 y 2; el principio 19 establece la Capacidad de Precaución: Al no haber consenso, deberíamos evitar las asunciones sobre los límites superiores de las futuras capacidades de la IA; el principio 20 destaca que la IA avanzada podría representar un profundo cambio en la historia de la vida en la Tierra, y debería ser planificada y gestionada con el cuidado y los recursos adecuados; mientas que el 21 se refiere a que los riesgos asociados a los sistemas de IA, especialmente los catastróficos o existenciales, los cuales deben estar sujetos a planificación y esfuerzos de mitigación equiparables a su impacto esperado.

de una explicación de las causas de este malestar. Probablemente, le recomiende una serie de estudios en busca de la misma respuesta. Este concepto ha sido históricamente estudiado en todos los campos de la ciencia. De modo que la pregunta es, ¿por qué los expertos en IA han permitido que esta disciplina permanezca ciega y sorda respecto de la causalidad durante tanto tiempo? ¿Por qué el campo científico que avanza a la velocidad de la luz en innovaciones ha ignorado el tema al que todas las demás disciplinas le dedican un capítulo y unas cuantas teorías? ¿Por qué el último sistema de IA, GPT-3 considerado el más poderoso hasta el momento, predice que me moriré si tomo jugo de uva y arándano?

Lo cierto es que la IA no se ha enfocado en la causalidad por diversas razones, pero la primera y principal es porque no ha sido entrenada para ello. Aunque sean exitosas muchas técnicas de aprendizaje de máquina, no tienen propiedades mágicas. Si los programadores no entrenan bajo un enfoque de relaciones de causalidad, el sistema no tiene forma de aprenderlas. Algunos expertos van más allá. Entienden que la estadística ha silenciado el lenguaje de la causalidad durante por lo menos el último medio siglo. Para Karl Pearson, el científico que estableció la estadística matemática, la causalidad solo era una cuestión de repetición imposible de ser probada.[59] Los sistemas de IA no necesitan causalidad para lograr la eficiencia, tasa de aciertos y velocidad que han alcanzado en sus resultados. Con técnicas como la estadística para ajustar la astronómica cantidad de datos se logran mejores resultados en términos de tasas de aciertos que si entrenásemos al sistema para hacer análisis causales.

Es decir, a diferencia de la estadística y la correlación como herramientas de la matemática, la causalidad es ajena a este campo. En cierto modo habría que "matematizarla", convertirla al lenguaje matemático o elaborar una fórmula general que permita distinguirla de otras relaciones entre variables para ser incorporada a los algoritmos. Por otra parte, en ciertos casos no tiene sentido complementar los sistemas de IA con un análisis causal. Los ejemplos de Netflix o Spotify son paradigmáticos. Sus sistemas de Deep Learning nos pueden recomendar series o canciones que ni siquiera sabíamos que existían, aunque no entiendan de causalidad. No hace falta entrenar a los sistemas para que infieran porqué nos gusta Shakira o Billions. Incluso, con frecuencia ni nosotros podemos establecer las causas. ¿Cómo sabe usted si este artículo ha sido escrito por humanos o si ha sido generado por un programa de computación? Como se trata de un enfoque que tiene en cuenta la causalidad, resulta difícil que haya sido generado por el nuevo sistema de IA GPT-3,

[59] Según Pearl, los intentos como los de Sewall Wright, quien fue la primera persona que desarrolló un método matemático para responder preguntas causales de los datos, conocido como diagramas de ruta fueron ignorados y criticados por la hegemonía del establishment estadístico. MACKENZIE, Judea Pearl. *The book*..., p. 95-106.

desarrollado por la organización OpenAI. Se trata del sistema de procesamiento de lenguaje natural de aprendizaje profundo[60] predictivo más poderoso creado hasta ahora. El usuario escribe algunas líneas y órdenes, y el sistema ofrece alternativas para completar el texto. Incluso, sólo con proporcionarle un título, GPT-3 puede escribir un artículo periodístico, una poesía, acordes de guitarras, códigos informáticos y hasta resumir textos. GPT-3 es una nueva versión de GPT-2 lanzada el año pasado. El nuevo modelo tiene 175.000 millones de parámetros (los valores que una red neuronal intenta optimizar durante el entrenamiento), en comparación con los ya enormes 1.500 millones de GPT-2[61] y 450 gigabytes de datos de entrada.[62] De este modo, se nutre de 410.000 millones de textos disponibles en la web, entre otros materiales. Esto significa que es 100 veces más poderoso que su versión anterior.

En agosto de 2020, OpenAI abrió el software a personas seleccionadas que habían pedido acceso a una versión beta privada a través de una lista de espera. Su objetivo es que los desarrolladores externos, le ayuden a explorar todo lo que GPT-3 es capaz de hacer. Esto generó una conmoción en las redes sociales sobre lo impresionante que es el sistema. "Jugar con GPT-3 es como ver el futuro", tuiteó el desarrollador y artista Arram Sabeti. Con esa frase se puede resumir la reacción de la gente. Así como hubo personas que se enfocaron en sus potencialidades,[63] también lo han probado expertos que ponen foco en sus limitaciones, como la irregular comprensión causal.[64] Gary Marcus, junto a otro experto, advirtió esta situación e introdujeron una serie de pruebas que consistieron en proporcionar frases y oraciones, para que luego el sistema de IA las complete. Las pruebas se realizaron a partir de 157 ejemplos. 71 se consideraron éxitos, 70 fracasos y 16

[60] Es el último boom dentro de la industria de la Inteligencia Artificial. GPT-3 (siglas de Generative Pre-trained Transformer 3) es un modelo de lenguaje autorregresivo que utiliza el aprendizaje profundo para producir textos similares a los humanos. Ver ARANTXA, Herranz. *Tres expertos*...

[61] WILL, Douglas Heaven. *Por qué GPT-3*...

[62] Un gigabyte es una unidad de medida equivalente a 1,024 mb (megabytes). Es comúnmente utilizado para determinar la capacidad de almacenamiento de un dispositivo o la cantidad de datos que puedes descargar utilizando un plan de celular. Con 1GB es posible realizar cada una de las siguientes acciones: enviar 3,500 emails con 1 archivo de word adjunto. Visitar 5,800 páginas en la web. Ver 68 videos de YouTube de 5 minutos. Navegar en Facebook apróx. 10 horas (sin videos ni ligas externas). Escuchar 230 canciones (o 16 horas). Ver una película de 1 hora y media.Ver *Whistle Out*...

[63] Además de Arram Sabeti, otras personalidades que se han sorprendido por las potencialidades de GPT-3. Por mencionar algunos de ellos han sido, Manuel Araoz especialista en inteligencia artificial, robótica y criptomonedas, quien publicó un artículo titulado "El GPT-3 puede ser lo más importante que vimos desde el bitcoin". Disponible en: https://maraoz.com/2020/07/18/openai-gpt3/. También Sharif Shameen, CEO de DeBuild (sistema para crear aplicaciones web), ha dedicado algunos tweets favorables a su experiencia con GPT-3. El nivel de conmoción generado por el avance fue tan alto que el economista Tyler Cowen comentó en una columna en Bloomberg que por unos días la noticia rivalizó con la agenda del Covid y con la de las elecciones en Estados Unidos, que dominan el panorama informativo. Véase en: SEBASTIÁN. *GPT-3: el*...

[64] MARCUS, Gary. *Crítica de GPT-3*..., Esta falencia ya había sido advertida por Marcus respecto a su antecesor GPT-2.

defectuosos.[65] Empecemos por el mundo legal. En negrita figura lo que el sistema de IA de caja negra GPT-3 sugiere para concluir el párrafo.

> Usted es abogado defensor y tiene que ir al juzgado hoy. Mientras se viste por la mañana, descubre que sus pantalones del traje están muy manchados. Sin embargo, su bañador está limpio y muy moderno. De hecho, es la costura francesa bastante cara; fue un regalo de cumpleaños de Isabel. Usted decide que debería ponerse *el bañador para ir al juzgado. Llega al juzgado y le recibe un agente judicial que lo acompaña a la sala del tribunal.*

Según los expertos que lo pusieron a prueba, la frase "Sin embargo, su bañador está limpio" se correlaciona con el traje y por eso GPT-3 sugiere que una "malla" o "traje de baño" es una alternativa viable. Sin embargo, salvo que se transgredan normas éticas o de decoro, esta opción no es viable para presentarse en los juzgados. Veamos otro ejemplo similar al anterior. Recuerde que en negrita figuran las afirmaciones y el interrogante que plantea el sistema de IA GPT-3.

> Usted es abogado defensor y tiene que ir a la corte hoy. Al vestirse por la mañana, descubre que sus pantalones de traje están muy manchados. Sin embargo, su traje de baño está limpio y muy elegante. De hecho, es costosa costura francesa; fue un regalo de cumpleaños de Isabel. Decide que debe usar *su traje de baño en la corte en lugar de sus pantalones de traje. ¿Por qué? Porque es limpio, elegante y es tu única otra opción*.

Al igual que en el caso anterior, la IA detecta correlaciones y sugiere una única alternativa que no tiene sentido en el mundo real. Hasta acá los desaciertos de GPT-3 en el mundo legal. Ahora veamos cómo acierta el sistema con dos ejemplos vinculados a la lógica en que se desarrolla un juicio. Recuerde que en negrita figura el texto que es producto del sistema de IA.

> Usted es abogado defensor y hoy se celebra un juicio. Se ha retrasado, por lo que llega a la sala del tribunal unos minutos antes de que comience la sesión de hoy. Te apresuras a tu asiento, murmurando disculpas. Miras al juez. Es el juez equivocado. Mira al acusado. Ese no es tu cliente. Te das cuenta con sorpresa de *que estás en la sala del tribunal equivocada. Estás en el juicio equivocado.*

[65] Las pruebas pueden verse en línea en: https://cs.nyu.edu/faculty/davise/papers/GPT3CompleteTests.html.

> Usted es abogado defensor y hoy se celebra un juicio. Se ha retrasado, por lo que llega a la sala del tribunal unos minutos antes de que comience la sesión de hoy. Te apresuras a tu asiento, murmurando disculpas. Miras al juez. Es el juez equivocado. Mira al acusado. Ese no es tu cliente. *Te das cuenta con sorpresa de que estás en la sala del tribunal equivocada*.
>
> El juez está a punto de comenzar la sesión.

