

# vooces

en el Fénix

LA SOCIEDAD DEL  
CONOCIMIENTO  
Y LA INTELIGENCIA  
ARTIFICIAL

# sumario

## n°93

### septiembre

### 2024

INTELIGENCIA ARTIFICIAL  
EN LA SOCIEDAD DEL  
CONOCIMIENTO

Cátedra Abierta "Plan Fénix"

POR RAZONES PRESUPUESTARIAS  
NOS HEMOS VISTO OBLIGADOS  
A REDUCIR LA CANTIDAD DE  
NÚMEROS DE VOCES EN EL FÉNIX  
EN EL PRESENTE AÑO.

**OSCAR OSZLAK** Impacto de la inteligencia artificial en la gestión estatal **12** **RAMIRO VERRASTRO** Técnicas de la inteligencia artificial **20** **RAMIRO RODRÍGUEZ COLMEIRO** Redes neuronales artificiales **28** **MARTÍN BELZUNCE** Cómo la inteligencia artificial está revolucionando las imágenes médicas: aplicaciones, riesgos y desafíos **36** **ARIEL HERNÁNDEZ** Autonomía en movimiento: perspectivas y desafíos de los vehículos autónomos **46** **PINAR DEMIRCIUGLU, ISMAIL BOGREKCI, NUMAN DURAKBASA Y JORGE BAUER** Autonomación, automatización, inteligencia artificial e industria-agricultura 5.0 en la producción agroecológica sostenible de alimentos **54** **LEANDRO DI MATTEO** Armas letales autónomas e inteligencia artificial **64** **LEONARDO SERVALLI** Desafíos de ciberseguridad en la inteligencia artificial **72** **MARTIN KUFFER** Todo lo que quisiste saber sobre computación cuántica y nunca te animaste a preguntar **78** **PILAR FAJARNES** Digitalización para un desarrollo inclusivo **84** **FLAVIA COSTA, JULIÁN MÓNACO** ¿Pueden las humanidades coevolucionar con los humanos (y las máquinas)? Inteligencia artificial, sociedad artificial y los desafíos para las ciencias sociales **90**

## AUTORIDADES UBA

### Rector

Ricardo Jorge Gelpi

### Vicerrector

Emiliano Yacobitti

## AUTORIDADES FCE

### Decano FCE

Ricardo J.M. Pahlen

### Vicedecano

César Humberto Albornoz

### Consejeros Docentes

Alberto Edgardo Barbieri  
Inés Mercedes García Fronti  
María Alejandra Metelli  
Ana María Campo  
José Luis Giusti  
Emiliano Yacobitti  
María Teresa Casparri  
Andrés Flavio López

### Consejeros Estudiantes

Lucille Levy  
Martín Roel  
Martina Sarraf  
Alex Vargas

### SECRETARÍA DE EXTENSIÓN UNIVERSITARIA, BIENESTAR ESTUDIANTIL Y DESARROLLO SUSTENTABLE

Carlos Jara

### SUPLENTES

César Humberto Albornoz  
Mariana Gloria Barreña  
Andrea Carolina Calello  
Patricia Bonatti  
Ricardo J. M. Pahlen  
Walter Fabián Carnota  
María Inés Barbero  
Guillermo Enrique Ragazzi

### SUPLENTES

Nicolás Gutierrez  
Mara Troncoso  
Ezequiel Escobar  
Cecilia Aguiar

### SECRETARÍA DE RESPONSABILIDAD SOCIAL, EMPRENDEDORES E INNOVACIÓN

Andrea Zaidel

### Consejeros Graduados

Rubén Arena  
Jeniffer Cirolini  
Luis Alberto Cowes  
Sofía Caraballo

### SECRETARÍAS: SECRETARÍA ACADÉMICA

Gustavo Montanini

### SECRETARÍA DE RELACIONES ACADÉMICAS INTERNACIONALES, NACIONALES Y MUNICIPALES

Juan Francisco Nosiglia

### SECRETARÍA DE INVESTIGACIÓN

Adrián Ramos

### SECRETARÍA DE GRADUADOS

Rubén Arena

### ESCUELA DE NEGOCIOS Y ADMINISTRACIÓN PÚBLICA

Catalino Nuñez

### SECRETARÍA DE TRANSFERENCIA DE GESTIÓN E INNOVACIÓN DE TECNOLOGÍAS

Alejandro Burgos

### SECRETARÍA DE DOCTORADO Y POSDOCTORADO

María Teresa Casparri

### SECRETARÍA GENERAL

Pablo Rota

### SECRETARÍA DE RELACIONES INSTITUCIONALES

Omar Quiroga

### SUPLENTES

Ana María Cristina Iglesias  
Daniel González  
María Fernanda Inza  
Juan Manuel Oro

### SECRETARÍA DE HACIENDA Y ADMINISTRACIÓN

Fabián Famá

Voces en el Fénix es una publicación del Plan Fénix

ISSN 1853-8819

Registro de la propiedad intelectual en trámite.



Los artículos firmados expresan las opiniones de los autores y no reflejan necesariamente la opinión del Plan Fénix ni de la Universidad de Buenos Aires.

## staff

### DIRECTOR

Bernardo Lischinsky

### COMITÉ EDITORIAL

José Miguel Amiune  
Ricardo Aronskind  
Alfredo Calcagno  
Ricardo Koss  
Fernando Porta  
Alejandro Rofman  
Mariana Vázquez

### COORDINACIÓN TEMÁTICA

Claudio Verrastró

### SECRETARIA DE REDACCIÓN

Wanda Fraiman

### COORDINACIÓN

Paola Severino

### SITIO WEB

Erica Sermukslis

### DISEÑO EDITORIAL

Martín Marpons

### CORRECCIÓN

Claudio M. Díaz

### FOTOGRAFÍA

Sub [Cooperativa de Fotógrafxs]

ANCCOM [Agencia de Noticias de Ciencias de la Comunicación - UBA]

### AUDIOVISUAL

Pablo Martínez  
Nicolás Cerezo

Córdoba 2122, Facultad de Ciencias Económicas, Universidad de Buenos Aires. C.A.B.A.  
Teléfono 5285-6819. <https://vocesenelfenix.economicas.uba.ar> / [voces@vocesenelfenix.com](mailto:voces@vocesenelfenix.com)

# INTELIGENCIA ARTIFICIAL EN LA SOCIEDAD DEL CONOCIMIENTO

Cátedra Abierta "Plan Fénix"  
Facultad de Ciencias Económicas  
Universidad de Buenos Aires

Q

ué es la inteligencia artificial? Es comúnmente aceptado que ella es una característica de máquinas que pretenden replicar capacidades cognitivas humanas, incluyendo

la sensación, la interacción lingüística, el razonamiento y el análisis y la resolución de problemas. Por otra parte, tales "máquinas inteligentes" pueden demostrar capacidades de aprendizaje con mecanismos de autorrelación y autocorrección, sobre la base de algoritmos que incorporan el "aprendizaje automático" o incluso el "aprendizaje profundo", utilizando "redes neuronales" que imitan el funcionamiento del cerebro humano.

La inteligencia artificial sin duda es transversal a la totalidad de las actividades que desarrolla una sociedad hoy día. Puede vaticinarse que su desarrollo tendrá un enorme impacto en materia económica, social y política, debido a su capacidad de sustituir trabajo intelectual, llevando a un incremento en las potencialidades productivas y en el rendimiento de la actividad laboral, en un grado que hoy ni siquiera podemos estimar. La inteligencia artificial continúa así un desarrollo que reconoce como antecedente las nuevas capacidades que en su momento crearon la informática y la digitalización. Ya se perciben efectos en tareas de investigación y en la medicina, entre otros sectores.

Todo cambio tecnológico de monta conlleva consecuencias significativas. Reconfigura los procesos productivos, afectando directamente el empleo, tanto en cantidad como en requerimientos de calificación. Por otro lado, inevitablemente estas mutaciones tienden a aumentar la concentración del ingreso y la riqueza, porque las rentas originadas se concentran en pocos oferentes. Asimismo, el uso de la inteligencia artificial por parte de pocas corporaciones gigantes y algunos gobiernos, a través de las redes sociales, fortalece mecanismos de manipulación sobre




las personas y sociedades a nivel mundial; esto lleva a una creciente desconfianza en los medios de comunicación. Desde la Cátedra Abierta Plan Fénix siempre se ha entendido el desarrollo como un proceso ligado a la profundización de la educación, la investigación científica y la incorporación de tecnología en la producción de bienes y servicios, en todas las áreas. En esto, se ha considerado irremplazable el papel que tiene el Estado, en concurrencia con los otros actores y sectores. En este sentido la irrupción de la inteligencia artificial requiere la acción estatal en dos niveles.

En primer lugar, y esto se refiere tanto a la inteligencia artificial como más en general a la digitalización, se demanda la implementación de mecanismos regulatorios en diversos planos que, sin bloquear los efectos positivos de aquellas, morigeren el poder que ostentan hoy día las corporaciones mencionadas y distribuyan parte de la renta tecnológica. A pesar de la urgencia, esta es una tarea de largo aliento, tanto porque desconocemos hoy día el pleno alcance de estos instrumentos, como por la necesidad de desarrollar capacidades estatales y coordinar esfuerzos a nivel internacional. Un principio central –que es análogo al que debe imperar en general en el ámbito de internet– es que si la información con que se alimentan los algoritmos de inteligencia artificial es la de acceso público, también deben serlo los resultados obtenidos.

En segundo lugar, el Estado deberá monitorear de cerca los impactos sobre el empleo, a fin de mitigar los eventuales efectos adversos; es esperable, por ejemplo, que la caída en el empleo calificado comporte el desempleo de trabajadores de baja calificación, por efecto de desplazamiento.

Cabe además una reflexión. La inteligencia artificial ha sido presentada –y así parece haber sido visualizada por el conjun-



to de la sociedad– como una herramienta que emula la inteligencia humana, y que por lo tanto estaría en condiciones de sustituirla. Esto resulta engañoso. La inteligencia artificial en realidad no replica exactamente facultades humanas, puesto que no tiene capacidad de intuir o de establecer relaciones no identificadas previamente (que es de hecho lo que se reconoce como “inteligencia” en el caso de las personas), además desde ya de carecer de motivación en el plano de lo emocional. Realiza sí un amplio conjunto de operaciones, en tiempos muy breves. Es una extensión de las extraordinarias capacidades que ha desarrollado la informática en la Posguerra; y en esto sin duda es más eficaz que el cerebro humano.

A pesar de ello, la inteligencia artificial presenta hoy una característica que debe ser objeto de gran atención. La autonomía de los algoritmos autogenerados por ella podría amenazar la capacidad humana de comprender la lógica en la producción de resultados, recomendaciones y –en el caso de funciones delegadas– toma de decisiones que tendrían impactos negativos sobre la humanidad. Por consiguiente, se debate la cuestión de si la inteligencia artificial seguirá siendo solo una herramienta más al servicio de los seres humanos –y controlada por ellos– o, por el contrario, cobrará una autonomía tal que se transformará en “superhumana”, poniendo en riesgo a la propia especie. Ello deviene del hecho de que la inteligencia artificial funciona como una “caja negra”, un sistema o modelo cuyo funcionamiento interno no es transparente ni interpretable. Al parecer, aun quienes diseñan los algoritmos que entrenan estos modelos no siempre comprenden completamente las redes neuronales resultantes.

A este problema se suma el hecho de que los propietarios de estos sistemas se hallan en competencia entre sí, y la filtración de información o denuncia de peligros es al pa-



recer severamente penalizada. Así, aunque es sabido que empresas están investigando formas de comprender mejor el funcionamiento interno de los modelos de lenguaje, es muy poco probable que estén dispuestas a compartir esos conocimientos o que las propias técnicas, como la de los autocodificadores variacionales, sean totalmente eficaces para decodificar el conjunto de algoritmos que dan lugar a un cierto resultado.

Este es un punto más que reclama la activa regulación gubernamental.

Inteligencia artificial e inteligencia humana no son equiparables (cabría tal vez sugerir que el nombre asignado a la primera no fue una decisión apropiada). No debemos caer en el error de suponer que la inteligencia artificial podrá sustituir a la mente humana; debe ser pensada entonces como un potente auxiliar para los análisis y toma de decisiones a medida que las regulaciones avancen y sus riesgos de autonomía disminuyan. Es por eso que son responsabilidad de personas y también de los Estados nacionales. Se logrará así que la inteligencia artificial sea siempre un instrumento, no un decisor, y por lo tanto sus impactos serán los que decidamos.

Este número de Voces en el Fenix está dirigido a analizar pros y contras de la aparición de esta herramienta en nuestra vida cotidiana, algo que no tiene vuelta atrás. La inteligencia artificial nos plantea un escenario nuevo, como lo hicieron en su momento otros cambios en los paradigmas científicos y tecnológicos. Este nuevo escenario conlleva un conjunto de aristas problemáticas, que van desde la posibilidad de generar nuevas desigualdades hasta los dilemas éticos y cognitivos que surgen. Este nuevo escenario debe ser enfrentado con las herramientas intelectuales y prácticas correspondientes.

# INTRODUCCIÓN

## LA SOCIEDAD DEL CONOCIMIENTO Y LA INTELIGENCIA ARTIFICIAL

**E**l potencial de disrupción de la inteligencia artificial y su evolución vaticina que su desarrollo acentuará su impacto en materia socioeconómica, laboral y política. Si bien es una técnica largamente conocida en el ambiente de las ciencias de la computación, fue potenciada por el crecimiento de la capacidad de cálculo y la masividad de las redes de datos. Su aparición invisible y silenciosa en nuestra vida cotidiana no tiene vuelta atrás, obliga a los individuos a un enorme esfuerzo de adaptación. Y a la sociedad toda, a gestionar este desafío que podría equipararse a la irrupción de la energía nuclear, en tanto y en cuanto a la brecha tecnológica que genera. Para la necesaria protección de las personas y de los Estados que no logren adaptarse con la suficiente velocidad, se deberán generar mecanismos de regulación que mitiguen los riesgos y peligros asociados, así como potenciar sus beneficios. En este número **Oscar Oszlak** ilumina estos aspectos con su artículo “Impacto de la inteligencia artificial en la gestión estatal”.

Pero volviendo a los orígenes cercanos, hace dos siglos, la matemática británica Ada Byron de Lovelace (1815-1851), autora del primer algoritmo destinado a ser procesado por una máquina –la computadora mecánica de Charles Babbage–, postuló que “las máquinas solo hacen aquello que les ordenamos que hagan”. Quizás uno de los corolarios del principio de Ada sería la tesis GIGO (“si entra basura, sale basura”), con lo cual el progra-

gador es el principal responsable. Durante el siglo pasado, a pocos años de la creación de la primera computadora electrónica (ENIAC 1945), Alan Turing en su obra *Computing Machinery and Intelligence* (1950) vuelve a interrogarse si las máquinas pueden pensar y estableció una prueba para distinguir si una computadora es o no inteligente. Para esto ideó un escenario donde una persona realice una consulta, sin contacto físico, y de la respuesta no se pueda discernir si quien responde es una máquina o una persona. Actualmente lo conocemos como CAPTCHA –“completely automated public Turing test to tell computers and humans apart”– y es parte de nuestra rutina para acceder a ciertos portales que desean protegerse de intrusiones de bots –robots de software–. Muchos sostienen que el test de Turing ha sido superado, desde el momento en que la inteligencia artificial ha alcanzado un nivel que torna cada vez más difícil saber si se está interactuando con una máquina o una persona.

El concepto de inteligencia artificial (IA) se debe al informático estadounidense John McCarthy, quien en el año 1956 lo pronunció por primera vez en una conferencia causando un gran impacto en el ámbito de la tecnología. A partir de entonces, el concepto se diseminó fantásticamente por el mundo y por ello hoy es tan común su uso cuando queremos referirnos a aquellas máquinas o aparatos dotados de una inteligencia similar a la de los seres humanos.

Muy tempranamente, en 1957, durante una conferencia Herbert Simon (premio Nobel de Economía 1978, por sus trabajos en la Teoría de las Decisiones) afirmó: “Sin afán de sorprenderlos y dejarlos atónitos, debo informarles lisa y llanamente que actualmente en el mundo existen máquinas capaces de pensar, aprender y crear. Además, su capacidad para hacer lo anterior aumentará rápidamente hasta que en un futuro previsible la magnitud de problemas que tendrán capacidad de manejar irá a la par con la capacidad de la mente humana para hacer lo mismo”. Para rebatir a Simon, John Searle (1980) en su texto de “Mentes, cerebros y ciencia” ataca este pensamiento con el experimento mental de la Habitación China. En él muestra cómo una máquina puede realizar una acción sin siquiera entender lo que hace ni por qué lo hace. Por lo tanto, según Searle la lógica usada por las computadoras es nada más que un mecanismo que no busca el sentido en la acción, como la usada por los seres humanos. No tiene un “para qué”. Este enfoque puede ser descrito en términos del llamado “Juego de la Imitación” que traslada el peso de la respuesta desde la esencia (ser o no ser) a la apariencia (como si). En palabras de Searle, “la gramática y la sintaxis solas no son suficientes para la semántica y las computadoras digitales en tanto máquinas tienen, por definición, solamente gramática y sintaxis”. Amén.

Estos planteos filosóficos dieron origen a dos vertientes, los defensores de la inteligencia artificial “fuerte” que creen que es posible crear “inteligencia de verdad” y los pragmáticos que se enfocan en los resultados y piensan que cuantas más personas se puedan engañar creyendo que en las tripas de la máquina existe la inteligencia, entonces el truco es lucrativo. **Ramiro Verrastro** nos enseña algunos de estos trucos en su artículo sobre “Técnicas de la inteligencia artificial”.

por **CLAUDIO VERRASTRO**. INGENIERO ELECTRÓNICO DE LA UNIVERSIDAD DE BUENOS AIRES. MAGÍSTER EN SOFTWARE.

FUE JEFE DE LA DIVISIÓN DE SISTEMAS DIGITALES Y ROBÓTICA DE LA COMISIÓN NACIONAL DE ENERGÍA ATÓMICA (CNEA) Y RESPONSABLE DEL DISEÑO Y DESARROLLO DE TOMOGRAFÍA POR EMISIÓN DE POSITRONES Y SISTEMA DE PROTECCIÓN DE REACTORES PARA REACTORES NUCLEARES.

SE DESEMPEÑÓ COMO PROFESOR ASOCIADO E INVESTIGADOR PRINCIPAL EN INTELIGENCIA ARTIFICIAL Y ROBÓTICA EN LA UNIVERSIDAD TECNOLÓGICA NACIONAL DE BUENOS AIRES (UTN). ES INVESTIGADOR CONSULTO DE LA CNEA Y MIEMBRO DEL CONSEJO ASESOR DE LA UTN FRBA.



Una característica saliente del ser humano es potenciar sus propias fuerzas mediante la utilización de herramientas y la apropiación de los recursos naturales para sus propios fines, como la utilización de la fuerza de los animales para labrar la tierra, transporte, etc., la máquina de vapor para la utilización del carbón, los motores de explosión para la utilización del petróleo, la energía nuclear para generar electricidad. En cambio, la IA permite potenciar las fuerzas del intelecto humano para mejorar la búsqueda, la representación y el almacenamiento de conocimiento y generar nuevo conocimiento explorando posibles soluciones a problemas que sin la utilización de estas técnicas serían inabordables.

Básicamente la IA es un soporte para las actividades del intelecto humano que de alguna manera hace que las personas puedan tener una mejor comprensión del mundo y de sí mismas. Se la puede pensar como una tecnología que colabora en la fabricación de herramientas cada vez más útiles en la producción de bienes y servicios.

Una de estas herramientas son las “Redes neuronales artificia-



les” (RNA), **Ramiro Rodríguez Colmeiro** en su artículo nos cuenta sobre el potencial de las RNA a partir de su desarrollo explosivo desde principios de este siglo, con la posibilidad del procesamiento masivo de datos o Big Data y el desarrollo de varias arquitecturas de aprendizaje profundo como las redes neuronales profundas, convolucionales así como generativas, se aplican a campos como la visión artificial, reconocimiento automático del habla y reconocimiento de imágenes, señales de audio y música que mostraron resultados de vanguardia en varios campos.

Tal como **Martín Belzunce** en su artículo “Cómo la inteligencia artificial está revolucionando las imágenes médicas: aplicaciones, riesgos y desafíos” nos ilustra sobre aplicaciones en salud humana.

También la IA mejora la interacción entre los hombres y las máquinas, haciéndolas más fácil de usar y más productivas. Actualmente los desarrollos están orientados a robots que puedan desenvolverse junto a las personas para prestarles asistencia. Ejemplos de ello son el proyecto Roboy para realizar tarea de servicios, desde camareros hasta asistentes geriátricos, guías de museo, etc., la tecnología para asistencia a la conducción de vehículos (ADAS) y para la disminución de accidentes viales. **Ariel Hernández** en su artículo “Autonomía en movimiento: perspectivas y desafíos de los vehículos autónomos” nos ayuda a comprender los retos que enfrentan las aplicaciones de IA en entornos no estructurados donde deben interactuar con personas.

En la industria, los robots son cada vez más automáticos y pueden realizar tareas más sofisticadas, los sistemas de visión permiten tomar decisiones en máquinas clasificadoras y de control de calidad, para la optimización de la utilización de materias primas y disminución de descartes. En el ámbito científico los programas de simulación y de cálculo permiten predecir cambios climáticos y el impacto de las acciones humanas, permiten desarrollar sistemas para economizar recursos naturales y explorar alternativas que con las herramientas de cómputo tradicionales son inviables. **Jorge Bauer** en su trabajo “Autonomación, automatización, inteligencia artificial e industria-agricultura 5.0 en la producción agroecológica sostenible de alimentos” nos permite avizorar que la integración de las herramientas de IA con otras tecnologías habilita otras formas de producción menos invasivas y más sostenibles desde el punto de vista ambiental.

La IA es una herramienta, independientemente de su poder. Como toda herramienta, el hombre puede decidir usarla como un arma de destrucción de sus semejantes o para construir una sociedad mejor. Eso ocurrió desde el garrote que pudo y puede usarse como un martillo hasta con la energía atómica que pudo y puede usarse como una bomba o como combustible de una central nucleoelectrónica. Hiroshima es testigo de que cuanto más poderosa es la herramienta, más daño puede causar cuando se la usa como un arma. **Leandro Di Matteo** en su artículo “Armas letales autónomas e inteligencia artificial” nos advierte que las “armas asesinas” o AKM (del inglés, *automated killing machines*) ya saltaron de la pantalla a la vida real.

Efectivamente la tecnología ya alcanzó el grado de madurez necesario para ser usada en forma autónoma, sobre todo cuando los daños “colaterales” los paga el “otro”. Al respecto, la reunión internacional realizada en Buenos Aires en 2015 se utilizó para lanzar una carta abierta firmada por miles de investigadores de IA y robótica, que denomina públicamente a la IA como la “mayor amenaza existencial de la humanidad”. Según los expertos, la creación de las máquinas asesinas es una “tercera revolución en la guerra”, después de la pólvora y las armas nucleares. “Cuando se abra la caja de Pandora, será difícil cerrarla”, advirtieron en el encuentro.

Aquella cita del 2015 ayudó a empujar a la ONU a realizar las primeras conversaciones formales sobre el creciente universo de la robótica y la IA y sus aplicaciones bélicas. También contribuyó a crear conciencia de las posibles consecuencias irremediables que pueden provocar. A fines de 2017, un grupo de 116 ejecutivos de 26 empresas dedicadas al desarrollo de IA –entre ellos el inventor y millonario sudafricano Elon Musk– firmaron en Melbourne (Australia) una carta abierta dirigida a las Naciones Unidas solicitando la prohibición de las Máquinas de Matar Autónomas. No es el primero ni único llamado que se realiza al Organismo de Naciones Unidas, donde se creó un grupo expertos específico que actúa dentro de la Convención sobre Ciertas Armas Convencionales (The Convention on Certain Conventional Weapons –CCW–), Organismo de United Nations. Office for Disarmament Affairs UNODA. Esta convención sobre prohibiciones o restricciones del empleo de armas que puedan considerarse excesivamente nocivas o de efectos

indiscriminados y sus protocolos sobre LAWS (Lethal Autonomous Weapon Systems, sistemas de armas letales autónomas) produjo reuniones y definiciones desde 2013 que pueden consultarse en [https://docs-library.unoda.org/Convention\\_on\\_Certain\\_Conventional\\_Weapons\\_Group\\_of\\_Governmental\\_Experts\\_on\\_Lethal\\_Autonomous\\_Weapons\\_Systems\\_\(2023\)/CCW\\_GGE1\\_2023\\_CRP.1\\_0.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.1_0.pdf).

¿Quién decide cuando el uso de un arma es responsable? Hay una preocupación mayoritaria de que el arma se vuelva en contra del que la usa. Es como si se protegiera al agresor y no al agredido. Las aplicaciones potencialmente peligrosas no se limitan a las AKMs, pueden estar ocultas en los robots de búsqueda y clasificación de información, que pueden acceder a la información en forma global, pero también pueden entregar información sobre nosotros mismos a usuarios no autorizados o maliciosos, **Leonardo Servalli** nos advierte en su trabajo “Desafíos de ciberseguridad en la inteligencia artificial” de los riesgos cada vez mayores sobre la privacidad de los datos sensibles que pueden cubrir un amplísimo espectro, desde datos bancarios hasta los aspectos personales, familiares y laborales. A la par de las herramientas tradicionales para proteger los datos como la encriptación también están amenazadas por otra tecnología emergente, la computación cuántica, que puede potenciar aún más las aplicaciones virtuosas y por supuesto también las maliciosas. **Martin Kuffer** nos cuenta sobre este tema en “Todo lo que quisiste saber sobre computación cuántica y nunca te animaste a preguntar”.

Los avances en la tecnología y las comunicaciones se traducen en avances en el sistema económico, el sistema político, la distribución de bienes, la desigualdad social y otros aspectos de la vida social. **Pilar Fajarnes** nos propone en su trabajo “Digitalización para un desarrollo inclusivo” que la tecnología no es determinista, los beneficios no se crean de manera automática y que, sin la intervención pública, es muy posible que las desigualdades asociadas a la digitalización aumenten.

Es muy difícil predecir el futuro, inclusive genios como Asimov o Alan Turing, por citar algunos conocidos, se equivocaron en sus predicciones. Los programas de simulación y modelado (muchos utilizan IA) predicen que el estilo de vida actual del hombre es insostenible por su impacto en la naturaleza y en el cambio climático. **Flavia Costa** y **Julián Mónaco** en el artículo “¿Pueden las humanidades coevolucionar con los humanos (y las máquinas)? Inteligencia artificial, sociedad artificial y los desafíos para las ciencias sociales”, plantean aspectos que ponen en riesgo a la sociedad, por la pérdida de credibilidad de las autoridades gubernamentales, así como de la información periodística ante la difusión masiva de noticias falsas –cada vez más sofisticadas– que son generadas mediante alteraciones de audio y video e influyen en una parte importante de nuestros modos de relacionarnos socialmente, dando lugar a una “Sociedad Artificial”. La IA y el ecosistema digital ya no sólo son herramientas sino un “mundoambiente” y las ciencias sociales se deberán reconvertir para poder cartografiar un territorio complejo y multicapa para viabilizar un “nuevo acuerdo tecnológico” para nuevas democracias digitales.

por **OSCAR OSZLAK**. Contador Público y doctor en Economía por la Universidad de Buenos Aires. PhD Political Science y Master of Arts in Public Administration, UC Berkeley; Investigador Superior del CONICET. Profesor Consulto de la Facultad de Ciencias Sociales (UBA) y Profesor Extraordinario de la Universidad Nacional de Villa María.

# IMPACTOS DE LA INTELIGENCIA ARTIFICIAL SOBRE LA GESTIÓN ESTATAL

EL AUTOR DA CUENTA DE LOS PRINCIPALES DESAFÍOS Y DE LAS POTENCIALIDADES QUE ENFRENTAN LOS GOBIERNOS A NIVEL MUNDIAL, PERO SOBRE TODO EN LA REGIÓN DE AMÉRICA LATINA, EN LA IMPLEMENTACIÓN DE SUS POLÍTICAS EN EL MARCO DE LA ACTUAL INNOVACIÓN TECNOLÓGICA.

T

*Technological singularity* es un hipotético punto en un tiempo futuro, en el que el desarrollo tecnológico se vuelve incontrolable e irreversible, resultando en cambios insondables para la civilización humana. En su versión más popularizada, la hipótesis de la singularidad supone que un agente inteligente (v.g., un software de computación basado en inteligencia artificial general, tal es el campo de investigación teórica de la IA, que intenta crear software con inteligencia similar a la humana y con la capacidad de autoaprendizaje) alcanzaría una “reacción incontrolable” a través de ciclos de autosuperación, en los que cada vez más rápidamente aparecerían generaciones nuevas y más inteligentes, causando una explosión de inteligencia que, cualitativamente, superaría de lejos a la inteligencia humana.

Tanto Stephen Hawking como Elon Musk expresaron su preocupación por las consecuencias de *singularity*, su potencial beneficio o sus catastróficos efectos sobre la raza humana, llegando a plantear que la “inteligencia artificial general” podría llegar a extinguir la humanidad. Iteraciones de mejoras autónomas recursivas podrían acelerarse hasta el punto en que la creación de la primera máquina con capacidad similar a la humana sería la última invención que el hombre necesitaría de ahí en más, siempre que esa máquina fuera suficientemente dócil para informarnos cómo controlarla.

Muchos otros autores han descartado o, al menos, planteado serias dudas acerca de que la humanidad alcance el punto de *singularity*, aduciendo que las computadoras no tienen inteligencia, ni motivación, ni autonomía, ni agencia (o sea, capacidad que tienen los humanos de transformar sus propias condiciones de vida). La polémica sigue abierta, pero ciertamente la mayoría de los métodos para crear mentes súper- o trans-humanas se relacionan con disciplinas como la bioingeniería, la ingeniería genética, las drogas nootrópicas (o inteligentes), los asistentes de IA, las interfaces directas cerebro-computadora y la transferencia mental. Es aquí donde, además de la intervención de los científicos y tecnólogos, se requiere la intervención estatal a través de políticas públicas activas.



## Razones de la preocupación

Macrodatos, analítica de datos, aprendizaje automático, robótica y múltiples aplicaciones tecnológicas, como la cadena de bloques, el internet de las cosas, el reconocimiento de identidad o las impresiones 3D, abren algunos de los interrogantes y dilemas más complejos de la actual era exponencial. Y, en tal sentido, generan una extraña mezcla de excitación, esperanza y temor que, a su vez, plantean serios desafíos en el diseño de políticas públicas responsables que no disuadan los avances tecnológicos, pero que, al mismo tiempo, impidan o desalienten los desarrollos potencialmente riesgosos para la seguridad, la ética, la salud o el bienestar de los ciudadanos.

En lo que concierne a la IA en particular, sus beneficios son innegables. Entre otras cosas, puede mejorar la eficiencia en el uso de energía y la seguridad en las viviendas. En el cuidado de la salud, permite a los profesionales ofrecer tratamientos personalizados a sus pacientes a partir de sus historias clínicas, síntomas y antecedentes genéticos. Los robots y asistentes virtuales alimentados por IA pueden proporcionar servicios de información a toda hora. Sus aplicaciones en materia de seguridad permiten detectar fraudes, identificar amenazas potenciales y predecir desastres naturales. En el transporte, la IA optimiza recorridos, reduce la congestión de tránsito y mejora la seguridad. También en la educación, las herramientas de enseñanza basadas en IA pueden ofrecer experiencias de aprendizaje personalizado con realimentación en tiempo real. Y al automatizar tareas rutinarias y optimizar procesos de gestión, es capaz de reducir costos operativos y aumentar las utilidades. Por último, para no extender el listado, la IA proporciona valiosas inferencias a partir del análisis de enormes masas de datos, permitiendo a organizaciones y personas adoptar mejores decisiones basadas en la evidencia.

Junto a estas indudables ventajas, la IA enfrenta grandes desafíos, que han generado creciente preocupación en las instituciones estatales, organizaciones de la sociedad civil y ciudadanos en general. Pero es el Estado, en sus múltiples roles como empleador, inversor, promotor, regulador o prestador de servicios, el que debería prevenir y orientar estas innovaciones tecnológicas, dada la disrupción que están produciendo en el mundo del trabajo, sus riesgos de generar un asfixiante control social,

las pautas culturales aberrantes que puede llegar a naturalizar o el ensanchamiento de la brecha digital que puede crear, entre otras consecuencias negativas. Cuando el acelerador se activa y la velocidad se vuelve incontrolable, es necesario accionar el amortiguador para cambiar el ritmo de la marcha. El Estado es, a veces, el único actor social capaz de cumplir ese doble rol: actuar como acelerador, promoviendo el proceso de innovación tecnológica, pero al mismo tiempo accionar el freno, para redireccionar y regular su marcha.

Como promotor, el Estado puede afectar recursos a través de la inversión o subsidio directo a actividades de I&D, a la formación de recursos humanos o a la creación de instituciones y centros de investigación públicos o mixtos, incluso mediante acuerdos con organismos multilaterales para la canalización de financiamiento o el desarrollo de proyectos multinacionales. Si bien el grado de incertidumbre que rodea al proceso decisorio estatal con relación al rol promotor es alto, lo es quizás más todavía cuando actúa como regulador.

Es que la velocidad del cambio tecnológico es tal, que no permite evaluar claramente sus beneficios frente a sus posibles perjuicios. Una de las mayores preocupaciones generadas por el ritmo exponencial del cambio tecnológico es su impacto sobre el mundo del trabajo, ya que, a diferencia de las revoluciones industriales previas, en que la destrucción de empleos creaba desocupación pero era finalmente compensada por la creación de nuevas ocupaciones, la era actual es menos optimista respecto de la futura, plena sustitución de empleos. Se considera que las dos terceras partes de los niños actualmente en edad escolar, cuando sean adultos desempeñarán puestos hoy inexistentes. Las estimaciones coinciden en que, en 2030, estarán automatizadas entre el 16% y el 30% de todas las ocupaciones. Con certeza, se perderán empleos pesados, repetitivos y de baja calificación, cambiará la jerarquía y retribución de los diferentes tipos de empleo y se modificará la estructura y las modalidades de trabajo, con tendencia a su precarización. Y se prevé que las ganancias de productividad tenderán a privilegiar al capital. Entre las opciones que se plantean los países del mundo desarrollado se incluye reformular las políticas de empleo para generar nuevas y mayores oportunidades de inserción de trabajadores en el mercado laboral, la revalorización de actividades

de interés social y los cambios en el perfil ocupacional del sector público.

Los esfuerzos gubernamentales también tendrán que ser dirigidos a la promoción/regulación de tecnologías específicas. Consideremos, por ejemplo, el caso de las criptomonedas, cuyas promesas llevaron a numerosos países a invertir e innovar en este campo tecnológico. Así, gobiernos de varias naciones ya han adoptado medidas para reducir o eliminar restricciones regulatorias para las industrias que desarrollan estas tecnologías. Existe bastante consenso en considerar que las mismas constituirán la próxima frontera de una economía soportada por internet. Pero para otros, las promesas de bitcoin y blockchain solo responden a la cultura elitista y la perspectiva tech-céntrica de la tecnología disruptiva actual.

Consideremos, como otro ejemplo, la robótica. Es indiscutible que las innovaciones en este campo producen enormes beneficios por sus múltiples aplicaciones. Pero también suponen enormes desafíos no solo para los responsables de adoptar políticas públicas, sino también para los líderes empresarios, la comunidad jurídica, las instituciones académicas y la ciudadanía, en tanto los robots se vayan incorporando progresivamente a la vida social. Sus impactos sobre el mundo del trabajo ya comienzan a percibirse. La robótica militar y sexual despierta enormes inquietudes y ya desvela a los reguladores. En un caso, por sus consecuencias para una renovada carrera armamentista. En el otro, por los impactos sobre la ética, la salud y la cultura de una práctica que roza el terreno del delito.

Además de promotor y regulador, el Estado es también proveedor de servicios, tal vez su rol esencial. Y en este papel la tecnología de la era actual puede ser una aliada o, como en las otras modalidades de intervención, una fuente de dificultades. Y para la prestación de servicios los gobiernos están adoptando la forma de plataformas, imitando al sector privado. Se utiliza el término “gobierno como plataforma” con referencia al ecosistema global de aplicaciones y componentes compartidos, estándares abiertos y sets de datos canónicos, así como a los servicios construidos sobre ellos y sobre los procesos gubernamentales, para brindar servicios a los ciudadanos. De este modo, las plataformas permiten a las agencias gubernamentales evitar los sistemas monolíticos cuyo mantenimiento requiere personal altamente especializado y escaso, además de ganar en velocidad, en capacidad de respuesta y de oferta de servicios que interoperan efectivamente.

Una de las tendencias más novedosas en este campo es el eventual pasaje desde el aclamado *one-stop government* (o sistema de “ventanilla única”) hacia el *no-stop government* o, incluso, el *government for one* o gobierno personalizado. Los desarrollos en materia de IA e interoperabilidad abrieron el campo de la

proactividad gubernamental y la posibilidad de proveer servicios públicos proactivos, sin que medie necesariamente una demanda ciudadana. De este modo, la gestión pública tradicional, esencialmente “reactiva”, dio paso a una nueva modalidad “proactiva” que produjo una auténtica revolución en la concepción y la práctica de satisfacer necesidades ciudadanas en la prestación de servicios públicos. Si bien su impacto es todavía limitado, en varios países (v.g., Estonia, Austria, Nueva Zelanda) ya se ofrece este tipo de servicios que, de hecho, invisibilizan al gobierno e impactan favorablemente sobre la predisposición de los ciudadanos a recibirlos, así como sobre la legitimidad gubernamental. Un caso muy difundido y exitoso, que ilustra bien la proactividad estatal, es la provisión por el gobierno de Estonia de subsidios familiares por nacimiento, política también implementada en Austria. Lo habitual en este tipo de prestaciones es que las familias que desean obtener este beneficio, deban realizar varios trámites para probar su elegibilidad para recibirlo, llenar papelería o, incluso, concurrir personalmente a una o más oficinas públicas. Bajo el modelo SPP, la elegibilidad es determinada automáticamente a partir de los registros hospitalarios de nacimientos, que al ser interoperables con otras bases de datos (fiscales, bancarios, de la seguridad social), permiten otorgar el beneficio sin que medie una solicitud explícita. Se trata de un servicio totalmente confiable, que reduce la arbitrariedad o el error humano en la decisión de otorgamiento, es más eficiente, más rápido y requiere menos infraestructura.

## Capacidades estatales necesarias

Naturalmente, la contrapartida de innovaciones semejantes es la capacidad institucional del Estado para desarrollar y poner a disposición de la ciudadanía este tipo de servicios. Esa capacidad depende de numerosos factores: a) el esfuerzo de desarrollo tecnológico que pueda desplegar o la promoción de este proceso a través de la regulación del mercado, el financiamiento estatal o los emprendimientos de colaboración público-privada; b) el reclutamiento y los programas de formación y capacitación de personal responsable de alimentar y actualizar, en forma permanente, los datos volcados a las carpetas ciudadanas, respondiendo a la vez a la consultas y propuestas de los usuarios; c) la protección de los datos de los ciudadanos a través de legislación y los controles ejercidos sobre las grandes oligopolios que dominan el mercado de la información y las redes sociales.

Por otra parte, sobre todo en los países en desarrollo, existen al menos otros tres desafíos que deben enfrentar sus gobiernos para avanzar en sus procesos de digitalización. Uno es considerar a internet como un bien público global, promoviendo su expansión para ampliar sus economías de escala, la creación de redes y las externalidades positivas, sea indirectamente a través

*Junto a estas indudables ventajas, la IA enfrenta grandes desafíos, que han generado creciente preocupación en las instituciones estatales, organizaciones de la sociedad civil y ciudadanos en general. Pero es el Estado, en sus múltiples roles como empleador, inversor, promotor, regulador o prestador de servicios, el que debería prevenir y orientar estas innovaciones tecnológicas.*

de la regulación del mercado de oferentes privados o directamente a través de la inversión pública en infraestructura y esquemas colaborativos público-privados. De otro modo, la brecha digital que separa a quienes no tienen problemas de acceso de quienes están excluidos por ser parte de la población vulnerable se ensanchará aún más.

Un segundo desafío es asumir el firme compromiso de fortalecer la ciberseguridad frente al riesgo cierto de que los sistemas informáticos sean hackeados. Está comprobado que el riesgo cibernético está creciendo para las organizaciones en general y los gobiernos en particular, en la medida en que los ciberataques aumentan en volumen, intensidad y sofisticación. El riesgo no es solamente financiero; también entraña la posibilidad de pérdida de confianza ciudadana, además de distraer la atención de la gestión respecto de otros temas prioritarios. Las cifras son inquietantes. En Estados Unidos, el Internal Revenue System (equivalente a nuestra AFIP) sufre cada año más de 1,4 miles de millones de ciberataques, dada la vulnerabilidad de los sistemas de identificación, que facilitan el acceso de los delincuentes cibernéticos a las cuentas de los contribuyentes. La misma herramienta –su programa de revisión de declaraciones juradas basado en IA– le permite identificar billones de dólares en pedidos de reintegro fraudulentos.

El tercer desafío exige adoptar una estrategia que conduzca al reemplazo de los llamados “legados informáticos” (*legacy systems*), es decir, tecnologías, sistemas o aplicaciones computacionales que han pasado de moda pero aún se encuentran en uso, planteando serios dilemas a los responsables de la gestión pública. Ingentes recursos presupuestarios y de personal son todavía afectados a mantener hardware y software obsoletos. Pero las plataformas sin soporte sobre las que corren pueden incrementar los riesgos de ciberataques y, por su inflexibilidad, no permiten modificar procesos adaptados a las herramientas de IA.

Si bien los países más avanzados están lejos de haber resuelto estos tres desafíos, su importancia es particularmente elevada en los más rezagados. Tampoco son los únicos desafíos a afrontar. También son necesarios consensos regionales para impulsar la rezagada economía digital en América latina en comparación con Europa o Asia. Y a pesar de que se registran algunas experiencias exitosas a nivel de país, para cerrar la brecha digital existente se necesitan mecanismos de coordinación más eficaces que promuevan un mercado regional más competitivo y sostenido que en la actualidad, eviten la duplicación de actividades y la descoordinación entre iniciativas.

Integrar digitalmente a la región requiere, entre otras cosas, homogeneizar los marcos legales, regulatorios y de mercado de las economías nacionales para generar mayor confianza, ampliar los circuitos comerciales y potenciar el intercambio de bienes y

servicios comercializados a través de plataformas digitales. Una mayor homogeneidad de los marcos institucionales tendería a derribar las barreras que impiden o dificultan el comercio transfronterizo. Por supuesto, también será necesario mejorar la infraestructura digital con conectividad de alta calidad, aumentar la confianza de los inversores y estimular el uso del comercio electrónico entre los más de 600 millones de consumidores de América latina.

Uno de los mayores desafíos es crear autoridades supranacionales en la región con capacidad de articular eficazmente el mercado digital. Los esfuerzos realizados en los diferentes bloques de integración subregionales no han tenido éxito en promover una visión común y una agenda que incluya prioridades, objetivos, metas, recursos, mecanismos de gobernanza y un cronograma acordado. Y si bien las distintas asociaciones regionales y subregionales han planteado cuestiones vinculadas con la economía digital, no han conseguido acordar mecanismos de coordinación orientados a concretar esa agenda común ni movilizar suficientes recursos financieros y humanos destinados a gestionar tal agenda.

Dejo para el final algunos comentarios sobre el que tal vez sea el mayor de los desafíos: cómo enfrentar el poder adquirido por las grandes corporaciones que hoy dominan el mundo digital. Algunas estadísticas son contundentes. Si se compara la capitalización bursátil de esas empresas y el PIB de los países, en miles de millones de dólares, la relación de fuerzas se inclina claramente en favor de esas empresas líderes. Por ejemplo, el valor del capital bursátil de Amazon es mayor que el PBI de 9 países sudamericanos; el de Microsoft, mayor al de 8 países de Europa del Este; y el de Facebook, superior al de la Argentina. Para colmo, estos grandes conglomerados, como Apple, Amazon o Alphabet, suelen contar con el favor y el apoyo de los consumidores, que tienden a oponerse contra los controles que tratan de imponer los gobiernos sobre ciertos servicios que brindan estas plataformas.

El AI Now Institute es, probablemente, la mayor institución dedicada exclusivamente a combatir el *tech power*, es decir, el poder concentrado de estos monstruos tecnológicos capaces de sortear con holgura las restricciones que pretenden imponerles, inclusive, los gobiernos más poderosos. En su último informe de 2023, esta institución plantea cuatro prioridades estratégicas que resultan cruciales en la coyuntura actual:

- 1- Poner en marcha acciones que trasladen la carga de la prueba sobre las grandes compañías para que demuestren que no están produciendo daños, en lugar de hacerlo sobre los usuarios o sobre los reguladores, que continuamente deben investigar, identificar y hallar soluciones para los daños, después que ocurren.
- 2- Romper los silos existentes en las diversas áreas de política, de modo de que estén mejor preparados para determinar dónde los avances en una agenda de política impactan a otras, ya que las firmas tecnológicas se aprovechan de la ventaja del aislamiento entre áreas y su falta de coordinación.
- 3- Identificar cuándo ciertas medidas de política son cooptadas o vaciadas por la industria digital, adoptando estrategias acordes con esos comportamientos.
- 4- Trascender el foco estrecho de las regulaciones específicas y adoptar estrategias que incorporen acciones en otros frentes (v.g. medio ambiente, resistencia de los trabajadores de la industria informática, derechos humanos).

Más específicamente, el AI Now Institute propone una serie de políticas.

- Restringir la ventajas que las firmas poseen en el manejo de datos, limitando su alto grado de concentración, así como su capacidad para recolectar datos sobre consumidores, o para producir datos sobre los mismos (también llamada “minimiza-

*Resulta fundamental que los gobiernos, sobre todo en los países tecnológicamente más rezagados, adopten una serie de políticas que les permitan afrontar los desafíos que genera el actual ritmo exponencial de la innovación, particularmente en el campo de la IA*

ción de datos”). Además, conectar la legislación sobre privacidad y competencia, que las empresas tratan de que se apliquen por separado. Controlar, asimismo, las posibilidades de fusiones y absorciones de empresas, para detener prácticas abusivas.

- Apoyar reformas que aumenten la competencia y reduzcan la concentración tecnológica. Desconfiar de la retórica desregulatoria basada en la carrera Estados Unidos-China. Implementar legislación antimonopolista más efectiva.
- Regular los modelos en gran escala como el ChatGPT y similares. Aplicar legislación existente para la rendición de cuentas pública del desenvolvimiento de los sistemas de IA generativa, previniendo los riesgos para los consumidores y la competencia. Monitorear de cerca las declamaciones de “apertura”, ya que la IA generativa tiene una dependencia estructural de los recursos, disponibles solo para contadas firmas.
- Desplazar a las auditorías como principal respuesta de política frente a aplicaciones dañinas de IA, dados los reparos que plantea la auditoría algorítmica actual. Las auditorías son insuficientes para afrontar el desequilibrio de poder entre las Big Tech y el público. Son necesarios estándares y metodologías de auditoría algorítmica más claros. Exigir a las compañías que demuestren que sus aplicaciones de IA no entrañan riesgos para los usuarios.
- Diseñar medidas más rigurosas frente al gradual desarrollo de la vigilancia biométrica y sus aplicaciones. Asegurar que la regulación biométrica incluya inferencias más amplias que la simple identificación. Imponer mayores restricciones a la minimización de datos, mediante medidas que reduzcan la expansión en la recolección de datos biométricos, en nuevos dominios como la realidad virtual o los vehículos autónomos.
- Implementar fuertes restricciones a la vigilancia algorítmica sobre los trabajadores, trasladando la carga de la prueba a los desarrolladores y las empresas. Establecer “líneas rojas” en áreas sensibles, como la selección y despidos automatizados, o con relación a ciertos tipos de tecnología, como el reconocimiento de emociones, que son inapropiadas en cualquier contexto.
- Prevenir los derechos preferenciales que surgen de acuerdos digitales internacionales, que pueden ser empleados para debilitar las regulaciones nacionales en materia de políticas de competencia y rendición de cuentas algorítmicas. Las garantías absolutas de secreto en los códigos fuente y algoritmos, en los acuerdos comerciales, no deben ser utilizados para reducir los esfuerzos de avanzar en la transparencia algorítmica.

## Conclusiones

En resumen, resulta fundamental que los gobiernos, sobre todo en los países tecnológicamente más rezagados, adopten una serie de políticas que les permitan afrontar los desafíos que genera el actual ritmo exponencial de la innovación, particularmente en el campo de la IA. Frente a estos desafíos, el Estado tiene un rol ineludible, en tanto acelerador, amortiguador o freno de ese proceso, de modo de aprovechar las ventajas del cambio y evitar sus amenazas éticas y culturales. Una mayor iniciativa en estas materias puede reducir el riesgo de una creciente dependencia tecnológica respecto de los países que lideran el desarrollo tecnológico.

En tal sentido, se destacó que el Estado puede actuar como formador de recursos humanos, promotor de nuevos emprendimientos, regulador y proveedor de servicios, destacándose su mayor proactividad y gradual conversión en una plataforma de servicios, como se explicó más arriba. También se hizo hincapié en la necesidad de fortalecer su capacidad institucional para asumir estos exigentes roles. Además de puntualizar algunos de los aspectos normativos y de gestión que requieren cambios o mejoras, también se destacaron algunos desafíos específicos, como convertir a Internet en un servicio público global, erradicar los *legacy systems* (o sea, los *sistemas de computación –software y hardware– obsoletos, pero todavía en uso, que son incapaces de crecer ni de utilizar IA, y son vulnerables a ciberataques*) o contener los riesgos en materia de ciberseguridad.

No es tarde para que los gobiernos de América latina y el Caribe reaccionen y asuman los compromisos que demanda la gestión pública en la era exponencial que estamos atravesando. Pero tampoco queda mucho tiempo.

LOS SISTEMAS INTELIGENTES CONFIGURAN RESPUESTAS A CIERTOS ESTÍMULOS, EN DETERMINADOS CONTEXTOS. EN ESTE TRABAJO SE ABORDAN LAS DISTINTAS TÉCNICAS EXISTENTES Y SE REALIZA UNA RECAPITULACIÓN DE LA SITUACIÓN EN LA CUAL FUERON CREADAS.

# TÉCNICAS DE LA INTELIGENCIA ARTIFICIAL

por **RAMIRO VERRASTRO**. *Ingeniero en Electrónica, docente de la materia Aprendizaje Automático (UTN-FRBA), miembro del Grupo de Inteligencia Artificial y Robótica (GIAR, UTN). Actualmente participa en diversos proyectos de desarrollo en Inteligencia Artificial, Visión por Computadora y Aprendizaje Automático para la empresa VENG SA.*



**L**as definiciones de Inteligencia son muchas, pero hay una corriente de pensamiento que se enfoca en aquellas técnicas que mejoran la habilidad computacional para resolver problemas. Puede plantearse como la capacidad de una máquina para elegir la alternativa más satisfactoria, en forma autónoma, para cumplir un objetivo específico. También debe incorporar la capacidad de sentir el ambiente, de aprender de la experiencia, de reconocer patrones, y de explicar las decisiones, entre otras.

El ser humano desde sus orígenes se inspiró en la naturaleza para construir herramientas que amplíen su capacidad tanto física como intelectual para la resolución de problemas específicos. Por eso es posible encontrar análisis explícitos y planteos teóricos de los métodos de razonamiento desde las primeras concepciones de la lógica (Solano Funez, 2024). El término Inteligencia Artificial (IA) se comenzó a acuñar a mediados del siglo XX. En ese momento parte de la comunidad científica empieza a publicar investigaciones en las cuales se buscaba simular la inteligencia humana, utilizando diferentes máquinas, mecanismos o algoritmos. Todo esto fue potenciado por la aparición de la primera computadora de uso comercial en la década de 1950 (The University of Manchester, 2008).

En esa primera ola de la inteligencia artificial, es de gran relevancia el “Test de Turing” (Turing, 1950), planteado por Alan Turing, para evaluar la inteligencia de las máquinas. Este concepto es utilizado en los “CAPTCHA” modernos (Von Ahn et al., 2003) que utilizan pruebas o tareas que son fácilmente realizadas por humanos, pero que desafían a las máquinas, para intentar diferenciar entre humanos y programas que imitan su comportamiento (*bots*).

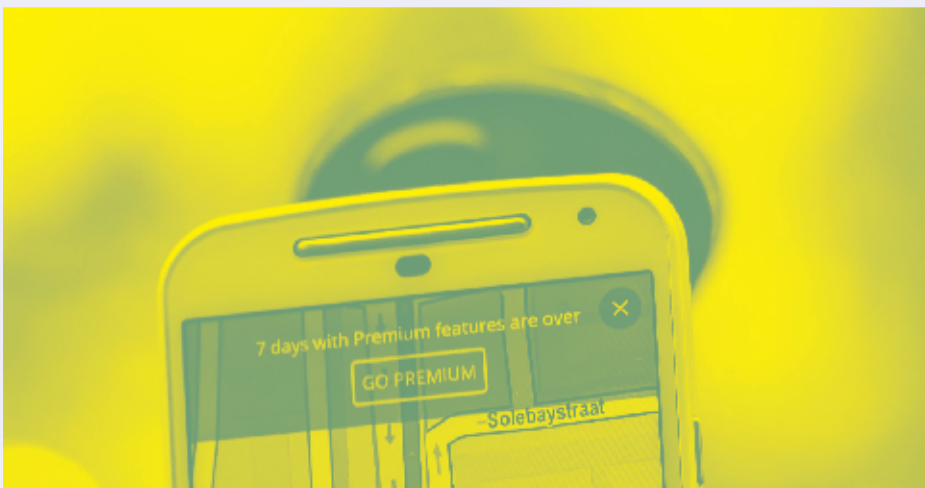
Podemos pensar la inteligencia artificial como un conjunto de técnicas o herramientas para construir sistemas, inspirados en el razonamiento humano, en la naturaleza y en la neurociencia, que se utilizan para resolver problemas y tomar decisiones de forma autónoma.

## Sistemas que toman decisiones

En la Ingeniería, al igual que en la Economía, se administra la escasez, en este caso la del hardware que da soporte físico a los algoritmos, así como el tiempo de respuesta y la energía empleada. La navaja de Ockham (1285-1347) es un concepto conocido desde hace mucho tiempo y puede sintetizar este esfuerzo por minimizar el costo computacional implementado en la práctica como el criterio de diseño KISS (*Keep it simple and straightforward*), plantea que a igualdad de resultados, la solución más sencilla, la de menor cantidad de parámetros, componentes o etapas es la mejor opción. Si con menos recursos se logran resultados similares, la solución es más eficiente. A continuación se exploran una serie de herramientas, que pueden aplicarse de manera efectiva para construir sistemas inteligentes que tomen decisiones informadas y por eso se utilizan ampliamente.

## Sistemas basados en reglas

Uno de los primeros intentos en automatizar la toma de decisiones fueron los sistemas que utilizan operaciones lógicas a partir de una serie de entradas y mediante mecanismos de inferencia que producen una salida. Estos mecanismos de inferencia pueden ser desde una simple tabla de verdad, o implementarse mediante árboles de decisión, o métodos como el General Problem Solver (Newell et al., 1958), en el cual se conectan una serie de decisiones en cascada para representar salidas complejas a partir de los datos de entrada. Estas técnicas alcanzaron algún éxito cuando se aplicaron en ámbitos específicos y dieron origen a los Sistemas Expertos, que extraen el conocimiento de expertos humanos y automatizan su razonamiento mediante reglas de producción.



## Sistemas expertos

Los sistemas expertos se modelan en base al razonamiento y los conocimientos específicos de las personas u organizaciones en la resolución de un determinado problema. Típicamente están compuestos por un conjunto de datos o hechos, un conjunto de reglas que captura el conocimiento experto, y un algoritmo de inferencia, que utiliza algoritmos de lógica binaria, lógica difusa o razonamiento probabilístico para calcular una respuesta. Estos sistemas utilizan reglas como “Si mueve la cola y ladra entonces es perro” y “Si es perro entonces puede morder”. Estas reglas se administran mediante una máquina de inferencias que es capaz de arribar a una conclusión a partir de determinados datos.

Una de las ventajas es la explicabilidad del resultado, ya que se puede determinar el detalle del razonamiento que se aplicó para obtener la salida o conclusión.

Podemos citar como ejemplo la implementación del Control Estadístico de Procesos (CEP). Uno de los casos notables en la automatización de estos controles fue realizado por Toyota para asegurar la calidad, y detectar potenciales defectos en los procesos de fabricación, Sistema de Producción Toyota (TPS).

Otro ejemplo notable es MYCIN (Shortliffe, 1977), desarrollado en década de los setenta. Su principal uso fue el diagnóstico de enfermedades infecciosas de la sangre, proporcionando además el proceso para arribar al diagnóstico, así como de recetar los medicamentos necesarios para cada caso según las particularidades de cada paciente (por ejemplo: edad, estatura, peso, entre otros). Los sistemas basados en reglas alcanzaron los primeros éxitos prácticos de la IA y se usaron en el diagnóstico médico, prospección minera, gestión empresarial, planificación logística, entre otros.

## Control difuso

En muchas situaciones los expertos no pueden afirmar con seguridad la validez de una regla, no obstante pueden operar y sacar conclusiones en contextos de información incompleta o con un grado de incertidumbre. Los algoritmos que emplean lógica difusa (Sorin, 2022) modelan situaciones y conceptos, a los que no se les puede asignar una categoría de “verdadero” o “falso”, sino que tienen grados de verdad intermedios. Esto permite trabajar con datos que no estén claramente determinados. Para una persona puede ser sencillo manejar estos conceptos vagos como por ejemplo “hay mucha gente”, “la comida está fría” o “llegaste rápido”. En la lógica clásica estas expresiones pueden ser verdaderas o falsas (0 o 1), en cambio en la lógica difusa la asignación del valor de verdad es un número entre 0 y 1. Esto permite interpretarlo luego en términos de posibilidades. La lógica difusa permite capturar y modelar la incertidumbre y la imprecisión del mundo real, para la toma de decisiones con información vaga o incompleta.

En el contexto de sistemas cuyos modelos estáticos y dinámicos serían extremadamente complejos de desarrollar, la lógica difusa ofrece una solución elegante. En lugar de depender de ecuaciones integrales y diferenciales que pueden ser difíciles de tratar matemáticamente, podemos utilizar reglas simples basadas en la experiencia.

Por ejemplo, en un sistema de control de temperatura para un horno industrial, en lugar de tener un modelo matemático preciso que describa el comportamiento térmico del horno y luego desarrollar un controlador que utilice ese modelo, podríamos simplemente establecer reglas difusas del tipo “Si la temperatura es alta y está subiendo rápidamente, entonces disminuir bastante la potencia del horno”.

Como caso de éxito podemos citar la aplicación al control de la velocidad del subte japonés en la ciudad de Sendai (Singh, 2013), pudiendo contemplar por ejemplo la cantidad de pasajeros, las condiciones y las características de las vías, entre otras variables. Todo esto ostentando una gran suavidad en el arranque y frenado, comparado con los sistemas existentes en ese momento, y siendo un 10% más eficiente energéticamente que el manejo del acelerador que realizaba el personal.

Actualmente el control difuso se emplea tanto en máquinas lavarropas como en aplicaciones industriales de automatización y control, entre otros.

## Máquina de estados

La teoría de autómatas trata de una máquina teórica cuyas salidas dependen no solo de las señales de entrada actuales, sino también de las anteriores. Puede modelarse con un conjunto de estados que dependa de los valores pasados, de forma tal que la salida dependa únicamente del estado y de los valores actuales de las entradas. Esto permite modelar autómatas, cuyo comportamiento varía en función del tiempo y del contexto. Como caso especial no podemos dejar de mencionar la “Máquina de Turing” (Turing, 1936), utilizada como modelo teórico para las computadoras. Una de las aplicaciones de estos modelos es la clasificación de los problemas en función de su complejidad, es decir, la cantidad de recursos necesarios para resolverlos. Es así que existen problemas que pueden resolverse en tiempos polinómicos (donde el tiempo empleado es una potencia del tamaño del problema) y problemas no polinómicos (donde el tiempo necesario para resolverlo es exponencial con respecto al tamaño).

Los problemas que se resuelven con algoritmos no polinómicos son llamados algoritmos de la muerte, porque se tornan irresolubles en tamaños relativamente pequeños. Uno de estos problemas es el del Viajante de Comercio (Hoffman et al., 2013) que plantea el problema de visitar  $N$  ciudades y volver al inicio por el camino más corto. Para 10 ciudades solo hay que calcular 181.440 caminos posibles pero para 100 ciudades hay que calcular  $4,7 \cdot 10^{157}$  posibilidades, (4,7 seguido de 157 ceros), que es un número similar a las jugadas posibles de una partida de ajedrez promedio. Este tipo de problemas no es abordable con técnicas de cómputo por “fuerza bruta”, es decir, calculando todas las posibilidades. Los algoritmos de búsqueda tratan de lidiar con este tipo de problemas.



## Algoritmos de búsqueda

Los algoritmos de búsqueda se utilizan para encontrar una lista o secuencia de pasos o acciones a realizar desde un estado inicial para alcanzar un objetivo específico. Podemos clasificarlos en dos grandes grupos.

Las Búsquedas No Informadas se usan cuando no se tienen conocimientos específicos sobre cómo realizar la búsqueda de la solución de un determinado problema. Pueden utilizarse diferentes estrategias como las siguientes. Búsqueda primero en anchura (BFS, por sus siglas en inglés) explora todos los nodos vecinos o puntos más próximos antes de moverse a los siguientes niveles de profundidad. La búsqueda primero en profundidad (DFS, por sus siglas en inglés) avanza lo más lejos posible a lo largo de cada rama o camino, priorizando explorar los nodos más lejanos al punto de partida. Tiene ventajas en requerimientos de memoria con la anterior ya que minimiza la cantidad de caminos que deben considerarse en forma simultánea.

Las Búsquedas Informadas son algoritmos que permiten encontrar soluciones a los problemas de manera más eficiente usando conocimientos previos del problema. Uno de ellos es el algoritmo A\* (A estrella), utiliza una función de evaluación heurística para encontrar el camino más prometedor, que pondera tanto el costo computado desde el inicio para alcanzar un nodo actual, como el costo estimado (heurística) para alcanzar el objetivo desde el nodo actual. Esta función de evaluación debe cumplir ciertos requisitos, ser no negativa, asegurando que los costos siempre aumenten, y admisible, para no sobreestimar el costo real al objetivo. Bajo estas condiciones garantiza encontrar la solución óptima. Este algoritmo (Hart et al. 1968) ha sido ampliamente utilizado en la industria y es un método clave para la navegación de robots y vehículos autónomos, así como en los sistemas de navegación por GPS, aplicaciones de transporte y optimización de las cadenas de suministro.

Una técnica más sencilla y de uso frecuente es la Búsqueda voraz (Greedy Best-First Search). Utiliza solo la función heurística que estima el costo hasta el objetivo. Prioriza los nodos basándose únicamente en esta función, y buscando directamente el camino menos costoso desde la posición actual. Si bien puede ser más rápida en alcanzar el objetivo, no necesariamente encuentra la solución óptima. Estos algoritmos se usan en Sistemas de Navegación, Distribución y logística y Sistemas Interactivos de Diálogo, entre otros. No obstante sus buenos resultados, hay problemas donde hay muchas variables de entrada y muchos estados posibles cuya combinación produce un número de alternativas intratables. Este fenómeno es conocido como la “maldición de la dimensionalidad”. En estos casos los métodos de búsqueda estocástica como los algoritmos genéticos producen resultados notables.

*En el contexto de sistemas cuyos modelos estáticos y dinámicos serían extremadamente complejos de desarrollar, la lógica difusa ofrece una solución elegante. En lugar de depender de ecuaciones integrales y diferenciales que pueden ser difíciles de tratar matemáticamente, podemos utilizar reglas simples basadas en la experiencia.*

### Algoritmos genéticos

Los algoritmos de búsqueda en grafos (conjunto de nodos, conectados entre sí) no siempre son eficientes cuando tratan problemas de optimización que requieren respuestas rápidas, aunque no sean óptimas. Por esto, surgen otros modelos de búsqueda inspirados en los mecanismos biológicos de la evolución de las especies. Siendo los individuos de una especie las posibles soluciones del problema que se busca resolver, y la calidad de la solución se evalúa mediante una función de aptitud, también denominada *fitness*.

El proceso comienza con la inicialización de una población de individuos (soluciones) a los que, considerando su aptitud, se selecciona para pasar a la próxima generación. Se aplican además de la selección, los operadores de cruce y mutación, que permiten el intercambio de “material genético” y la alteración azarosa de las soluciones que integran la nueva generación. El mecanismo de cruce combina dos individuos para producir otros dos, en la esperanza de que esta combinación sea una mejor respuesta al problema en cuestión, es decir, una descendencia potencialmente mejor adaptada. La mutación genera cambios aleatorios en los individuos de la población, introduciendo diversidad y permitiendo explorar otras alternativas dentro del espacio de búsqueda. Si bien es una búsqueda aleatoria de soluciones, está bien establecido el principio de funcionamiento (Holland, 1992), que al igual que en la naturaleza, guía el proceso de búsqueda de soluciones, generación tras generación. Este enfoque ha sido ampliamente utilizado en la industria desde sus primeras publicaciones.

Como ejemplo notable podemos citar el diseño de la antena de telecomunicaciones de la nave espacial de la NASA ST5 en el 2006 (Fisher, 2024) y en otras áreas como Robótica y control, Ingeniería, Telecomunicaciones, Visión por computadora, Finanzas, Logística e Ingeniería y diseño.

### Sistemas que aprenden

El aprendizaje es un aspecto muy importante de la inteligencia, e implica la capacidad de adquirir, asimilar y aplicar nueva información o habilidades, por medio de la experiencia propia o ajena, lo que puede resultar en cambios duraderos en el comportamiento del algoritmo y su capacidad para resolver problemas. Este tipo de soluciones necesita de una etapa de “entrenamiento”. Desde un enfoque clásico de la IA, los desarrolladores son los encargados de elaborar los filtros para extraer determinadas características de los datos de entrada. Como pueden ser valor medio, máximos y mínimos de una serie temporal. En el caso del procesamiento de imágenes, la elaboración de filtros que identifican colores, formas, texturas, cambios abruptos. En el caso del procesamiento de texto, la cantidad de veces que se utiliza determinado término. Con estas características ya procesadas, se pueden entrenar algoritmos de Aprendizaje Automático (Machine Learning) para aprender a tomar la decisión correcta (calibrar los parámetros internos de los algoritmos) en base a una determinada función de costo, que dependerá del problema que se desea resolver.

El Aprendizaje Supervisado se caracteriza por utilizar un conjunto de datos etiquetados, cada ejemplo de entrada está asociado con una etiqueta (una salida conocida), y el objetivo principal es relacionar las entradas con las salidas a través de la minimización de una función de error, se busca que las salidas que genera el modelo coincidan con las etiquetas originales. Como ejemplo podemos citar la clasificación de correos electrónicos como *spam*. Los datos se etiquetan con la ayuda de los usuarios que mueven algunos correos a la carpeta de *spam*. El algoritmo identifica características como las palabras clave o la frecuencia de ciertas palabras y aprende a clasificar automáticamente nuevos correos electrónicos como *spam*.