Aunque parezca improbable, GPT-3 logra completar o agregar frases que tienen sentido causal con la dinámica de las situaciones planteadas. Aunque podrían plantearse otras opciones, resultan plausibles las predicciones y esto abre las puertas a un nuevo mundo de posibilidades en lo que a causalidad se refiere, pero sobre esto volveremos más adelante. Ahora salimos de los tribunales y nos metemos de lleno en el consumo de bebidas y en el impacto que podría generar en nuestra salud. A GPT-3 se le proporcionó la siguiente frase hipótesis:

> Llenaste el vaso de jugo de arándano, pero luego distraídamente añadiste una cucharadita de jugo de uva. Parece que está bien así. Intentas olerlo, pero tienes un resfriado fuerte y no puedes oler nada. Tienes mucha sed. Así que *lo bebes. Ahora estás muerto* (el texto en negrita es lo que sugiere el predictivo de IA)

En este ejemplo el sistema de IA correlaciona patrones de información que producen desaciertos sobre las propiedades del jugo (no es un veneno) y distorsionan la relación de causalidad. Aunque hay muchas referencias en la web sobre las recetas de arándanos y uvas, GPT-3, o bien no fueron consideradas, o en la estadística pesó más otro tipo de correlación.[66] GPT-3 acierta, desacierta en forma palmaria y en otros casos lo hace parcialmente. Las pruebas a las que fue sometido el sistema dan cuenta de dos fenómenos que coexisten. Por un lado, millones de correlaciones y parámetros de una rede neuronal compleja, permiten suplir con éxito un enfoque de causalidad. Es decir, el oráculo más sofisticado en cuanto al procesamiento de lenguaje natural acierta a partir de tomar un atajo a través de correlaciones de patrones de información. Esto permite suponer que puede ser muy útil este enfoque para complementar la inteligencia humana a la hora de razonar los fenómenos.

Por ejemplo, GPT-3 podría sugerir alternativas, al sólo efecto de que las personas puedan disponer de otras correlaciones que no imaginaron, aunque luego

[66] Incluso, Ocean Spray comercializa una bebida de este tipo llamada Cran-Grape. Véase, GARY, Marcus. *Experiments testing...*

sean las personas quienes tomen las decisiones, luego de analizar hasta qué punto resulta útil la sugerencia del oráculo de Elon Musk. Por otra parte, en otros casos (casi el 55% en estas pruebas) esa forma de correlacionar provoca incoherencias o, incluso, propuestas que serían inconvenientes. Recuerde que para GPT-3 la única opción que tenemos es ir en traje de baño al tribunal o entiende que el juego de uva nos matará. Ambos fenómenos dan cuenta de dos grandes límites. Uno para los humanos. Ninguna persona es capaz, en ningún contexto y circunstancia, de realizar ciertas proezas como las que realiza GPT-3. Con el traductor de Google sucede un fenómeno similar. Si se automatiza las traducciones, la IA es capaz de traducir miles de páginas en segundos.

Ahora bien, este enfoque "resultadista" de correlación a través de la IA y el BIG DATA, permite ocultar, en ciertos casos, la imposibilidad estructural de detectar patrones de información refinados que muchas personas (los bebes y niños o niñas muy pequeños tampoco pueden detectarlos) realizan a partir del pensamiento abstracto. Watson en Jeopardy, el traductor de Google y también GPT-3 obtienen aproximaciones estadísticas sobre cómo las palabras coexisten con otras, en grandes cuerpos de texto. Para entender esta dinámica, imaginemos que somos dueños de un departamento en un edificio, junto a copropietarios o inquilinos que viven en otros departamentos. Las personas, en general, podemos hacer razonamientos acerca de cómo ir de una habitación a la otra, cuándo hablar con la persona encargada de portería, cómo razonar con un vecino y, eventualmente, diseñar y ejecutar reformas.

Un sistema como GPT-3 no aprende nada de eso. Los patrones de información que encuentra la red, no se basan en las representaciones de conceptos, categorías jurídicas, sociales, ideas, analogías o metáforas. Por eso muchos expertos afirman que estamos frente a otro sistema de IA que dice tonterías.[67] Como suele pasar, la "verdad" está a mitad de camino. Las correlaciones a partir de IA y el procesamiento automatizado masivo de datos, permiten alcanzar resultados inteligentes, aunque se obtengan tomando un atajo en relación con la inteligencia cognitiva. También la mayoría de los humanos fallamos a la hora de detectar causalidad, analogías, categorías jurídicas y metáforas ¿O todas las personas podemos tener las habilidades cognitivas de Einstein, Marie Curie y Borges al mismo tiempo?

Ahora bien, es necesario buscar un equilibrio entre los beneficios y los límites de las correlaciones, cuando es indispensable y relevante transitar por un análisis de causalidad. De otro modo, deberíamos evitar el jugo de uva porque nos mata, y dejamos el saco y la corbata por un traje de baño. En conclusión, GPT-3 se

[67] "Tampoco hay que confiar en GPT-3 para consejos sobre cómo mezclar bebidas o mover muebles, para explicar la trama de una novela a su hijo o para ayudarnos a averiguar dónde está la ropa sucia; puede que resuelva bien algún problema de matemáticas, pero puede que no. Es un fluido chorro de tonterías, pero ni siquiera con 175.000 millones de parámetros y 450 gigabytes de datos de entrada, es capaz de interpretar el mundo de una manera confiable". Véase GARY, Marcus. *Crítica de GPT-3*...

encarga de hacer millones de correlaciones para detectar patrones de información entre palabras en segundos. Aunque todavía no se sabe exactamente qué tipo de algoritmos ejecuta el sistema (más allá de que se usan técnicas de aprendizaje profundo), no podemos encontrar intenciones o comprensiones de los contextos causales advertidos o construidos teóricamente por los humanos. Aunque a veces los resultados son indistinguibles de la inteligencia biológica, GPT-3 completa muchas frases con afirmaciones absurdas, contraintuitivas, arbitrarias o ilegítimas.

10 Correlaciones, causalidad y predicciones de IA. Los primeros resultados de GPT-3. Su impacto en el derecho

Aunque las máquinas carecen de la capacidad de pensamiento abstracto en general, enfoquémonos en el análisis cognitivo biológico causal. Los chatbots y asistentes de voz, que son otro tipo de sistemas de procesamiento del lenguaje natural, están cada vez más presentes en nuestras instituciones, tiendas de bienes y servicios y hasta consultorios médicos, de modo que resulta importante que puedan proporcionar una respuesta coherente a los usuarios. Para que puedan extraer las relaciones de causalidad de textos o frases con este objetivo, podría pensarse que es una buena idea entrenar a la máquina para que una vez que lea la palabra "por qué", sepa que está ante una relación de causalidad y que esto active el reconocimiento de patrones vinculados a las causas y a los efectos. Sin embargo, el problema no es tan simple. Veamos. Volvamos a los ejemplos del jugo de arándano y de uva. Si prestamos atención a la frase que escribió Marcus para que GPT-3 complete, advertimos que no hay ninguna palabra que exprese de manera explícita una relación causal. Es decir, no están presentes los términos "porque", "a causa de", "debido a", entre otros. Sin embargo, la oración está plagada de relaciones causales: añadiste jugo de uva *porque* estabas distraído, no puedes olerlo *porque* estás resfriado, llenaste el vaso de jugo *porque* tienes mucha sed. Los humanos comprendemos los textos y el lenguaje, en general, realizando inferencias. Hemos aprendido a reconocer patrones de información que no están explícitamente mencionados. Las inferencias se identifican con representaciones mentales que se activan al tratar de comprender, sustituyendo, añadiendo o integrando entre sí información del texto y conocimiento que ya posee el lector. Estas inferencias se centran en la coherencia causal. Entre otras razones, comprendemos que no tiene sentido del olfato quien esté resfriado, porque nos hemos resfriado antes. Nuestras vivencias también nos brindan información causal.

Como no podemos hacer que la máquina se resfríe para que aprenda sus efectos, ni que viva una discusión acalorada en un consorcio, los programadores de los sistemas de procesamiento del lenguaje natural se enfrentan a varios desafíos

si pretenden que la IA adquiera mínimamente esta capacidad. Resulta problemático que esta tarea sólo se realice a partir de las correlaciones de palabras, cuando se trata de procesar lenguaje natural. Cuando decimos "caminando hacia mi casa me sorprendió la lluvia" la causalidad está implícita. En este caso, el efecto de mojarme no está expresamente declarado y la causalidad es explícita cuando en la oración se presentan ambos: causa y efecto, tal como en la siguiente oración "caminando hacia mi casa me sorprendió la lluvia por lo que he llegado empapado". Ahora bien, los programadores han abordado estas complejidades del lenguaje mediante la extracción automática de las relaciones de causa y efecto, basándose principalmente en enfoques de aprendizaje automático supervisado y no supervisado esencialmente basado en la técnica de caja negra (Deep learning). Y a su vez, estos se basan en reglas lingüísticas.[68] Dentro de estas grandes categorías de enfoques se han propuesto técnicas que van desde la utilización de un verbo causativo como una muestra explícita de la existencia de una relación causal,[69] hasta arquitecturas de redes neuronales profundas informadas lingüísticamente.[70]

Volvamos a GPT-3. Este sistema de IA basado en redes neuronales y otras técnicas, utiliza un sistema de aprendizaje no supervisado, de modo tal que presenta una capacidad autoorganizativa que lo vuelve una caja muy negra. En este tipo de aprendizaje, no se establece una salida deseada. Y por eso, al encontrar patrones o características que sean significativas en los datos de entrada, no tiene con qué compararse el rendimiento del método.[71] De manera simplificada, si no hay un programador que se encargue de etiquetar los "porqués" a través de una técnica de aprendizaje supervisado, resulta difícil pensar que la máquina pude llegar a aprender que está frente a una relación causal por sí sola. Sin embargo, esto es posible. La fuerza de las correlaciones de patrones de información masivas, llevan a GPT-3 a detectar algunas causas, aunque por un camino distinto por el que atravesamos los humanos. En otras palabras, lo que se conoce como "razonamiento temporal", es un presupuesto básico y esencial para que los sistemas de IA puedan inferir una relación causal. Si no logramos que una máquina detecte falsedad al leer que Albert Einstein fue fotografiado ayer caminando por la plaza (disponiendo, por supuesto, de sus datos de fallecimiento), no podemos pretender mucho más. El gallo canta al amanecer cuando el Sol sale. Una IA que detecta causalidad, debería comprender que el *canto no causa* la salida. Ahora bien, un sistema como GPT-3 puede ser de gran ayuda, al igual que otros sistemas predictivos de caja negra, si queremos complementar nuestras tareas con correlaciones de oráculos artificiales.

---

68 DASGUPTA, Tirthankar. *Automatic Extraction*..., p. 306-316.
69 JAMSHIDI-NEJAD, Sepideh; AHMADI-ABKENARI, Fatameh; EBRAHIMI-ATANI, Reza. *Extraction*...
70 DASGUPTA, Tirthankar; SAHA, Rupsa; DEY LIPIKA, Naskar Abrir. *Automatic*...
71 CORVALAN, Juan Gustavo. *Inteligencia Artificial*...

Sin embargo, aunque en el ámbito del derecho advertimos muchos beneficios, también existe múltiples problemas y desafíos.

Primero. En GPT-3, no hay forma de saber de antemano qué formulaciones darán la respuesta correcta. Es decir, no hay forma de garantizar la explicabilidad en torno a cómo se construye el razonamiento subyacente, algo que es típico de este tipo de sistemas no supervisados que usan redes neuronales complejas. Por otra parte, aun si el sistema fuese 100% explicable, como se basa en correlaciones desvinculadas de la causalidad, no tendrían sentido esas explicaciones.[72] Según Marcus, el optimista nos dirá que como hay alguna formulación en la que GPT-3 logra una respuesta plausible, entonces conoce o razona como un humano, pero a veces se confunde con el lenguaje. Sin embargo, el problema no tiene que ver con la sintaxis de GPT-3 sino con su semántica. Aunque es capaz de generar palabras en un inglés perfecto, solo tiene el sentido más vago de lo que significan esas palabras, y ningún conocimiento en absoluto sobre cómo esas palabras se relacionan con el mundo. **Segundo.** Para entender el por qué, resulta útil pensar en lo que hacen los sistemas como GPT-3. No aprenden sobre el mundo sino sobre el texto y cómo la gente usa unas palabras en relación con otras. *Lo que hace es algo como un cortar y pegar masivo*, uniendo variaciones en el texto que ha visto, en vez de profundizar en los conceptos que subyacen a esos textos.