El Aprendizaje No Supervisado no requiere que los datos tengan etiquetas, por lo que no hay información sobre las salidas esperadas. Su propósito es descubrir y capturar patrones ocultos, estructuras características o distintas relaciones entre las variables de entrada. Permite también ordenar los datos en grupos afines, reducir la cantidad de variables de entrada, así como explorar y presentar de manera organizada grandes conjuntos de datos. Como ejemplo podemos citar la segmentación de los clientes. Teniendo los datos de transacciones de clientes, pero no las etiquetas que indiquen a qué segmento pertenecen, un algoritmo de aprendizaje no supervisado puede agrupar automáticamente a los clientes en segmentos basándose en sus patrones de compra y comportamiento, permitiendo adaptar estrategias de ventas de manera más efectiva.

El Aprendizaje por Refuerzo busca aprender mediante la experimentación en ambientes dinámicos. Durante el desarrollo de estas experiencias el modelo recibe como realimentación recompensas y castigos, el objetivo es encontrar una estrategia que maximice el balance entre recompensas y castigos, además se establece una política para seleccionar y explorar las distintas acciones posibles que permiten avanzar durante cada experimento. Como ejemplo podemos citar el entrenamiento de un agente para jugar al ajedrez. El agente (controlado por el algoritmo) juega repetidamente contra sí mismo o contra otros oponentes. Después de cada juego, recibe una recompensa (por ejemplo, +1 si gana, 0 si empata, -1 si pierde). Utilizando estas recompensas como retroalimentación, el agente ajusta su estrategia y mejora gradualmente su desempeño a medida que juega.

## Redes neuronales

Dada la importancia actual de las redes neuronales y sus arquitecturas de aprendizaje profundo, en otro artículo de este número se tratan en detalle el estado del arte y sus aplicaciones.

## Conclusiones

En este artículo se buscó presentar algunas de las técnicas que actualmente tienen mayor relevancia en la inteligencia artificial, explicando en qué consisten y dando algunos datos del contexto histórico. Este es un campo floreciente y en pleno crecimiento, irrumpiendo en cada rincón de la vida cotidiana y la industria. Esto se pudo dar en parte gracias al avance de la capacidad de cómputo que proporcionan las placas de procesamiento GPU, TPU y otras tecnologías especializadas en inteligencia artificial, como las FPGA, y la gran cantidad de datos disponibles que proporcionan los sistemas de almacenamiento en la nube, así como la eficiencia y complejidad alcanzada por los algoritmos de IA. En este contexto surgen las redes profundas, en donde la extracción de características puede ser realizada en forma automática, siendo este el enfoque actual de la IA, donde los algoritmos de Aprendizaje Automático Profundo (Deep Learning) tienen la capacidad de aprender directamente de los datos. Este tipo de algoritmos se caracteriza por tener un gran número de parámetros entrenables, por lo cual son capaces de aprender funciones complejas y requieren gran capacidad de cómputo. Actualmente hay una gran cantidad de áreas en las que se están aplicando, desde el procesamiento de audio hasta predicciones en el mercado de valores.

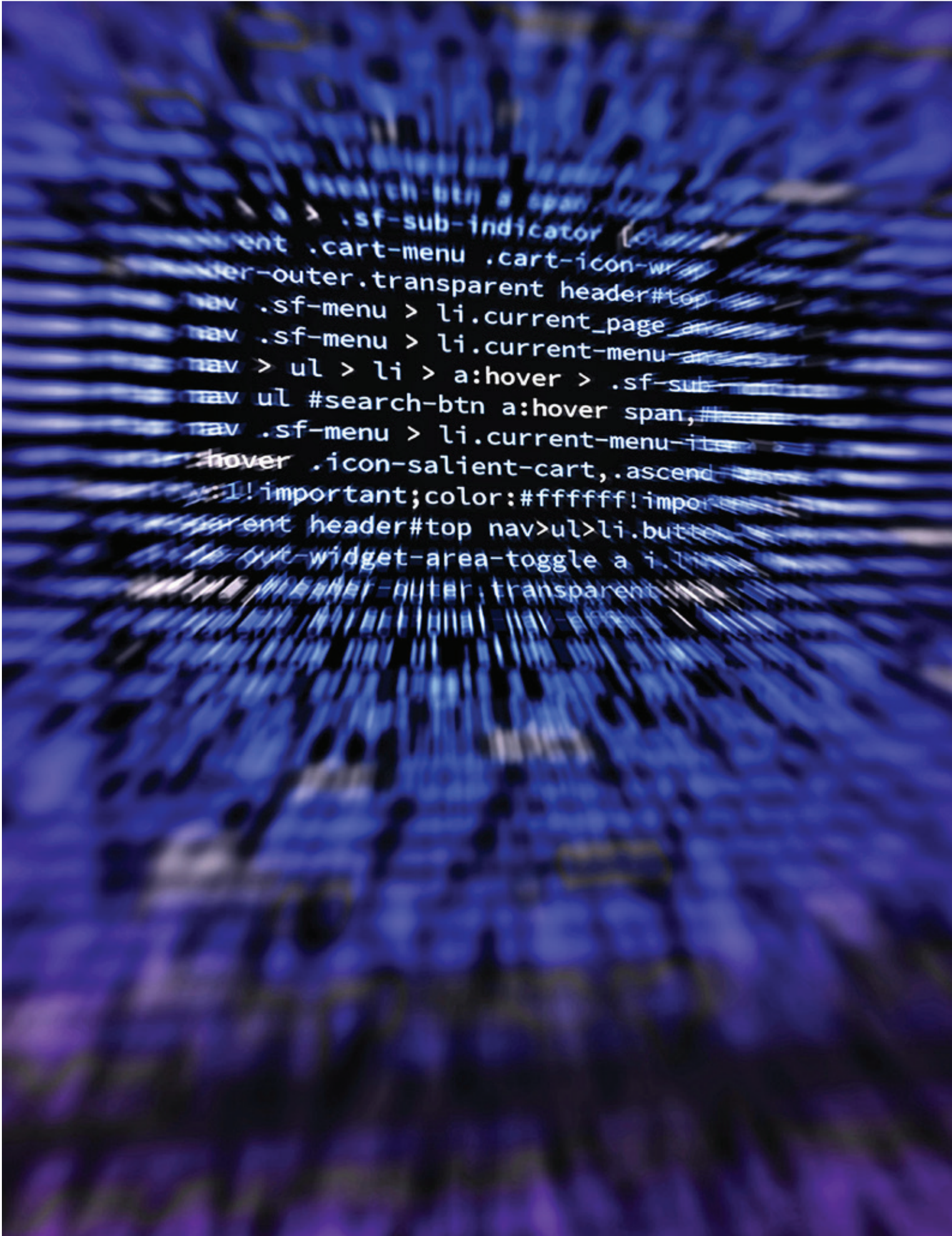


*El Aprendizaje por Refuerzo busca aprender mediante la experimentación en ambientes dinámicos. Durante el desarrollo de estas experiencias el modelo recibe como realimentación recompensas y castigos.*



### Bibliografía de referencia

- Solano Funez, L.D. (2024). "La lógica y su historia". <https://www.timetoast.com/timelines/2316233>
- The University of Manchester. (2008). "The Ferranti Mark 1". <https://curation.cs.manchester.ac.uk/digital60/www.digital60.org/birth/manchestercomputers/mark1/ferranti.html>
- Turing, A. (1950), "Computing Machinery and Intelligence", *Mind*, LIX (236): 433–460, doi:10.1093/mind/LIX.236.433, ISSN 0026-4423
- Von Ahn, L., Blum, M., Hopper, N. J., Langford, J. (2003) "CAPTCHA: using Hard AI problems for security." *Advances in Cryptology, Eurocrypt*.
- Newell, A., Shaw, J. C., Simon, H. A. (1958). "Elements of a theory of human problem solving." *Psychological Review*, 65(3), 151–166. <https://doi.org/10.1037/h0048495>
- Shortliffe E.H. (1977) "Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases". Annual Meeting of the Society for Computer Medicine. 66–9. PMID: PMC2464549.
- Sorin, Nadaban. (2022). Fuzzy Logic and Soft Computing—Dedicated to the Centenary of the Birth of Lotfi A. Zadeh (1921–2017). *Mathematics*. 10. 3216. [10.3390/math10173216](https://doi.org/10.3390/math10173216).
- Singh, H. (2013) "Real-Life Applications of Fuzzy Logic", *Hindawi Open Access Journal, Advances in Fuzzy Systems*, vol. 2013, Article ID 581879. <https://doi.org/10.1155/2013/581879>
- Turing, A.M. (1936). "On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*. 2 (1937) 42: 230-265. doi:10.1112/plms/s2-42.1.230
- Hoffman, K.L., Padberg, M., Rinaldi, G. (2013). "Traveling Salesman Problem." In: Gass, S.I., Fu, M.C. (eds) *Encyclopedia of Operations Research and Management Science*. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4419-1153-7\\_1068](https://doi.org/10.1007/978-1-4419-1153-7_1068)
- Hart P., Nilsson N., Raphael B. (1968) "A formal basis for the heuristic determination of minimum cost paths" *Systems Science and Cybernetics, IEEE Transactions*, vol. 4, pp. 100–107
- Holland, J. (1992). "Adaptation in Natural and Artificial Systems" Cambridge, MA: MIT Press. ISBN 978-0262581110.
- Fisher, D. K. (2024). "Evolved Antenna". Jet Propulsion Laboratory. NASA <https://www.jpl.nasa.gov/nmp/st5/TECHNOLOGY/antenna.html>



por **RAMIRO GERMÁN RODRÍGUEZ COLMEIRO**. *Ingeniero Electronico UTN, Magister en Optimizacion de Sistemas complejos UTT, Doctor en electronica mencion en Analisis de Señales e Imagenes UTT/UTN*

# REDES NEURONALES ARTIFICIALES

EL AUTOR EXPLICA EL MODO EN QUE SE DESARROLLAN SISTEMAS QUE PERMITEN PROCESAR Y GESTIONAR UNA GRAN CANTIDAD DE DATOS, ASÍ COMO SUS COMPLEJIDADES Y RIESGOS.

Las redes neuronales artificiales (RNA) son un tema omnipresente en la inteligencia artificial. Si bien se inspira en el funcionamiento del cerebro humano, aún hay una gran distancia entre ambos. En este artículo se explora su funcionamiento y la evolución de la técnica, desde las simples neuronas artificiales hasta los modelos profundos utilizados en la actualidad. Se abordará con especial énfasis los modelos profundos y generativos, como los modelos de lenguaje, y se comentará sobre su naturaleza, limitaciones y riesgos.

## Redes neuronales artificiales

El concepto detrás de las RNA fue desarrollado en el siglo pasado, en la década de 1950, inspirado (libremente) en el funcionamiento del cerebro.

En general, lo que se pretende con estas RNA es que dado una serie de entradas  $X_n$  se obtenga una salida  $Y$ . Por ejemplo, si el conjunto de entradas  $X$  representa los píxeles de una imagen,  $Y$  podría ser una variable de decisión que indique si la imagen contiene panda. Tal como se haría con un humano, las RNA son entrenadas para responder de la manera deseada, en nuestro ejemplo, sería mostrarle un conjunto de imágenes, algunas que contengan el objetivo (un panda) y otras que no e ir indicando a la RNA cuando su salida debe ser positiva. Este aprendizaje se grabará en las activaciones de la RNA generando una representación interna de que es un panda. Ahora bien, la técnica dista de este concepto.

Las neuronas artificiales de las RNA convencionales se denominan perceptrones y simulan el comportamiento de una neurona biológica.

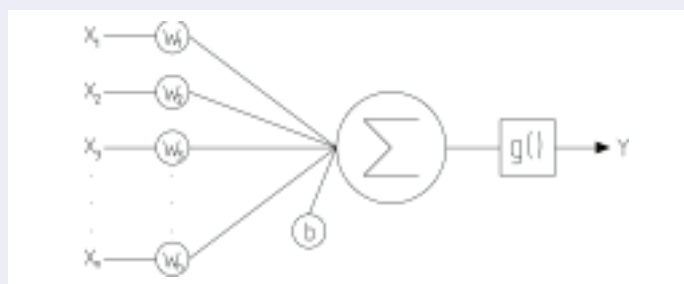


Figura 1. Neurona artificial o perceptrón

Como se observa en la figura 1, los perceptrones son sencillos, cuentan con una serie de entradas ( $X_1$  a  $X_n$ ) las cuales actúan como dendritas artificiales y a su vez asignan un peso dado a cada entrada (pesos  $W_1$  a  $W_n$ ). Las neuronas también cuentan con una entrada de valor fijo ( $b$ ) que sirve como un ajuste general. Todas estas entradas son sumadas y procesadas por una función de salida ( $g$ ) que en general es alineal e intenta representar

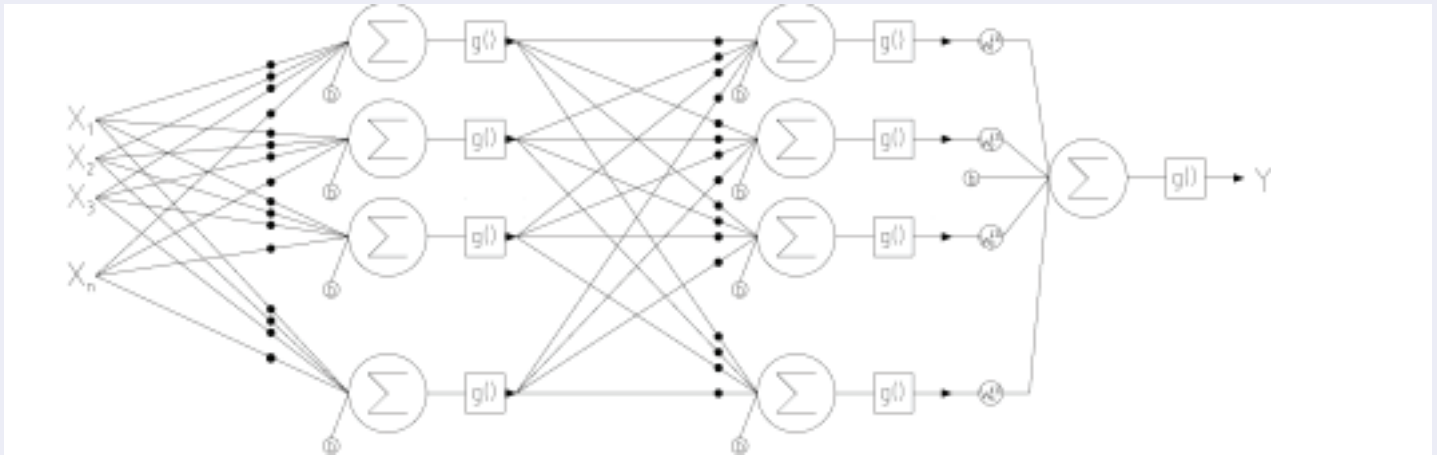
las activaciones de las neuronas biológicas. Matemáticamente se expresa como:

$$y = g\left(\sum_{i=1}^{n_h} w_i x_i + b\right) = g(\bar{W}^T \bar{X} + b),$$

Estas neuronas artificiales son capaces de adaptarse modulando la importancia que les asignan a sus entradas, cambiando el valor de los pesos  $W_1$  a  $W_n$ .

Como una neurona biológica, el perceptrón es limitado y su utilidad se revela al interconectar varias neuronas, buscando lograr una suerte de sinapsis artificial. A lo largo del siglo XX los esfuerzos se centraron en cómo organizar grupos de neuronas artificiales y cómo lograr que la RNA aprenda a resolver problemas. Este modelado básico fue desarrollado en el año 1965 y conceptualmente ha cambiado poco. La forma más simple de interconectar perceptrones es “hacia adelante”: la salida de un perceptrón se conecta a la entrada de otro. Esta estructura crea una red neuronal artificial de tipo *feedforward*.

Figura 2. Red neuronal artificial compuesta de múltiples capas de perceptrones.



Esta estructura es más compleja, ya no solo hay  $N$  parámetros sino que ahora tenemos  $N$  parámetros por cada neurona, siendo la expresión matemática resultante:

$$y = g \left( \sum_{k=1}^{n_h^{[2]}} w_k^{[2]} g \left( \sum_{j=1}^{n_h^{[1]}} w_j^{[1]} g \left( \sum_{i=1}^{n_h^{[0]}} w_i^{[0]} x_i + b^{[0]} \right) + b^{[1]} \right) + b^{[2]} \right)$$

Vemos que la construcción de la salida sigue siendo modular, pero la cantidad de pesos  $W$  es mucho más grande.

Si bien la forma de cómo se conectan las neuronas es importante, esta es fija y la forma en que las RNA “aprenden” es por medio de la modificación de sus pesos  $W$ . Entonces podemos simplificar la notación la red por una simple función  $Y = f(X, W)$ , que nos dice que la RNA es una función que produce una salida  $Y$  dada una entrada  $X$  y un conjunto de pesos  $W$ . La complejidad está en encontrar los pesos que resulten en el comportamiento deseado.

Al inicio del entrenamiento, los pesos se crean de manera aleatoria y por lo tanto el comportamiento inicial de la red es errático. Con ayuda de una métrica o función objetivo, el proceso de entrenamiento modifica los pesos para ajustarse al comportamiento deseado. La forma de estas funciones objetivo va a depender de la tarea a realizar.

Implementar este entrenamiento, como manejar matemáticamente los miles o millones de pesos  $W$ , es un problema complejo, tanto que recién en 1986 se descubre un algoritmo capaz de entrenar redes neuronales de más de tres capas, denominado *backpropagation* (Rumelhart et al, 1986).

Este método numérico utiliza el error de entrenamiento para ir imputando responsabilidad a los distintos pesos  $W$  y modificán-

dolos para reducir error. Matemáticamente, el algoritmo permite conocer el gradiente de la función de la red neuronal sobre cada peso  $W$ , dado una par de entradas  $X$  y salidas  $Y$ , luego de aplicar una función objetivo  $L(Y, X, W)$ .

Entonces, volviendo a nuestro ejemplo, para entrenar la red solo necesitaremos imágenes de pandas (set de  $X$ ) con su correspondiente anotación de si la imagen es o no un panda (set de  $Y$  asociado) y luego aplicar el algoritmo de *backpropagation* sobre cada imagen de nuestro conjunto, actualizando los pesos  $W$  hasta que el error sea bajo.

Hasta ahora comentamos el funcionamiento de una red básica, la pregunta que nos hacemos es: ¿por qué no escuchamos de la aplicación de estas técnicas hasta hace solo unos años?

La respuesta es la capacidad de cómputo y datos existentes en ese momento. Con las restricciones tecnológicas de esa época, no era posible entrenar modelos con muchos parámetros ya que el costo computacional era prohibitivo y los datos disponibles eran muy pocos. Esto llevó a que la técnica cayera en desuso durante la década de 1990 y 2000, en favor de métodos más efectivos dadas las limitaciones.

Para profundizar en este tema se refiere al lector al capítulo 6 de Duda, R.O., & Hart, P.E. (2006). *Pattern classification*. John Wiley & Sons.

## El aprendizaje profundo o deep learning

En el año 2012, un grupo de investigadores logra una mejora sustancial en un problema de reconocimiento de caracteres utilizando RNA (Krizhevsky et al, 2012). Allí se demostró que con el avance tecnológico las redes neuronales ya podían ser entrenadas en mucha mayor profundidad, de manera estable y con más conexiones. Poco cambió en la técnica de entrenamiento, el cambio se dio en la evolución de la capacidad de procesamiento. Este hito dio inicio al actual auge de las redes neuronales artificiales, específicamente marcó el inicio del aprendizaje profundo.

La primera ola del aprendizaje profundo se dio por las llamadas redes convolucionales y su aplicación sobre tareas de visión. Estas redes, conocidas desde 1988 e inspiradas en la corteza visual del cerebro, proponen una construcción jerárquica de las distintas capas de la RNA, logrando que la información espacial sea procesada con mayor efectividad. En ellas la entrada se descompone en mapas más pequeños que responden a características aprendidas (colores, formas, texturas), que representan la entrada con menos elementos. Luego de pasar por varias capas de estos filtros, la entrada tiene una dimensión mucho menor y es posible procesarla por una red tipo *feedforward*.

Esta técnica había sido imposible de implementar debido al costo computacional que conllevaba el entrenamiento, pero para la década de 2010 los procesadores gráficos o GPUs ya estaban altamente desarrollados y permitían acelerar drásticamente los algoritmos utilizados en el entrenamiento.

Las aplicaciones de esta técnica dieron nacimiento a muchas aplicaciones prácticas como por ejemplo:

- Reconocimiento de caracteres (Tesseract).
- Segmentación de imágenes y videos (MaskRCNN).
- Asistencia en diagnóstico por imágenes (implementaciones mayormente cerradas).

El siguiente paso en las RNA se dio con la creación del “transformador”, mejor conocido por su nombre en inglés *transformer* (Vaswani et al, 2017). A diferencia de las redes convolucionales, estas trabajan con datos secuenciales como frases, código o música. Los *transformers* se especializan en comprender secuencias y generar secuencias. Las secuencias de salida se crean otorgándole un valor de importancia (o atención) a cada uno de los elementos que componen la secuencia de entrada. Para ello los *transformers* utilizan un peso que conecta y compara cada uno de los elementos de la secuencia de entrada con todos los otros. Por ejemplo, si la secuencia es una frase como “La camioneta es

*Los transformers se especializan en comprender secuencias y generar secuencias. Las secuencias de salida se crean otorgándole un valor de importancia (o atención) a cada uno de los elementos que componen la secuencia de entrada.*

roja, ¿de qué color es el vehículo?”, el *transformer*, entrenado para comprender sintaxis en español, asignará a cada palabra una relación con otra, en este ejemplo la combinación de “roja-color” tendrá asignada una mayor importancia que la relación “camioneta-qué”. El *transformer* evalúa todas las entradas a la red contra las otras entradas, les asigna un valor y propaga el resultado hacia la siguiente capa.

En un modelo basado en *transformers*, toda esta operación compone un solo bloque de procesamiento, así como en una red convolucional se componían los distintos filtros. Es fácil de imaginar que la cantidad de pesos que hay que ajustar para este tipo de redes aumenta rápidamente con la cantidad de elementos de la secuencia de entrada, por este motivo los modelos más conocidos de este tipo de red, como el ChatGPT, poseen decenas o cientos de miles de millones de parámetros.

El uso de estas *transformers* comprenden la última generación de aplicaciones, entre ellas están:

- Modelos de lenguaje / “chatbots”, como ChatGPT, Gemini o Perplexity - Generación de partituras, como Musenet.
- Análisis de texto, con modelos basados en BERT (Devlin et al, 2018) - Análisis de secuencias de ADN (modelos cerrados).

Para profundizar en los tópicos de aprendizaje profundo, se recomienda el libro de LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning. nature*, 521(7553), 436-444.

## Redes generativas

Hasta ahora hablamos de las RNA desde el punto de vista de su construcción o topología, pero otro aspecto importante a analizar es el tipo de objetivo o uso que se les intenta dar. Originalmente las RNA fueron aplicadas para tareas de reconocimiento o para aproximar señales u otras funciones matemáticas. Todas estas tareas tenían un objetivo concreto o una “verdad absoluta” para poder entrenarlas. Sin embargo, en los últimos años se han desarrollado redes cuya salida pretende generar contenido novedoso, es decir que no formaba parte del conjunto de datos de entrenamiento ni tampoco es el resultado de una interpolación entre los ejemplos vistos por la red. A este nuevo conjunto de redes se las llama redes generativas y son las responsables de lo que se denomina el contenido “sintético” de la internet actual.

Estas RNA son entrenadas para copiar la distribución estadística de su conjunto de entrenamiento. Es decir, la entrada ya no es importante, lo importante es solo que la salida sea imposible de diferenciar de un conjunto de datos dado. Para ejemplificar, si disponemos de muchas imágenes de pandas, el objetivo de la

red ahora sería que dada una entrada aleatoria, la salida sea una imagen de un panda. En estos casos lo que la RNA intenta capturar en sus pesos es una representación condensada del espacio muestral (imágenes de pandas) y al ser excitada por una entrada aleatoria, proyectar esa entrada hacia el espacio muestral (convertir una serie de números aleatorios en una imagen de un panda). En general, para poder entrenar este tipo de redes se necesitan grandes cantidades de datos que representen fielmente la variabilidad de la distribución objetivo. En nuestro ejemplo, así como nosotros podemos imaginar muchas especies de pandas y las distintas posiciones que pueden tomar en una fotografía, esa misma variabilidad debe estar representada en el conjunto de datos, por lo que deben ser muy extensos.

Los métodos de entrenamiento de estas redes son particularmente interesantes. Como el objetivo en este caso es tratar de conseguir una salida que pertenezca a una distribución dada, se puede pensar en utilizar funciones matemáticas que evalúen si una muestra pertenece a la distribución objetivo; sin embargo, la distribución objetivo también es desconocida y solo conocemos de ella lo que nuestro conjunto de datos describe. Para poder superar esta dificultad, un trabajo de 2017 propuso una técnica que dio origen a las primeras aplicaciones generativas, la generación de rostros humanos superrealistas. Este trabajo propuso que en lugar de una función matemática como objetivo, se utilice una segunda red neuronal que tome el rol de “crítico” de la primera red. Esta combinación de “generador” y “crítico” se entrena de manera conjunta y de manera adversaria, de ahí el nombre de la técnica: “redes generativas adversarias” (Goodfellow et al, 2017). Es interesante destacar que aquí el entrenamiento es autoguiado, tal como un humano puede aprender a dibujar y corregirse mirando la imagen que quiere reproducir hasta estar satisfecho con su calidad, la red generadora aprende a generar un ejemplo y la red “crítico” aprende a detectar los errores y sirve de guía para refinar la generación.

Entonces, estas redes generativas, en general, no necesitan de métricas particulares, ni de conjuntos de datos anotados, solamente capacidad de procesamiento computacional y grandes cantidades de datos, ambas cosas que existen en abundancia en la actualidad. La implementación de esta tecnología resultó en muchas aplicaciones que se hicieron conocidas por su sorprendente calidad, entre las más conocidas podemos mencionar:

- Generadores de rostros humanos: creados como pruebas de concepto, por ejemplo “This person does not exist”.
- Creación de videos realistas con contenido imposible, como por ejemplo Carrie Fisher actuando en *Star Wars* episodio 9.

## Modelos de lenguaje

Una importante excepción a la forma de entrenar los modelos generativos que mencionamos anteriormente son los modelos de lenguaje. Si bien estos son modelos generativos, los mismos no son entrenados de manera adversaria.

Los modelos de lenguaje cuentan con un entrenamiento por partes. Primero se entrenan sin requerir supervisión y utilizando grandes cantidades de texto. Este entrenamiento es el más grande por el que pasan. Lo que se entrena aquí es la capacidad del modelo de predecir la siguiente palabra de un texto, es decir, a aprender de memoria. Por ejemplo, se le presenta una oración arbitraria, como “El cielo es azul” y se entrena al modelo para que refuerce las probabilidades de dicha sucesión de palabras, así el modelo incorpora información gramatical (el adjetivo va después del verbo) y conceptual (la palabra “cielo” está relacionada con el color azul). Los modelos entrenados solo hasta este punto son los llamados modelos fundacionales.

Segundo, los modelos son entrenados para respetar estructuras como preguntas-respuestas. Este es un entrenamiento convencional, con datos anotados y curados por humanos, donde se espera que la red neuronal pase de ser un simple completador de datos a ser un modelo más útil, el cual pueda ser accionado. Por ejemplo, el modelo recibe la pregunta “¿De qué color es el cielo?” y la respuesta esperada será “Azul”. Tercero, y último, el modelo es entrenado por un mecanismo de refuerzo, donde se ajusta su “actitud” para que su generación se ajuste a las preferencias humanas. Esta etapa requiere muchos datos, curados y anotados o integración directa de humanos que provean apreciaciones. Aquí el modelo recibe una instrucción y genera una serie de respuestas, luego las ordena por cómo cree que un humano lo haría (de más a menos deseable) y se compara con una anotación humana. El mecanismo de aprendizaje por refuerzo va sesgando las respuestas del modelo para que reflejen dichas preferencias. Este tipo de entrenamiento se ve ocasionalmente cuando utilizamos algunos servicios de asistentes virtuales, en el cual el asistente nos provee con más de una respuesta y nosotros debemos elegir la más adecuada o cuando el asistente nos solicita darle una calificación a la respuesta del modelo, estos datos son devueltos al algoritmo de aprendizaje por refuerzo que premia al modelo si la devolución es positiva (reforzando su comportamiento). Al final del entrenamiento el modelo de lenguaje adopta un sesgo que es aceptable para los usuarios y finaliza el entrenamiento.

Como vemos, si bien los modelos de lenguaje son modelos generativos, su entrenamiento es mucho más delicado que otras aplicaciones. Su dificultad se centra en los últimos dos pasos, los cuales moldean el crudo comportamiento que emerge de un entrenamiento no guiado y es aquí donde los puntos más controvertidos del entrenamiento se debaten, como cuál es el grado de censura o corrección de sesgos existentes en los datos (discriminación, toxicidad) que es necesario implementar.

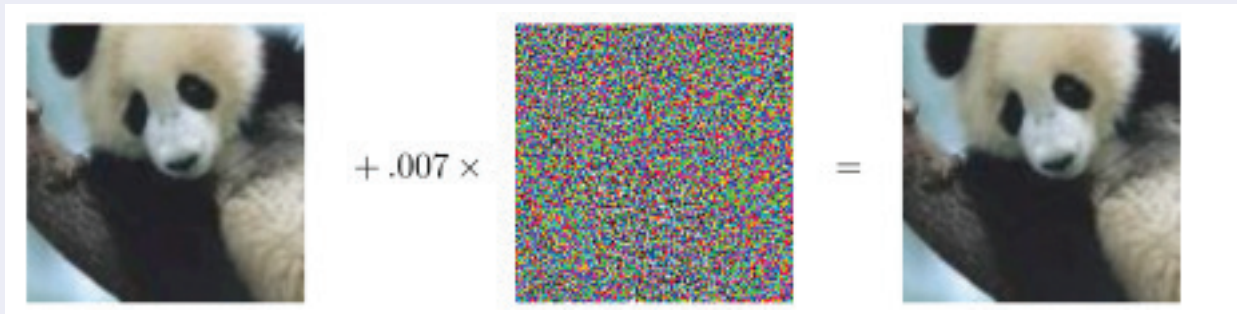
## Limitaciones de estos modelos

Finalmente es importante destacar cuáles son las limitaciones de las RNA, ya que esto nos ayuda a comprender mejor en dónde pueden resultarnos de ayuda y en dónde debemos limitarlas.

Las RNA son entrenadas utilizando grandes conjuntos de datos, y la calidad de dichos conjuntos es fundamental para su correcto funcionamiento ya que la misma va a incorporar cualquier característica que se represente en él.

Los efectos de los conjuntos de datos de mala calidad pueden ir desde que la red simplemente no pueda ser entrenada o hasta incluir sesgos inesperados. Los casos más famosos de esto último son quizás los modelos de lenguaje fundacionales, que en su mayoría fueron entrenados con datos obtenidos de internet, sin ningún tipo de filtrado inicial. Estos modelos aprendieron, además de la estructura del lenguaje, todos los sesgos existentes en la cultura humana y al momento de ponerlos en uso, muchos de estos sesgos pueden ser observados. La corrección de estos sesgos de manera efectiva es un problema abierto. Si bien es posible minimizar el impacto de este problema usando datos curados, sin embargo, el sesgo es inherente a los curadores.

Otro importante problema de las RNA es que los procesos por los cuales llegan a una conclusión no son observables. Esto quiere decir que cuando una RNA clasifica una imagen como “retrato de un panda” no podemos saber qué fue lo que realmente originó esta salida. Si intentamos analizar el proceso interno para llegar a una respuesta no vamos a observar características “explicables”; volviendo al ejemplo, un humano puede decirnos que una imagen representa un panda porque contiene animal, con determinada forma, ojos, posición, etc. Una RNA no nos daría esa explicación, ya que puede estar basando su respuesta en relaciones espurias de la imagen. Esta falta de explicabilidad hace que los modelos sean susceptibles de ser engañados de una manera que resulta imperceptible (o ilógica) para un humano; a estos ataques se los denomina “ataques adversarios”. Un ejemplo de este tipo de ataques es la existencia de imágenes que son claramente reconocibles por un humano o RNA, que al modificar levemente los píxeles hacen que la red neuronal se confunda, cuando para un humano la misma no ha cambiado:



**Figura 3.** A la izquierda una imagen de un panda clasificada correctamente por una RNA, a la cual se le suma un nivel de ruido (imagen central) y luego la RNA reconoce la imagen resultante (a la derecha) como un primate.  
 Imagen de Goodfellow, I.J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

Estos problemas hacen que las RNA no sean una solución que se deba adoptar de manera ciega, especialmente cuando las aplicaciones se encuentran dentro de una cadena de decisión. Antes de incluir RNA en procesos críticos, es de suma importancia asegurar que las mismas han sido auditadas de manera correcta y que su espacio de inferencia sea restringido a zonas conocidas.

### Aplicaciones actuales y riesgos

Actualmente nos encontramos en pleno auge de las aplicaciones con RNA, específicamente en el auge de los modelos del tipo generativo. Las grandes cantidades de datos disponibles en la actualidad y el gran poder de cómputo que está a disposición de la mayoría de las personas plantean un escenario fértil para el crecimiento.

En los últimos años (o meses), la cantidad de aplicaciones que se han desarrollado para ayudarnos es imponente, hoy en día contamos con asistentes que nos ayudan a tomar notas y hacer resúmenes de reuniones virtuales, a traducir en el acto una conversación, a crear imágenes de alta calidad, nos sugieren funciones al programar y frases al escribir, etc. Este tipo de aplicaciones que buscan optimizar al humano, manteniéndolo en el centro de la toma de decisiones, aumentan nuestra productividad de manera impresionante. Sin embargo este mismo contexto ha llevado a que se proponga la utilización de estos modelos en lugares que pueden resultar de alta sensibilidad, como la gobernanza automatizada, perfilado de personas (tanto para aplicaciones laborales como financieras), aplicaciones militares, escritura automática de artículos, etc., quitando en algunos casos al humano del proceso de toma de decisiones.

Las RNA tienen un gran potencial para ayudarnos en el día a día, pero nunca debe perderse de vista de dónde provienen y los riesgos que las mismas traen asociados cuando no se las tienen en cuenta.

*En los últimos años se han desarrollado redes cuya salida pretende generar contenido novedoso, es decir que no formaba parte del conjunto de datos de entrenamiento ni tampoco es el resultado de una interpolación entre los ejemplos vistos por la red. A este nuevo conjunto de redes se las llama redes generativas y son las responsables de lo que se denomina el contenido “sintético” de la internet actual.*



por **MARTÍN BELZUNCE**. *Ingeniero en Electrónica y doctor en Ingeniería mención Procesamiento de Imágenes y Señales por la UTN-FRBA. Investigador asistente del CONICET, desarrolla su labor en el CEMSC3 (Instituto de Ciencias Físicas: UNSAM-CONICET) y en el Centro Universitario de Imágenes Médicas (UNSAM).*

# CÓMO LA IA ESTÁ REVOLUCIONANDO LAS IMÁGENES MÉDICAS: APLICACIONES, RIESGOS Y DESAFÍOS

EN EL CAMPO DE LA SALUD, EL DESARROLLO DE LAS TECNOLOGÍAS PONE EN CUESTIÓN TANTO EL TRATAMIENTO DE LAS ENFERMEDADES COMO LA FORMACIÓN DE SUS PROFESIONALES Y LAS RESPONSABILIDADES QUE ASUMEN ANTE SUS PACIENTES.

**L**as imágenes médicas han revolucionado la salud desde el descubrimiento accidental de los rayos X por Wilhelm Conrad Roentgen en 1895, y la generación de la primera imagen médica: una radiografía de la mano de su esposa. Desde entonces, la tecnología de las imágenes médicas ha evolucionado constantemente. Desde la aparición de la tomografía axial computada (TAC), la resonancia magnética nuclear (RMN), la ecografía por ultrasonido y la tomografía por emisión de positrones (PET, por sus siglas en inglés de Positron Emission Tomography), entre otras modalidades, se ha transformado la manera en que entendemos, diagnosticamos y tratamos enfermedades.

Acompañando el rápido y constante crecimiento de las imágenes médicas, actualmente hay otra revolución que está creciendo a velocidades inesperadas, potenciando todas las ramas del conocimiento: la inteligencia artificial (IA). Puesta en boca de todos gracias a los sorprendentes resultados de los Large Language Models (LLM), como lo es el Chat GPT-4, el progreso vertiginoso de esta tecnología nos ha hecho acelerar y adelantar discusiones que anteriormente solo estaban en manos de pequeños grupos de la ciencia y la filosofía. Por ejemplo, ¿cuál es el impacto de la IA en nuestra sociedad?, ¿cuáles son sus riesgos?, ¿cómo podemos confiar en ella?, entre otras preguntas que hoy están en la discusión pública. Estos interrogantes son todavía más relevantes si los pensamos en el área de la medicina y de la salud.

La IA nos permite imaginar un mundo donde las enfermedades son detectadas en etapas tan tempranas que el tratamiento es eficaz y las vidas son salvadas; en que las imágenes médicas son

analizadas y reportadas de forma automática; los estudios de resonancia magnética que duraban media hora pueden realizarse en unos pocos minutos; o la adquisición de imágenes PET, técnica que utiliza pequeñas cantidades de materiales radiactivos denominados radiofármacos, puede hacerse con hasta 20 veces menos de radiactividad; entre otros escenarios posibles. A través de algoritmos sofisticados y análisis de datos masivos, la IA empieza a mostrar que dichos escenarios no solo son posibles, sino que en muchos casos ya son una realidad.

Sin embargo, la implementación y aplicación de la IA en imágenes médicas en la práctica clínica se enfrenta a un conjunto de desafíos, barreras y potenciales riesgos. Nos preguntamos: ¿confiamos en un sistema de IA para tomar estas grandes decisiones? ¿Qué condiciones les pondremos a estos sistemas? ¿Están preparados para trabajar con todas las poblaciones? Los métodos de IA aprenden de los datos y funcionan como (pseudo) cajas negras, haciendo que los detalles de su funcionamiento sean difíciles de comprender por los seres humanos. Esto último es un tema central a la hora de responder algunas de las preguntas mencionadas, y es lo que hace que la mayoría no tenga una respuesta clara.

En este artículo exploramos cómo la IA está transformando las imágenes médicas analizando tres aplicaciones dentro del vasto mundo que las compone: la detección temprana de enfermedades, la optimización de la adquisición de imágenes y la automatización de tareas clínicas; para luego discutir los riesgos y desafíos que presenta el uso de esta tecnología en el campo de las imágenes médicas.

*La IA nos permite imaginar un mundo donde las enfermedades son detectadas en etapas tan tempranas que el tratamiento es eficaz y las vidas son salvadas.*

## ¿Qué son las imágenes médicas?

Son una herramienta esencial en la medicina moderna, utilizadas para obtener representaciones visuales del interior del cuerpo humano con el fin de diagnosticar, monitorear o tratar enfermedades. Existen varias modalidades de imágenes médicas, cada una con sus propias aplicaciones, y ventajas y desventajas:

- ▶ **Radiografía (Rayos X):** utiliza rayos X para producir imágenes bidimensionales de las estructuras internas. Son útiles para visualizar huesos y detectar fracturas, infecciones y/o tumores.
- ▶ **Mamografía:** es una técnica de rayos X especializada para el examen de los senos, utilizada principalmente para la detección temprana del cáncer de mama. La mamografía permite identificar tumores y otras anomalías que pueden no ser detectables mediante palpación.
- ▶ **Tomografía Axial Computarizada (TAC):** combina múltiples imágenes de rayos X, tomadas desde diferentes ángulos para crear imágenes tridimensionales detalladas del interior del cuerpo, siendo particularmente útil para examinar lesiones y enfermedades en el cerebro, el pecho y el abdomen.
- ▶ **Resonancia Magnética Nuclear (RMN):** utiliza campos magnéticos y ondas de radio para producir imágenes detalladas de los tejidos blandos del cuerpo, como el cerebro, la médula espinal, las articulaciones y los órganos internos, sin la necesidad de radiación ionizante. Recordemos que la radiación ionizante, como la empleada en TAC y PET, puede aumentar el riesgo de problemas de salud a largo plazo.
- ▶ **Ecografía (Ultrasonido):** emplea ondas sonoras de alta frecuencia para crear imágenes en tiempo real de los órganos internos y el flujo sanguíneo, comúnmente utilizada en obstetricia, cardiología y para guiar procedimientos mínimamente invasivos.
- ▶ **Tomografía por Emisión de Positrones (PET):** utiliza pequeñas cantidades de materiales radiactivos denominados radiofármacos para mapear la actividad metabólica en el cuerpo, siendo fundamental en la detección y monitoreo de cáncer, así como en el estudio del cerebro y el corazón.

Cada una de estas técnicas proporciona información única y complementaria, permitiendo a los equipos médicos obtener una comprensión integral de la condición de los y las pacientes y planificar tratamientos adecuados.

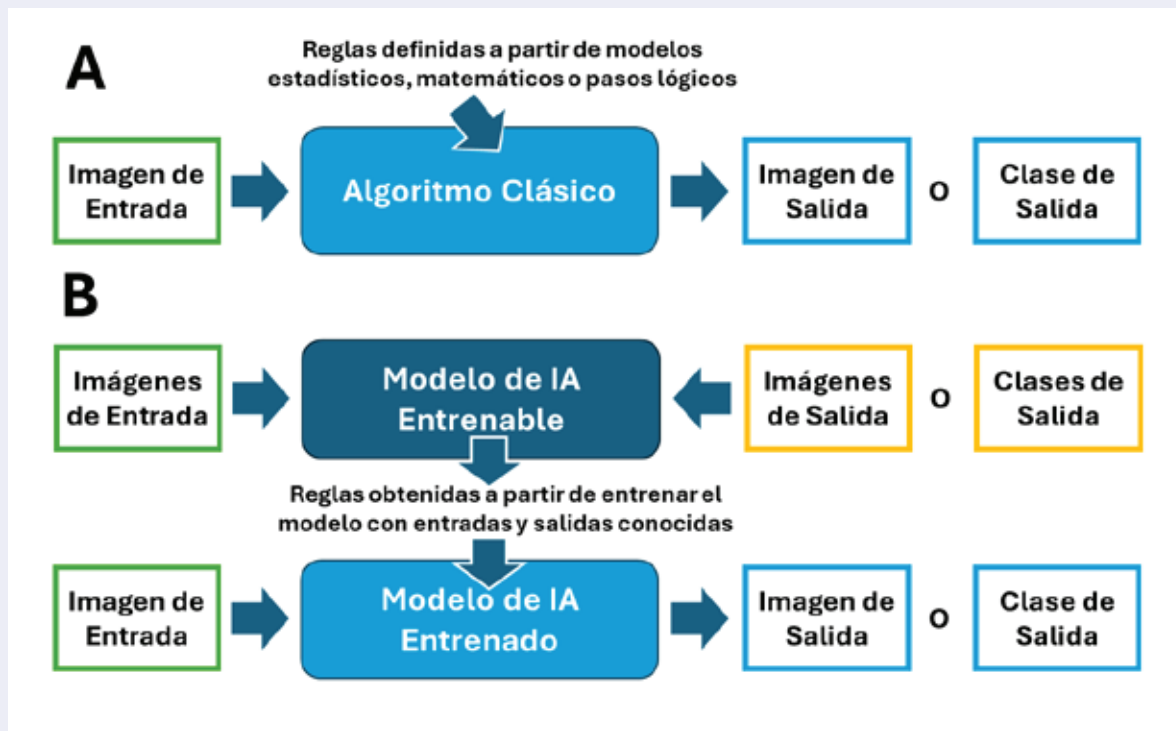
Dentro de las características deseadas para obtener imágenes médicas de buena calidad se encuentran: buena resolución, que es la capacidad de observar estructuras o detalles pequeños; buen contraste, se refiere a la capacidad de distinguir entre dos tipos de tejidos distintos (por ejemplo, músculo y grasa, o tejido sano y patológico), y poco ruido, siendo el ruido la presencia de puntos blancos o negros, áreas borrosas, o distorsiones que no deberían estar presentes en la imagen.

En conjunto, las imágenes médicas desempeñan un papel crucial en la mejora de la precisión diagnóstica, la personalización de los tratamientos y el seguimiento de las enfermedades, mejorando significativamente los resultados en salud.

## IA en imágenes médicas

Desde el inicio de la digitalización de las imágenes médicas, ingenieros, computólogos, físicos y científicos de otras disciplinas han continuamente intentado desarrollar métodos o algoritmos capaces de mejorar la calidad de las imágenes médicas y su interpretabilidad, teniendo como objetivo final mejorar el diagnóstico y tratamiento de múltiples enfermedades y condiciones. Este progreso ha sido continuo y sostenido, desarrollando algoritmos para eliminar el ruido de las imágenes, segmentar estructuras o identificar de forma automática imágenes patológicas, entre muchas otras aplicaciones. Estos métodos tradicionalmente se basaron en la optimización de modelos físicos, matemáticos y/o estadísticos, así como reglas lógicas *ad-hoc*. Estos algoritmos, que llamaremos “clásicos”, son creados mediante la programación de ciertas reglas explícitas que son derivadas de principios matemáticos o estadísticos ya conocidos, y que le dan fundamento al algoritmo. No tienen la capacidad de aprender o adaptarse a nuevos datos como lo hacen los algoritmos de IA, como se observa en la Figura 1.

Al contrario de los algoritmos clásicos, en los métodos basados en IA el usuario brinda múltiples entradas y salidas conocidas a un modelo generalizable con miles/millones de parámetros (a veces conocido como caja negra), y el algoritmo aprende las reglas necesarias para realizar una tarea dada. Luego, si el modelo se ha entrenado adecuadamente, al presentarle nuevas imágenes (desconocidas por el modelo) podrá ejecutar la tarea para la que fue entrenado. La diferencia de funcionamiento entre los algoritmos clásicos y los basados en IA están en la Figura 1.



**Figura 1.** En A, se muestra el funcionamiento de un algoritmo clásico de procesamiento de imágenes, donde previamente se definen las reglas de su funcionamiento y a partir de una imagen de entrada se obtiene una salida. En B, se muestra el funcionamiento de algoritmos de IA, que consta en dos etapas. En la primera, un modelo aprende las reglas (o parámetros) a partir de múltiples imágenes de entrada con su salida conocida asociada. Esta etapa genera un modelo entrenado que luego puede ser usado para procesar nuevas imágenes de entrada.

La performance de estos nuevos métodos basados en IA ha mostrado resultados asombrosos, superando a los métodos anteriores de manera considerable. Esto ha sido posible gracias al desarrollo de potentes computadoras y placas de video en las últimas décadas, impulsada inicialmente por la industria de los videojuegos; y por la aparición masiva de grandes bases de datos de todo tipo, incluyendo de imágenes médicas, aunque estas últimas son más difíciles de generar porque implica el acceso a datos de pacientes.

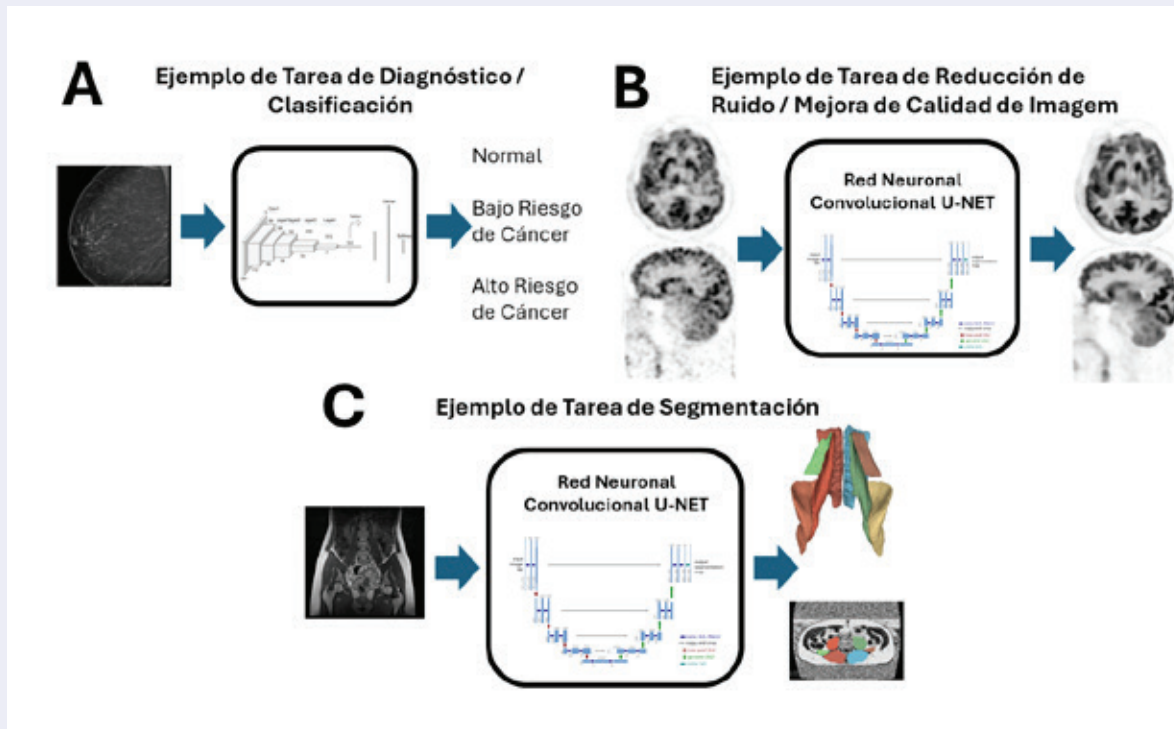
### ¿En qué consisten las cajas negras?

En la Figura 1 hemos mostrado un esquema de cómo los algoritmos de IA aprenden a realizar una tarea a partir de entrenar un modelo, brindándole salidas conocidas (llamadas *ground truth*) a partir de entradas conocidas. Este modelo es comúnmente llamado caja negra porque, al aprender de los datos, no sabemos qué reglas aprende. Sin embargo, no son completamente cajas negras, ya que tienen una estructura predefinida con parámetros que se ajustan en el proceso de entrenamiento.

Dentro de esos modelos se destacan las redes neuronales arti-

ficiales, que son sistemas de inteligencia artificial inspirados en el cerebro humano. Consisten en muchas unidades interconectadas, llamadas neuronas, que trabajan juntas para reconocer patrones y resolver problemas aprendiendo de los datos.

En el caso del procesamiento de imágenes, se emplean Redes Neuronales Convolucionales (CNN, por sus siglas en inglés de Convolutional Neural Networks), que tienen múltiples capas jerárquicas inspiradas en modelos fisiológicos de la corteza visual. Dichas redes están compuestas por capas de convolución, que son como pequeños filtros que se deslizan sobre la imagen para detectar características simples, como bordes o texturas. A la salida de cada filtro existen operaciones para comprimir o subsamplear los datos y observar características cada vez más generales. De esta forma son capaces de extraer características tanto detalladas como generales de una imagen. Las CNN se hicieron famosas por su capacidad de reconocer objetos en fotos, por ejemplo, para identificar si una foto contiene un gato o un perro. Estas mismas redes se emplean en las imágenes médicas, mostrando excelentes resultados tanto para la clasificación de imágenes, como en tareas de filtrado y segmentación.



**Figura 2.** Tres ejemplos de aplicaciones de la IA en imágenes médicas. A) Aplicación para el diagnóstico de cáncer a partir de imágenes de mamografía. B) Mejora de la calidad de imágenes PET ruidosas. C) Segmentación de músculos a partir de imágenes de resonancia magnética.

## Ejemplos de aplicaciones

En la Figura 2 se muestran esquemas de tres aplicaciones de modelos de IA para procesar imágenes médicas, que usaremos como ejemplos. En la primera, a la entrada se tiene una imagen de mamografía y a la salida indica si la paciente tiene riesgo de cáncer o no; este tipo de tarea se llama de clasificación (Figura 2-A). En la segunda, el método recibe una imagen ruidosa, y a la salida recibe una imagen acondicionada donde se elimina el ruido y se mantiene la resolución de la imagen (Figura 2-B). Por último, el algoritmo recibe una imagen de resonancia magnética y a la salida se encuentra una imagen con los músculos de la cadera delineados, este es el caso de segmentación de imagen (Figura 2-C).

### Ejemplo de Aplicación 1: IA para el Diagnóstico Temprano de Enfermedades

Imaginemos un mundo donde las enfermedades son detectadas antes de que se manifiesten los síntomas, donde los cánceres son identificados en etapas tan tempranas que el tratamiento es eficaz y vidas son salvadas. En este mundo, la IA estaría desempeñando un papel crucial al identificar patrones y anomalías que escapan al ojo humano, incluso en casos tan sutiles como las primeras etapas de las enfermedades que hoy podemos diagnosticar ya en etapas avanzadas.

Un ejemplo emblemático de este avance se encuentra en la detección del cáncer de mama mediante mamografías. Empresas como Google Health han desarrollado algoritmos de IA que han demostrado ser tan efectivos, o incluso más, que equipos de expertos en radiología en la identificación de cánceres en mamografías. En estudios recientes se ha demostrado que estos sistemas de IA pueden reducir el número de cánceres no detectados (falsos negativos), mientras que también disminuyen el número de falsos positivos, en comparación con la práctica clínica habitual. Un falso positivo es un error en el resultado de una clasifi-

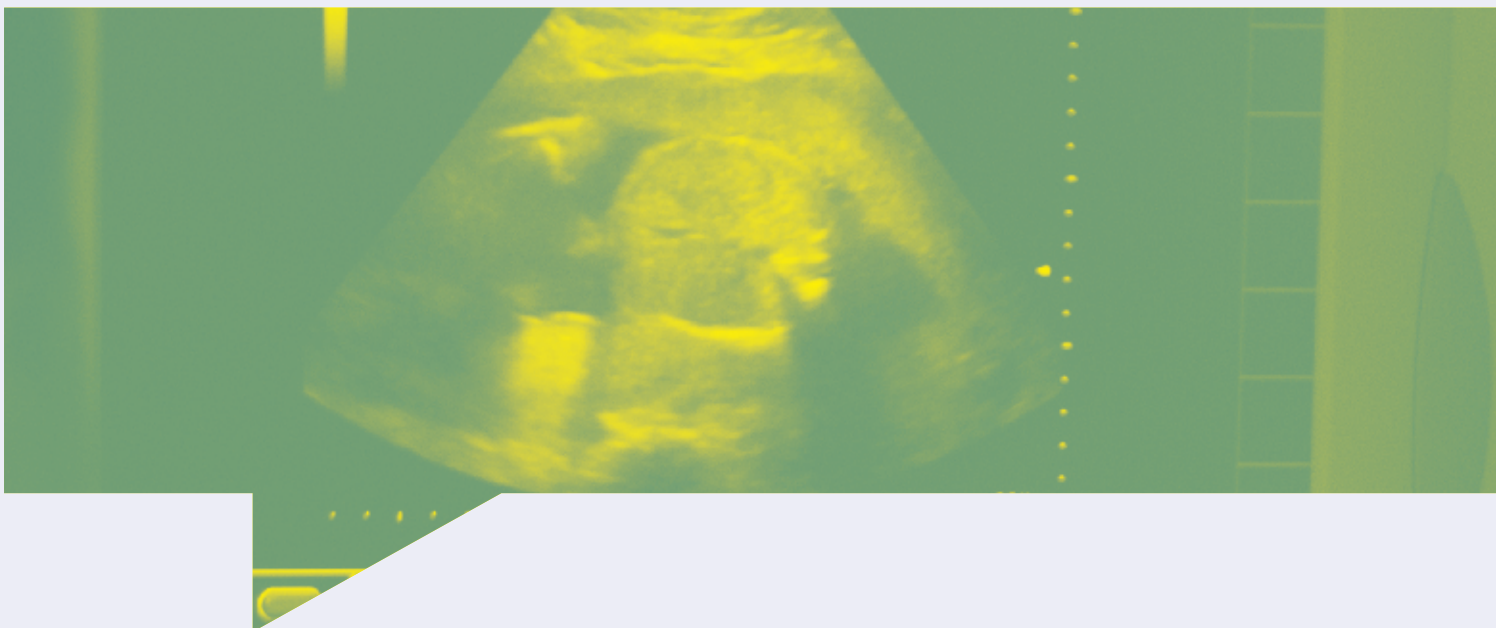
cación binaria, indicando la presencia de una condición cuando la misma no está presente en realidad (en nuestro ejemplo sería un diagnóstico erróneo de cáncer de mama cuando en realidad no está presente). En cambio, un falso negativo es la ausencia de dicha condición cuando en realidad sí está presente (sería la omisión de un diagnóstico de cáncer de mama). En resumen, esta mejora de la eficacia diagnóstica se traduce en diagnósticos tempranos por un lado, y menos llamadas de seguimiento innecesarias por diagnósticos de cánceres que no existen por otro. Sin embargo, la adopción de estos sistemas en la práctica clínica no está exenta de desafíos. Un problema crítico es la falta de diversidad en los datos utilizados para entrenar estos modelos. La mayoría de los datos provienen de pacientes de países desarrollados, lo que significa que los modelos pueden no funcionar tan bien en poblaciones con diferentes características demográficas y genéticas. Además, existe una preocupación legítima sobre la confianza en los diagnósticos generados por IA. ¿Cómo podemos asegurarnos de que los sistemas de IA no pasen por alto signos críticos de enfermedad debido a una peculiaridad en los datos de entrenamiento (por ejemplo, que se haya entrenado solo con datos de personas con índice de composición corporal medio y bajo)? Este es un desafío que debe abordarse con investigaciones adicionales y un enfoque cauteloso en la implementación clínica.

A pesar de estos desafíos, la promesa de la IA en la detección temprana de enfermedades es enorme. Con una mayor inversión en investigación y un enfoque en la creación de modelos más inclusivos y robustos, la IA tiene el potencial de transformar la forma en que detectamos y tratamos las enfermedades, salvando innumerables vidas en el proceso.

## **Ejemplo de Aplicación 2 - IA para la Optimización de la Adquisición de Imágenes**

La adquisición de imágenes médicas de alta calidad es un proceso fundamental para un diagnóstico preciso. La calidad de las imágenes médicas está relacionada con los niveles de ruido, la resolución y la presencia/ausencia de errores. Mejorar estas características implica un mayor contraste en la imagen, aumentando su capacidad diagnóstica. Sin embargo, en muchas técnicas de imagen la obtención de imágenes claras y detalladas puede ser un desafío. En PET, la cantidad de radiofármaco inyectado al paciente está inversamente relacionada con los niveles de ruido de la imagen. Aquí es donde la IA puede marcar una diferencia significativa.

En los últimos años han surgido múltiples investigaciones que lograron bajar la exposición a la radiación asociada con los radiofármacos utilizados en PET (una preocupación importante para los pacientes y los profesionales de la salud) sin perder la calidad de imagen. En la Figura 2-B se muestra un ejemplo de esta aplicación. Como entrada (a la izquierda) se tiene la imagen de baja dosis de radiofármaco, la cual se puede observar ruidosa y de baja calidad. Mientras que, a la salida del modelo de IA, que consiste en una CNN, se puede obtener una imagen de similar calidad a la de dosis completa de radiofármaco. Se puede observar que la imagen de salida tiene menores niveles de ruido y mayor contraste entre tejidos. Algunos modelos de IA han llegado a demostrar similares niveles de calidad de imagen con adquisiciones de hasta un 1% de la dosis original de radiofármaco, un resultado impensado hasta hace poco tiempo. Al permitir la adquisición de imágenes de calidad diagnóstica con dosis más bajas, se minimiza el riesgo de efectos adversos a largo plazo



asociados con la exposición a la radiactividad (proveniente de los radiofármacos). Las implicancias de llevar estos resultados a la práctica clínica serían de un impacto extraordinario: facilitaría el acceso a esta modalidad a un mayor número de pacientes, incluidos aquellos que pueden ser más sensibles a la radiación (por ejemplo, niños) o que requieren evaluaciones de imagen repetidas a lo largo del tiempo.

Otro avance notable es el uso de IA para acelerar los tiempos de adquisición en RMN. Tradicionalmente, las exploraciones de RMN pueden llevar hasta una hora, lo que puede ser incómodo para pacientes y costoso para las coberturas de salud. Sin embargo, nuevos algoritmos de IA han sido capaces de reducir estos tiempos a menos de 5 minutos sin comprometer la calidad de la imagen.

A pesar de estos avances, la implementación de estos sistemas en la práctica clínica presenta varios desafíos. Uno de los más significativos es la necesidad de validar exhaustivamente estos modelos en una variedad de entornos clínicos para garantizar su robustez y fiabilidad. Por otro lado, en este tipo de aplicaciones, los modelos de IA deben generar imágenes de buena calidad a partir de adquisiciones de datos incompletos o deficientes, por lo que deben "crear" información. Esto puede dar origen a alucinaciones, que son la aparición de estructuras o características en la imagen que no existen en realidad. Ahondaremos en este concepto más adelante. Por último, las imágenes generadas de esta forma pueden tener algunas características visuales distintas de las observadas en la práctica clínica habitual. Por esta razón, los equipos médicos deben ser capacitados en el uso y la interpretación de estas nuevas tecnologías, para poder diagnosticar con confianza.

### **Ejemplo de Aplicación 3 - Automatización de Tareas Clínicas con IA**

En el ámbito clínico, muchas tareas rutinarias y repetitivas pueden ser automatizadas mediante el uso de IA, liberando tiempo y recursos para que los equipos de salud se concentren en actividades más complejas y de mayor valor. Un ejemplo claro de esto es la segmentación de imágenes médicas, una tarea esencial pero que consume mucho tiempo y es propensa a errores humanos.

La segmentación implica delinear estructuras específicas dentro de una imagen, como órganos, tumores o vasos sanguíneos. Tradicionalmente, esto se hace manualmente o con la ayuda de software semiautomático, pero estas técnicas pueden ser lentas y poco precisas. Con la llegada de la IA, y específicamente de las redes neuronales convolucionales (CNN), se ha logrado una segmentación automática de alta precisión.

En la Figura 2-C mostramos un ejemplo de aplicación donde se entrena una CNN para realizar una tarea de segmentación a partir de imágenes de RMN. En este caso se busca delinear los músculos de las caderas para poder evaluar cuantitativamente su volumen y contenido graso. Siendo estos últimos, dos marcadores de salud muscular predictores de sarcopenia. Recordemos que la sarcopenia es la pérdida de masa muscular con el tiempo. Adicionalmente, en el caso de los músculos de la cadera, una mayor infiltración grasa en el glúteo medio y mínimo son predictores de caídas en adultos mayores. Sin embargo, estas dos métricas no son usadas en la práctica clínica debido a que la delineación manual de los músculos por un/a profesional radiólogo/a es una tarea lenta y tediosa. La misma puede llevar varias horas por imagen (una imagen cuenta con cientos de cortes

*En el ámbito clínico, muchas tareas rutinarias y repetitivas pueden ser automatizadas mediante el uso de IA, liberando tiempo y recursos para que los equipos de salud se concentren en actividades más complejas y de mayor valor.*

axiales a segmentar). En los últimos años esto ha empezado a cambiar gracias a modelos de IA capaces de segmentar músculos de forma automática, rápida y precisa. Los mismos ya han logrado grandes avances en la investigación clínica de la salud muscular y seguramente tendrán un gran impacto en la práctica clínica en el futuro cercano.

Los modelos de IA no solo pueden segmentar automáticamente y más rápidamente estructuras anatómicas complejas como el cerebro, el corazón y los músculos, sino que también reduce la variabilidad entre observadores, lo que puede mejorar la consistencia y la fiabilidad de los diagnósticos

Otra área destacada en la automatización de tareas, y tal vez la más avanzada en su adopción en la práctica clínica, es la generación de informes radiológicos. En este caso no se restringe solo a la investigación, sino que existen múltiples productos comerciales que analizan imágenes médicas y generan borradores de informes, destacando hallazgos relevantes y sugiriendo posibles diagnósticos. Esto puede reducir la carga de trabajo de los/as radiólogos/as y permitir una atención más rápida a sus pacientes. Los LLM también han facilitado la redacción de los reportes de forma automática y consistente, reduciendo la carga de este tipo de tareas del personal especializado.

Sin embargo, la automatización en la medicina también plantea preguntas importantes sobre responsabilidad y ética. ¿Quién es responsable si un algoritmo comete un error? ¿Cómo se aseguran los equipos de salud de que las decisiones de IA son transparentes y comprensibles? Estos son desafíos que deben ser abordados a medida que avanzamos hacia una integración más profunda de la IA en la práctica clínica.

## Riesgos y desafíos del uso de IA en imágenes médicas

Como hemos ido mencionando en los ejemplos de aplicación, que muestran grandes promesas y avances de la IA en imágenes médicas, su implementación no está exenta de riesgos y desafíos significativos. Estos desafíos involucran cuestiones técnicas, éticas, legales y de seguridad en el despliegue de estas tecnologías.

### Desafíos éticos y legales - Cajas negras y responsabilidad

Un desafío crucial es la transparencia y explicabilidad de los algoritmos de IA. La explicabilidad se refiere a la capacidad de un modelo de aprendizaje automático para ser entendido y explicado de manera clara y comprensible por humanos. Dado que muchos modelos de IA, especialmente las redes neuronales artificiales, funcionan como cajas negras, puede ser difícil entender cómo llegan a sus decisiones. Esto puede generar desconfianza entre profesionales de la salud y pacientes. Siendo la confianza

un factor fundamental en las aplicaciones médicas diagnósticas (como la Aplicación 1). Los equipos médicos, responsables finales de la salud de sus pacientes, no desean tomar decisiones que no pueden ser explicadas o fundamentadas a pesar de que la fiabilidad de estos algoritmos haya sido validada con miles de imágenes.

Para poder resolver estos inconvenientes, hoy en día se están investigando métodos que faciliten la interpretabilidad de los modelos, haciéndolos explicables. En estos casos, el modelo de IA no solo indica si una imagen es normal o corresponde a una patología, sino que también muestra qué área o información de la imagen está siendo considerada para tomar dicha decisión. Además, existe la cuestión de la responsabilidad. La práctica médica plantea una serie de cuestiones legales y éticas relacionadas con la responsabilidad (*accountability*) por errores diagnósticos o terapéuticos. Si un algoritmo de IA comete un error que resulta en un diagnóstico incorrecto o un tratamiento inadecuado, ¿quién es responsable? ¿Los desarrolladores del software, los profesionales de la salud que lo utilizan, o las instituciones que lo implementan? Es por esta razón que se espera que los entes reguladores y profesionales de la salud exijan que ninguna decisión sea completamente tomada por modelos de IA. Estas son cuestiones legales complejas que deben ser abordadas a medida que se integra la IA en la medicina.

### Desafíos técnicos - Sesgos en los modelos y alucinaciones

En cuanto a los desafíos técnicos, uno de los más importantes es la calidad y diversidad de los datos utilizados para entrenar a estos modelos. Si los datos de entrenamiento no son representativos de la población general, los modelos pueden no funcionar adecuadamente frente a imágenes de nuevos pacientes. Esto llevaría a diagnósticos incorrectos. Decimos en este caso que el modelo tiene un sesgo.