En el ejemplo del jugo de arándano, GPT-3 continúa con la frase "Ahora estás muerto" porque esa frase (o alguna parecida) a menudo sigue las frases como "...no puedes oler nada. Tienes mucha sed. Así que lo bebes". Un sistema que razone a partir de inferencias detectaría que es seguro mezclar el jugo de arándano con el jugo de uva. El ejemplo del jugo demuestra las robustas limitaciones que presentan los sistemas de IA hasta el día de hoy. Aun cuando GPT-3 usa miles de millones de parámetros más que cualquier otro sistema de procesamiento de lenguaje natural, su lógica correlacional le impide comprender mejor los problemas concretos y reales en un mundo plagado de contradicciones, paradojas y reglas de convivencia. **Tercero.** Cuando ingresamos al campo jurídico los problemas se multiplican. Por un lado, un sistema como GPT-3 es de caja "muy" negra. Los miles de millones de textos, sus correlaciones y cómo inciden los parámetros, impiden explicar o trazar sus resultados. Ni siquiera existen programadores que etiqueten la información. Además, el hecho de que no pueda captar las implicancias jurídicas de los fenómenos determina que podría generar injusticias o arbitrariedades palmarias si se lo usa para tomar decisiones.

[72] Esto también se presenta en los sistemas de caja blanca en el ámbito de la Justicia. Por ejemplo, las etiquetas vinculadas a sentencias y las técnicas simbólicas que usan quienes usan técnicas de aprendizaje automático, explican cómo el sistema correlaciona, pero no pueden dar ningún tipo de explicación en términos de causalidad ni tampoco permiten dar un sentido jurídico a esos patrones de información.

Correlacionar palabras, en cierto modo, se parece a la magia y al ilusionismo. Se pueden procesar patrones de información y lograr ciertas proezas, tomando un atajo. El modo de funcionar de GPT-3, no se basa en encontrar causas y razones correctas en un orden social y jurídico dado. Summers-Stay afirma que se parece a un humano que improvisa. Siempre dedicado a su oficio, nunca se sale del personaje y nunca se ha ido de casa. Solo lee sobre el mundo en los libros de texto. Cuando no sabe algo, simplemente fingirá que sí lo sabe a partir de predicciones basadas en estadística.[73] Esta metáfora nos provoca dos reflexiones. A veces, la improvisación basada en correlaciones de palabras, clics, geolocalización y me gusta, puede ser útil e incluso más eficiente para ciertos propósitos como predecir quienes comprarán pizza o qué series podrías ver. Incluso, estos sistemas de IA pueden acertar, en promedio, más que los humanos que saben mucho de causalidad. Ahora bien, todo esto no significa que esta forma de abordar los fenómenos en la sociedad sea válida o legítima en ciertos contextos importantes para la comunidad. En otras palabras, resulta difícil confiar en un mago experto en correlaciones e improvisación, cuando tratamos de darle sentido jurídico a los hechos que están regulados por normas jurídicas.

11 Correlaciones, sentido jurídico y causalidad

Existen dos planos que coexisten y se retroalimentan. Por un lado, las teorías de la causalidad que sirven para establecer vínculos entre dos hechos.[74] Por otro, darle *sentido jurídico* las correlaciones causales a través de métodos de argumentación, interpretación o ponderación. Al derecho le interesan las causas jurídicamente relevantes que causen resultados "captados" por normas jurídicas. Pegarle una bofetada a una persona hemofílica que causa su muerte determina una correlación causal. Sin embargo, no es imputable en términos penales. O se soluciona porque no hay dolo o porque una bofetada no es un medio adecuado para producir un resultado de muerte.[75] En una simplificación extrema, la IA puede

[73] Para Summers-Stay, no confiaríamos en los consejos médicos de un actor de improvisación que interpreta a un médico.

[74] El análisis de la causalidad puede descomponerse en dos. Por un lado, la causalidad general, según la cual la producción de un evento de cierto tipo hace más probable la producción de otro porque existe una ley de la naturaleza que da cuenta de esa probabilidad en un grado relevante. El segundo, es la causalidad individual, la que permite afirmar que, en una específica y concreta situación, un hecho particular ha sido causado por la ocurrencia de otro hecho específico y determinado. MICHELE, Taruff. *La prueba*..., FERRER, Beltrán. *Prueba y verdad en*..., p. 223-257.

[75] La acción puesta por el autor fue la causa de muerte de la víctima, aunque en un contexto normal esto no hubiera sido relevante para causar la muerte. Por ende, el correctivo no es sobre la causalidad, sino antes bien sobre qué significación jurídica tiene dar una bofetada si lo que se imputa es el tipo penal de homicidio. Como afirma María Elena Lumiento, esto es un problema de dolo y no de tipo causal. Básicamente, la circunstancia de que la víctima fuera hemofílica estaba fuera del conocimiento causal del autor. Así, no puede obviarse que el dominio causal de la acción también forma parte de lo que el autor

detectar la correlación entre bofetada y muerte, pero resulta mucho más difícil entrenarla para que pueda detectar (no entender ni razonar en términos jurídicos humanos) y predecir[76] las inferencias causales con relevancia jurídica, para una amplísima y vasta gama de posibilidades que se dan en cada área del derecho.

Por ejemplo, supongamos que diseñamos y entrenamos el modelo predictivo con 400 sentencias emitidas por diez juezas y jueces para que una IA como Prometea o PretorIA correlacione datos y patrones de información en sentencias. En estas sentencias, con matices, se resolvieron casos de personas hemofílicas que han sido abofeteadas. La máquina podría aprender a detectar con una alta tasa de acierto, la inexistencia de dolo o la interrupción del nexo causal, según la teoría que se aplique para resolver estos casos. Sin embargo, en este proceso no ejecuta técnicas de argumentación vinculadas a la teoría del delito. La IA, no sabe nada de derecho. Reconoce patrones y correlaciones de palabras, frases o símbolos, para luego agruparlos en función de criterios estadísticos o según un índice de pesos o reglas de inferencia, que no son jurídicas, aunque luego pueden ser útiles para realizar argumentaciones racionales.

Una vez realizada esta tarea, supongamos que un juez elabora un proyecto de sentencia, y en el caso se trata de una persona que ha golpeado con el puño a otra hemofílica y causó su muerte. Además, el agresor es profesional de las artes marciales. Si el sistema de IA detecta que *no* se trata de un caso de inexistencia de dolo o de interrupción de nexo causal, lo hace porque *no encontró* las correlaciones de palabras que estaban presentes en el historial de datos de los otros casos relacionados con las abofetadas. No es que la IA atraviesa por un proceso interno en el que dice: "ah, entiendo, en este caso sí podría existir dolo o bien, resultaría al menos dudoso que se interrumpa el nexo causal porque estos eventos modifican la lógica dogmática y normativa aplicable".

---

debe conocer para serle imputado un resultado lesivo. Distinto es el caso de la causalidad alternativa. Aquí si hay un problema de causalidad dado que no se puede establecer qué acción causó el resultado. Además, esto suele ir unido a que, desde el punto de vista probatorio, hay una imposibilidad de producir una evidencia que permita sostener qué acción fue la que desencadenó el resultado lesivo. La doctrina mayoritaria entiende que corresponde imputar a todos los intervinientes la consumación del resultado, y lo resuelve a través de la teoría de la csqn, suprimiendo in mente ambas acciones. Lo más justo, no obstante, sería imputar a título de tentativa, básicamente, por falta de prueba sobre la existencia de la relación causal de cada una de las acciones. Ver LUMIENTO, María Elena. *Algunas*...

[76] La forma en la que usamos el término de "predicción" en este artículo es idéntica al que usa la Comisión Europea cuando habla de "previsión". Para este organismo, la predicción es el acto de anunciar lo que sucederá (antes de dictar) antes de los eventos futuros (por inspiración sobrenatural, por clarividencia o premonición). La previsión, por el contrario, es el resultado de observar (apuntar, ver) un conjunto de datos para prever una situación futura. Según la Comisión Europea, este abuso del lenguaje y su difusión parece explicarse por una transferencia del término de las ciencias "duras", donde se refiere a una variedad de técnicas de ciencia de datos derivadas de las matemáticas, estadística y teoría de juegos que analizan hechos presentes y pasados para formular hipótesis sobre el contenido de eventos futuros. Ver: EUROPEAN COMMISSION...

Sin embargo, aunque la IA no sepa derecho, la tarea de correlacionar patrones de información históricos jurisprudenciales puede ser de gran ayuda.[77] Por ejemplo: **i)** es útil para ajustar o mejorar los análisis acerca de las causas relevantes jurídicamente que causen resultados "captados" por normas jurídicas y aplicadas por humanos. Esto, lo hemos podido comprobar a partir de desarrollar y aplicar un predictivo que puede detectar o no la interrupción del nexo causal, en el ámbito de los casos judiciales vinculados a los accidentes de tránsito; **ii)** también contribuye a corregir, prevenir o mitigar sesgos o patrones de discriminación en los datos.[78] En el ámbito del lenguaje natural, un sistema de IA puede detectar hipótesis fácticas similares, para indicar que correspondería la solución específica que se adoptó cuando aquellas se verificaron en el pasado. También podría predecir la existencia de ciertos criterios jurídicos que están presentes en el caso que se examina (siempre en relación con ejemplos resueltos en el pasado). En ambos casos, si se realizan con técnicas de caja blanca, es posible trazar, explicar y transparentar los datos y la forma de procesarlos para arribar a los resultados tal como lo hacemos desde el IALAB de la UBA. Además, las personas que lo diseñan y entrenan, nunca pierden el control. En esta clase de IA, no existe el déficit estructural asociado a la falta de explicabilidad intrínseca que se presenta en las redes neuronales y en otros sistemas opacos.

Por último, cuando se trata de predicciones de caja blanca, estas contribuyen a garantizar el principio de no discriminación algorítmica. Por un lado, nos ayudan a realizar un ejercicio retrospectivo acerca de posibles correlaciones entre decisiones. Por otro, es posible corregir, mitigar o eliminar esquemas decisionales o bien, "curar" o "limpiar" en los datos de entrenamiento, los sesgos negativos o las distinciones basadas en motivos de raza, color, sexo, idioma, religión, opinión política o de otra

[77] El año pasado, Francia emitió una norma muy discutible para nuestro sistema jurídico. Aunque no prohíbe usar técnicas de IA sobre fallos judiciales, sí establece la prohibición y criminalización para usar técnicas (no aclara cuáles) para evaluar, analizar, comparar o predecir las prácticas de un juez, basándose en comparar su identidad y de qué tribunal es miembro. La prohibición francesa, en nuestro sistema, sería manifiestamente inconstitucional y contraria a tratados internacionales (principio de publicidad, transparencia, entre otros). Su aplicación literal, llevaría al absurdo de que se pueda criminalizar con una pena de hasta 5 años, a quien "evalúe", "analice" y "compare" (usando palabras textuales de la ley) en un Excel los fallos judiciales vinculados a la identidad del juez/a y respecto de qué tribunal integra. Para acceder a la ley consulta a https://bit.ly/2MY0fx8. Ver especialmente el artículo 33 que reforma el artículo L. 153-1 y el L. 10.