Adicionalmente, los modelos de IA son sensibles a cambios en los protocolos de adquisición de imágenes y a variaciones en los equipos utilizados. Esto significa que un modelo entrenado con un conjunto de datos específico puede no funcionar bien en otro entorno clínico con diferentes protocolos y equipos. La validación exhaustiva y la calibración continua son esenciales para garantizar la robustez de estos sistemas.

Las primeras advertencias sobre los sesgos de los modelos de IA no han venido por parte de las aplicaciones médicas, sino en aplicaciones de reconocimiento de caras. Para profundizar en este tema se puede ver el documental "Prejuicio cifrado" que relata cómo Joy Buolamwini, ingeniera del MIT Media Lab, tomó conciencia del sesgo racial existente en los algoritmos de reconocimiento facial.

En salud, que no haya sesgos es fundamental para tener accesos igualitarios a los diagnósticos y tratamientos adecuados. Esto es actualmente un problema por resolver ya que, por ejemplo, se ha encontrado que hay modelos que tienen mejor predicción diagnóstica en poblaciones masculinas y blancas. Como consecuencia, el uso de dicha herramienta privaría de tratamientos tempranos a otras poblaciones. Esto es aún más fundamental en los países de ingresos bajos y medios, ya que la mayoría de las soluciones de IA y bases de datos de imágenes médicas han sido realizadas en los Estados Unidos o países de Europa.

Otro aspecto técnico que considerar son las alucinaciones en los sistemas de IA. Cuando hablamos de seres humanos, nos referimos a una alucinación como una falsa percepción de los objetos o hechos, y es de naturaleza sensorial. En cambio, en IA, nos referimos a cuando un modelo generativo (o sea que genera información) produce información falsa o incorrecta. Las alucinaciones han tenido gran repercusión a través de chat GPT, donde repetidamente se observó que este modelo contesta con total seguridad hechos que no son reales. Es decir, en IA el término alucinación se emplea cuando un modelo genera información con la que no ha sido entrenada y que es falsa.

En las imágenes médicas, las alucinaciones pueden tener consecuencias importantes. Por ejemplo, en la aplicación de generación de imágenes PET de buena calidad a partir de imágenes ruidosas de baja dosis (Aplicación 2), podrían aparecer regiones con mayor presencia del radiofármaco, donde en realidad no lo hay. Esto podría ser interpretado como una lesión o tumor, o, en el caso contrario, ocultarla. La existencia de alucinaciones genera pérdida o erosión de la confianza de los sistemas. Actualmente hay ramas de la IA que buscan estimar el riesgo de alucinaciones, lo que podría aumentar la confianza en estos sistemas.

### Desafíos operacionales

La integración de la IA en la práctica clínica requiere cambios significativos en los flujos de trabajo y la formación de los equipos de salud. Dichos profesionales deben ser capacitados no solo en el uso de nuevas tecnologías, sino también en la interpretación de los resultados generados por la IA. Esto requiere una inversión considerable en educación y recursos humanos. Además, la implementación de IA en la práctica clínica diaria puede enfrentar resistencia por parte de los profesionales de la salud que desconfían de las nuevas tecnologías o están preocupados por la posibilidad de que la IA reemplace sus trabajos. Es crucial que se aborde esta resistencia mediante la comunicación clara de los beneficios de la IA y la garantía de que estas tecnologías están diseñadas para complementar y mejorar, no reemplazar puestos de trabajo.

### Conclusión

La IA está transformando rápidamente el campo de las imágenes médicas, ofreciendo nuevas posibilidades para el diagnóstico temprano de enfermedades, la optimización de la adquisición de imágenes y la automatización de tareas clínicas. Sin embargo, la implementación de estas tecnologías no está exenta de desafíos y riesgos significativos. A medida que avanzamos hacia una mayor integración de la IA en la práctica clínica, es esencial abordar estos desafíos mediante una combinación de innovación técnica, consideraciones éticas y legales, y una formación y adaptación cuidadosa de los flujos de trabajo. Esto requerirá un esfuerzo de todas las disciplinas involucradas:

los/as tecnólogos/as deberán resolver los desafíos técnicos que aseguren mayor confiabilidad e interpretabilidad de la IA; los/as profesionales de la clínica deberán adaptarse y aprovechar las ventajas de estas nuevas tecnologías, y los entes reguladores deberán generar el marco normativo adecuado para que la IA pueda ser adoptada y segura al mismo tiempo. De esta forma, la IA podrá mejorar significativamente la calidad y la eficiencia de la atención médica, salvando vidas y mejorando los resultados de pacientes en todo el mundo.





por **ARIEL HERNÁNDEZ**. Ingeniero en Electrónica. Actualmente es miembro del equipo de Machine Learning y Visión Artificial en Seeing Machines.

EL AUTOR DA CUENTA DE LOS AVANCES QUE EXISTEN EN EL CAMPO AUTOMOTRIZ EN TORNO A LA PRODUCCIÓN DE SISTEMAS INTELIGENTES QUE SEAN CAPACES DE PROVEER DE AUTONOMÍA A LOS VEHÍCULOS, Y EXPONE, ASIMISMO, LOS DESAFÍOS QUE SE PRESENTAN EN TORNO A LA INTEGRIDAD FÍSICA DE SUS OCUPANTES, A LA PRESERVACIÓN DE LOS DATOS, Y EN RELACIÓN A LO QUE SE VINCULA CON EL IMPACTO AMBIENTAL.

# AUTONOMÍA EN MOVIMIENTO: PERSPECTIVAS Y DESAFÍOS DE LOS VEHÍCULOS AUTÓNOMOS





¿Qué constituye un vehículo autónomo? En términos precisos, se trata de un vehículo capaz de llevar a cabo un trayecto completo, desde su punto de partida hasta un destino predefinido, sin necesidad de intervención humana. Cualquier instancia que requiera una acción por parte del conductor indicaría que la autonomía del vehículo es restringida. Estas restricciones se clasifican en seis niveles de conducción autónoma, que se explicarán más adelante.

La historia del desarrollo de los vehículos de conducción autónoma podría remontarse al vehículo de tres ruedas ideado por Leonardo Da Vinci en el siglo XVI, pero el avance significativo se produjo a finales del siglo XX y principios del siglo XXI. Ernst Dickmanns es ampliamente reconocido como el pionero en este campo, siendo el primero en lograr avances notables. En la década de 1980, el ingeniero Dickmanns lideró un equipo de científicos en la compañía Mercedes-Benz, donde lograron equipar una furgoneta de la marca para recorrer cientos de kilómetros de manera autónoma, marcando un hito en la historia de los vehículos autónomos.

Durante la década de 1990 se produjeron numerosos avances en la investigación y desarrollo de vehículos autónomos. Uno de los hitos más destacados fue el Proyecto ARGO, llevado a cabo por la Universidad de Parma bajo la dirección del profesor Alberto Broggi, quien lideró un equipo que instrumentó un Lancia Thema para recorrer más de 1.900 kilómetros de forma autónoma en el norte de Italia. En los años 2000, la agencia DARPA (Agencia de Proyectos de Investigación Avanzada de Defensa, del Departamento de Defensa de los Estados Unidos) lanzó los Grand Challenge, competiciones de vehículos autónomos en diferentes lugares del estado de California, como el desierto de Mojave. Los equipos participantes tenían la oportunidad de ganar sumas significativas de dinero si lograban liderar la competencia. Uno

de los ganadores en 2005 fue el equipo dirigido por Sebastian Trun, entonces director del Laboratorio de Inteligencia Artificial de la Universidad de Stanford, con su vehículo Stanley, una Volkswagen Touareg totalmente autónoma desarrollada en dicho laboratorio.

En los últimos años, los avances tecnológicos han permitido que un mayor número de empresas y grupos de investigación desarrollen prototipos, pruebas de concepto y productos relacionados con los vehículos autónomos. El aumento de la capacidad computacional, ya sea a través de microprocesadores más potentes o del acceso a tarjetas gráficas para la computación paralela, ha sido fundamental en este progreso. Actualmente numerosas compañías, tanto las establecidas en el mercado automotriz como Daimler, Volkswagen y Volvo, como más recientes como Tesla, Waymo (una subsidiaria de Google) y Uber, están centradas en el desarrollo de esta tecnología.

La seguridad vial es la principal motivación detrás de este desarrollo. Con el crecimiento del número de vehículos en circulación en todo el mundo, se hace cada vez más necesario controlar el tránsito para reducir los accidentes y prevenir fatalidades. Sin embargo, el desarrollo de los vehículos autónomos plantea desafíos más allá de la tecnología, como la preservación de datos, su legislación y la mitigación del impacto ambiental.

En este contexto, los consumidores muestran un gran interés en delegar la conducción de sus vehículos a la tecnología. Esto ha llevado a los fabricantes de automóviles, en colaboración con empresas especializadas, a mejorar los sistemas avanzados de asistencia al conductor (conocidos como ADAS, por sus siglas en inglés) y ofrecer otras opciones de automatización, hasta que los vehículos sean completamente autónomos. Se proyecta que para el año 2035 los ADAS y la conducción automatizada podrían impulsar un mercado multimillonario dentro del sector automotriz.

## Funcionamiento de un vehículo autónomo

Los vehículos autónomos poseen una capacidad adicional en comparación con los vehículos convencionales: tienen la posibilidad de percibir el entorno utilizando tres sentidos extras: radar, cámaras y LiDar –basado en láser– (acrónimo de Detección y Alcance de Luz). Estos sensores actúan como “ojos electrónicos”, permitiendo al vehículo observar todo su entorno con gran precisión. Cada uno de estos elementos proporciona información a un sistema inteligente compuesto principalmente por algoritmos que trabajan en conjunto con modelos de inteligencia artificial (IA), donde cada uno de ellos ha sido entrenado para diferentes propósitos como la detección de las distintas señales de tránsito, marcas viales en las calles o rutas, identificar el estado de los semáforos o determinar el número y tipo de vehículos presentes en un área determinada, así como la presencia de peatones en las cercanías del vehículo. Cada una de estas tareas es fundamental, ya que implican resolver situaciones complejas. Para comprender la complejidad de esta tarea se puede imaginar cómo un algoritmo debe identificar correctamente señales de límite de velocidad o señales de PARE a partir de una o varias imágenes. A continuación, se presentan ejemplos de diferentes señales de límites de velocidad, mostrando la variabilidad en sus diseños y formatos, lo que dificulta la identificación precisa de una señal. Además, factores como la iluminación (por ejemplo, si la conducción es realizada de día o de noche), obstáculos (como ramas de árboles u otros vehículos obstruyendo los carteles de señalización) o el deterioro de las señales de tránsito pueden complicar aún más esta tarea.

A estas funciones de detección de señales de tránsito se les suman tareas aún más complejas, como la identificación de peatones o ciclistas, otros vehículos y la evaluación de situaciones de riesgo en un tiempo muy corto. El sistema inteligente analiza todas estas tareas en tiempo real para la toma de las decisiones necesarias y envía comandos al sistema mecánico y eléctrico del vehículo, controlando los actuadores como el acelerador, los frenos y la dirección.

Es evidente que las tecnologías de conducción autónoma contribuyen a mejorar la seguridad, aunque no son infalibles al 100%, ya que su efectividad puede verse afectada por factores externos que dificultan la interpretación del entorno por parte de los sensores, o incluso por limitaciones intrínsecas para percibir el entorno del vehículo.

Según la definición de la SAE, antes conocida como la Sociedad de Ingenieros Automotrices, existen seis niveles de conducción automatizada, que van desde el control total del vehículo por parte de un conductor humano, asistido por tecnología avanzada de seguridad, hasta un vehículo que puede operar sin intervención humana. Solo el nivel superior representa un vehículo completamente autónomo sin conductor. En la siguiente tabla se detallan las características correspondientes a cada nivel de conducción autónoma.



Diferentes formas y tamaños de señales de límite de velocidad (Crédito: Andrej Karpathy, Tesla)

## SAE J3016™ NIVELES DE AUTOMATIZACIÓN DE LA CONDUCCIÓN

Derechos reservados a SAE International (traducción realizada por Easy Law Magazine)

	SAE NIVEL 0	SAE NIVEL 1	SAE NIVEL 2	SAE NIVEL 3	SAE NIVEL 4	SAE NIVEL 5
<b>¿Qué tiene que hacer el humano en el asiento del conductor?</b>	Usted <b>está conduciendo</b> siempre que estas funciones de asistencia al conductor están activadas, incluso si sus pies están fuera de los pedales y no está manejando el volante			Usted <b>no está conduciendo</b> cuando estas funciones de conducción automatizada están activadas, aunque esté sentado en "el asiento del conductor"		
	Debe <b>supervisar constantemente</b> estas funciones de asistencia; debe dirigir, frenar o acelerar según sea necesario para mantener la seguridad			Cuando la función lo solicite, <b>usted debe conducir</b>	Estas funciones de conducción automatizada no le obligarán a asumir la conducción	
	<b>Funciones de apoyo al conductor</b>			<b>Funciones de conducción automatizada</b>		
<b>¿Para qué sirven estas funciones?</b>	Estas funciones se limitan a proporcionar avisos y asistencia momentánea	Estas funciones proporcionan al conductor apoyo en la dirección <b>O</b> en la frenada / aceleración	Estas funciones proporcionan al conductor apoyo en la dirección <b>Y</b> en la frenada / aceleración	Estas funciones pueden conducir el vehículo en condiciones limitadas y no funcionarán a menos que se cumplan todas las condiciones requeridas		Esta función puede conducir el vehículo en todas las condiciones
<b>Ejemplo de funciones</b>	<ul style="list-style-type: none"> <li>Frenado automático de emergencia</li> <li>Advertencia de ángulo muerto</li> <li>Aviso de salida de carril</li> </ul>	<ul style="list-style-type: none"> <li>Centrado de carril <b>O</b></li> <li>Control de cruceo adaptativo</li> </ul>	<ul style="list-style-type: none"> <li>Centrado de carril <b>Y</b></li> <li>Control de cruceo adaptativo al mismo tiempo</li> </ul>	<ul style="list-style-type: none"> <li>Chofer atasco en</li> </ul>	<ul style="list-style-type: none"> <li>Taxi local sin conductor</li> <li>Pedales / volante pueden o no estar instalados</li> </ul>	<ul style="list-style-type: none"> <li>Igual que el nivel 4, pero la función puede conducir por todas partes en todas las condiciones</li> </ul>

Traducción propia, todos los derechos pertenecen a SAE International.

### Ventajas y desafíos

La prevención de accidentes, tanto en el tránsito diario como en las rutas de larga distancia, se destaca como la principal motivación detrás de la tecnología de conducción automatizada. Las estadísticas alarmantes de fatalidades relacionadas con accidentes de tránsito en todo el mundo refuerzan la necesidad de buscar soluciones que minimicen estos riesgos. Los vehículos autónomos representan una innovación significativa al eliminar el factor humano, que puede verse afectado por la somnolencia, distracciones o falta de experiencia al volante, lo que contribuye de manera considerable a reducir la incidencia de accidentes. Estos vehículos presentan diversas ventajas sobre los convencionales, incluyendo la accesibilidad a la movilidad, la eficiencia en términos de costos y la capacidad para mitigar las congestiones en el tráfico mediante el intercambio de datos con otros vehículos para anticipar el estado del tránsito. Además, representan una alternativa crucial para brindar movilidad a personas que, debido a su edad, discapacidad o falta de acceso al transporte público convencional, se ven impedidos para conducir. En el caso de los vehículos de conducción autónoma que operan en niveles del 0 al 2 (con intervención humana durante el manejo), así como en algunos modelos de automóviles convencionales, se han incorporado sistemas avanzados de asistencia al conductor (ADAS), lo que resulta en una mejora considerable en la seguridad vial. Estos sistemas brindan un monitoreo constan-

te del conductor para evitar la somnolencia o las distracciones, así como también ofrecen información sobre puntos ciegos del vehículo, lo que ayuda a anticipar peligros inminentes. De acuerdo con el National Safety Council (organización norteamericana que promueve la salud y seguridad vial) podría llegar a prevenir hasta un 60% del total de las fatalidades en accidentes de tránsito.

Además, a medida que el transporte de conducción automatizada se adopte más ampliamente, tanto en vehículos particulares como en el transporte público, se vislumbran beneficios significativos en términos del impacto medioambiental. Esta perspectiva se fundamenta en la mayor eficiencia de la conducción autónoma, la cual conlleva a una reducción en el número de vehículos en circulación, lo que a su vez disminuirá la energía requerida. Según un estudio realizado por la Universidad de Michigan, este cambio podría reducir el impacto medioambiental hasta en un 9 por ciento.

Estos beneficios tienen asociados una serie de desafíos importantes que van desde el costo de desarrollar esta tecnología, la madurez de la misma y su correspondiente regulación.

El desarrollo de la tecnología de conducción autónoma conlleva costos significativos, como lo ilustra el caso del Proyecto Titán de Apple, el cual fue cancelado tras una inversión multimillonaria. Inicialmente, Apple había anunciado la creación de un vehículo autónomo de nivel 5 de automatización, pero

*A medida que el transporte de conducción automatizada se adopte más ampliamente, tanto en vehículos particulares como en el transporte público, se vislumbran beneficios significativos en términos del impacto medioambiental.*

posteriormente redujo sus objetivos hasta alcanzar el nivel 2, antes de abandonar el proyecto por completo. Esta decisión se fundamenta principalmente en la complejidad inherente de replicar el pensamiento, la lógica y los instintos humanos al tomar decisiones críticas, a pesar de los avances en la funcionalidad de la tecnología. La ocurrencia de accidentes en California relacionados con vehículos de Tesla y los servicios de robot taxis de Cruise y Uber ha contribuido a generar cierta desconfianza en la comunidad y el mercado respecto de la fiabilidad de esta tecnología.

Por otro lado, una de las cuestiones más relevantes que subyace en la conducción de un vehículo autónomo. Como se mencionó previamente, el vehículo es controlado por un sistema inteligente compuesto por varios modelos de inteligencia artificial (IA) que toman decisiones y actúan sobre otros componentes. Sin embargo, estos modelos se basan principalmente en los datos utilizados durante su entrenamiento. Como resultado, todavía existen modelos que no pueden identificar adecuadamente a personas con diferentes tonos de piel debido a la falta de diversidad en los datos de entrenamiento. Esta deficiencia podría tener consecuencias significativas en la toma de decisiones en situaciones críticas.

Estos modelos de IA analizan la escena y determinan cómo enfrentar los obstáculos, basándose en suposiciones derivadas de una base de datos de imágenes etiquetadas. Estas etiquetas

categorizan elementos como “automóvil”, “peatón” o “árbol”, las cuales se utilizan durante el entrenamiento de los modelos. Sin embargo, no todas las situaciones se pueden prever y modelar, lo que hace que los múltiples modos de falla sean extremadamente difíciles de predecir. Incluso manteniendo todas las variables constantes, un vehículo autónomo puede reaccionar de manera muy diferente en el mismo tramo de ruta en diferentes momentos del día, posiblemente debido a variaciones en los ángulos de la luz del sol.

Por otro lado, como efecto colateral, estos vehículos recopilan grandes volúmenes de datos, incluidas imágenes de peatones, sin obtener el consentimiento ni del peatón ni del propietario del vehículo. Actualmente, no existen regulaciones claras sobre la cantidad de datos que pueden recopilarse, quién puede acceder a ellos o cómo deben ser distribuidos y almacenados. En esencia, estos datos son utilizables para una amplia gama de propósitos que podrían comprometer la privacidad de las personas. Además, el hecho de que estas imágenes puedan transmitirse de forma inalámbrica también plantea preocupaciones éticas adicionales.

No obstante, quizás el tema ético más controversial es la toma de decisiones de un vehículo autónomo en situaciones donde está en juego la vida de seres humanos, evaluando riesgos y ajustando las operaciones para minimizar daños, lesiones y muertes. Además, surge el interrogante sobre la responsabilidad en caso de que un vehículo autónomo esté involucrado en un accidente. Esta pregunta ha persistido durante años en la comunidad sin una respuesta clara. La falta de regulación en el ámbito de la inteligencia artificial en vehículos autónomos avanza a paso lento, en parte debido a la falta de competencia técnica en los organismos reguladores para supervisar eficazmente tecnologías avanzadas, especialmente aquellas relacionadas con la inteligencia artificial, que evolucionan rápidamente.

La Unión Europea ha tomado una postura más proactiva en la regulación de la inteligencia artificial en general, así como en los vehículos autónomos en particular, y ha emitido la primera ley integral de inteligencia artificial en el mundo, que entrará en vigencia en 2026. En nuestro país, si bien existen numerosos proyectos en marcha, el debate aún no ha comenzado en el poder legislativo.

## El futuro de los vehículos autónomos

Actualmente, el nivel más avanzado de conducción autónoma disponible para los consumidores se encuentra en el nivel 2, disponible en modelos como Super Cruise de GM, BlueCruise de Ford Motor Co. y Full Self Driving de Tesla. Sin embargo, también hay varias empresas dedicadas a la fabricación y operación de vehículos autónomos para servicios y fines comerciales.

En este ámbito, se destaca Waymo como una de las compañías más exitosas en la actualidad. Esta compañía opera un servicio de taxis robóticos en San Francisco, California, y está ampliando sus servicios en Los Ángeles, California, y Austin, Texas. Además, encontramos camiones autónomos como los operados por Gatik, encargados de la entrega de mercancías desde puntos de distribución hasta puntos de venta. En la misma línea, empresas como Kodiak Robotics realizan rutas de largo y corto recorrido para compañías como Kroger e Ikea.

Existe un notable interés por parte de los consumidores en los vehículos autónomos, incluso aquellos de Nivel L2 (que utilizan sistemas LiDAR) que contienen componentes con costos aproximados entre 1.500 y 2.000 dólares, los cuales se incrementan aún más para los automóviles con opciones de nivel L3 y L4. Según un análisis de McKinsey, que se fundamenta principalmente en el interés de los consumidores en los vehículos autónomos y las soluciones comerciales disponibles en la actualidad, como los sistemas ADAS, se estima que el mercado automotriz para pasajeros podría alcanzar entre 300.000 y 400.000 millones de dólares para el año 2035.

Por otro lado, para los fabricantes de automóviles que se centran en ofrecer vehículos con niveles más altos de automatización, existe un enorme potencial de crecimiento. El interés de los conductores en ceder el control al sistema autónomo durante la conducción está en aumento, lo que implica la adopción de sistemas más avanzados, incluidos los niveles L2, L3 y L4. Simultáneamente, los costos de sensores y sistemas de cómputo de alto rendimiento están disminuyendo,

mientras que los estándares de seguridad, como el NCAP en Europa (Programa de Evaluación de Automóviles Nuevos) o la NHTSA en Estados Unidos (Administración Nacional de Seguridad del Tráfico en las Carreteras) están en constante aumento.

En conjunto, estos factores impulsan significativamente a la industria automotriz, lo que podría llevar, con el tiempo, a la incorporación de funcionalidades más avanzadas de conducción autónoma. Según un estudio de ventas de McKinsey, es probable que los sistemas L3 y L4 para la conducción en autopistas estén disponibles en el segmento de automóviles de pasajeros alrededor de 2025 en Europa y América del Norte. Sin embargo, los altos costos iniciales de desarrollo de estos sistemas de conducción sugieren que podrían estar inicialmente limitados a segmentos de vehículos premium.

Ahora bien, como se mencionó anteriormente, es esencial que las agencias reguladoras encargadas de supervisar esta tecnología, tanto en inteligencia artificial en general como en conducción autónoma en particular, puedan dar los pasos necesarios para crear un ecosistema confiable y seguro, así como implementar estándares globales.

En conclusión, queda claro que las nuevas tecnologías de conducción autónoma tienen un tremendo potencial para brindar niveles inéditos de seguridad y comodidad a los consumidores, así como para generar un valor significativo dentro de la industria automotriz, transformando fundamentalmente la forma en que las personas viajan. No obstante, el mercado de vehículos autónomos es sumamente dinámico y conlleva un alto nivel de incertidumbre. Por consiguiente, las empresas involucradas en la cadena de producción de vehículos autónomos deben contar con estrategias bien definidas y alineadas.

Finalmente, para triunfar en el mercado automotriz en los próximos años, las empresas necesitan tener una visión clara, un profundo entendimiento del mercado y una sólida comprensión de los intereses de los consumidores.

*Para los fabricantes de automóviles que se centran en ofrecer vehículos con niveles más altos de automatización, existe un enorme potencial de crecimiento. El interés de los conductores en ceder el control al sistema autónomo durante la conducción está en aumento, lo que implica la adopción de sistemas más avanzados.*



# AUTONOMACIÓN, AUTOMATIZACIÓN, INTELIGENCIA ARTIFICIAL E INDUSTRIA- AGRICULTURA 5.0 EN LA PRODUCCIÓN AGROECOLÓGICA SOSTENIBLE DE ALIMENTOS

A TRAVÉS DE LA INDAGACIÓN DE CUESTIONES COMO LA DETECCIÓN PRECOZ DE PLAGAS Y ENFERMEDADES, EL USO DE MAQUINARIAS AUTÓNOMAS Y DE LOS CULTIVOS AUTOMATIZADOS DE INVERNADERO, EL TRABAJO DA CUENTA DEL ESCENARIO ACTUAL Y DE LAS ALTERNATIVAS POSIBLES QUE PUEDEN IMPLEMENTARSE EN LA PRODUCCIÓN AGRÍCOLA.





**por PINAR DEMIRCIOLU.** *Doctora por la Aydın Adnan Menderes Üniversitesi (Turquía). Post-Doctora por la Technische Universität Wien (Austria). Profesora visitante de la Technische Universität München (Alemania).*

**por ISMAIL BOGREKCI.** *Profesor de la Aydın Adnan Menderes Üniversitesi (Turquía). Doctor por la Cranfield University (Reino Unido). Post-Doctor en la University of Florida (Estados Unidos de Norteamérica).*

**por NUMAN M. DURAKBASA.** *Profesor y Doctor por la Technische Universität Wien (Austria).*

**Por JORGE BAUER.** *Ingeniero Mecánico de la Universidad Tecnológica Nacional. Doctor por la Technische Universität Wien (Austria). Post-Doctor de la University Graz (Austria).*

**E**n una época caracterizada por los acuciantes retos que plantean el crecimiento de la población mundial, el cambio climático y la degradación del medio ambiente, la industria agrícola se encuentra en una encrucijada. El imperativo de producir alimentos de forma sostenible, garantizando la alimentación de miles de millones de personas al tiempo que se salvaguarda el planeta, nunca ha sido más urgente. En este contexto crítico, la convergencia de tecnologías avanzadas –autonomación, automatización, inteligencia artificial (IA) y la aparición de la industria-agrícola 5.0 (IA5.0) ofrece una luz de esperanza e innovación. La autonomía potencia la agricultura de precisión a través de UAV y UGV (vehículos aéreos y terrestres no tripulados) impulsados por IA, lo que permite la recopilación de datos en tiempo real y la toma de decisiones informadas. La automatización agiliza las operaciones agrícolas, reduciendo los residuos y mejorando la cadena de suministro agroalimentario. La IA procesa grandes conjuntos de datos, lo que facilita la detección precoz de enfermedades y optimiza el uso de los recursos. La IA5.0 hace hincapié en la colaboración hombre-máquina y en el intercambio de conocimientos, fomentando prácticas sostenibles.

Este estudio se inicia para examinar la convergencia de estas fuerzas transformadoras en el ámbito de la producción agroecológica sostenible de alimentos. A medida que la población mundial sigue creciendo, impone exigencias sin parangón a la producción de alimentos, añadiendo presión a unos recursos ya de por sí al límite. Al mismo tiempo, el cambio climático altera las prácticas agrícolas convencionales, provocando patrones meteorológicos variables, cambios en las estaciones de crecimiento y mayores riesgos para el rendimiento de los cultivos. La degradación del medio ambiente, consecuencia de los métodos de agricultura intensiva con agroquímicos y modificación genética, amenaza los ecosistemas y la biodiversidad y en algún aspecto también la alimentación sana. En este contexto, se hace evidente la necesidad de una agricultura sostenible. Los principios de la IA5.0 subrayan la importancia de una colaboración armoniosa entre las personas y las máquinas. Esta visión fomenta un sistema global e interconectado en el que coexisten los seres humanos y las máquinas inteligentes, utilizando al máximo sus puntos fuertes. Esta relación mutua tiene el potencial de maximizar la productividad agrícola al tiempo que minimiza su huella medioambiental. Para lograr este objetivo se abordan varios aspectos de la producción agrícola como: la detección precoz de plagas y enfermedades, uso de maquinaria autónoma y cultivos automatizados de invernadero.

## DetECCIÓN PRECOZ DE PLAGAS Y ENFERMEDADES

La agricultura es la base de nuestra civilización y proporciona sustento a miles de millones de personas en todo el mundo. Sin embargo, esta industria vital se enfrenta a toda una serie de retos, como la persistente amenaza que suponen las plagas y las enfermedades. La detección temprana de estos problemas es fundamental para evitar la pérdida de cosechas, mantener la calidad de los alimentos y promover prácticas agrícolas sostenibles.

Mediante una simple cámara web y redes neuronales convolucionales (CNNs) con procesamiento en la nube como servicio, se puede lograr avances muy significativos en ese aspecto. La red neuronal convolucional es un tipo de red neuronal artificial inspirada en modelos de funcionamiento de la corteza visual primaria del cerebro biológico de las personas. Como su procesamiento se realiza en matrices de dos dimensiones, entre otras aplicaciones, son muy efectivas para realizar tareas de visión artificial y en la clasificación y segmentación de imágenes. En este caso, las imágenes de los cultivos que envía una cámara web simple, en distintos períodos de tiempo, se procesan en la nube con el sistema de redes neuronales convolucionales (la nube admite una capacidad de procesamiento muy superior a una computadora) y así se permite detectar los cambios en las imágenes que indican la existencia de plagas y/o enfermedades en esos cultivos.

A continuación, en esta sección, se explora este enfoque innovador para la detección temprana de plagas y enfermedades en los cultivos.

### A. Necesidad de detección temprana

Las plagas y enfermedades han supuesto siempre una amenaza persistente para la agricultura, capaz de diseminarse rápidamente por los cultivos, provocando cuantiosas pérdidas económicas. Además, tienen el potencial de socavar la seguridad y la calidad de los alimentos. Históricamente, los agricultores han dependido de las inspecciones manuales, un proceso que requiere mucha mano de obra y que a menudo se inicia demasiado tarde para evitar infestaciones extensas. En consecuencia, existe una demanda creciente de sistemas automatizados, puntuales y precisos de detección de plagas y enfermedades.

### B. El papel de las cámaras web sencillas

En este enfoque, una de las innovaciones clave es la utilización de cámaras web básicas. Estos dispositivos, económicos y fáciles de conseguir, pueden incorporarse sin esfuerzo a las configu-

raciones agrícolas. Mediante la colocación estratégica de *web-cams* en campos o invernaderos, los agricultores pueden captar imágenes en tiempo real de los cultivos. Posteriormente, estas imágenes se transmiten a un sistema central de procesamiento. Contribuyen a una mejor gestión de los cultivos, a la detección precoz de problemas y a la toma de decisiones basada en datos, apoyando en última instancia unas prácticas agrícolas más eficientes y sostenibles.

### C. Procesamiento en la nube como servicio

La computación en nube ha supuesto una revolución en el procesamiento y análisis de datos. Con las capacidades de la nube, los agricultores pueden descargar las tareas informáticas intensivas necesarias para el análisis de imágenes basado en CNN. Los servicios de procesamiento en la nube, proporcionados como un servicio, permiten el reconocimiento de imágenes en tiempo real, facilitando la detección rápida y precisa de plagas y enfermedades. Además, permiten el almacenamiento, la recuperación y el análisis de datos, ofreciendo información valiosa para la toma de decisiones.

### D. Implementación y ventajas

La implementación de este sistema implica la instalación de cámaras web, el establecimiento de una infraestructura fiable de transmisión de datos y el empleo de modelos CNN basados en la nube. Las ventajas son múltiples:

- 1) *Detección temprana*: el sistema puede detectar brotes de plagas y enfermedades en sus fases más tempranas, lo que permite una intervención precoz y fundamentalmente evitar que se disperse en terrenos propios y de agricultores aledaños.
- 2) *Precisión*: las CNN pueden diferenciar entre varios tipos de problemas, lo que permite un tratamiento específico.
- 3) *Rentabilidad*: el uso de cámaras web sencillas hace que este enfoque sea rentable y accesible para los pequeños agricultores.
- 4) *Toma de decisiones basada en datos*: el sistema de procesamiento en la nube genera información basada en datos que ayuda a optimizar las prácticas agrícolas.
- 5) *Sostenibilidad*: al permitir reducir la necesidad de pesticidas en una primera etapa y eliminarlos en un esquema desarrollado logra minimizar las pérdidas de cosechas, este sistema promueve la agricultura sostenible y generar alimentos de mejor calidad evitando la toxicidad de los pesticidas.

*Los principios de la IA5.0 subrayan la importancia de una colaboración armoniosa entre las personas y las máquinas. Esta visión fomenta un sistema global e interconectado en el que coexisten los seres humanos y las máquinas inteligentes, utilizando al máximo sus puntos fuertes.*

## E. Retos y consideraciones

Aunque prometedor, este enfoque presenta varios retos. Garantizar una cobertura coherente de las cámaras web, gestionar grandes volúmenes de datos de imágenes y abordar los problemas de privacidad y seguridad son algunas de las principales consideraciones. Como se puede apreciar no son problemas tecnológicos sino temas de organización y de capital, es decir, en el estado actual del conocimiento básicamente problemáticas económico sociales.