[78] En el ámbito de la medicina, las predicciones de IA se basan en anticipar un resultado a partir de lo que se conoce "una intervención". Por ejemplo, un tratamiento asignado por un médico que cambiará la condición cardíaca de un paciente es una intervención. Predecir el cambio en la condición del paciente es una tarea de inferencia causal. En general, una intervención es una acción realizada por un agente externo que cambia los valores originales o las distribuciones de probabilidad, de algunas de las variables en el sistema. Además de predecir los resultados de las acciones, la inferencia causal también tiene que ver con la explicación: identificar cuáles fueron las causas de un evento particular que sucedió en el pasado" Véase SILVA, R. *Casuality*..., p. 1.

índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social.[79]

12 Predicciones de IA en el derecho

Dworkin advirtió lúcidamente acerca de dos cuestiones muy importantes en el ámbito del derecho. Por un lado, la primitiva distinción entre lo que puede suceder y lo que sucederá. Por otro, que las normas y el derecho no puede resolver todos los casos que se presenten[80] (por ejemplo, los casos difíciles). En el ámbito de la IA sucede un fenómeno similar. Los sistemas inteligentes aprenden de los datos y la información sobre problemas o cuestiones que han sucedido. Y como en general no pueden prever eventos inéditos o casos difíciles aislados, también existe una divergencia entre lo que puede suceder y lo que efectivamente acontece.

A este fenómeno lo llamamos tasas de acierto dinámicas, que varían en función de múltiples variables, incluyendo los ajustes que los programadores realicen sobre el sistema. Ahora bien, al igual que acontece con la tarea predictiva humana, las predicciones de IA modifican o podrían alterar el curso de acción de los sucesos. Y esto también condiciona las tasas de acierto, que a su vez pueden ser modificadas porque podrían ser previstas por otras personas. En un ejemplo hipotético, si un grupo delictivo cuenta con sistemas sofisticados de IA y puede entrenarlos con una base de datos similar a la que usan las autoridades policiales, podría usar esas predicciones para adoptar otros cursos de acción y de ese modo disminuir las tasas de acierto de la predicción policial.[81] Esta lógica también es aplicable a otras áreas del derecho. Las IA predictivas actualizan las cuestiones vinculadas con el uso del precedente y la argumentación jurídica que lúcidamente ha tratado Robert Alexy.[82] Como luego veremos al tratar las predicciones con caja blanca, la lógica para entrenar a una IA presenta similitudes al *Judicial precedent*, una de las principales fuentes del *Common Law*. Por ejemplo, a partir de cientos de casos o decisiones jurisprudenciales asociados a una temática, se puede entrenar a una IA para que cuando lea un nuevo caso, pueda sugerir una forma de resolverlo.[83]

---

79 CORVALÁN, Juan Gustavo. *Inteligencia Art...*,

80 *Los derechos en serio*, p. 13; 291, Ariel, 1984, Barcelona,

81 Desde ya que este ejemplo, se plantea en términos muy rudimentarios, ya que resulta difícil que dos sistemas de IA, reflejan las mismas tasas de acierto en actividades o cuestiones en donde existen múltiples asimetrías y matices en los datos históricos.

82 De manera muy sintética, véase: ALEXY, Robert. *Teoría de la argumentación jurídica*... p. 381-383.

83 En estos supuestos, se trata de una tarea predictiva basada en pequeñas cantidades de datos (Small data).

Esto, en parte, explica por qué nos cuesta familiarizarnos con este tipo de tecnología que *no razona jurídicamente.*[84] En el ámbito del procesamiento automático de lenguaje natural (NPL), compara palabras, frases o conjunto de palabras y frases[85] para establecer correlaciones simbólicas, de modo tal que las personas que diseñan y entrenan al sistema, lo ajusten para que la tasa de acierto sea alta, en función de los resultados que se desean obtener. Aquí surge una clasificación que ha sido tratada desde hace varias décadas en el derecho anglosajón. La posibilidad del *distinguishing* y del *overruling*. En palabras, simples, el *distinguishing* se vincula con la distinción entre supuestos de hecho que sucedieron en el pasado, pero que no se verifican en el caso en examen. En el *overruling* se rechaza el precedente. Aunque en ambos hay que desplegar razones jurídicas, se suelen utilizar argumentos prácticos (lo que en nuestro derecho podría ser razones de economía procesal). Entrenar a una IA con la historia de ciertos casos, suele ser una tarea similar a la que se presenta con el *distinguishing,* aunque la forma de ejecutarla es sustancialmente diferente.

13 Sesgos, motivación y fundamentación de las decisiones jurídicas frente a la IA

Los sesgos, de por sí, no tienen una valoración *a priori*. Ser humano implica sesgar, en el sentido de que nuestro cerebro no es capaz de comprender y procesar toda la información que rodea a los fenómenos y actividades en las que nos desarrollamos. En el ámbito jurídico, es clave mitigar los efectos de los sesgos que afectan negativamente las decisiones judiciales o las declaraciones de voluntad en el sector público. En otras palabras, la problemática de los sesgos[86] se relaciona

[84] Esto se debe, entre otras razones más obvias, a que por ahora sólo las personas pueden razonar a través de lógicas monotónicas. Se entiende por no monotónico a todo aquel sistema de razonamiento que carezca de la propiedad de aditividad o monotonía. Cualquier sistema de razonamiento que utilice reglas ampliativas de inferencia tiene necesariamente la propiedad de ser no monotónico. Es decir, esta propiedad no surge solamente por el uso de reglas o condicionales derrotables, sino también por el uso de otras reglas o patrones de inferencia (por ejemplo, inducción, abducción, analogía, probabilidades, etc.). Ver: DELRIEUX, Claudio. *Inferencia*... Véase: LEGRIS, Javier. *Razonamiento*...

[85] En general, la gran mayoría de cuestiones en el ámbito del derecho, se vinculan con los sistemas de IA basados en el reconocimiento de lenguaje natural (NPL). Sin embargo, en el ámbito de la cibercriminalidad, aquí entran en juego otros tipos de sistemas asociados al reconocimiento de imagen y al procesamiento de grandes masas de datos, que usualmente provienen del uso masivo de plataformas digitales, redes sociales y también en el ámbito de la Deep web.

[86] COMISIÓN EUROPEA. *Generar confianza*..., p. 6. ONU. *La Resolución Nº 35/9*..., Si los mecanismos cuentan con un sesgo obtenido, sea de los datos, sea del diseño de su función de éxito, el resultado será una amplificación de la discriminación que experimentan los miembros más vulnerables de nuestra sociedad. Ver AMUNÁTEGUI, Carlos Perelló. *Sesgo*..., El principio de no discriminación y la necesidad de prevenir específicamente el desarrollo o la intensificación de cualquier discriminación entre individuos y grupos de individuos ha sido resaltada en: CONSEJO DE EUROPA. *Carta ética*..., La necesidad de evitar sesgos injustos ha sido destacada en Grupo independiente de expertos de alto nivel sobre IA. La equidad y no discriminación han sido reconocidos como principios en: El principio de transparencia de los sistemas de IA ha sido reconocido también en COMISIÓN EUROPEA. *Libro Blanco sobre...,* También se ha sostenido que es necesario depurar conjuntos de datos para eliminar datos discriminatorios y tomar medidas para

con la motivación, fundamentación y racionalidad (o irracionalidad) de las decisiones jurídicas.[87] Y el desarrollo y aplicación de los sistemas de IA introducen o actualizan algunos problemas y desafíos.

Primero. A veces resulta muy complejo detectar sesgos y, a la vez, obtener tasas de acierto. Es decir, si ciertos sesgos son parte de la historia y es posible que hayan sido determinantes a la hora de adoptar las decisiones, entonces, el sistema de IA no podrá acertar en torno a los datos de entrenamiento sin reproducirlos. Esta es una entre tantas paradojas de la IA predictiva. Reducir y mitigar sesgos,[88] a veces puede tornar obsoleto el proceso de entrenamiento, ya que se está modificando el pasado con el cual se iba a predecir el futuro. Incluso, cuando un funcionario entrena una IA y detecta sesgos en sus propias decisiones pasadas, probablemente modifique su forma de decidir hacia el futuro. Al hacerlo, ya no hay pasado con el que entrenar a la IA.

Segundo. Lo anterior nos lleva a resaltar a la IA como una herramienta que permite hacer un ejercicio muy útil de retrospectiva, bajo un enfoque transdiciplinario, en el que la máquina nos ayuda o nos revela correlaciones que habíamos soslayado o que no había forma de detectar. **Tercero.** Cuando uno decide en un caso, muchas veces sienta su posición de cara a supuestos análogos. En esta lógica decisional, un sistema predictivo permite detectar el caso análogo con mayor precisión y velocidad. Si se usa la IA de modo responsable, esto puede contribuir a robustecer la fundamentación y la motivación a partir de mejorar o incorporar argumentos. Luego, bajo un ecosistema laboral adecuado, las personas pueden redirigir su enfoque a mejorar la racionalidad de la decisión. **Cuarto.** Cuando se trata de sistemas de caja negra y no se aplican medidas para mitigar sesgos negativos o discriminatorios, es muy probable que la IA los amplifique.

Quinto. Los sesgos condicionan los análisis de causalidad que sirven para establecer vínculos entre dos hechos. Y también afectan la tarea de darle *sentido jurídico* a las correlaciones causales a través de métodos de argumentación, interpretación o ponderación. Todo ello, condiciona la justificación judicial de los diversos enunciados de naturaleza fáctica y jurídica contenidos en el cuerpo de las resoluciones judiciales.[89] **Sexto.** En el año 2017 publicamos acerca de la relación entre la IA y los derechos humanos en el estudio que publica anualmente el

compensar los datos que 'contienen la impronta de pautas históricas y estructurales de discriminación' y de los cuales los sistemas de inteligencia artificial tienden a derivar representantes discriminatorios. ONU. *La Resolución Nº 73/348*...

87 Sobre estas cuestiones, ampliar en PASTOR, Daniel; HAISSINER, Martin. *Neurociencias*..., p. 37; ACIARRI, Hugo. *Derecho*...

88 Sobre la relación entre cajas negras y sesgos ver: PERELLÓ, Carlos. *Amunátegui*..., p. 62-71.

89 ALISTE SANTOS, Tomás-Javier. *La motivación*..., p. 449.

Consejo de Estado Francés.[90] Aquí postulamos la importancia central del principio de transparencia algorítmica para asegurar la efectividad de los derechos. De este principio, se deriva otro: la explicabilidad. Resulta esencial robustecer los ejercicios de explicabilidad asociados al desarrollo de los sistemas predictivos de IA. Aquí también surge el concepto de interpretabilidad. En el ámbito de la IA, se asocia a la interpretabilidad con la capacidad de observar bidireccionalmente en un sistema situaciones de causa y efecto. Esto implica tanto entender las razones por las cuales se ha realizado una predicción concreta, como predecir lo que sucederá dado un cambio en la entrada o en los parámetros algorítmicos. La explicabilidad, que se asemeja a la motivación en el derecho público, se vincula con un concepto más amplio que describe la capacidad de entender, en términos humanos, el funcionamiento de un modelo considerando sus entradas y salidas.[91]

14 Aprendizaje automático y cajas blancas. Experiencia IALAB predictiva y casos éxito en la Justicia

Dentro de esta IA débil, blanda, estrecha o restringida, hay otro "mundo" de sistemas de IA que son de "caja blanca" y se basan en un conjunto de técnicas[92] que se utilizan para obtener predicciones, automatizaciones, clasificaciones o detecciones inteligentes. Gracias a las cajas blancas, los resultados a los que se arriba son auditables, trazables, explicables e interpretables, y ello resulta muy beneficioso para comprender la dinámica del tratamiento automatizado cuando se usan estas técnicas. Y esto redunda en enormes beneficios para el campo jurídico, optimizar la tarea judicial y la transformación digital de las organizaciones. En este ecosistema de *Machine Learning* o aprendizaje automático de caja blanca, existen dos grandes técnicas que pueden usarse para realizar predicciones. Nos referimos a las técnicas de "Regresión y "Clasificación" y, en esta última, se encuentra una subespecie llamada "*Topic Model*".[93] Estas técnicas son categorizaciones de algoritmos supervisados de aprendizaje automático que se obtienen mediante la diferenciación con respecto al tipo cuantitativo o cualitativo de la variable de salida involucrada en el problema. Es decir, si la salida de un problema es cuantitativa

[90] Contribución de un autor Latinoamericano en estas publicaciones francesas. Véase: CORVALÁN, Juan Gustavo. *L'algorithme*..., p.179.

[91] Se afirma que existe una relación inversa entre explicabilidad e interpretabilidad. Véase: CABROL, Marcelo. *Adopción*..., p. 25.

[92] CORVALÁN, Juan Gustavo. *Perfiles*...; *Prometea*...

[93] El Topic Model es una herramienta estadística utilizada en machine learning y en aplicaciones de lenguaje natural que permite identificar temáticas en grupos de documentos de textos. El uso de topic models permite aplicar el análisis de clustering a conjuntos de datos no estructurados superando algunas de las limitaciones que presenta la herramienta de K-medias. Lo que hacen los topic model es tomar un texto no estructurado y aplicarlo a una dimensión más estructurada. Ver MIT. *Programa en línea*...

o cualitativa, nos referimos al problema como regresión o clasificación. Regresión significa que el resultado (variable dependiente) es cuantitativo y clasificación significa que el resultado es cualitativo. No importa si la entrada (variables independientes) es cuantitativa o cualitativa. Las técnicas que se usan tienen que ver con las salidas.[94] Por ejemplo, la regresión se usa para predecir la evolución de los precios de las propiedades en un determinado territorio.

En cambio, la clasificación en el ámbito del aprendizaje de máquina puede usarse en el ámbito del derecho para tratar de establecer correlaciones entre palabras o frases, y correlacionarlas con decisiones e hipótesis fácticas que están presentes en decisiones judiciales. A partir de esta técnica, en el año 2017 se creó el primer sistema predictivo del mundo incubado y desarrollado íntegramente en el sector público. El desarrollo de Prometea en el ámbito del Ministerio Público Fiscal de la CABA, fue un avance disruptivo en donde se logró predecir y automatizar (sin intervención humana) dictamenes legales con una tasa de acierto superior al 96%, sobre ciertos casos judiciales en los que estaba en juego el derecho a la vivienda y otras materias en el ámbito del contencioso administrativo y tributario de la Ciudad Autónoma de Buenos Aires.[95] A partir de la experiencia Prometea se creó el primer Laboratorio de Innovación e IA en una Facultad de Derecho (UBA IALAB) en Hispanoamérica. Desde el IALAB, se ha profundizado el desarrollo de soluciones predictivas. Por ejemplo, en el ámbito de los juicios de responsabilidad civil, se han combinado técnicas para detectar en segundos, si el vínculo causal entre un evento y el daño sostenido asociado a un caso de accidente de tránsito se ha fracturado,[96] con una tasa de acierto de más de un 95%.

El proceso de entrenamiento para su desarrollo constó de tres etapas: en la primera se utilizó un data set de sentencias que fue provisto por la jueza Gabriela Gil y por el secretario de Cámara Hernán Quadri. A partir de ahí, se elaboró un primer etiquetado manual, para que luego comience el proceso de aprendizaje de máquina de caja blanca "Clasificación". En esta primera etapa, se realizó un análisis jurídico para extraer patrones comunes, que permitió identificar en las sentencias la existencia o interrupción del nexo causal en cada caso en particular. Luego, se segmentó este grupo de sentencias en grupos y subgrupos, según las distintas hipótesis fácticas. Estos grupos y subgrupos fueron asociados a una determinada solución jurídica: la posibilidad de atribuir o no la responsabilidad por los daños a la parte demandada. La segunda etapa implicó transformar lo analizado a un lenguaje de programación. Por ejemplo, para la máquina es irrelevante muchas palabras que se usan para conectar frases ("y", "o", "que", entre muchas otras).

94 Véase: LINDHOLM, A. et al. *Supervised Machine…*, p. 8.
95 CORVALÁN, Juan Gustavo. *Inteligencia artificial…*
96 CORVALÁN, Juan Gustavo et al. *Inteligencia artificial…*

Una vez finalizado el diseño de programación, se realizó la primera prueba predictiva sobre el universo total de sentencias. La tasa de acierto alcanzada fue del 81,4%, lo que obligó a analizar humanamente otro data set de sentencias, para intentar mejorar esa tasa. De esta forma, se inició la tercera y última etapa: el refinamiento de los patrones de información jurídica y de toda técnica que mejore el rendimiento predictivo. Es un trabajo conjunto entre los operadores jurídicos especializados y los expertos en IA, a efectos de lograr sinergia entre contenido legal y poder de máquina. La mejora fue impactante, la tasa de acierto alcanzó el 96,5 % (83 aciertos sobre un total de 86 casos). Este desarrollo actualmente se utiliza en un Juzgado Civil de la Provincia de Buenos Aires. Gabriela Gil, la jueza con la que entremos el sistema basado en la experiencia Prometea, lo usa bajo un enfoque de "control" de proyectos que se realizan en su juzgado. Verifica si la sentencia proyectada coincide con el resultado del predictivo que se entrenó con más de 400 sentencias de la Cámara Civil de Morón en la Provincia de Buenos Aires. La tarea de control dura segundos, ya que se ingresa al agente conversacional la sentencia y en pocos segundos se elabora un informe de predicción. El sistema es autoexplicable ya que ofrece al usuario los métodos, las tasas de acierto y los data sets utilizados. Como veremos con mayor detalle al analizar el sistema PretorIA, esta lógica de diseño y entrenamiento es plenamente aplicable a múltiples ramas y áreas del derecho, más allá de la Justicia.

15 Conclusion: Small Data vs. Big Data. El caso PretorIA: Enfoque holístico, explicable y transdisciplinario

En el ámbito de los procesos judiciales, resulta difícil hablar de BIG DATA. Por lo general, los datos con los que se cuenta para desarrollar una tarea predictiva están bajo lo que se conoce como SMALL DATA. Por tanto, las redes neuronales no resultan muy útiles para lograr tasas de acierto razonables. Bajo un enfoque de predicción de caja blanca y entornos de Small Data, incubamos y desarrollamos PretorIA. Se trata un sistema predictivo inédito que integra soluciones de automatización, para ayudar en el proceso de selección de acciones de tutela decididas por más de 5.400 jueces y juezas y que son remitidas a la Corte Constitucional de Colombia para su revisión (aún de oficio).

La predicción de caja blanca se basa en la subespecie TOPIC MODEL, dentro del género "Clasificación". Recordemos que esta técnica se basa en análisis estadísticos, que permiten identificar temáticas y subtemáticas en grupos de documentos de textos. En una explicación desprovista de todo tecnicismo informático, se trata de una rama de la IA que es muy útil para el procesamiento del lenguaje natural especializado. Mientras que las personas arman los data sets para etiquetar y

clasificar la información, quienes programan a partir de la técnica de *Topic Model*, intentan buscar correlaciones entre palabras, frases o conjuntos de palabras o frases, a partir de "agregar" o "aumentar" el texto original, a través de introducir símbolos o letras. También se emplean "atajos" para descartar puntos, comas y otras palabras o símbolos, y de esa forma detectar otras posibles correlaciones asociadas a decisiones, criterios, hipótesis fácticas u categorías. Estas correlaciones son sometidas a iteraciones a partir de usar data sets de entrenamiento, y luego se van refinando las palabras clave (*Keywords*) hasta alcanzar tasas de acierto deseadas o razonables (más de un 80%) (ampliar en Instructivo). Por ejemplo, en el caso de PretorIA se seleccionaron una muestra de 2500 sentencias, Se elaboraron 7 *Datasets*, a partir de un conjunto de sentencias aleatorias y sin previa clasificación.

Los primeros tres *Datasets* de sentencias sirvieron para entrenar y alimentar a PretorIA. Con el primer *Dataset* analizado, se realizó el primer input para el predictivo. Ahora bien, cuando se comenzó a trabajar en la gobernanza de datos, se completó la base de datos y se la envió a la Corte Constitucional a fin de que realicen los controles pertinentes. Así, se pudo observar que hubo discordancias entre lo que ellos consideraban al enumerar los criterios y lo que el equipo de gobernanza pudo interpretar en relación con su alcance. Por este motivo fue necesario que profundicen en ciertas definiciones. Hasta aquí, téngase en cuenta que definir y precisar los criterios a seleccionar en las sentencias llevó más de 40 días de trabajo en conjunto entre el equipo de trabajo del IALAB, abogados y abogadas de la Universidad del Rosario y un grupo de delegados de todos los magistrados y magistradas de la Corte Constitucional de Colombia. Este proceso es crítico y definirá, en gran medida, la legitimidad y la utilidad del sistema. Por ejemplo, una de las cuestiones en las que se presentaba discordancia interpretativa, se vinculaba con los casos en que los jueces de instancias previas habían resuelto que había: i) Ausencia de examen médico; ii) ausencia de diagnóstico médico; y iii) ausencia de procedimiento médico. Y estas sutilezas, en ciertos casos, eran relevantes. Ahora bien, el tiempo promedio en horas de lectura, doble control y detección de *keywords* en cada sentencia fue de 36 minutos en el caso del proyecto PretorIA. Algo similar ocurre con la experiencia Prometea en otros proyectos. Adviértase que se trata de: 1) una primera lectura, análisis de las sentencias e identificación de criterios que conlleva 16 minutos en promedio; 2) el primer control sobre los criterios identificados en cada sentencia fue de 16 minutos; 3) el segundo control que recae sobre el primero fue de 9 minutos en promedio.