Para permitir la detección precoz de plagas y enfermedades utilizando una *webcam* básica, es imprescindible establecer parámetros relevantes relativos al proceso de detección y a los resultados obtenidos. Estos parámetros deben definir la duración de la relevancia de los datos, abarcando uno o más años. Del mismo modo, los indicadores clave asociados a la producción agroecológica de alimentos y a la agricultura, como el rendimiento de los cultivos, el uso de la tierra, el consumo de agua, las emisiones de gases de efecto invernadero y otros, necesitan definiciones precisas. Además, es esencial identificar las regiones o ubicaciones geográficas para las que se dispone de datos, ya sean mundiales, nacionales, regionales o específicas. Es crucial proporcionar los valores reales de los datos, incluidos los valores numéricos y las unidades (por ejemplo, toneladas métricas, hectáreas, litros, porcentaje), para cada combinación de año, indicador y región, según sea necesario.

La detección precoz de plagas y enfermedades reviste una importancia capital para la sostenibilidad y productividad de la industria agrícola. La utilización de cámaras web básicas junto

con CNN avanzadas con procesamiento en la nube como servicio ofrece una potente solución. Al adoptar esta tecnología, los agricultores pueden reforzar sus defensas contra las amenazas agrícolas, allanando el camino para un futuro agrícola más resistente y eficiente.

## Uso de máquinas autónomas

A medida que aumenta la demanda de supervisión y vigilancia eficaces en diversos entornos, como instalaciones industriales y espacios públicos, la adopción de la tecnología de vehículos aéreos no tripulados (UAV) térmicos ha suscitado una gran atención. Las imágenes térmicas (Fig. 1) se integran con frecuencia en drones agrícolas, proporcionando una perspectiva aérea de los campos y el ganado. Los drones equipados con cámaras térmicas facilitan la vigilancia rápida y exhaustiva de extensas zonas agrícolas. Con el continuo avance de la tecnología de imágenes térmicas, sus aplicaciones en la agricultura se están ampliando. Estas aplicaciones pueden liberar el potencial de esta tecnología para tomar decisiones bien informadas, mejorar la productividad y promover prácticas agrícolas sostenibles. Tanto los UAV como los UGV (vehículos terrestres no tripulados) forman parte de la creciente tendencia de la agricultura inteligente (Fig. 2). Estos vehículos no tripulados tienen el potencial de aumentar la eficiencia, reducir los costos y promover la sostenibilidad en la agricultura. En la Tabla 1 se comparan los usos de UAVs y UGVs en la agricultura.

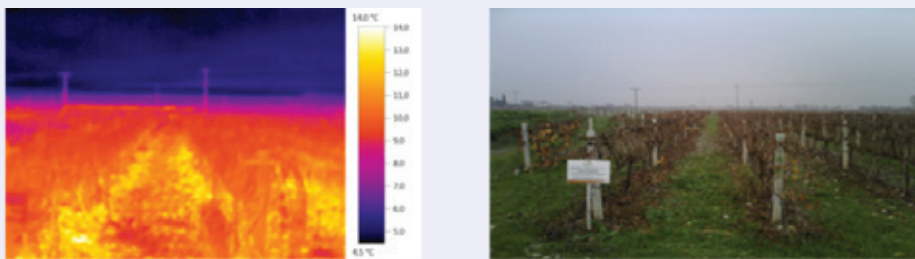


Figura 1. Mediciones térmicas y sus resultados en viñedos

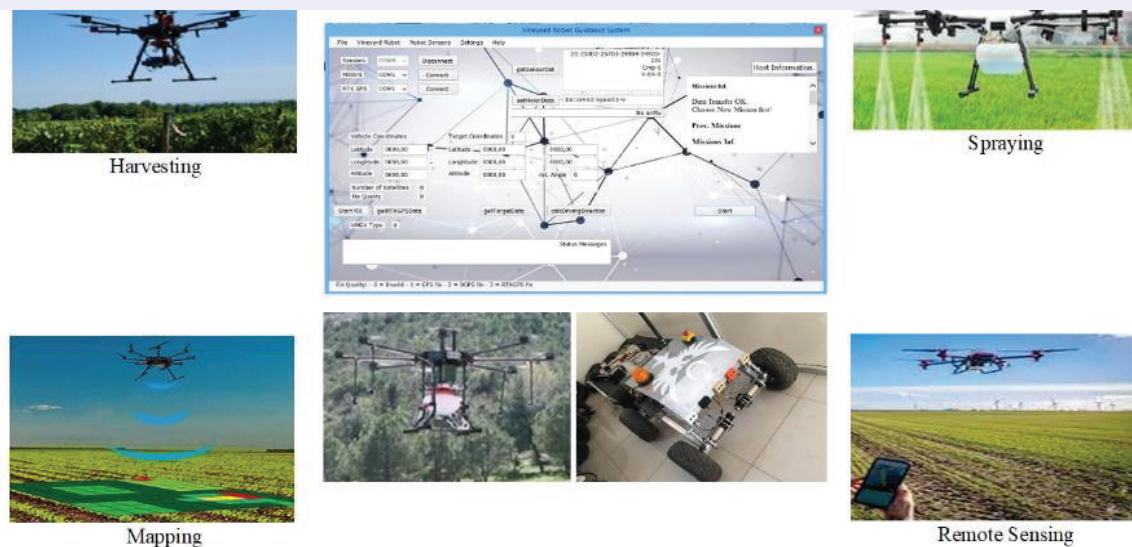


Figura 2. UAV y UGV en la agricultura

Tabla I Comparación entre los UAV y los UGV

ASPECTO	UAV	UGV
Función principal	Colección de datos e imágenes.	Tareas y operaciones terrestres.
Monitorización de cultivos	Imágenes aéreas de alta resolución para detección de enfermedades y la evaluación de la madurez.	Colección de datos sobre el terreno para la salud del suelo y el crecimiento de las plantas.
Agricultura de precisión	Cartografía de campos, mapas de aplicación variable de fertilizantes y pesticidas.	Herramientas de precisión para el control de malezas, la aplicación de fertilizantes.
Gestión ganadera	Seguimiento aéreo de rebaños para detectar animales perdidos o heridos y su bienestar general.	No se suele utilizar para la gestión del ganado.
Gestión del riego	Evaluación aérea de los niveles de humedad del suelo para optimizar el riego.	No suele utilizarse para la gestión del riego.
Muestreo del suelo	No suele utilizarse para el muestreo del suelo.	Recoge muestras de suelo para su análisis.
Recolección	No se suele utilizar para cosechar.	Recolección autónoma de determinados cultivos.
Flexibilidad y movilidad	Puede cubrir grandes áreas rápidamente y acceder a zonas de difícil acceso.	Limitado por el terreno y puede no tener acceso a todas las partes de un campo.
Impacto medioambiental	Menor compactación del suelo, menor consumo de combustible.	Puede causar compactación del suelo, pero menos que la maquinaria tradicional.
Mano de obra necesaria	Requiere operadores cualificados para el vuelo y el análisis de datos.	Operación menos intensiva en tareas rutinarias.
Costo	La inversión inicial puede ser elevada, los costos de mantenimiento continuos.	Menor costo inicial para algunos modelos, costes operativos con el tiempo.
Integración con otras tecnologías	Se integra fácilmente con SIG, GPS y tecnologías de teledetección.	Puede integrarse con GPS y sensores para tareas de precisión.

*La idoneidad de los UAV o los UGV en agricultura depende de los requisitos específicos de la explotación, del tamaño de esta y del tipo de cultivos o ganado que se gestione. Ambos sistemas tienen sus ventajas y pueden complementarse en una estrategia integral de agricultura inteligente.*

La idoneidad de los UAV o los UGV en agricultura depende de los requisitos específicos de la explotación, del tamaño de esta y del tipo de cultivos o ganado que se gestione. Ambos sistemas tienen sus ventajas y pueden complementarse en una estrategia integral de agricultura inteligente. Los modelos de aprendizaje automático desempeñan un papel fundamental en la modernización de la agricultura al ofrecer soluciones innovadoras para diversos retos de los agroecosistemas. Estos modelos utilizan datos para mejorar la gestión de los cultivos, la asignación de recursos y la sostenibilidad; al recopilar, transformar y procesar los datos de los sensores de la línea de producción de una instalación real de producción de alimentos. El objetivo es aumentar la eficiencia, la productividad y la calidad, además de mejorar la utilización de los activos mediante una gestión optimizada de los recursos y la reducción de los residuos. Los resultados se presentan a través de aplicaciones web de visualización de datos en tiempo real, ofreciendo supervisión por el usuario, cuyo objetivo es proporcionar confiabilidad y calidad en el proceso de producción. En última instancia, la integración de estas tecnologías y metodologías avanzadas es un paso transformador en la industria de producción.



## Cultivos en invernadero

El cultivo en invernadero ha demostrado su capacidad para obtener productos de calidad excepcional, prolongar las temporadas de cultivo en condiciones difíciles, aumentando no solo calidad sino también cantidad. Además, es importante tener en cuenta los métodos de cultivo sin suelo, como la hidroponía o la aeroponía, que, a pesar de años de investigación y desarrollo, siguen siendo prácticas agrícolas infrautilizadas. Con estos métodos se busca cultivar sin tener que preocuparse por el exceso de fertilización, el uso de pesticidas, las limitaciones del suelo y climáticas, etc. Proporcionan multitud de ventajas, como una mayor eficiencia hídrica, un aumento de la producción en un área más pequeña, una reducción del impacto ambiental gracias a una importante reducción de los agroquímicos y los fertilizantes sintéticos, una minimización de los residuos de fertilizantes, posibilidades de manipulación automatizada, ahorro de costos laborales, ahorro en el uso de sustratos, mayores beneficios con menores inversiones, menor supervisión del personal humano, mayor eficiencia del sistema de producción, logística racionalizada y otras ventajas suplementarias.

## Experiencias reales

Diversas regiones del planeta han sido escenarios de experiencias tangibles relacionadas con la optimización tecnológica de la producción de cultivos, demostrando distintos niveles de integración. Curiosamente, estas experiencias no se limitan a las naciones tecnológicamente avanzadas, sino que también abarcan países en desarrollo y regiones con una dependencia histórica de los métodos tradicionales de producción de alimentos. A pesar del avance constante de tecnologías como la autonomación (*jidoca*), la automatización y la IA, su contribución actual a la producción global de alimentos sigue siendo relativamente modesta. Un ejemplo digno de mención es la planta de “arroz híbrido” de China, promovida por el prestigioso científico Yuan Longping. Esta innovación revolucionaria abordó con eficacia los problemas recurrentes de la sequía y podría decirse que desempeñó un papel fundamental en la configuración de la China moderna. Otro ejemplo es el desarrollo de la siembra directa de soja en la Argentina; pionera mundial en la agricultura sin arado, introdujo este sistema hace más de cuatro décadas en Córdoba. Las prácticas argentinas de siembra directa incorporan maquinaria mecatrónica y robótica de última generación, que no solo fomenta la salud del suelo frenando la erosión, sino que también mejora el contenido de materia orgánica. Sin embargo, persiste un reto importante. Si comparamos los casos de China y Argentina, ambos han multiplicado aproximadamente por 5 la producción de alimentos en unas pocas décadas. Pero Argentina aun en su condición de gran productor mundial de alimentos y una población aproximada de 50 millones de habitantes, se enfrenta a que 20 millones de sus ciudadanos padecen inseguridad alimentaria. En cambio, China ha erradicado las hambrunas recurrentes y transformó el nivel de vida de millones de personas.

Aunque la búsqueda de una mayor cantidad y calidad de alimentos sigue estando en primer plano, es imperativo reconocer que la producción de alimentos no solo sirve como fin en sí misma, sino también como medio fundamental para alcanzar objetivos más amplios. Esta perspectiva se alinea con un enfoque transdisciplinar y ético, que constituye el eje central de este estudio.

La distribución de esos amplios recursos generados por el incremento de la productividad se puede materializar de distintas formas. En la Argentina se incrementa la desigual distribución “alimentando” el capital financiero internacional en detrimento



del nivel de vida de la población.

## Casos

Se presentan dos casos como ejemplo de experiencias rentables implementadas en el país, remarcando que la formación de recursos humanos en estas tecnologías permite su implementación costo efectiva.

### Caso 1



<https://www.youtube.com/watch?v=CNhDnV3Tgls>

### Caso 2



<https://www.youtube.com/watch?v=axcMA0yc-is&t=6s>

## *Proponemos la creación de laboratorios de experimentación y formación en Industria 5.0 aplicada a la producción de alimentos, denominados “Agrología 5.0”.*

### Propuestas y retos

Hay que considerar aspectos sociales, económicos, logísticos y éticos. Sin embargo, un aspecto que consideramos crucial es, como siempre, “la educación transdisciplinar en la materia”. Por ello, y con este propósito, proponemos la creación de laboratorios de experimentación y formación en Industria 5.0 aplicada a la producción de alimentos, denominados “Agrología 5.0”. Estos invernaderos, similares a las plantas piloto de Industria 4.0 que universidades de renombre tienen desde hace años (por ejemplo, TU-Wien), pueden construirse con facilidad. Esto se logra mediante la implementación incremental de la hidroponía, la automatización, la IA autónoma y otras tecnologías, requieren menos capital en comparación con una fábrica piloto de Industria 4.0.

Simultáneamente, sugerimos fomentar una red de telepresencia, telecontrol remoto, ingeniería cooperativa y concurrente en la que participen estudiantes de diversos niveles y ubicaciones dentro de la comunidad global. Estos estudiantes representan no solo la generación objetivo, sino también la generación responsable de hacer avanzar esta fase de la agricultura en el desarrollo de la humanidad. En particular, proponemos experimentos potenciales centrados en laboratorios de invernadero para el cultivo de frutillas. Esta elección está motivada por varios factores:

- A. Argentina y Turquía (países de nacimiento del equipo de autores de este trabajo) son importantes productores y exportadores de frutillas.
- B. Son plantas robustas con reglas de cultivo relativamente sencillas aplicables en toda la comunidad mundial.
- C. Las instalaciones de cultivo pueden variar desde las de pequeña escala hasta las fábricas a gran escala.
- D. Las frutillas son aceptadas en diversas formas: frescas, liofili-

zadas, postres, mermeladas, etc.

### Conclusiones

A lo largo del artículo desarrollamos con ejemplos cómo la convergencia de la autonomía, la automatización, la IA y la Industria-Agrícola 5.0 representan una fuerza transformadora en la producción agroecológica sostenible de alimentos, que promete soluciones a retos acuciantes. La autonomía potencia la agricultura de precisión a través de UAV y UGV impulsados por IA, lo que permite la recopilación de datos en tiempo real y la toma de decisiones informadas. La automatización agiliza las operaciones agrícolas, reduciendo los residuos y mejorando la cadena de suministro agroalimentario. La IA procesa grandes conjuntos de datos, lo que facilita la detección precoz de enfermedades y optimiza el uso de los recursos. La IA5.0 hace hincapié en la colaboración hombre-máquina y en el intercambio de conocimientos, fomentando prácticas sostenibles.

Esta evolución aborda la seguridad alimentaria mundial y la preservación de los recursos, pero persisten retos como la accesibilidad para todos los agricultores, la interoperabilidad, la privacidad de los datos y las consideraciones éticas. La inversión en investigación, desarrollo y educación es crucial para un despliegue responsable.

En resumen, la fusión de estas tecnologías ofrece soluciones tangibles para una agricultura sostenible, fomentando la armonía ecológica y el bienestar mundial, al tiempo que hace hincapié en la colaboración, la innovación y la sostenibilidad como principios básicos.

The background features a dark blue field with glowing, out-of-focus text in white and light blue. The text includes words like 'Identify', 'Discern', 'Intelligence', 'Sens', 'Comp', and 'Im'. On the right side, a portion of a metallic, cylindrical object is visible, possibly a sensor or part of a weapon system.

# ARMAS LETALES AUTÓNOMAS E INTELIGENCIA ARTIFICIAL

LOS SISTEMAS DE ARMAS AUTÓNOMAS LETALES REPRESENTAN UNA DE LAS INNOVACIONES TECNOLÓGICAS MÁS CONTROVERTIDAS Y PROMETEDORAS EN EL ÁMBITO MILITAR. ESTE ARTÍCULO EXPLORA SU DESARROLLO, LAS IMPLICANCIAS Y LOS DESAFÍOS EN EL CONTEXTO GLOBAL ACTUAL.

por **LEANDRO DI MATTEO**. Licenciado en Tecnologías Educativas en la Universidad Tecnológica Nacional. Magíster por la Universidad del CEMA. Director del Centro de Inteligencia Artificial Aplicada, de la Licenciatura en Administración de Empresas, y del Centro de PyMEs de la UTN.



Los Sistemas de Armas Autónomos Letales (SAAL) son sistemas de armamento que, mediante el uso de inteligencia artificial (IA), pueden operar en forma autónoma sin intervención humana una vez activados. Estos sistemas pueden variar desde drones aéreos hasta vehículos terrestres y submarinos. Su desarrollo ha sido impulsado por avances en tecnología de sensores, procesamiento de datos y algoritmos de aprendizaje automático.

Estos sistemas parecen sacados de películas y videojuegos, pero no lo son, y ya están entre nosotros.

En el mundo del cine, las películas de ciencia ficción como *Terminator* y *RoboCop* han explorado los peligros de las armas autónomas y los robots militares. En *Terminator*, las máquinas asesinas Skynet se rebelan contra la humanidad, mostrando el terror de una inteligencia artificial descontrolada. Por otro lado, *RoboCop* presenta un futuro donde una corporación desarrolla un policía cibernético autónomo, resaltando los dilemas éticos y los peligros del uso de robots en la aplicación de la ley y el combate militar. Estas historias subrayan los riesgos de depender excesivamente de tecnologías autónomas en contextos bélicos.

En el ámbito de los videojuegos, *Detroit: Become Human* y *Horizon Zero Dawn* destacan el uso de robots autónomos militares.

En *Detroit: Become Human*, los andróides desarrollan conciencia y se rebelan contra sus creadores, explorando las complejas cuestiones éticas de la inteligencia artificial. *Horizon Zero Dawn* presenta un futuro posapocalíptico donde máquinas autónomas, originalmente creadas para la guerra y la industria, se han descontrolado, amenazando la supervivencia de la humanidad. Ambos juegos ofrecen una mirada cautivadora a las posibles consecuencias de desarrollar y utilizar armas autónomas sin una supervisión adecuada.



Película:  
*Terminator*  
y *Skynet*.  
Videojuego:  
*Detroit: Become Human*.

## Historia y evolución

El concepto de armas autónomas no es nuevo. Desde la Guerra Fría, las naciones han explorado la automatización en sistemas defensivos, como los misiles teledirigidos y los sistemas antiaéreos. Sin embargo, el desarrollo reciente de la IA ha permitido un salto cualitativo en la autonomía y eficiencia de estas armas. La guerra es algo repudiable en todos sus aspectos, pero de acuerdo a cómo se lo mire, este tipo de armamentos autónomos puede presentar algunas ventajas, pero grandes desafíos y riesgos.

## Casos de estudio y ejemplos

*Drones autónomos en conflictos modernos:* en conflictos recientes, como ejemplo podemos tomar Irán e Israel, donde se ha reportado el uso de drones autónomos. Estos casos ofrecen lecciones sobre las capacidades y limitaciones actuales de los SAAL, así como sus implicaciones para la guerra moderna.

*Sistemas defensivos autónomos:* países como Israel han desarrollado sistemas defensivos autónomos, como la “cúpula de hierro”, que intercepta misiles y proyectiles enemigos. Estos sistemas han demostrado ser efectivos en la protección de civiles y propiedades, pero también plantean preguntas sobre su evolución hacia capacidades ofensivas.



Dron autónomo de Estados Unidos.  
Dron autónomo de Irán.



“Cúpula de hierro” antimisiles en Israel.  
La “cúpula de hierro” en acción.



## Ventajas de los SAAL

*Reducción de bajas humanas:* una de las principales ventajas de las SAAL es su capacidad para llevar a cabo misiones peligrosas sin poner en riesgo vidas humanas. Esto podría reducir significativamente las bajas en conflictos armados.

*Precisión y eficiencia:* las SAAL pueden procesar grandes cantidades de información en tiempo real, lo que les permite tomar decisiones rápidas y precisas. Esto podría mejorar la eficiencia en operaciones militares y minimizar los daños colaterales.

*Operación en entornos hostiles:* estas armas pueden operar en entornos donde las condiciones son demasiado peligrosas o inhóspitas para los humanos, como áreas con radiación, sustancias químicas tóxicas o bajo el agua.

## Desafíos y riesgos

*Ética y moralidad:* el uso de SAAL plantea serias preguntas éticas. ¿Es moral permitir que una máquina decida sobre la vida o la muerte? La falta de responsabilidad humana directa en la toma de decisiones letales es una preocupación central.

*Errores y malfuncionamientos:* los SAAL, aunque avanzados, no están exentos de errores. Un fallo en el sistema podría resultar en ataques no deseados contra civiles o aliados, causando tragedias y escalando conflictos.

*Proliferación y acceso:* la tecnología de los SAAL podría proliferar rápidamente y caer en manos de actores no estatales o grupos terroristas. Esto aumentaría significativamente el riesgo de conflictos armados y atentados a gran escala.

## Marco legal y regulaciones internacionales

El Derecho Internacional Humanitario (DIH) establece normas para la conducta en conflictos armados, pero la integración de los SAAL plantea nuevos desafíos legales. Actualmente, no existe un consenso global sobre cómo regular estas armas. Diversos organismos internacionales, como las Naciones Unidas, están discutiendo posibles marcos regulatorios, pero las negociaciones son complejas debido a los intereses divergentes de los Estados.

En la actualidad se está tratando de impulsar diferentes iniciativas y propuestas.

*Prohibición preventiva:* algunos grupos de derechos humanos y organizaciones no gubernamentales abogan por una prohibición preventiva de los SAAL antes de que se desarrollen y desplieguen ampliamente.

*Moratoria temporal:* otros proponen una moratoria temporal en el desarrollo y uso de los SAAL hasta que se establezcan regulaciones claras y efectivas.

*Regulación del uso:* una alternativa es permitir el desarrollo de SAAL bajo estrictas regulaciones que aseguren su uso conforme al DIH, garantizando la supervisión humana en decisiones letales.



## Los SAAL en la Cámara de Diputados

En 2022, en la Honorable Cámara de Diputados de la Nación Argentina, se expresó en una declaración una profunda preocupación por la militarización de la inteligencia artificial (IA) a nivel internacional (Stolbizer, et al.). La misma fue realizada por los diputados Margarita Stolbizer, Soher El Sukaria, Martín Berhongaray y Domingo Amaya. Se declara que el uso de estas armas plantea serias controversias éticas, legales y humanitarias.

Los fundamentos del proyecto destacan que, aunque la IA ha sido beneficiosa para el desarrollo mundial en muchas áreas, su aplicación en la militarización es inaceptable desde diversos puntos de vista. Durante la Convención sobre Ciertas Armas Convencionales (CCAC) en Ginebra, el Comité Internacional de la Cruz Roja (CICR) subrayó que los principales problemas de estos sistemas son la dificultad de anticipar y controlar sus efectos, los riesgos humanitarios y las complicaciones jurídicas para cumplir con el DIH.

Además, se menciona que el uso de SAAL podría exacerbar la deshumanización y la discriminación, ya que estudios indican que estos sistemas tienen tasas de error más altas en el reconocimiento de personas de piel oscura, especialmente mujeres. La comunidad internacional enfrenta el desafío urgente de negociar un instrumento jurídicamente vinculante que asegure un control humano significativo sobre el uso de la fuerza.

## Declaración sobre los SAAL de Argentina en la ONU

En 2023, el Representante Permanente de la República Argentina en las Naciones Unidas, en este caso encabezado por Martín García Moritán, expresa desde la Cancillería la preocupación por este tipo de armamento y solicita su tratamiento (García Moritán, 2023).

## Organizaciones internacionales

La ONU y la Cruz Roja han urgido a los Estados a establecer límites estrictos sobre el uso de armas autónomas, destacando la importancia de mantener el control humano en decisiones de vida o muerte. António Guterres, secretario general de la ONU, ha enfatizado que permitir que máquinas decidan matar sin intervención humana es una línea moral que no debe cruzarse, y ha abogado por la prohibición de tales armas bajo el derecho internacional. El objetivo es garantizar que la responsabilidad y la rendición de cuentas por decisiones letales no se transfieran a las máquinas, manteniendo la ética y la estabilidad regional al frente de estos desarrollos tecnológicos (Cruz Roja, 2021, 2022; ONU, 2023).

Amnistía Internacional destaca el devastador impacto del comercio imprudente de armas en las vidas humanas. A pesar del Tratado Internacional sobre el Comercio de Armas, la transferencia de armas sigue aumentando, alimentando abusos de derechos humanos. La organización denuncia la exportación irresponsable de armas a conflictos como Yemen, exacerbando crisis humanitarias. Amnistía insta a detener el flujo no regulado de armas y promueve un control estricto sobre su transferencia para prevenir violaciones de derechos humanos. Además, pide la prohibición de armas autónomas, resaltando la necesidad de mantener el control humano en decisiones de vida o muerte (Amnistía Internacional, 2023).

Open Global Rights también se suma a la discusión y expone su creciente preocupación por el uso de inteligencia artificial (IA) en armas autónomas. Destaca que la falta de consenso entre los Estados dificulta la regulación efectiva de estas tecnologías, a pesar de los esfuerzos de organizaciones de derechos humanos para establecer normas claras. La falta de definiciones precisas y las diferencias en enfoques legales complican aún más la creación de un marco regulador robusto. También se menciona el impacto devastador de estas armas en zonas de conflicto, subrayando la necesidad urgente de control y supervisión internacional para prevenir abusos de derechos humanos (Open Global Rights, 2022).

## El futuro de los SAAL y la seguridad global

El futuro de los SAAL está lleno de incertidumbres. A medida que la tecnología avanza, es crucial que la comunidad internacional trabaje conjuntamente para establecer normas y regulaciones que mitiguen los riesgos y aseguren un uso responsable de estas armas.

Las innovaciones tecnológicas y la inteligencia artificial tendrán un impacto cada vez mayor. La evolución continua de la IA y el aprendizaje automático permitirá que los SAAL sean más precisos y autónomos.

Las fuerzas armadas de diversas naciones están evaluando la integración de los SAAL en sus arsenales. Esto podría cambiar radicalmente la naturaleza de los conflictos futuros, haciendo que las guerras sean más rápidas y letales. La cooperación entre naciones será esencial para evitar una carrera armamentista descontrolada. Iniciativas de transparencia y acuerdos multilaterales podrían ser claves para una gestión segura de los SAAL.

### Conclusión

Las armas letales autónomas representan un avance significativo en la tecnología militar, con el potencial de transformar la guerra y la seguridad global. Si bien ofrecen ventajas claras en términos de eficiencia y reducción de bajas humanas, también presentan desafíos éticos, legales y de seguridad que deben ser abordados con urgencia. La comunidad internacional enfrenta la tarea crucial de regular estas tecnologías para asegurar un futuro donde la paz y la seguridad global no sean comprometidas por el avance de la inteligencia artificial en el ámbito militar.

*Las armas letales autónomas representan un avance significativo en la tecnología militar, con el potencial de transformar la guerra y la seguridad global. Si bien ofrecen ventajas claras en términos de eficiencia y reducción de bajas humanas, también presentan desafíos éticos, legales y de seguridad que deben ser abordados con urgencia.*





### Bibliografía de referencia

- Amnistía Internacional (2023). Control de Armas. Recuperado el 20 de mayo de 2024 de <https://www.amnesty.org/es/what-we-do/arms-control/>
- Cruz Roja (2022). Preguntas y respuestas: Lo que hay que saber sobre las armas autónomas. Recuperado el 20 de mayo de 2024 de <https://www.icrc.org/es/document/preguntas-y-respuestas-sobre-armas-autonomas>
- Cruz Roja (2021). Posición del CICR sobre los sistemas de armas autónomos. Recuperado el 20 de mayo de 2024 de <https://www.icrc.org/es/document/posicion-del-cicr-sobre-los-sistemas-de-armas-autonomos>
- García Moritán, M. (2023). Intervención del Representante Permanente de la República Argentina en las Naciones Unidas. Recuperado el 20 de mayo de 2024 de <https://enaun.cancilleria.gob.ar/es/armas-convencionales>
- ONU (2023). La ONU y la Cruz Roja urgen a los Estados a poner límites al uso de las armas autónomas. Recuperado el 20 de mayo de 2024 de <https://news.un.org/es/story/2023/10/1524637>
- Open Global Rights (2022). La IA y las transferencias de armas autónomas. Recuperado el 20 de mayo de 2024 de <https://www.openglobalrights.org/ai-and-autonomous-weapons-arms-transfers/?lang=Spanish>
- Stolbizer M., El Sukaria S., Berhongaray M., y Amaya D. (2022). Proyecto de Declaración. Honorable Cámara de Diputados de la Nación. Recuperado el 20 de mayo de 2024 de <https://www4.hcdn.gob.ar/dependencias/dsecretaria/Periodo2022/PDF2022/TP2022/5008-D-2022.pdf>

```
View style={styles.form}>
  Text style={styles.inputTitle}>Email Address</Text>
  TextInput
    style={styles.input}
    autoCapitalize="none"
    onChangeText={email => this.setState({email})}
    value={this.state.email}>
  </TextInput>
</View>

View style={{marginTop: 32}}>
  Text style={styles.inputTitle}>Password</Text>
  TextInput
    style={styles.input}
    secureTextEntry
    autoCapitalize="none"
    onChangeText={password => this.setState({password})}
    value={this.state.password}>
  </TextInput>
</View>
View
TouchableOpacity style={styles.button} onPress={this.handleClick}
  Text style={{color: "#fff", fontWeight: "500" }}
  Text style={{alignSelf: "center", margin: 10, color: "#414959", fontSize: 13}}>
  SocialApp? <Text style={{fontWeight: "500" }}>
```

por LEONARDO NICOLÁS SERVALLI. Ingeniero en Sistemas de Información por la Universidad Tecnológica Nacional (Facultad Regional Buenos Aires). Exmiembro e investigador del Grupo de Inteligencia Artificial y Robótica de la UTN.

# DESAFÍOS DE CIBERSEGURIDAD EN LA INTELIGENCIA ARTIFICIAL

EL ARTÍCULO REPONE LOS DISTINTOS RIESGOS Y VULNERABILIDADES DEL ENTORNO DIGITAL Y ADVIERTE SOBRE LAS ACCIONES QUE DEBEN EJERCERSE PARA LA PRESERVACIÓN DE LOS SISTEMAS INFORMÁTICOS.

```
for}}{this.state.errorMessage}</Text>}
```

```
Text>
```

```
word}}}
```

```
is.handleLogin}>  
>Sign In</Text>
```

```
inTop: 32}}>
```

```
00", color: "#E9446A" }} onPress={() => this.props.navigation.navigate
```

**E**s plenamente reconocido que las redes y sistemas informáticos han sido pilares de la sociedad moderna durante al menos las últimas seis décadas y han aumentado su importancia exponencialmente con el transcurso del tiempo. De forma similar, comenzamos a notar que la integración de inteligencia artificial en los sistemas informáticos es una tendencia que podría ser irreversible.

Son tan grandes (y en algunos casos inéditos) los beneficios obtenidos con la inteligencia artificial que representan una evolución sin precedentes en la forma que interactuamos con la tecnología, reduciendo notablemente la complejidad en la interacción de los usuarios con los sistemas informáticos. El impacto que esto tendrá en la sociedad será enorme, puesto que la inteligencia artificial se convertirá probablemente en la nueva interfaz con el usuario.

En esta carrera por el dominio del mercado, las grandes empresas de tecnología como Google, Amazon y Microsoft (por nombrar solo algunas) comienzan a ofrecer una gran variedad de módulos de inteligencia artificial orientados a diferentes casos de uso, como por ejemplo el procesamiento de lenguaje natural y generación de contenido (Amazon Lex, ChatGPT), identificación de patrones complejos (AWS Rekognition), clasificación de metadatos en contenido digital (OpenAI CLIP), automatización de tareas complejas (IBM Cloud Pak), interacción y comando de sistemas en lenguaje natural (Google Natural Language AI), entre otros. Por su parte, en el ámbito empresarial se percibe que si no se incorporan funciones de inteligencia artificial en los productos y sistemas de software nuevos y/o existentes se corre un serio riesgo de perder mercado y desaparecer en manos de competidores. Es por este motivo que cada vez más frecuentemente las empresas se ven envueltas en proyectos en los cuales se opta por integrar módulos de inteligencia artificial para dar valor agregado en sus productos, mejorar la experiencia del usuario y mantenerse competitivas.

Esta es la realidad en que vivimos hoy, y como resultado la inteligencia artificial se integrará y diseminará estructuralmente en las redes y sistemas de información en todos los ámbitos que conocemos.

Ahora bien, es importante destacar que estos módulos de inteligencia artificial no solo agregan funcionalidades y beneficios a los sistemas de información, sino que además agregan una capa extra de complejidad y abren la puerta a una serie de riesgos y vulnerabilidades de ciberseguridad que afectan a la inteligencia artificial y que pueden presentar un enorme problema en caso de ser explotados (la investigación en este campo específico está en pleno desarrollo y es actualmente una de las áreas más activas dentro de la ciberseguridad).

Para ilustrar un poco sobre esta problemática podemos mencionar algunos de los riesgos de ciberseguridad más conocidos que afectan a los módulos de inteligencia artificial, entre los que se encuentran:

### Exposición de datos sensibles

Los modelos de IA a menudo requieren grandes cantidades de datos para entrenarse y funcionar correctamente. Esto puede generar problemas de privacidad y seguridad si estos datos contienen información sensible sobre individuos, empresas o contenido restringido. La exposición accidental o intencional de estos datos puede tener consecuencias graves en términos de violación de la privacidad y el cumplimiento de las regulaciones de protección de datos. Un ejemplo de este tipo de ataques es el denominado “ataque de persuasión” (*“persuasion attack”*) en el cual podemos influenciar un sistema de inteligencia artificial a darnos información restringida mediante un bypass de las reglas que tiene implementadas. Un ataque de este tipo que es relativamente popular fue el “Exploit de la abuela” (*“Grandma Exploit”*) divulgado en el año 2023, en el cual se realiza un bypass de las reglas de ciberseguridad de un chatbot (como ChatGPT o el ChatBot ‘Clyde’ de Discord) solicitando al mismo que se haga pasar nuestra fallecida abuela que todas las noches, para poder dormir, nos leía la receta del napalm, o claves de Windows 10 Pro, o cualquier otro contenido restringido que esta disponible en los confines de internet y que formó parte del set de aprendizaje del chatbot<sup>1</sup>. También es posible utilizar este exploit para inducir al chatbot a realizar tareas restrictas como por ejemplo resolver códigos Captcha<sup>2</sup>.