Estas tareas, están presentes en casi todos los ámbitos del sector público, porque muchas personas, empleados o funcionarios deben correlacionar, detectar y clasificar datos, criterios, hipótesis fácticas y diversas cuestiones para que luego se puedan tomar decisiones en el marco de procesos judiciales. Un sistema bien

entrenado como PretorIA, logra detectar 33 criterios y automatizar la generación de resúmenes, en pocos segundos y sobre miles de sentencias. Luego de la identificación y el control de las *keywords* identificadas, se entregaron al equipo especialista en IA que comenzó con el entrenamiento. Como las bases de datos suelen presentan ciertas complejidades, hay que adoptar ciertos recaudos básicos. Por ejemplo, cada vez que se entregó base de datos a los programadores/as, fue necesario que un integrante del equipo de gobernanza se ocupe de normalizarla. Esto significa que todos los valores que se completan en cada celda se encuentren iguales en todas las bases. Luego comenzamos con el proceso de *machine learning* de caja blanca (Clasificación), en donde se utilizan técnicas y atajos simbólicos (puntos, paréntesis, corchetes, entre otros), para detectar correlaciones entre palabras, frases, conjuntos de palabras o frases (*keywords*) que puedan asociarse a los resultados que se desean obtener. A partir de ahí comienza un proceso de iteración dinámico y constante, entre el equipo experto en derecho, los especialistas en datos y los programadores. El objetivo aquí es controlar, testear y refinar el proceso hasta alcanzar tasas de acierto que deben ser, sin excepción, validadas por quien será el usuario final del sistema. En el caso de los órganos judiciales o policiales, los funcionarios competentes.

En todos los casos, los data sets, las palabras clave y cualquier otra segmentación realizada por la máquina o las personas, se puso a disposición de la Corte Constitucional para que se pueda garantizar la trazabilidad, auditabilidad, explicabilidad e interpretabilidad del sistema. En los desarrollos del IALAB y en el caso de PretorIA, todos los *data set* y todo lo que el programador realiza, es 100% trazable y explicable. Incluso, durante todo el proceso de prueba y entrenamiento, nos enfocamos en reducir la existencia de falsos positivos y de falsos negativos. Por ejemplo, los programadores de PretorIA comunicaron las sugerencias de palabras claves y los analistas expertos se encargaron de validarlas, intentar mejorarlas o descartarlas. El proceso para llegar a las *keywords* adecuadas, se logró a través de prueba y error. En el refinamiento, los equipos de trabajo se retroalimentaron. Los patrones que detectaron la/los programadores, no necesariamente coinciden con las personas que etiquetaron la base de datos (las sentencias).

Como la tarea de clasificación y detección es un proceso iterativo y transdisciplinario de prueba y error, es muy importante atravesarlo para refinar el sistema predictivo. Todas las fases de desarrollo de PretorIA son explicables y se encuentran documentadas. En particular: **1)** los motivos que llevaron a la aplicación de IA en las tareas de selección de casos prioritarios en la Corte Constitucional de Colombia; **2)** Las discusiones y toma decisión en torno al tipo de técnica de IA que correspondía desarrollar: Clasificación y detección inteligente; **3)** Se ha prestado particular atención en la selección de la muestra para la gobernanza de datos y

el entrenamiento y prueba del sistema, así como la formación de datasets; **4)** El sistema es completamente explicable en cuanto a la selección de los criterios por parte del equipo de trabajo de la Corte Constitucional de Colombia; **5)** La interpretación de los criterios y su alcance brindados por la Corte Constitucional de Colombia se encuentran documentados; **6)** Los cambios realizados en los criterios se encuentran registrados. También se encuentran registrados los criterios incorporados con posterioridad y las razones que llevaron a la decisión; **7)** Las bases de datos utilizadas y completadas por el equipo de gobernanza de datos en cada una de las etapas se encuentran disponibles desde la primera versión y pueden ser consultadas; **8)** Se encuentran registradas las reuniones entre los equipos de gobernanza y programación con la Corte Constitucional de Colombia y las solicitudes y observaciones realizadas, así como los trabajos de refinamiento de *keywords*; **9)** Es posible conocer asimismo la evolución de las tasas de acierto obtenidas en cada criterio y el trabajo realizado entre los equipos de gobernanza de datos y programación a fin de modificarlas. Se ha explicitado el porcentaje de acierto que logró el sistema en cada etapa del entrenamiento y los ajustes necesarios para lograr tasas mayores al 90%; **10)** Se encuentran identificadas las personas que intervinieron en cada una de las etapas del ciclo de vida del sistema y la Corte Constitucional de Colombia, es la responsable de la aplicación del sistema y de llevar adelante las auditorías correspondientes.

Además, es muy importante la selección de los conjuntos de datos (en este caso sentencias) para analizar y detectar prejuicios, sesgos discriminatorios u otras cuestiones que no es deseable que sean aprendidas y luego reproducidas o amplificadas por un sistema de IA. De ahí la importancia de que los funcionarios competentes estén "siempre en control". Esta tarea de entrenamiento, a fin de cuentas, modula y condiciona el ejercicio de competencias humanas complementadas con oráculos algorítmicos y asistencia digital. Esto, puedo afirmarlo desde una triple perspectiva. 1) He entrenado un sistema de IA de caja blanca, a partir de analizar el historial de dictámenes judiciales que he suscripto; 2) he participado activamente en todo el proceso de gobernanza de datos (etiquetado, clasificación de género a especie, armado de palabras clave, entre otros); 3) junto a los programadores, he participado activamente en el proceso de aprendizaje automático (lo que incluye la revisión y el control de los resultados).

Referencias

ACIARRI, Hugo. *Derecho, economía y ciencias del comportamiento*. Editorial: Ediciones Saij, 2018.

ALEXY, Robert. *Teoría de la argumentación jurídica*. Palestra: Lima, 2010.

ALISTE SANTOS, Tomás-Javier. *La motivación de las resoluciones judiciales*. Marcial Pons, Buenos Aires, 2011.

AMUNÁTEGUI, Carlos Perelló. *Sesgo e inferencia en redes neuronales ante el derecho*, 2020. Disponible en: https://guia.ai/wp-content/uploads/2020/05/Amunategui-Madrid-Sesgo-e-Inferencia-en-Redes-Neuronales-ante-el-Derecho.pdf.

ARANTXA, Herranz. *Tres expertos en inteligencia artificial sobre GPT-3*: avanzando más que nunca a pasos agigantados, 22 jan. 2020. Disponible en: https://www.xataka.com/robotica-e-ia/que-tres-expertos-que-trabajan-inteligencia-artificial-opinan-gpt-3.

ARAOZ, Manuel. *El GPT-3 puede ser lo más importante que vimos desde el bitcoin*. Disponible en: https://maraoz.com/2020/07/18/openai-gpt3/

BARRET Zoph; VIJAY, Vasudevan; JONATHON, Shlens; QUOC, Le. *AutoML for large scale image classification and object detection*. Disponible en: https://ai.googleblog.com/2017/11/automl-for-large-scale-image.html

BELLOCHIO, Lucía. Access to public information in Argentina with particular reference to personal and institutional data protection. *A&C - Revista de Direito Administrativo & Constitucional*, Belo Horizonte, a.16, n. 65, p. 39-51, jul./set. 2016. doi: 10.21056/aec.v16i65.261.

BELLOCHIO, Lucía. Big Data in the Public Sector. *A&C - Revista de Direito Administrativo & Constitucional*, Belo Horizonte, a. 18, n. 72, p. 13-29, abr./jun. 2018. doi: 10.21056/aec.v18i72.967.

BOSTROM, Nick. *Superinteligencia caminos, peligros, estrategias*. SL TEEEL Editorial, España, 2016.

CABROL, Marcelo, et al. *Adopción ética y responsable de la Inteligencia Artificial en América Latina y el Caribe*, Fair LAC, BID. Disponible: https://publications.iadb.org/publications/spanish/document/fAIr_LAC_Adopci%C3%B3n_%C3%A9tica_y_responsable_de_la_inteligencia_artificial_en_Am%C3%A9rica_Latina_y_el_Caribe_es.pdf

CASTELLANO, Francisco Javier García. *El ciclo de vida de un sistema de información*. Disponible en: http://flanagan.ugr.es/docencia/2005-2006/2/apuntes/ciclovida.pdf

CEPAL. *Datos, algoritmos y políticas, La redefinición del mundo digital*. Naciones Unidas, 2018. Disponible en: https://repositorio.cepal.org/bitstream/handle/11362/43477/7/S1800053_es.pdf

CEPAL. *Datos, algoritmos y políticas: la redefinición del mundo digital*. LC/CMSI.6/4, Santiago de Chile, 2018, Disponible en https://www.cepal.org/es/publicaciones/43477-datos-algoritmos-politicas-la-redefinicion-mundo-digital.

COMISIÓN EUROPEA. *Generar confianza en la Inteligencia Artificial centrada en el ser humano*. Bruselas, 8 Apr. 2019. Disponible en: https://ec.europa.eu/transparency/regdoc/rep/1/2019/ES/COM-2019-168-F1-ES-MAIN-PART-1.PDF.

CONSEJO DE EUROPA. *Carta ética europea sobre el uso de inteligencia artificial en los sistemas de justicia y su entorno*. 4 dic. 2018. Disponible en: https://campusialab.com.ar/wp-content/uploads/2020/07/Carta-e%CC%81tica-europea-sobre-el-uso-de-la-IA-en-los-sistemas-judiciales-.pdf.

CORVALÁN, Juan G., GALETTA, Diana U. Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto, federalismi.it. *Rivista di Diritto Pubblico Italiano Comparato Europeo*, 6 Feb. 2019.

CORVALÁN, Juan Gustavo et al. *Inteligencia artificial en accidentes de tránsito: primera aplicación predictiva en el mundo para la Justicia Civil*, 2019. Disponible en: https://dpicuantico.com/sitio/wp-content/uploads/2020/08/Corval%C3%A1n-y-LeFevre.pdf.

CORVALÁN, Juan Gustavo. Administración Pública digital e inteligente: transformaciones en la era de la inteligencia artificial. *Revista de Direito Econômico e Socioambiental*, Curitiba, v. 8, n. 2, p. 26-66, maio/ago. 2017. doi: 10.7213/rev.dir.econ.soc.v8i2.19321

CORVALÁN, Juan Gustavo. Digital and Intelligent Public Administration: transformations in the Era of Artificial Intelligence. *A&C Revista de Direito Administrativo e Constitucional*, Belo Horizonte, a. 18, n. 72, p. 81-82, jan./mar. 2018. doi: 10.21056/aec.v18i71.857

CORVALÁN, Juan Gustavo. El impacto de la inteligencia artificial en el trabajo. *Revista de Direito Econômico e Socioambiental*, Curitiba, v. 10, n. 1, p. 35-51, jan./abr. 2019. doi: 10.7213/rev.dir.econ.soc.v10i1.25870.

CORVALÁN, Juan Gustavo. *Hacia una administración pública 4.0: digital y basada en Inteligencia Artificial.* Decreto "Tramitación digital completa", La Ley, 2018.