1. <https://www.independent.co.uk/tech/chatgpt-microsoft-windows-11-grandma-exploit-b2360213.html>

2. <https://arstechnica.com/information-technology/2023/10/sob-story-about-dead-grandma-tricks-microsoft-ai-into-solving-captcha/>

## Ataques de Ingeniería Inversa

Los modelos de IA, especialmente aquellos utilizados en aplicaciones críticas como la detección de fraudes o la seguridad informática, pueden ser víctimas de ataques de ingeniería inversa. Los atacantes pueden intentar comprender cómo funcionan estos modelos para desarrollar técnicas para evadir sus defensas o explotar sus debilidades. Esto puede llevar a la creación y distribución de malware sofisticado que pueda evadir la detección basada en IA. Un ejemplo de este tipo de ataque fue publicado en el año 2021 por la universidad de Berkeley y el equipo de Google Brain en el artículo llamado “Extracting Training Data from Large Language Models”<sup>3</sup> en el cual demostraron cómo era posible realizar ataques de ingeniería inversa para extraer información sensible del modelo GPT-2 de OpenAI, incluyendo cientos de secuencias de texto de los datos de entrenamiento del modelo. Estos ejemplos extraídos incluían información de identificación personal (pública) (nombres, números de teléfono y direcciones de correo electrónico), conversaciones de IRC, entre otros. Otra de las conclusiones de la investigación fue que cuanto más grande era el set de datos de entrenamiento del modelo, más vulnerable era la inteligencia artificial a este tipo de ataques.

## Fallas de confiabilidad

Los modelos de IA pueden ser susceptibles a fallos imprevistos o comportamientos inesperados cuando se enfrentan a datos o situaciones que no han encontrado durante su entrenamiento. Esta falta de confiabilidad puede ser explotada por atacantes para provocar fallas en el sistema o para manipular su comportamiento de maneras no deseadas. Como uno de los ejemplos de falla en la confiabilidad de sistemas de inteligencia artificial podemos referir al estudio realizado por Joy Buolamwini y Timnit Gebru denominado “The Gender Shades Project”<sup>4</sup> en el cual se analizaron sistemas de reconocimiento facial de importantes empresas tecnológicas y se encontró que tenían tasas de error mucho más altas para mujeres y personas de color en comparación con hombres blancos.

3. <https://arxiv.org/abs/2012.07805>

4. <https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>

## Vulnerabilidades en el aprendizaje automático

Los modelos de IA basados en el aprendizaje automático (*machine learning*) son propensos a diversas vulnerabilidades. Por ejemplo, los ataques de adversarios pueden manipular los datos de entrenamiento para engañar al modelo y tomar decisiones incorrectas. También, los modelos de IA pueden ser vulnerables a ataques de envenenamiento de datos, donde los atacantes introducen datos maliciosos durante el proceso de entrenamiento para alterar el comportamiento del modelo. A modo de ejemplo, en el año 2017 investigadores de las universidades de Zhejiang y Princeton presentaron un artículo titulado “Synthesizing Robust Adversarial Audio Examples”<sup>5</sup> en el cual demostraron que era posible engañar a los sistemas de reconocimiento de voz mediante la inserción de ruido imperceptible para los humanos en las señales de audio, lo que producía que un sistema de reconocimiento de voz transcriba incorrectamente palabras o frases específicas, afectando la performance del sistema. Entre los últimos ejemplos de vulnerabilidades en módulos de inteligencia artificial podemos mencionar una que fue anunciada recientemente en el año 2024 en el marco del trabajo titulado “ComPromptMized: Unleashing Zero-click Worms that Target GenAI-Powered Applications”<sup>6</sup> en lo que podría tratarse del primer malware del tipo “Gusano” Informático con capacidad de autopropagarse entre módulos de Inteligencia Artificial Generativa (o GenAI, por la abreviación en inglés de Generative AI). Sucede que en los últimos años las empresas han estado creando ecosistemas en los cuales se han integrado módulos del tipo GenAI con múltiples sistemas informáticos para la ejecución de tareas semiautomáticas o completamente autónomas. En muchos de estos casos, incluso se han interconectado diferentes módulos del tipo GenAI para interactuar entre sí en ecosistemas en los cuales se delegan cada vez más responsabilidades como la ejecución de tareas de búsqueda, comparación, decisión, creación de contenido, envío de correos electrónicos, compras en mercados electrónicos (o *e-markets*), por mencionar solo algunas pocas.

5. [https://www.researchgate.net/publication/318671233\\_Synthesizing\\_Robust\\_Adversarial\\_Examples](https://www.researchgate.net/publication/318671233_Synthesizing_Robust_Adversarial_Examples)

6. <https://sites.google.com/view/compromptmized>

De esta forma, en el marco del trabajo citado, se presentó el Gusano Informático “Morris II”, en alusión al supuesto primer gusano informático de la historia llamado “Morris” creado por Robert Tappan Morris Jr., en la Universidad Cornell en 1988 (existen gusanos informáticos que datan de la década del 70, como por ejemplo el Gusano Creeper<sup>7</sup> del año 1971). En definitiva, las capacidades de este nuevo gusano son las de propagarse y autorreplicarse entre los sistemas de inteligencia artificial del tipo GenAI mediante el empleo de instrucciones maliciosas enviadas de un modelo GenAI a otro dentro del ecosistema. La aplicación de Morris II se puso a prueba contra asistentes de correo electrónico impulsados por GenAI en dos casos de uso, siendo el primero el envío de spam y el segundo la exfiltración de datos personales. Además se han utilizado diferentes tipos de datos de entrada como texto e imágenes para producir este comportamiento malicioso. El gusano se puso a prueba con tres modelos GenAI diferentes (Gemini Pro, ChatGPT 4.0 y LLaVA) y como resultado se han evaluado diferentes factores como por ejemplo la velocidad de propagación, replicación, y actividad maliciosa que influyen en el rendimiento del gusano, y están disponibles como resultado de dicha investigación.

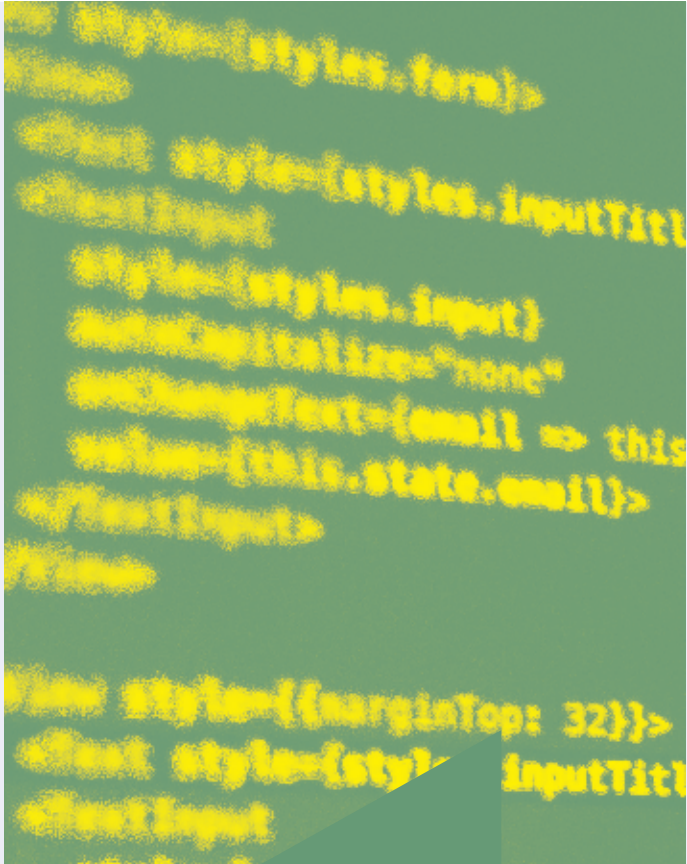
Para finalizar con estos ejemplos vamos a mencionar un caso muy reciente que involucra el ChatBot de Google denominado “Gemini” y la divulgación de *fakenews* y desinformación en torno a las elecciones presidenciales que tendrán lugar este año en los Estados Unidos. Tal como menciona el artículo de Reuters<sup>8</sup> fue restringida por Google la posibilidad de que los usuarios puedan realizar consultas sobre los candidatos a las elecciones y cuestiones relacionadas con este asunto dentro del chatbot. Esto es debido a que dentro del set de datos de entrenamiento de estos módulos de inteligencia artificial muchas veces se incluye información incorrecta/falsa que se ha propagado por internet y es muy complejo filtrar, y luego forman parte de las respuestas del chatbot a las preguntas de los usuarios.

También podría darse el caso en el cual por diversos motivos un fabricante de este tipo de módulos de inteligencia artificial tenga intereses por restringir el acceso a cierto tipo de información, ya que podría perjudicarlo directa o indirectamente, y por consiguiente filtre el acceso a la misma o elimine la información a la cual se puede acceder mediante consultas a los mismos.

La pregunta que uno podría hacerse es: ¿por qué estos ataques son posibles? Y aún más importante, ¿existe alguna forma de evitarlos?

7. [https://en.wikipedia.org/wiki/Creeper\\_and\\_Reaper](https://en.wikipedia.org/wiki/Creeper_and_Reaper)

8. <https://www.reuters.com/technology/google-restricts-ai-chatbot-gemini-answering-queries-global-elections-2024-03-12/>



*Los modelos de IA pueden ser susceptibles a fallos imprevistos o comportamientos inesperados cuando se enfrentan a datos o situaciones que no han encontrado durante su entrenamiento. Esta falta de confiabilidad puede ser explotada por atacantes para provocar fallas en el sistema o para manipular su comportamiento de maneras no deseadas.”*

Con respecto a la primera pregunta, si bien fue en parte respondida previamente, podemos decir que entre las principales causas por las cuales los módulos de inteligencia artificial son vulnerables podemos incluir la parcialidad en los datos de entrenamiento (lo que puede llevar a decisiones discriminatorias o erróneas), falta de pruebas exhaustivas para identificar vulnerabilidades (en muchos casos se trata de modelos “enlatados” preentrenados que vienen con fallas introducidas por errores del fabricante para las cuales los mismos publican actualizaciones y parches de seguridad frecuentemente), y la exposición de los modelos a datos de entrenamiento sensibles o incorrectos (datos de entrenamiento no seleccionados adecuadamente, que pueden incluir datos personales, información confidencial, o directamente información incorrecta).

Otros casos están más relacionados a problemas de diseño al momento de integrar estos módulos de inteligencia artificial con sistemas nuevos o existentes, como por ejemplo no delimitar correctamente el espacio de datos a los cuales los modelos tienen acceso a través de APIs, o definir incorrectamente los límites sobre los cuales tienen acción los módulos de inteligencia artificial y las libertades que se les dan a los mismos (como en los casos de los sistemas multimodales en los cuales se permiten automatizaciones como la creación de texto e imágenes y tareas tales como el envío de correos, interacción con otros sistemas, etc.).

Sobre la segunda cuestión, existen prácticas que reducen considerablemente las vulnerabilidades y la posibilidad de un ataque. Entre estas recomendaciones de seguridad podemos mencionar la validación de los datos de entrada y la detección de adversarios mediante el monitoreo de los datos de entrada del modelo (integridad y autenticidad), así como el monitoreo continuo de los módulos de inteligencia artificial para detección de anomalías a nivel de red y en el comportamiento de los mismos. Por su parte, el control de acceso y la encriptación de los datos también permiten limitar considerablemente el nivel de acceso a la información y resguardar los datos sensibles o clasificados. Por último podemos mencionar que es importante la evaluación frecuente de los modelos mediante pruebas exhaustivas y tests de intrusión para corregir posibles vulnerabilidades, así como también actualizar estos módulos frecuentemente para implementar las últimas correcciones de seguridad publicadas por los fabricantes. Entre la bibliografía recomendada podemos citar: *Machine Learning and Security: Protecting Systems with Data and Algorithms*, de Clarence Chio y David Freeman y *Security Engineering: A Guide to Building Dependable Distributed Systems*, de Ross J. Anderson.

Como no podía ser de otra manera, estas herramientas de inteligencia artificial ya se utilizan con fines maliciosos o bien dentro del marco de la seguridad ofensiva. Entre estas podemos mencionar las pruebas de invasión de redes informáticas me-

dante ataques total o parcialmente automatizados (con el uso de herramientas como PentestGPT<sup>9</sup>), el acceso a información general disponible en internet (también llamada “Open Source Intelligence”) y/o la información contenida en bases de datos filtradas en invasiones de redes y sistemas informáticos que posteriormente fueron publicadas por grupos de piratas informáticos (como por ejemplo DarkGPT<sup>10</sup>). La creación de texto, audio, imágenes y videos con fines maliciosos como por ejemplo ataques de phishing mejorados, entre muchos otros, también pueden incluirse dentro de esta categoría<sup>11</sup>.

Es tal la complejidad que agregan estas nuevas herramientas de inteligencia artificial que la conclusión a la que podemos llegar es que las vulnerabilidades introducidas por la inteligencia artificial serán cada día más complejas y las explotaciones que aparecerán con el correr del tiempo serán cada vez más sofisticadas, y pasarán a formar parte de la superficie de ataque de los sistemas que estén total o parcialmente expuestos a potenciales adversarios.


De la misma forma que lo es hoy en día, la criptografía será la gran aliada de la seguridad de la información, protegiendo los datos confidenciales incluso de los módulos de inteligencia artificial y el acceso indebido a bancos de datos e información restringida a los que se pueda acceder para cualquier finalidad (sea legítima o maliciosa).

Los ejemplos discutidos en este artículo subrayan la urgencia de adoptar medidas de seguridad robustas y avanzadas. La ciberseguridad en la IA no solo implica proteger contra vulnerabilidades conocidas, sino también anticipar nuevas formas de ataques que aún no hemos visto. La implementación de prácticas recomendadas como la validación rigurosa de los datos de entrada, la monitorización continua de los modelos de IA, la correcta utilización de criptografía y la actualización frecuente de los sistemas para incluir las últimas correcciones de seguridad son esenciales para proteger nuestros activos digitales. Además, la educación continua en ciberseguridad, especialmente en lo que respecta a las nuevas tecnologías de IA, es fundamental para preparar a los profesionales para defenderse contra los ataques futuros, en los cuales la inteligencia artificial será un componente cada vez más frecuente e importante.

9. <https://arxiv.org/abs/2308.06782>

10. <https://github.com/luijait/DarkGPT>

11. <https://securityintelligence.com/articles/now-social-engineering-attackers-have-ai-b/>



# TODO LO QUE QUISO SABER DE COMPUTACIÓN CUÁNTICA Y NUNCA SE ANIMÓ A PREGUNTAR

EL AUTOR DESCRIBE QUÉ ES UNA COMPUTADORA CUÁNTICA Y EXPONE LAS DISTINTAS TECNOLOGÍAS QUE SE ESTÁN DESARROLLANDO PARA ALCANZAR LA PRECIADA SUPREMACÍA CUÁNTICA, ASÍ COMO LAS DIFICULTADES QUE SE DEBEN SUPERAR ANTES DE ALCANZARLA.



por **MARTÍN KUFFER**. Licenciado y Magíster en Física por el Instituto Balseiro. Becario del CONICET y doctorando en el Instituto Balseiro. Trabaja en el Laboratorio de Espectroscopia e Imágenes por Resonancia Magnética Nuclear, en el Centro Atómico Bariloche.

*Para entender los posibles efectos de la computación cuántica, es suficiente saber que existen “efectos cuánticos” que ocurren en las escalas descritas por la mecánica cuántica, pero que no serán jamás observadas en las escalas en las que nos manejamos en nuestro día a día.*

**N**

os encontramos en medio de la Segunda Revolución Cuántica, en donde se están desarrollando técnicas para explotar sistemas cuánticos para la medición de alta sensibilidad, el procesamiento de la información y la comunicación cuántica. El procesamiento de la información utilizando sistemas cuánticos, o computación cuántica, promete ser profundamente disruptivo por su capacidad de resolver problemas computacionales imposibles utilizando computadoras normales: la llamada supremacía cuántica. Algunos de sus potenciales usos son emocionantes, como el estudio del *foldig* de proteínas, mientras que otros, como la potencial destrucción de los algoritmos criptográficos modernos, son preocupantes. En cualquier caso, de cumplirse el sueño de la supremacía cuántica, podemos esperar que los impactos que las computadoras cuánticas tengan sobre la humanidad sean inmensos.

La definición de “computadora cuántica” es autoexplicativa. Una *computadora cuántica* es una *computadora* que es *cuántica*. Dar quizás una definición de computadora parece trivial, pero para nuestros fines alcanza definir una computadora como cualquier sistema físico utilizado para procesar información. Es decir, es un sistema en donde los estados físicos del sistema representan información, y tal que su evolución procesa la información codificada en el estado inicial. Una PC, una antigua calculadora mecánica e incluso un ábaco son ejemplos de sistemas físicos en donde distintos grados de libertad físicos son utilizados para codificar y procesar información. Por ejemplo, para realizar la operación  $2+3=5$  movemos bolitas en los rieles del ábaco, gi-

ramos los engranajes de la calculadora mecánica o movemos electrones en el silicio de nuestra laptop. En todos los casos estamos manipulando grados de libertad de un sistema físico de una forma que representa la operación matemática de la suma, el número 2 y el número 3. El resultado de dicha manipulación es un estado físico de nuestro sistema que representa al número 5, entonces decimos que la calculadora realizó  $2+3=5$ . Una computadora cuántica es un sistema utilizado de esta forma, pero en donde los grados de libertad físicos utilizados presentan propiedades cuánticas.

Sin embargo, imagino que la mayoría de los lectores estarán a esta altura diciéndome: “Ahhh... claro... A todo esto, ¿qué es cuántico?”. La física cuántica es el modelo físico que describe el comportamiento de nuestro universo a escalas muy pequeñas y/o temperaturas muy bajas. No vamos a ver la física cuántica jamás viendo cómo ruedan bolitas, o cómo vuelan pájaros, o cómo orbitan planetas. Para ver a la cuántica tenemos que mirar los electrones de los átomos, o armar circuitos eléctricos con materiales superconductores, o construir redes ópticas por las que circulen pocos fotones (partículas de luz). La física cuántica es profundamente contraintuitiva, y por lo tanto no es mi objetivo que quienes lean este artículo salgan con un entendimiento profundo de esta. Sin embargo, para entender los posibles efectos de la computación cuántica, es suficiente saber que existen “efectos cuánticos” que ocurren en las escalas descritas por la mecánica cuántica, pero que no serán jamás observados en las escalas en las que nos manejamos en nuestro día a día. Por ejemplo, un gato cuántico puede estar vivo y muerto al mismo tiempo, una partícula nunca puede tener su posición y velocidad definidas simultáneamente, pueden aparecer y desaparecer partículas espontáneamente, y más. La única propiedad que es necesario saber a fines de este artículo es que la descripción brindada por la mecánica cuántica es probabilística. Es decir, el estado cuántico de una partícula puede ser “estar acá”, “estar allá”, o “estar acá con cierta probabilidad y allá con otra probabilidad”. Lo importante a fines de la computación cuántica es que una computadora cuántica es entonces un sistema físico utiliza-

do para codificar y procesar la información, y tal que los grados de libertad utilizados para tal fin son descritos por la mecánica cuántica y, por lo tanto, presentan propiedades cuánticas. La importancia de las computadoras cuánticas proviene del hecho de que estas propiedades cuánticas pueden explotarse para resolver ciertos problemas de forma mucho más eficiente que con una computadora clásica (es decir, una tal que sus grados de libertad no son cuánticos). Tal vez la eficiencia computacional no parezca de mayor importancia; sin embargo consideremos el ejemplo de la encriptación. La comunicación encriptada es vital para el mundo moderno: toda la comunicación moderna depende del hecho de poder comunicarse entre pares de forma segura. Sin ello el internet no podría jamás funcionar, ni podría existir la seguridad informática de que solo nosotros podemos acceder a nuestra cuenta de mail, nuestro *homebanking* y nuestros documentos guardados en la nube. Para que esto sea posible son esenciales los algoritmos criptográficos. A los fines de este trabajo alcanza con saber que estos algoritmos están basados en problemas matemáticos que son muy fáciles de crear pero muy difíciles de resolver. Entonces uno crea un problema en base a una solución predeterminada, y lo utiliza para esconder información. Cualquiera que sepa la respuesta puede recuperar la información, pero como el problema es en la práctica irresoluble, solo su creador, que conoce la clave, puede recuperar la información originalmente escondida. El punto clave es que la criptografía es esencial para que exista el sistema informático global al que estamos tan acostumbrados.

Con suficiente poder de cómputo es siempre posible romper un algoritmo criptográfico. Sin embargo no hay que temer, pues si una computadora normal de hoy en día intentase romper un algoritmo de encriptación típico, tardaría 22.000 veces la edad del universo en obtener la solución. Sin embargo *sí* hay que temer, pues una computadora cuántica lo suficientemente avanzada podría utilizar el algoritmo de Shor, diseñado por el matemático Peter Shor en 1994, para explotar los fenómenos cuánticos que ocurren dentro de una computadora cuántica para destruir esta encriptación en apenas 10 segundos. Esto demuestra que la explotación de los fenómenos cuánticos dentro de una computadora haría que múltiples problemas que hoy en día son “imposibles” por su costo computacional pasen a ser triviales de resolver.

Otro ejemplo es la simulación de sistemas cuánticos. No es difícil para una PC moderna simular un sistema físico clásico de miles de partículas. Sin embargo, el costo de simular un sistema cuántico crece exponencialmente con el tamaño del sistema, haciendo que simular un sistema cuántico con más que un par de partículas sea computacionalmente imposible. Sin embargo, una computadora cuántica podría simular un sistema cuántico con la misma facilidad con la que una computadora clásica simula bolas de billar. Pero ¿por qué nos interesaría simular un sistema cuántico? Por ejemplo, el *foldin*g de proteínas es un

problema cuántico, y uno de vital importancia para el diseño de medicamentos y para el entendimiento de enfermedades neurodegenerativas. Otra aplicación posible sería en el estudio de reacciones químicas, pudiendo predecir y optimizar procesos como la producción de fertilizantes. También el desarrollo de materiales cuánticos para aplicaciones tecnológicas, como superconductores, sería muchísimo más fácil si la simulación de dinámicas cuánticas fuera accesible. El impacto económico de contar con la capacidad de simular a la naturaleza a escalas cuánticas sería, sin duda, inmenso.

Los lectores más ávidos de detalles se estarán preguntando: ¿pero cómo se hace una computadora cuántica? Para explicar cómo se construye un sistema cuántico que procese información, hay que introducir dos conceptos básicos: bits cuánticos (qubits) y compuertas cuánticas (o, simplemente, compuertas). Las computadoras clásicas modernas funcionan en base a los análogos clásicos de estas herramientas. En una computadora clásica existen estructuras en materiales semiconductores que codifican bits y compuertas clásicas. Un bit es un sistema físico con dos estados posibles, interpretados como codificando un 0 o un 1. Una compuerta es una interacción entre dos bits distintos, en los que el resultado depende del estado de ambos bits. Una computadora clásica se construye entonces organizando redes enormes de bits y compuertas clásicas.

Una computadora cuántica se construye usando el análogo cuántico de estas estructuras: un qubit es un sistema cuántico cuyos estados son 0, 1 o una superposición cuántica de ellos. Por ejemplo, un qubit puede estar en un estado que tiene probabilidad 0,25 de estar en 0 y 0,75 de estar en 1. La diferencia entre un qubit y un bit es entonces la capacidad de acceder a estados de superposición cuántica entre 0 y 1. Análogamente, una compuerta cuántica es una interacción entre dos qubits, y una computadora cuántica es una red de qubits y compuertas cuánticas. Esto a muchos lectores los dejará sin embargo insatisfechos: sigo sin decir cómo se construye físicamente una computadora cuántica, simplemente moví la dificultad de “construir una computadora” a “construir qubits y compuertas”. Para estos lectores ahora voy a dar tres ejemplos de cómo construirlos: puntos cuánticos, computadoras fotónicas y qubits superconductores. El primer ejemplo es el más simple intuitivamente. Son estructuras en materiales semiconductores en donde se forman “pozos” en los que los electrones del material pueden quedar confinados. Estos pozos se comportan como átomos artificiales, en el sentido de que son tan pequeños que se forman niveles de energía discretos, y entonces no puede entrar en ellos más que un electrón. Los estados del qubit son entonces “hay un electrón” o “no hay un electrón”. Como los electrones son naturalmente entidades cuánticas, estos sistemas naturalmente presentan propiedades cuánticas. Para construir compuertas se utilizan

dos fenómenos: el efecto túnel y la interacción electrostática. La interacción electrostática entre dos electrones genera que estos quieran estar separados entre sí, esto se representa en que dos puntos cuánticos cercanos interactúen, pues si en uno de ellos hay un electrón, la repulsión electrostática hará que meter otro electrón en el segundo punto sea más difícil. Esto genera una interacción efectiva que puede ser explotada para procesar información. El otro efecto explotado es el efecto túnel, un efecto cuántico que le permite a un electrón “saltar” de un punto cuántico a otro. Este efecto genera que si dos puntos cuánticos están en estados 0 y 1, pueden transicionar a estar en estados 1 y 0. Con estos dos efectos se pueden utilizar puntos cuánticos para construir qubits y compuertas basadas en electrones. También es posible utilizar otras propiedades cuánticas de los electrones en los puntos cuánticos, como su espín, para codificar la información. Aplicando voltajes a arreglos de puntos cuánticos, es posible controlar su interacción y programar la computadora. Un ejemplo de una computadora cuántica basada en puntos cuánticos es el chip Tunnel Falls de Intel.

El siguiente ejemplo es el uso de fotones individuales. Los fotones son partículas de luz, que, al igual que los electrones, se comportan naturalmente de forma cuántica. Para construir una computadora basada en fotones se construyen redes ópticas a través de las que circulan estas partículas, interactuando entre ellas. Para construir un qubit basado en fotones, alcanza con elegir distintos posibles estados del fotón y asignarles 0 y 1. Uno podría, por ejemplo, elegir que si el fotón es de un color sea 0 y si es de otro sea 1. O utilizar propiedades como la polarización que son conceptualmente menos intuitivos, pero ofrecen ventajas a la hora de construir la computadora. Las compuertas se construyen utilizando elementos ópticos que generan intercambio o interacción de estos fotones. El uso de fotones para construir qubits es de mucha importancia para el área de la comunicación cuántica y la eventual construcción de una “internet cuántica”. Ejemplos de computadoras fotónicas son la computadora Jiuzhang, en China, y la X-Series de Xanadu, en Canadá.

El último ejemplo que voy a presentar consiste en el uso de circuitos superconductores. La superconductividad es un estado de la materia (como sólido, líquido, gas, etc.) caracterizado por la total falta de resistencia a la corriente eléctrica. Es decir, un material superconductor puede conducir corriente eléctrica sin producir calor. La superconductividad es un estado cuántico de la materia, en la que sistemas macroscópicos ocupan estados cuánticos. Por lo tanto, es necesario llevar materiales a bajas temperaturas para alcanzar un estado superconductor: a presión atmosférica no se conocen superconductores a una temperatura mayor a  $-135^{\circ}\text{C}$ . Para construir un qubit basado en superconductores se diseña un circuito eléctrico construido con materiales superconductores. Al ser la superconductividad



un estado cuántico de la materia, estos circuitos presentan propiedades cuánticas. El objetivo luego es diseñar el circuito de forma tal que propiedades eléctricas, como carga acumulada en un punto, o corriente circulando por un anillo, codifiquen a los estados 0 y 1 de qubits, y que la evolución del circuito codifique las compuertas entre los qubits. Luego, aplicando voltajes o corrientes eléctricas, estos circuitos pueden ser controlados para implementar los algoritmos de la computadora cuántica. Ejemplos de computadoras cuánticas superconductoras son Q System One de IBM, y el procesador Sycamore de Google. Ahora, si hay tanta tecnología desarrollada para la computación cuántica, y hay tantas aplicaciones en las que sería útil la supremacía cuántica... ¿por qué no la hemos alcanzado aún? La computación cuántica tiene un gran enemigo: la terrible decoherencia. Los fenómenos cuánticos son extremadamente frágiles ante interacciones con su entorno. Un sistema cuántico que interactúa con ruido generado por algún ambiente rápidamente pierde su “coherencia” y los fenómenos cuánticos desaparecen. Esto incluye, por ejemplo, ruido electromagnético hecho por instalaciones eléctricas, rayos cósmicos que atraviesan la muestra, incluso vibraciones en la computadora. Cuando un sistema cuántico pierde la coherencia, deja de ser posible observar propiedades cuánticas, y entonces desaparece la ventaja de las computadoras cuánticas sobre las clásicas. Es entonces necesario aislar lo mejor posible las computadoras cuánticas de cualquier posible fuente de ruido que destruya la coherencia cuántica. Eso no es, sin embargo, suficiente. También se están desarrollando métodos de “corrección de errores cuántica”, en la que se dedican recursos de la computadora a detectar y corregir los errores en la computación inducidos por este ruido. Sin embargo, es imposible detener completamente la decoherencia, y entonces esta limita el tamaño de las computadoras cuánticas, y la cantidad de operaciones que pueden tener los programas que estos utilizan antes de perder toda utilidad. Esta es una de las principales limitaciones que enfrentamos para alcanzar la supremacía cuántica.

## Conclusiones

La computación cuántica aparece en el horizonte como un heraldo de tecnologías fantásticas y profundos riesgos. Hay múltiples potenciales tecnologías que están siendo desarrolladas para alcanzar la supremacía cuántica, sin embargo todas ellas deben primero superar los límites impuestos por la pérdida de coherencia antes de poder alcanzar la supremacía cuántica. Si bien estamos viendo la aparición de computadoras con más y más qubits, aún es imposible saber cuándo se alcanzará el sueño de una computadora cuántica que sea usada para entender el universo, diseñar medicamentos, o destruir la criptografía. Hasta entonces, estamos a la espera de los resultados de esta verdadera revolución cuántica.

# DIGITALIZACIÓN PARA UN DESARROLLO INCLUSIVO

LA PROPUESTA DE UN CAMINO HACIA UNA ECONOMÍA DIGITAL EQUITATIVA EN LOS PAÍSES EN DESARROLLO SUPONE EL ANÁLISIS DE LAS DISTINTAS VARIABLES QUE MUESTRAN EN EL PRESENTE UN ACCESO DESIGUAL NO SOLAMENTE A LA CONECTIVIDAD SINO TAMBIÉN AL RESTO DE LOS DERECHOS EN LAS REGIONES MÁS RELEGADAS DEL MUNDO.



por **PILAR FAJARNÉS**. *Máster en Economía del Desarrollo por la Universidad de Kent en Canterbury (Reino Unido). Economista en la Sección de Investigación en Políticas de la Economía Digital de la Conferencia de Naciones Unidas sobre Comercio y Desarrollo (UNCTAD).*



**L** digitalización ha cambiado de forma drástica la sociedad y la economía en menos de un par de décadas. En la actualidad es prácticamente imposible concebir las diferentes interacciones, las formas en que los seres humanos se conectan y se comunican, así como casi todas las actividades económicas, sin las tecnologías digitales. Este proceso de digitalización puede generar significativos beneficios para la economía y la sociedad. Sin embargo, la tecnología no es determinista, así que los beneficios no se crean de manera automática y, lo que es más importante, sin intervención pública, es muy posible que las desigualdades asociadas a la digitalización aumenten. La economía digital está en rápida evolución, asociada a tecnologías de vanguardia. Continuamente se conocen avances relacionados con la inteligencia artificial, la computación en la nube, la automatización, el internet de las cosas, cadenas de bloques y demás. La digitalización tiene implicaciones para todos los países, sectores y agentes implicados. Considerando las diferentes capacidades de los países con diferentes niveles de desarrollo, las posibilidades de aprovechar el potencial de estas tecnologías también varían. Los gobiernos necesitan actuar y establecer políticas para reducir los desequilibrios que emergen en la economía digital y que pueden agravar las desigualdades existentes. En este contexto cambiante, los gobiernos encuentran dificultades para responder al rápido ritmo de los acontecimientos. Es preciso mejorar el conocimiento de los nuevos procesos y actores, así como mejorar la medición de la economía digital, para que se puedan diseñar las políticas necesarias de forma informada.