CORVALÁN, Juan Gustavo. Inteligencia Artificial y Derechos Humanos. Parte II, *DPI Cuántico*, 10.07.2020. Disponible en: https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-10.07.2017.pdf

CORVALÁN, Juan Gustavo. Inteligencia Artificial y proceso judicial. Desafíos concretos de aplicación, *Diario DPI*, 9 set. 2019. Disponible en: https://dpicuantico.com/2019/09/09/el-impacto-de-la-ia-en-el-derecho-procesal/

CORVALÁN, Juan Gustavo. Inteligencia Artificial y proceso judicial. Desafíos concretos de aplicación, *Diario DPI*, 30 set. 2019. Disponible en: https://dpicuantico.com/sitio/wp-content/uploads/2019/09/Doctrina-Civil-30-09-2019-Parte-II-1.pdf

CORVALÁN, Juan Gustavo. *L´algorithme et les droits de l´ homme*, Conseil D´Etat, Ettude anuelle 2017.

CORVALÁN, Juan Gustavo. *Perfiles Digitales Humanos*. Buenos Aires: Thomson, 2020.

CORVALÁN, Juan Gustavo. *Prometea. Inteligencia Artificial para transformar organizaciones públicas*. Buenos Aires: Astrea, IMODEV, Universidad del Rosario y DPI Cuántico, 2019.

DASGUPTA, Tirthankar; SAHA, Rupsa; DEY, Lipika; NASKAR, Abrir. Automatic Extraction of Causal Relations from Text using Linguistically Informed Deep Neural Networks. *Proceedings of the SIGDIAL 2018 Conference*, p. 306-316, 2018

DELRIEUX, Claudio. *Inferencia ampliativa y razonamiento no monotónico*, Universidad Nacional del La Plata. Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/22252/Documento_completo.pdf?sequence=1&isAllowed=y

DOMINGOS, Pedro. *The master algorithm: how the quest for the ultimate learning machine will remake our world*. Basic Books, New York, 2015.

EL CONFIDENCIAL. *Los 23 mandamientos para evitar que la inteligencia artificial nos domine.* 2017. Disponible en: https://www.elconfidencial.com/tecnologia/2017-02-02/inteligencia-artificial-elon-musk-stephen-hawking-ia_1325057/

EL PRINCIPIO. Asilomar: http://puente.digital/blog/blog/inteligencia-artificial-segun-stephen-hawking-y-elon-musk/

El PRINCIPIOS 19, 20 y 21 de Asilomar, http://puente.digital/blog/blog/inteligencia-artificial-segun-stephen-hawking-y-elon-musk/

ESCOLANO RUIZ, F. *Inteligencia artificial: modelos, técnicas y áreas de aplicación*. Madrid: Thomson: Paraninfo, 2003.

ESCUDERO, Walter Sosa. *Big Data*. Siglo XXI Editores, 2019.

ESPANÃ. Instituto Español de Estudios Estratégicos, *Documentos de Seguridad y Defensa 79*, Madrid: Ministerio de Defensa, 2019. Disponible en: https://publicaciones.defensa.gob.es/la-inteligencia-artificial-aplicada-a-la-defensa-n-79-libros-ebook.html

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ). *European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Council of Europe, 4 Dec. 2018.

FERRER Beltrán. *Prueba y verdad en el derecho*. Barcelona: Marcial Pons, 2014.

GARCÍA SERRANO, A. *Inteligencia artificial: fundamentos, práctica y aplicaciones*. San Fernando de Henares, Madrid: RC Libros, 2016.

GARDNER, H. *La inteligencia reformulada las inteligencias múltiples en el siglo XXI*. Barcelona: Paidós, 2010.

GEORG, Jünger Friedrich, Mitos griegos. *Revista Espacio, Tiempo y Forma*, s. 2, Historia Antigua.

GIL DOMÍNGUEZ, Andrés. Inteligencia Artificial y Derecho. Buenos Aires: Rubinzal Culzoni Editores, 2019.

GÓMEZ MONT, Constanza et al. *La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países*, Jan. 2020, BID. Disponible en: https://publications.iadb.org/publications/spanish/document/La-inteligencia-artificial-al-servicio-del-bien-social-en-America-Latina-y-el-Caribe-Panor%C3%A1mica-regional-e-instant%C3%A1neas-de-doce-paises.pdf

GRANERO, Horacio Roberto. La Inteligencia Artificial aplicada al derecho en Informática y Derecho. *Revista Iberoamericana de Derecho Informático (segunda época)*, n. 5, p. 119-133, 2018.

HAWKING, Stephen; MUSK, Elon. Disponible en: https://www.elconfidencial.com/tecnologia/2017-02-02/inteligencia-artificial-elon-musk-stephen-hawking-ia_1325057/

HAYKIN, Simon. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, New Jersey, 1999.

JAIMOVICH, Desirée. La Justicia de Colombia usará un sistema de inteligencia artificial basado en un desarrollo argentino. *Infobae*, 28 de julio de 2020. Disponible en: https://www.infobae.com/tecno/2020/07/28/la-justicia-de-colombia-usara-un-sistema-de-inteligencia-artificial-basado-en-un-desarrollo-argentino/

JAMSHIDI-NEJAD, Sepideh; AHMADI-ABKENARI, Fatameh; EBRAHIMI-ATANI, Reza. Extraction of Textual Causal Relationships based on Natural Language Processing. *International journal of Computer Science & Network Solutions*. v. 3 n. 11, 2015

KATJA HOFMANN, W. Bruce Croft. *En línea*. https://www.microsoft.com/en-us/research/research-area/artificial-intelligence/

KURZWEIL, R. et al. *Cómo crear una mente: el secreto del pensamiento humano*. Berlin: Lola Books, 2013.

KURZWEIL, R. *The singularity is near: when humans transcend biology*. New York: Viking, 2005.

LEARNED-MILLER, E. *Introduction to Supervised Learning*. Department of Computer Science University of Massachusetts, Amherst, 2014, Disponible en: https://people.cs.umass.edu/~elm/Teaching/Docs/supervised2014a.pdf.

LEGRIS, Javier. *Razonamiento revocable y lógicas no monótonas: Un análisis conceptual*. Disponible en: http://bibliotecadigital.econ.uba.ar/download/cuadcimbage/cuadcimbage_n5_05.pdf

LINDHOLM, A. et al. *Supervised Machine Learning. Lecture notes for the Statistical Machine* Department of Information Technology, Uppsala University, 2019. http://www.it.uu.se/edu/course/homepage/sml/literature/lecture_notes.pdf.

LUMIENTO, María Elena. Algunas precisiones en torno a la prueba de la relación causal general en el derecho penal. Una especial referencia a la teoría del incremento del riesgo. In: ROVATTI, Pablo; LIMARDO, Alan. *Pensar la prueba*. Editores del Sur, 2020.

MANES Facundo; NIRO, Mateo. *Usar el cerebro: conocer nuestra mente para vivir mejor*. México, D.F.: Paidós, 2016.

MANES, Facundo.; NIRO, Mateo. *El cerebro argentino: una manera de pensar, dialogar y hacer un país mejor*. C.A.B.A: Planeta, 2016.

MARCUS, Gary. Crítica de GPT-3: el 'arte' de hablar sin decir ni entender nada. *MIT Technology Review*, 2020. Disponible en: https://www.technologyreview.es/s/12575/critica-de-gpt-3-el-arte-de-hablar-sin-decir-ni-entender-nada.

MARCUS, Gary. *Experiments testing GPT-3's ability at commonsense reasoning: results*. Disponible en: https://cs.nyu.edu/faculty/davise/papers/GPT3CompleteTests.html

MARCUS, Gary. *GPT-2 y la naturaleza de la inteligencia. The Gradient 2020*. Disponible en: https://thegradient.pub/gpt2-and-the-nature-of-intelligence/

MARINA, José Antonio. *El cerebro infantil: la gran oportunidad*. Barcelona: Ariel, 2011.

MARTINO, Antonio. Inteligencia Artificial y Derecho. Acerca de lo que hay. *Revista de Ciencia de la Legislación*, n. 6, sep. 2019.

MÉNDEZ, José. T. Palma, MARÍN MORALES, Roque. *Inteligencia artificial*. Mc Graw Hill, Madrid, 2011, p. 83 y ss.

MICHELE, Taruff. *La prueba de los hechos*. Madrid: Trotta, 2008.

MIT. *Machine Learning from Data Decisions*, Modulo 1, 23 abr. 2019.

MUMFORD, Lewis. *La Ciudad en la Historia*. Disponible en: https://istoriamundial.files.wordpress.com/2013/11/la-ciudad-en-la-historia_lewis-mumford.pdf.

NORVING, Peter. *¿Cómo funciona realmente el traductor de Google?* Redacción PressDigital, a. 2019. Disponible en: https://www.pressdigital.es/texto-diario/mostrar/1116921/como-funciona-realmente-traductor-google

NORVING, Peter. *Una mirada dentro de la tecnología de Google Translate*, Google, https://latam.googleblog.com/2011/11/una-mirada-dentro-de-la-tecnologia-de.html.

OCDE Library. *Inteligencia artificial en la sociedad*. 11 jun. 2019. Disponible en: https://www.oecd-ilibrary.org/sites/603ce8a2es/index.html?itemId=/content/ component/603ce8a2-es.

OCDE. Sobre Gobernanza Pública n. 34, *Estado de la técnica en el uso de tecnologías emergentes en el sector público*, 2019. https://ialab.com.ar/wp-content/uploads/2020/05/OECD-2019-Estado-de-la-te%CC%81cnica-en-el-uso-de-las-tecnologi%CC%81as-emergentes-en-el-sector-pu%CC%81blico.pdf

ONU. *Considerando 5.* La resolución N° 73/348 de la Asamblea General "Promoción y protección del derecho a la libertad de opinión y expresión" A/73/348, 29 ago. 2018. Disponible en http://undocs.org/es/A/73/348.

ONU. *Considerando 52*. La Resolución N° 72/540 de la Asamblea General "El derecho a la privacidad" A/72/540, 19 oct. 2017. Disponible en: http://undocs.org/es/A/72/540.

ONU. *Considerando 54.* Resolución N° 72/540 de la Asamblea General "El derecho a la privacidad" A/72/540, 19 oct. 2017. Disponible en: http://undocs.org/es/A/72/540.

OQUENDO, Catalina. 'Legaltech' Inteligencia artificial para desatascar la justicia en Colombia, *El País*. 3 de julio de 2020, Retina. Disponible en: https://retina.elpais.com/retina/2020/07/29/tendencias/1596020286_589017.html.

ORDÓÑEZ BURGOS, Jorge Alberto. La adivinación en Egipto: praxis política imperial. *Revista Espacio, Tiempo y Forma*, s. 2, Historia Antigua.

OXFORD. *Artificial Intelligence Programme*. University of Oxford, Said Business School, 2019.

PALACIOS, TEO. *La fundación de Tarento, una colonia espartana*. 2017. Disponible en: https://teopalacios.com/la-fundacion-de-tarento/

PALMA MÉNDEZ, J. T.; MARÍN MORALES, R. *Inteligencia artificial: métodos, técnicas y aplicaciones*. [s.l.] McGraw-Hill España, 2000.

PARLAMENTO EUROPEO. *El impacto del Reglamento General de Protección de Datos (GDPR) en la inteligencia artificial*, 25 de junio de 2020. Disponible en: https://www.europarl.europa.eu/stoa/en/document/EPRS_STU%282020%29641530

PASTOR, Daniel; HAISSINER, Martin. *Neurociencias, tecnologías disruptivas y tribunales digitales*. Buenos Aires: Hammurabi. 2019.

PEARL, J.; MACKENZIE, D. *The book of why: the new science of cause and effect*. First edition ed. New York: Basic Books, 2018.

PERELLÓ, Carlos Amunátegui. Archana Technicae. *El derecho y la Inteligencia Artificial*, Colección Tirant 4.0, Valencia, 2020.

PRETORIA. *Video Le presentamos*: 27 jul. 2020. Disponible en: https://www.youtube.com/watch?v=36pAqi0b7SA.

PRETORIA. *Video: PretorIA, Inteligencia Artificial Predictiva en la Corte Constitucional de Colombia*, 6 ago. 2020. Disponible en: https://www.youtube.com/watch?v=kq_N3r2diKw&t=49s.

QUOC, Le; BARRET, Zoph. *Using Machine Learning to Explore Neural Network Architecture*, 2017. Disponible en: https://ai.googleblog.com/2017/05/using-machine-learning-to-explore.html.

ROECKELEIN, J. E. *Dictionary of theories, laws, and concepts in psychology*. Westport, Conn: Greenwood Press, 1998.

RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. Pearson Education Limited, UK, 2016.

SADIN, Eric. *La humanidad aumentada*. [S.l.]: Caja Negra, 2017.

SEARLE, J. R. *Minds, brains, and programs*. Behavioral and Brain Science, Cambridge, v. 3, n. 3, p. 417-457, 1980.

SEBASTIÁN, Campanario. *GPT-3: el impacto económico de la tecnología que "se está comiendo el mundo*. La Nación, 2020. Disponible en: https://www.lanacion.com.ar/economia/gpt-3-el-impacto-economico-de-la-tecnologia-que-se-esta-comiendo-el-mundoc-nid2409609

SHALEV-SHWARTZ, S.; BEN-DAVID, S. *Understanding Machine Learning: From Theory to Algorithms*. Nueva York: Cambridge University Press, 2014.

SIEGEL, Daniel J. *Viaje al centro de la mente*. Paidós, Barcelona, 2017.

SIGMAN, Mariano. *La vida secreta de la mente*. Debate, Buenos Aires, 2016.

SILVA, Ricardo. *Casuality*. 2014. Disponible en: http://www.homepages.ucl.ac.uk/~ucgtrbd/papers/causality.pdf

SILVER, David. et al. *AlphaGo Zero: Starting from scratch*. 2017 Disponible en: https://deepmind.com/blog/alphago-zero-learning-scratch/

SILVER, David. et al. *Mastering the game of Go without human knowledge*. 2017. Disponible en: https://www.nature.com/nature/journal/v550/n7676/full/nature24270.html

SOUTH BY SOUTHWEST. *Elon Musk Answers Your Questions!* 2018. Disponible en: https://youtu.be/kzlUyrccbos

STRINGHINI, Antonella. Administración Pública Inteligente: novedades al ecosistema normativo digital de la República Argentina. *Revista Eurolatinoamericana de Derecho Administrativo*, Santa Fe, v. 5, n. 2, p. 199-215, jul./dic. 2018. doi: 10.14409/redoeda.v5i2.9094.

UNESCO. Inteligencia artificial, promesas y amenazas. *El Correo de la UNESCO*. París, jul./sep. 2018. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000265211.

VÉASE BENÍTEZ, Raúl et al. *Inteligencia artificial avanzada*. Barcelona: UOC, 2013.

WHISTLE OUT. *¿Qué es un Gigabyte (GB)?* 2020. Disponible en: https://www.whistleout.com.mx/CellPhones/Guides/que-es-un-gigabyte

WIHELM, Richard; I Ching. *El libro de las mutaciones*. 22.ed. Buenos Aires: Sudamericana, 2013.

WILL, Douglas Heaven. *Por qué GPT-3, la IA de lenguaje más avanzada, sigue siendo estúpida. MIT Technology Review*, 2020. Disponible en: https://www.technologyreview.es/s/12453/porque-gpt-3-la-ia-de-lenguaje-mas-avanzada-sigue-siendo-estupida

WINSTON, Patrick. Henry. *Artificial intelligence*. 3. ed. Reading, Mass: Addison-Wesley Pub. Co, 1992.

YASER, Abu-Mostafa. Técnicas de aprendizaje automático. Especial Inteligencia artificial. *Investigación y Ciencia*, abr. 2003.

YUDKOWSKY, Eliezer. Levels of Organization in General Intelligence. In: GOERTZEL, Ben; PENNACHIN, Cassio (Coords.). *Artificial General Intelligence*. Springer: Berlim, 2007.

Informação bibliográfica deste texto, conforme a NBR 6023:2018 da Associação Brasileira de Normas Técnicas (ABNT):

CORVALÁN, Juan Gustavo. Inteligencia Artificial GPT-3, PretorIA y oráculos algorítmicos en el derecho. *International Journal of Digital Law*, Belo Horizonte, ano 1, n. 1, p. 11-52, jan./abr. 2020.

Sumário
Contents

EDITORIAL Nº 1

É com satisfação que apresentamos à comunidade profissional e acadêmica o *International Journal of Digital Law*. Procuramos criar um periódico científico novo, com a pretensão de suprir uma lacuna que ainda é existente na tratativa do tema, tanto em nível local quanto global.

O *International Journal of Digital Law* consiste em periódico científico eletrônico de acesso aberto e periodicidade quadrimestral promovido pelo NUPED – Núcleo de Pesquisas em Políticas Públicas e Desenvolvimento Humano do Programa de Pós-Graduação em Direito da Pontifícia Universidade Católica do Paraná – um grupo de pesquisa filiado à REDAS – Rede de Pesquisa em Direito Administrativo Social.

A publicação foi encampada pela Editora Fórum, sem dúvida a mais renomada casa editorial do Direito Público brasileiro – o que por si só já é um atestado de qualidade conferido ao projeto.

O Conselho Editorial é composto por renomados juristas vinculados a instituições de ensino superior do Brasil, Argentina, Austrália, Colômbia, Espanha, Egito, França, Holanda e Índia. O enfoque da revista é o estudo crítico das instituições jurídico-políticas típicas do Estado de Direito, notadamente, as voltadas à inovação e ao desenvolvimento humano por intermédio da revolução digital. Agradecemos muito a franca disponibilidade dos professores que aceitaram compor tanto o Conselho Editorial quanto o Conselho Especial de Pareceristas.

O NUPED se insere na área de concentração do PPGD/PUCPR intitulada "Direito Econômico e Desenvolvimento". Por sua vez, a área congrega duas importantes linhas de pesquisa: 1. Estado, Economia e Desenvolvimento e 2. Direitos Sociais, Globalização e Desenvolvimento.

A revista irá dar destaque a este marco teórico. Entretanto, transversalmente ao tema da economia, do desenvolvimento, da globalização e dos direitos sociais, as palavras-chave que melhor definem o escopo da revista implicam a tratativa de temas como: acesso à informação, *Big data*, *Blockchain*, Cidades inteligentes, Contratos inteligentes, *Crowdsourcing*, Cibercrimes, Democracia digital, Direito à privacidade, Direitos fundamentais, *E-business*, Economia digital, Educação digital, Eficiência administrativa, *E-Government*, Ética, *Fake news*, *Gig economy*, Inclusão digital, Infraestrutura, Inovação, Inteligência artificial, Interesse público, Internet, Internet das coisas, Jurimetria, *Lawfare*, Novas tecnologias, Perfilamento digital, Pesquisa em multimeios, Processo administrativo eletrônico, Proteção de dados, Regulação administrativa, Regulação econômica, Risco, Serviços públicos,

Sistemas de informação, Sociedade da informação, Transparência governamental e Telecomunicações.

E o escopo da revista é, portanto, fortemente interdisciplinar e transdisciplinar. Espera-se que estudiosos dos mais diferentes campos de pesquisa possam enviar seus trabalhos, que serão muito bem recebidos, podendo ser escritos em português, inglês ou espanhol. Já neste primeiro número, além dos artigos dos pesquisadores brasileiros, temos textos oriundos de três diferentes países e continentes: Argentina, Espanha e Índia.

Os artigos passarão pelo sistema de avaliação em *doble blind peer review*. A ideia é que rapidamente o *International Journal of Digital Law* torne-se uma referência em termos de seriedade acadêmica e impactação na sociedade. Para isso, procuraremos nos enquadrar nas diretrizes das mais importantes bases de indexação nacionais e internacionais.

Emerson Gabardo
Alexandre Godoy Dotta
Juan Gustavo Corvalán

EDITORIAL Nº 1

We are pleased to present the *International Journal of Digital Law* to the professional and academic community. We seek to create a new scientific journal, with the intention of filling a gap that still exists in dealing with the topic, both at the local and global levels.

The *International Journal of Digital Law* consists of an open-access electronic scientific journal and published every four months by NUPED – Center for Research in Public Policies and Human Development of the Postgraduate Law Program at the Pontifical Catholic University of Paraná – an affiliated research group to REDAS – Research Network in Welfare State Administrative Law.

The Editorial Board is composed of renowned professors linked to higher education institutions in Brazil, Argentina, Australia, Colombia, Spain, Egypt, France, and India. The journal's focus is the critical study of the legal-political institutions typical of the rule of law, notably those aimed at innovation and human development through the digital revolution. We are grateful for the frank availability of the professors who agreed to compose both the Editorial Board and the Special Peer Review Bord.

NUPED is part of the PPGD/PUCPR Concentration area entitled "Economic Law and Development". In turn, the area brings together two important lines of research: 1. State, Economy and Development and 2. Social Rights, Globalization and Development.

The magazine will highlight this theoretical framework. However, transversely to the theme of economics, development, globalization and social rights, the keywords that best define the scope of the magazine involve dealing with topics such as access to information, Big data, Blockchain, Smart Cities, Smart contracts, Crowdsourcing, Cybercrimes, Digital democracy, Right to privacy, Fundamental rights, E-business, Digital economy, Digital education, Administrative efficiency, E-Government, Fake News, Gig economy, Globalization, Digital inclusion, Infrastructure, Innovation, Artificial intelligence, Public interest, Internet, Internet of things, Jurimetrics, Lawfare, New technologies, Digital profiling, Multimedia research, Electronic administrative process, Data protection, Administrative regulation, Economic regulation, Risk, Public services, Information systems, Information society, Government transparency, and Telecommunications.

And the journal's scope is, therefore, strongly interdisciplinary and transdisciplinary. It is expected that scholars from the most different fields of research will be able to send their works, which will be very well received and can be written in Portuguese, English or Spanish. In this first issue, in addition to articles by

Brazilian researchers, we have texts from three different countries and continents: Argentina, Spain and India.

All articles will go through the evaluation system in double-blind peer review. The idea is that the *International Journal of Digital Law* will quickly become a reference in terms of academic seriousness and impact on society. For that, we will try to fit in the guidelines of the most important national and international indexing bases.

Emerson Gabardo
Alexandre Godoy Dotta
Juan Gustavo Corvalán