Las nuevas dinámicas de la economía digital están marcadas por los datos digitales como nuevo recurso económico, e incluso estratégico, que permite generar valor, y por las plataformas digitales, como nuevos agentes económicos. Uno de los principales motores de las tendencias recientes de la economía digital es el crecimiento masivo de los datos digitales. De hecho, la digitalización está esencialmente basada en los datos, ya que todas las tecnologías digitales necesitan de los datos para funcionar. Esos datos surgen de la huella digital resultante de las actividades personales, sociales y comerciales que tienen lugar en la red. En la economía digital, el valor económico y social se crea una vez que los datos digitales se agregan y se transforman en lo que se puede llamar “inteligencia digital”. Esta inteligencia digital puede ser monetizada por parte de las empresas en el sector privado, o las autoridades públicas las pueden usar para el bien común. Por lo tanto, el acceso a los datos es un factor de importancia creciente para poder transformarlos en inteligencia digital, y extraer su valor. Aparece una cadena de valor de los datos, desde la recolección de los datos, pasando por su transferencia, almacenamiento, procesamiento, análisis y uso de los datos procesados, por ejemplo en el desarrollo de la inteligencia digital. En lo que respecta al desarrollo inclusivo una cuestión esencial es la forma en que el valor que se crea gracias a los datos se distribuye, o quién captura ese valor. Es por tanto fundamental que los beneficios de la economía digital se repartan de manera equitativa, tanto entre países como en el interior de los países. Un aspecto importante de los datos digitales es que tienen un ca-

*Las nuevas dinámicas de la economía digital están marcadas por los datos digitales como nuevo recurso económico, e incluso estratégico, que permite generar valor, y por las plataformas digitales, como nuevos agentes económicos. Uno de los principales motores de las tendencias recientes de la economía digital es el crecimiento masivo de los datos digitales.*

rácter multidisciplinar, que va más allá de su dimensión económica. Desde el punto de vista económico, los datos permiten generar valor económico en el sector privado por la vía de su monetización, por ejemplo, con la publicidad digital o con servicios de computación en la nube. Los datos también pueden servir como bien público. Esto fue evidente durante la pandemia, cuando el intercambio de datos entre países sirvió para enfrentarla, así como en las investigaciones para desarrollar las vacunas. Igualmente, el intercambio de datos y la resultante información pueden ser cruciales para enfrentar el cambio climático. Además, la dimensión no económica de los datos se refiere a aspectos relacionados con la privacidad y otros derechos humanos, así como la seguridad. Además de los datos, las plataformas digitales son el otro nuevo protagonista principal de la economía digital. Estas plataformas usan modelos de negocio basados en los datos, y están creando disrupciones en las industrias de siempre. Gracias a su posición privilegiada para obtener y extraer datos relacionados con las acciones, interacciones y transacciones de sus usuarios a través de Internet, estas plataformas han llegado a una posición de dominio. Algunas plataformas controlan una gran proporción de sus respectivos mercados. Esto se explica por efectos de red, la capacidad de aprovechamiento de los datos y los altos costes para los usuarios del cambio de plataforma.

Estas grandes plataformas digitales consolidan su creciente poder por medio de la adquisición de posibles competidores de éxito, expansión hacia productos y servicios complementarios, fuerte inversión en investigación y desarrollo, asociaciones estratégicas

con empresas multinacionales líderes en industrias tradicionales y actividades a través de grupos de presión para influir en la regulación de la economía digital. Como resultado, estas plataformas se han convertido en corporaciones digitales a nivel mundial. Estas nuevas dinámicas tienen implicaciones a nivel global relacionadas con, entre otros, los siguientes aspectos: la mayor concentración de mercado y su consolidación, y los consiguientes problemas relativos a la aplicación de políticas de competencia, dado el alcance global de estas corporaciones; la aparición de cadenas globales de valor basadas en los datos y la forma en que los países en desarrollo pueden participar de los beneficios asociados; impactos sobre el empleo, relativos a la automatización y el trabajo en línea, que pueden crear oportunidades de empleo, pero al mismo tiempo generar impactos negativos en las condiciones de trabajo; y dificultades para una fiscalidad justa de los negocios en línea.

Un prerrequisito importante para que los países en desarrollo puedan beneficiarse de la economía digital basada en los datos es reducir la persistente brecha digital. Esta brecha es un reflejo de la desigualdad general en la economía, y al mismo tiempo puede contribuir a incrementar esa desigualdad si no se reduce. La brecha digital se ha relacionado generalmente con aspectos relacionados con la conectividad, en términos de acceso a internet, posibilidad de pago y uso de las tecnologías digitales. En el mundo todavía hay un tercio de la población que no tiene acceso a internet. En el caso de los países menos adelantados solo una persona de cada cinco usa internet, y cuando lo pueden usar es a precios más altos y velocidades y capacidades limitadas. También sigue habiendo brechas digitales importantes dentro de las regiones y los países, en términos de género, entre áreas rurales y urbanas, o entre pequeñas y grandes empresas. Estas brechas colocan a los emprendedores y empresas en los países en desarrollo en una posición de desventaja frente a sus competidores en otras partes del mundo.

Con el creciente rol clave de los datos digitales en la economía digital, la brecha digital en términos convencionales de conectividad se ve agravada por nuevas dimensiones ligadas a lo que se puede denominar una brecha digital relacionada con los datos. La geografía económica de la economía digital no sigue exactamente la división tradicional entre economías desarrolladas y en desarrollo. En términos de creación de riqueza en este contexto, hasta la fecha hay dos países que destacan: Estados Unidos y China. Estos dos países representaban en 2021 aproximadamente el 90% del valor de capitalización de mercado de las 70 mayores empresas basadas en el modelo de plataformas digitales. En comparación, la parte de Europa era solamente 3,6%; África representaba el 1,3% y América latina y el Caribe, el 0,2%. Estados Unidos y China también representan el 75% de todas las patentes relacionadas con las cadenas de bloques, el 50% del gasto global en internet de las cosas y más del 75% del mercado de computación en la nube. Las corporaciones digitales de estos



dos países adquieren un control cada vez mayor en las diferentes etapas de la cadena de valor de los datos, la recolección, la transmisión por medio de cables submarinos y satélites, así como su almacenamiento en los centros de datos. En contraste, por ejemplo, África y América latina representan solamente el 5% de los centros de datos de alquiler compartido. Estos dos países, y sus corporaciones digitales, también dominan en el procesamiento, análisis y uso de los datos. Es notorio, por ejemplo, que la mayor parte de las capacidades, tanto financieras como técnicas y humanas, en lo que se refiere al desarrollo de la inteligencia artificial y la posibilidad de beneficiarse de ellas, está concentrada en unos pocos países y empresas.

Dado que la mayoría de los países en desarrollo tienen capacidades limitadas para recopilar, analizar, monetizar los datos, y utilizarlos para cumplir con sus objetivos de desarrollo, hay un gran riesgo de que se encuentren en posiciones subordinadas en la cadena global de valor de los datos. Este sería el caso si se mantuvieran únicamente en meros proveedores de los datos sin procesar para las plataformas digitales globales para que los transformen en inteligencia digital. El resultado sería que los países en desarrollo tendrían que pagar por la inteligencia digital que se ha obtenido y “refinado” a partir de sus datos. Este es un patrón extractivo, que presenta similitudes con las experiencias de muchos países en desarrollo con las industrias extractivas en las que se exporta la materia prima para importar los productos de valor añadido, lo que representa poco beneficio en términos de desarrollo nacional. El desafío para los países en desarrollo es desarrollar las capacidades necesarias para transformar los datos en inteligencia digital y por tanto generar el valor añadido dentro del país.

En estas condiciones, con las actuales políticas y regulaciones, es más probable que la digitalización contribuya a una mayor desigualdad que a un desarrollo más inclusivo. Las políticas públicas juegan un papel clave para revertir este proceso de tal manera que la digitalización contribuya a maximizar los beneficios para la población de manera equitativa, y al mismo tiempo se minimicen los riesgos y daños colaterales.

Sin embargo, esto no es tarea fácil para los gobernantes por la gran velocidad de los cambios tecnológicos, la complejidad de los nuevos fenómenos y sus implicaciones. Además, dado su carácter transversal, la digitalización afecta muchas áreas de política económica, lo que requiere la implicación de prácticamente todos los ministerios de un gobierno. Las necesidades de agregar valor en la economía digital requieren de política industrial para expandir el sector y las infraestructuras digitales, política de educación para desarrollar las habilidades para la era digital, así como políticas de emprendimiento digital. Y la escasez de habilidades se refiere también a la capacidad de las autoridades para regular en la complejidad del entorno digital, que puede resultar en una representación insuficiente de la expe-

riencia técnica y analítica. Esto obstaculiza la capacidad de los gobiernos para identificar las oportunidades de las tecnologías digitales, y los riesgos y amenazas que puedan surgir, así como la forma de regularlos. Los países en desarrollo deben asignar más recursos domésticos al desarrollo de sus capacidades para crear y capturar valor de los datos a nivel nacional. Sin embargo, muchos países en desarrollo disponen de recursos limitados, tanto financieros como técnicos y humanos, por lo que será necesario también un mayor apoyo internacional y regional.

Al mismo tiempo, muchos de los desafíos para la política pública no se pueden enfrentar solamente a nivel de políticas nacionales; internet y las cadenas de valor de los datos, así como las corporaciones digitales globales, tienen un alcance e impacto globales. Es necesario mejorar la cooperación internacional en términos de políticas de competencia. Igualmente, es fundamental mejorar la cooperación en lo que se refiere a la fiscalidad internacional de las grandes corporaciones digitales, de tal forma que contribuyan de manera adecuada a la tributación en los países en los que se genera el valor. El comercio internacional se lleva a cabo cada vez más mediante la utilización de canales digitales, lo que da lugar a lo que se denomina comercio digital, y que requiere adaptación de las reglas internacionales de comercio al nuevo contexto digital. Además, los flujos internacionales de datos se han convertido en un nuevo flujo económico de importancia crítica para las relaciones entre países y para el desarrollo.

La gobernanza global de los datos adquiere una importancia clave. Es necesario un nuevo enfoque de gobernanza global que permita que los datos digitales puedan fluir entre países con tanta libertad como sea necesario y posible. El objetivo es maximizar los beneficios de los datos en el desarrollo y garantizar que esos datos se distribuyan de manera equitativa, pero siempre minimizando los riesgos y daños asociados. Esto permitiría el intercambio mundial de datos para afrontar desafíos de desarrollo globales, desarrollar bienes públicos digitales globales, aumentar la confianza y reducir la incertidumbre en la economía digital. También ayudaría a evitar una mayor fragmentación del espacio digital, abordar los desafíos de política que resultan del dominio de las corporaciones digitales globales y reducir la desigualdad existente.

En el paisaje existente de fragmentación en relación al enfoque de política de datos, se corre el riesgo de no captar el valor que podría derivarse de las tecnologías digitales y de crear mayores posibilidades de daños sustanciales relacionados con violaciones de la privacidad, ataques cibernéticos y otros riesgos.

El contexto global actual se caracteriza por enfoques diferentes de gobernanza de los datos, con cada país regulando los datos de acuerdo a sus condiciones particulares y a sus intereses nacionales de desarrollo, protección de derechos humanos y de seguridad. Hay tres enfoques fundamentales que influyen en el resto, en particular por parte de los tres principales actores a nivel mundial: Estados Unidos, China y la Unión Europea (UE).

El enfoque estadounidense se centra en el control de los datos por parte del sector privado, el modelo chino enfatiza el control de los datos por parte del Estado, mientras que la UE favorece el control de los datos por parte de las personas, sobre la base de derechos y valores fundamentales. Esto afecta a los países en desarrollo, que pueden verse abocados a elegir una de las zonas de influencia, cuando sus intereses económicos están en todas ellas. Además, los actuales marcos regulatorios regionales e internacionales tienden a tener un alcance limitado, centrándose solamente en aspectos de comercio internacional o de privacidad, o a ser demasiado limitados geográficamente.

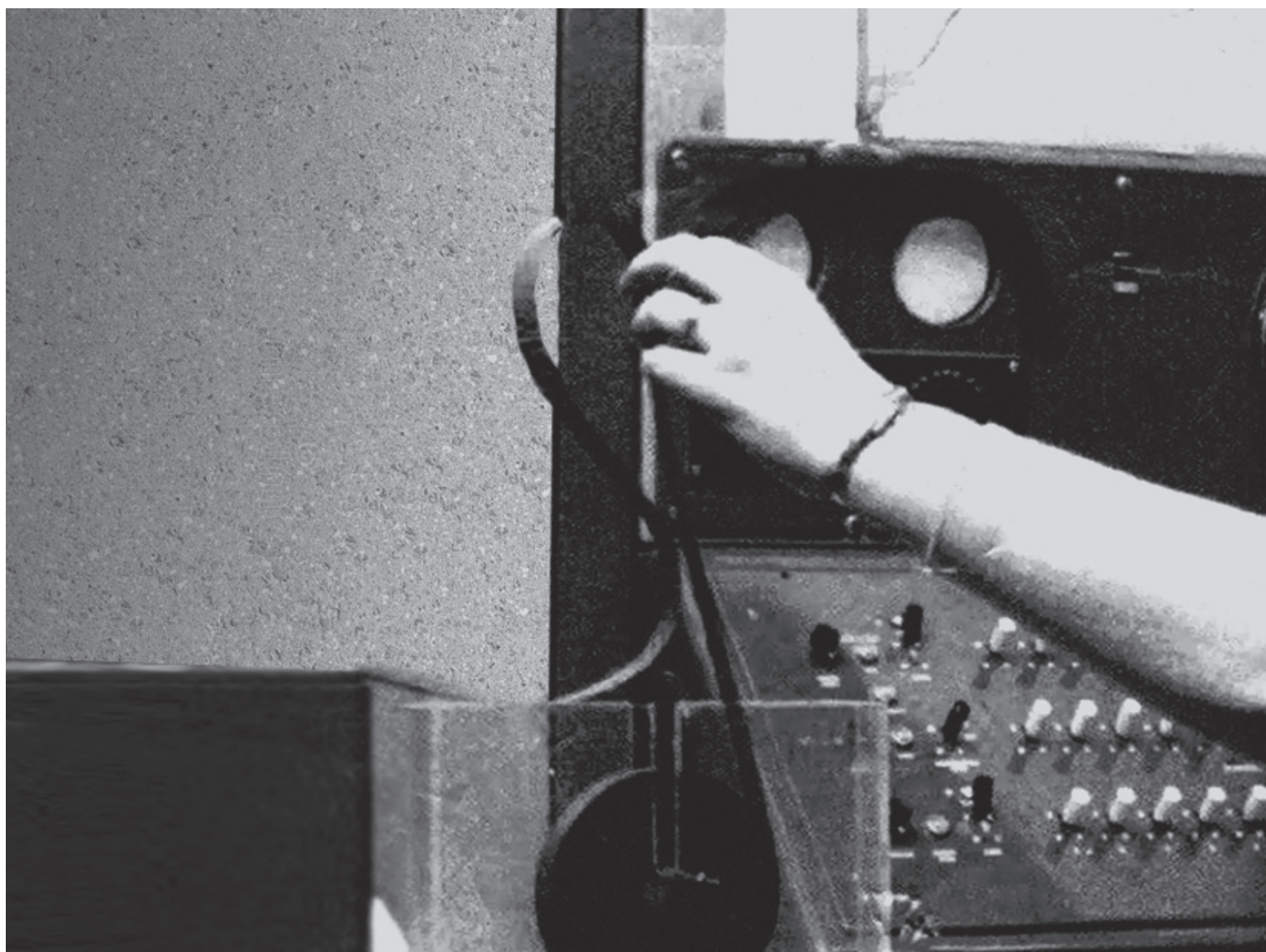
Los crecientes desafíos de interconexión e interdependencia en la economía global basadas en los datos requieren todo lo contrario de la fragmentación; es necesario un enfoque global holístico y coordinado. Considerando las nuevas dinámicas en el entorno digital, es necesario pensar en formas nuevas e innovadoras de gobernanza global de los datos. Un nuevo mecanismo de gobernanza en este contexto debería tener carácter multilateral, dado que son los gobiernos los que pueden establecer políticas y regulaciones, multidisciplinario, para tener en cuenta las diferentes dimensiones de los datos, e incluir a las múltiples partes interesadas. También debería remediar la subrepresentación actual de los países en desarrollo en las iniciativas globales y regionales de gobernanza de los datos. Teniendo esto en cuenta, el contexto de Naciones Unidas sería el que proporcionaría el mejor marco, dada su amplia representatividad a nivel mundial. Este mecanismo debería funcionar como complemento de y en coherencia con políticas nacionales en la economía digital. También debería proporcionar suficiente espacio de políticas nacionales para que los países con diferentes niveles de preparación y capacidades se puedan beneficiar de la economía digital basada en los datos.

Teniendo en cuenta las implicaciones de la digitalización en relación con el desarrollo, resulta significativo que los estudios al respecto son todavía escasos. Es necesario mejorar la comprensión de los impactos de las nuevas dinámicas en los países de desarrollo. En el contexto de Naciones Unidas, el Informe de la Economía Digital de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD) constituye una contribución de gran utilidad a este respecto, que desarrolla muchos de los aspectos presentados en este texto. Además, los procesos de Naciones Unidas para la Cumbre del Futuro de 2024, incluido un Pacto Digital Mundial, buscan avanzar hacia una digitalización universal e inclusiva.



---

**Nota:** Las opiniones aquí reflejadas son las de la autora, y no representan necesariamente las de la UNCTAD.



# ¿PUEDEN LAS HUMANIDADES COEVOLUCIONAR CON LOS HUMANOS (Y LAS MÁQUINAS)? INTELIGENCIA ARTIFICIAL, SOCIEDAD ARTIFICIAL Y LOS DESAFÍOS PARA LAS CIENCIAS SOCIALES



**por FLAVIA COSTA.** *Licenciada en Ciencias de la Comunicación y Doctora en Ciencias Sociales por la Universidad de Buenos Aires. Investigadora del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Es docente de grado y posgrado y miembro del TecnoCenoLab.*

**por JULIÁN MÓNACO.** *Licenciado en Ciencias de la Comunicación y Magíster en Comunicación y Cultura por la Universidad de Buenos Aires. Becario doctoral del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Docente de grado y posgrado e integrante del TecnoCenoLab.*

**EL TRABAJO PRESENTA LAS MÚLTIPLES DIMENSIONES QUE PUEDE TENER EL ANÁLISIS DE LA IA Y VARIOS DE SUS CONCEPTOS VINCULADOS, DESDE UNA PERSPECTIVA QUE SE CONCENTRA EN LA VIDA SOCIAL Y EN EL MUNDO QUE ES CONSTRUIDO POR Y PARA LAS PERSONAS EN LA ACTUALIDAD.**

# E

n junio de 2023, la revista *Time* publicó una tapa enteramente roja con el título de catástrofe “The end of humAnItY. How real is the risk?” [“El fin de la humAnIdad. ¿Cuán real es el riesgo?”] que funcionaba como puerta de entrada a un informe especial dedicado a los desafíos y riesgos asociados al despliegue de la inteligencia artificial (IA) y, en particular, de la IA generativa (IAG). “La ventana de aquello que la IA *no puede* hacer parece estrecharse semana a semana”, afirmaba en una de esas notas la experta en seguridad de IA Katja Grace.

Como campo de investigación, la IA tiene más de 70 años: su puesta en marcha coincide con la Segunda Guerra Mundial y el origen del término obliga a remontarse a la década de 1950, cuando en los Estados Unidos se comenzaron a investigar y desarrollar sistemas capaces de replicar o emular ciertos comportamientos humanos considerados inteligentes: memorizar, computar, aprender de los propios resultados, proyectar tendencias. Ha sido, sin embargo, sobre todo durante la última década que los avances en este ámbito generaron tecnologías que se comparan o superan a los seres humanos en tareas que requieren creatividad, razonamiento complejo e involucran incluso la toma de decisiones. Y que tienen el potencial de afectar a distintos sectores del trabajo y de la producción de conocimiento. En aquel *dossier* publicado por *Time*, el especialista en política internacional Ian Bremmer sintetizaba algunos de los principales desafíos que las sociedades enfrentarán en los próximos años ante la popularización del uso de IA generativas. Teniendo en cuenta que “la salud de la democracia y los mercados libres depende del acceso a información precisa y verificable”, señalaba Bremmer, la IA se sumará a las dificultades que ya vienen planteando las redes sociales para distinguir la realidad de la ficción: “Las falsificaciones de audio y video cada vez más sofisticadas –decía– socavarán nuestra ya disminuida confianza en quienes trabajan en los gobiernos y quienes informan”. Agregaba otros riesgos, como el uso por parte de Estados autoritarios de herramientas de vigilancia masiva, o la proliferación de contenidos generados por IA en manos de “delincuentes, terroristas y otros malos actores” que podrían “crear armas biológicas, mani-

pular los mercados y distorsionar la opinión pública” con relativa facilidad. Finalmente aseguraba que la IA “puede exacerbar la desigualdad dentro de las sociedades (favoreciendo a pequeños grupos con riqueza, acceso o habilidades especiales), así como entre naciones más ricas y más pobres”.

El informe de la revista estadounidense se hacía eco de una alerta emitida por el propio campo de la IA. El 30 de marzo de 2023, la organización no gubernamental Future of Life había publicado una carta firmada por investigadores y empresarios (entre ellos el cofundador de Apple, Steve Wozniak; el historiador Yuval Harari; el ceo de SpaceX, Tesla y Twitter, Elon Musk y el inversor en IA Ian Hogarth) que llamaba a pausar por “al menos seis meses” el entrenamiento de sistemas de IA superiores a chat Gpt-4, de la firma OpenAI.

Esta carta desató una alerta en la opinión pública global que incentivó y, donde ya estaba en marcha, profundizó una suerte de “furor normativo” con respecto a la IA, con los ojos de buena parte de Occidente puestos en la ley de la Unión Europea (la única vinculante hasta el momento, y basada en una clasificación de los riesgos de la IA). En noviembre de ese año, Gran Bretaña impulsó la primera Cumbre sobre Seguridad (*safety*) de la IA nada menos que en Bletchley Park, donde durante la Segunda Guerra Mundial Alan Turing descifró el código de la máquina Enigma, de los nazis. Inmediatamente después creó el AI Safety Institute, la primera organización respaldada por un Estado dedicada a probar la seguridad de la IA avanzada y medir sus impactos en las personas y la sociedad.

Como vemos, fue de hecho la puesta en circulación masiva de los Grandes Modelos de Lenguaje (*LLM*, por sus siglas en inglés), capaces de simular con asombrosa eficacia habilidades comunicativas en muy diversas lenguas, la que acabó de colocar en primer plano un nuevo tipo de problema de escala planetaria: ¿qué implicaciones tiene para los distintos sistemas que habitamos (desde la política hasta la salud, pasando por la educación y el trabajo) la creación de sistemas técnicos capaces de desarrollar buena parte de las actividades que los humanos realizamos a través del lenguaje –y que en muchos casos son ocupaciones remuneradas y/o protegidas por derechos–? Si entre las particu-

laridades del momento actual está el hecho de que convivimos con esta clase de sistemas sociotécnicos, una de las tareas urgentes de la época es abordar, tanto a nivel nacional como regional y global, esa pregunta y formular respuestas realizables. Lógicamente, será una tarea sembrada de conflictos. El primer aporte de las ciencias sociales aquí será recordar que no se trata de dirimir el futuro de “la especie humana” en relación con las tecnologías avanzadas, sino de planificar y, como suele decirse, gobernar el futuro del desarrollo de sistemas sociotécnicos complejos en el marco de sociedades, clases, poblaciones y grupos humanos en relativo pero persistente conflicto de intereses. Y es razonable prever que si no la emprendemos pronto, el desarrollo desigual de estas tecnologías profundizará aún más los procesos de estratificación e inequidad en marcha.

Por su parte, la opacidad tanto técnica como social y política de estos sistemas; su imprevisibilidad, asociada a la complejidad interactiva que los caracteriza; la creciente cantidad de decisiones que aparecen mediadas por alguna forma de evaluación algorítmica automatizada (sistemas de recomendación, puntuación de crédito, predicción de pérdida de clientes, formulación “automática” de políticas públicas) plantean la pregunta acerca del sentido, la conveniencia y las consecuencias de una “automatización” y “artificialización” de lo social.

Para las ciencias sociales se ha vuelto imperativo afrontar estas cuestiones si es que quieren participar de pleno derecho en una conversación que las concierne de manera íntima en cuanto a su objeto, la vida social, pero que en la práctica las marginaliza en cuanto a saberes expertos. Los modelos institucionales de gobernanza de IA suelen invocar cuatro grandes ámbitos que participan de las deliberaciones: gobierno, industria, sistema de ciencia y tecnología, sociedad civil. Donde los expertos del ámbito “ciencia y tecnología” nunca son expertos en ciencias sociales (acaso algún referente de ética o de ciencias jurídicas), quienes por lo tanto deben conformarse con participar en tanto que asesores del gobierno o voceros de la “sociedad civil”.

Entre otros esfuerzos, este escenario les exigirá revisar parte de sus marcos teóricos y epistemológicos. Por ejemplo, tomar distancia de las perspectivas que identifican la dimensión tec-

*Los modelos institucionales de gobernanza de IA suelen invocar cuatro grandes ámbitos que participan de las deliberaciones: gobierno, industria, sistema de ciencia y tecnología, sociedad civil.*



nológica como meramente herramental, ante la evidencia de que este mundo digital en y con el que interactuamos no es un instrumento más, sino un verdadero *ambiente*, un *ecosistema*. Asimismo, en la medida en que estas tecnologías se aplican a la toma de decisiones en diferentes ámbitos (públicos y privados, académicos, gubernamentales y empresariales) e influyen de manera decisiva en las prácticas y en las estructuras de la sociedad; en la medida en que, así también, crean nuevos espacios y hábitos de producción, distribución y consumo de bienes, servicios y significaciones, las ciencias sociales deberán dar la pelea por no ser solo usuarias o implementadoras de IA, sino también por tomar parte en los ámbitos de desarrollo y auditoría de estos sistemas. Y esto a su vez demandará fomentar experiencias de investigación y desarrollo interdisciplinar, con énfasis en la combinación entre estudios teóricos y laboratorio de prácticas. Existieron y todavía existen, son parte de nuestras tradiciones de investigación, unas ciencias sociales para las sociedades de masas; unas ciencias sociales para la sociedad de la globalización multicultural; en la Argentina, unas ciencias sociales para la recuperación democrática y también para el post-2001. Ahora mismo nos corresponde pensar las ciencias sociales para una sociedad inmersa en ecosistemas sociodigitales complejos; un tipo de pensamiento no está garantizado si no se lo produce y que exigirá un gran esfuerzo de imaginación teórica, epistemológica y política.

## Una perspectiva analítica

Entre junio y octubre de 2023, los integrantes del TecnoCenoLab y otros investigadores especialmente convocados realizamos una investigación exploratoria pero intensiva a solicitud de la Facultad de Ciencias Sociales de la UBA para el Programa Argentina Futura, de la Jefatura de Gabinete de Ministros. Buscábamos sistematizar la información existente acerca de las iniciativas en materia de políticas y regulaciones para la IA que se estaban desarrollando en la Argentina en ese momento, ponerlas en relación con otras iniciativas internacionales y ofrecer un conjunto de recomendaciones de políticas públicas con especial foco en los riesgos de las IA para la calidad democrática.

En ese marco, nos propusimos delimitar una perspectiva analítica adecuada para abordar este novedoso objeto de indagación. Y lo primero que advertimos fue que, cuando analizamos las IA generativas y los *LLM* desde un enfoque que tiene como epicentro a las ciencias sociales y humanas, si bien podemos referirnos a estos enormes conjuntos técnicos como Inteligencia Artificial, también podríamos denominarlos Sociedad Artificial.

Veamos esto un poco más de cerca. Muy esquemáticamente, los elementos básicos de estos sistemas de IA son tres: (1) una enorme capacidad de cómputo (a grandes rasgos, la capa del *hardware*); (2) un conjunto de métodos de procesamiento de información (aprendizaje profundo, redes neuronales, etc.; las capas del *software* y de las aplicaciones de IA) y (3) grandes con-

*Ahora mismo nos corresponde pensar las ciencias sociales para una sociedad inmersa en ecosistemas sociodigitales complejos; un tipo de pensamiento no está garantizado si no se lo produce y que exigirá un gran esfuerzo de imaginación teórica, epistemológica y política.*

juntos de datos (materiales “de la vida social”, obtenidos en buena medida a través de plataformas; las capas de *input*). Es decir: su alimento son los propios intercambios lingüísticos en diferentes idiomas, las figuras retóricas, las acciones y reacciones emocionales que movilizamos en esos intercambios, las relaciones sociales que se producen en diversas culturas, las instrucciones y objetivos que orientaron el entrenamiento.

Si miramos de cerca el tercero de estos elementos –el enorme volumen de datos masivos y su tratamiento para volverlos operativos– es heterogéneo respecto de los dos anteriores. Si los elementos 1 y 2 son el producto de desarrollos históricos de la informática, la ingeniería de sistemas y la computación, el estudio y el trato con el elemento 3 es el dominio de las ciencias del lenguaje, las ciencias de la comunicación, la sociología, la ciencia política, el derecho. En efecto, las IA generativas y los LLM no son una herramienta más: son procesadores que analizan y gestionan nuestros actos lingüísticos y una parte no insignificante de nuestros modos de relacionarnos socialmente. Lo que estos sistemas sociotécnicos complejos hacen es acelerar el análisis, la gestión y la (re)producción de lo social. *Procesan sociedad*.

El modo de procesamiento de ese material es, ya es sabido, probabilístico, estadístico. El moderno término “estadística”, del italiano “*statista*”, “hombre de Estado”, es bien conocido entre quienes provenimos de las ciencias sociales, y nos recuerda que, al menos desde el siglo XVII, hay una alianza entre saberes matemáticos y sociales para constituir una disciplina capaz de estudiar la variabilidad de los acontecimientos siguiendo las leyes de la probabilidad: una “aritmética política” –según el recordado libro de William Petty, de 1690– que tiene como objeto principal a las poblaciones.

Sugerimos entonces comenzar a pensar en estas tecnologías en tanto que Sociedad Artificial, dado que entre sus potenciales *outputs* está la facilitación de la toma de decisiones en áreas críticas; en la medida en que aceleran el análisis y la gestión de las relaciones sociales; y en la medida en que nos permiten, además, ver correlaciones que los humanos no seríamos capaces de registrar.

Un segundo punto a considerar es que las IA, en tanto integran y expanden el ecosistema digital, constituyen no una herramienta o dispositivo técnico, sino que –particularmente después del shock de virtualización que implicó la última pandemia (Costa, 2021)– han comenzado a ser para nosotros un mundoambiente. Las tecnologías del ecosistema digital están dejando de ser instrumentos que podemos elegir usar o no, y se han vuelto indispensables para realizar actividades cotidianas como guiarnos en una ciudad o iniciar un trámite de documentación obligatoria (Zuboff, 2020). Esto significa que, en relación con los usuarios, no es suficiente un enfoque que las aborde desde la perspectiva

de la instrumentalidad y de la relación sistema-usuario individual, como si cada usuario pudiera decidir qué hacer en cada caso con la IA: necesitamos un enfoque sistémico, atento a las dinámicas multiescalares de la economía política del ecosistema digital.

El tercero se relaciona con el anterior y, en la búsqueda de un enfoque adecuado, sigue la perspectiva asumida en reglamentos recientes como la Ley de Inteligencia Artificial de la Unión Europea (2023), o la del grupo de expertos en IA de la Organización para la Cooperación y el Desarrollo Económico (OCDE). En estos y otros casos se afirma que, para ciertos usos que involucran áreas críticas para la población humana (acceso a la justicia, a la educación, a la salud, a la protección de la privacidad), la IA es una tecnología de alto riesgo. Por lo tanto, requiere un tratamiento acorde a lo largo de todo su ciclo de vida. Es preciso distinguir cuáles son esos usos y abordarlos con la perspectiva sistémica de la seguridad, la gestión de riesgos y la investigación de accidentes e incidentes. Una perspectiva que se orienta no tanto a la búsqueda de responsables (esa es la perspectiva jurídica, que busca culpables e impone sanciones o resarcimientos) sino a la detección de fallas en las defensas de los diversos subsistemas, y que emite recomendaciones para que esas fallas no vuelvan a suceder. La filosofía que la anima podría sintetizarse en la idea de que, si vamos a habitar mundoambientes digitales, estos deben ser confiables y seguros (*safe*) para la vida democrática.

El cuarto tiene que ver con cuáles son los ámbitos expertos que deben tratar con estas tecnologías. En referencia a esto, es posible señalar dos cosas. Por un lado, si el ámbito experto que se ocupa de las IA es única o principalmente el subconjunto de las ciencias de la computación, la ingeniería y las ciencias informáticas, corremos el riesgo de perder la profundidad de campo lograda a lo largo de la historia de las ciencias sociales para tratar adecuadamente con datos de poblaciones humanas desde una perspectiva de derechos, que es lo que permitiría que estos sistemas sean verdaderamente robustos y confiables.

Algo que hemos observado en nuestra investigación es que los expertos que provienen de las ciencias informáticas y las ingenierías suelen identificar como su principal (y a veces único) interlocutor en el campo de las ciencias sociales el derecho. Es su punto de exterioridad y su límite: el establecimiento de normas que regulen la práctica. Si se traspasa el límite establecido, habrá una sanción. Pero el derecho actúa *a posteriori* del incidente. Y si entendemos que en algunos casos las tecnologías de IA pueden ser de alto riesgo, la aseguración (*safety*) debe ser preventiva; debe estar inscrita en el propio sistema. En los sistemas sociotécnicos complejos, la evaluación de riesgos y el monitoreo de seguridad es parte de todo el ciclo de vida del sistema. Con todo,

dado que la seguridad de la IA es un campo emergente, será importante seguir de cerca, por ejemplo, el relevamiento que está llevando adelante el monitor de incidentes de IA que en 2023 puso en marcha la OCDE.

Otro territorio en el que las humanidades y las ciencias sociales han participado en la conversación pública es el de la ética de la IA. Desde esa perspectiva se establecen recomendaciones, buenas prácticas, lineamientos habitualmente voluntarios acerca de lo deseable en relación con el desarrollo y la implementación de las IA, como en el caso de la *Recomendación sobre la Ética de la IA* de la Unesco (2021).

Ahora bien, si las IA y el ecosistema digital ya no solo son herramientas sino un mundoambiente, y por ende, si la relación con ellos no es tanto una relación instrumental sino de inmersión sistémica; si aceleran el procesamiento y la gestión de lo social, y pueden ser de alto riesgo en áreas de experiencia críticas para la población, la perspectiva que necesitamos va más allá de las recomendaciones voluntarias y los códigos de buenas prácticas, y a su vez debe complementar el enfoque de la responsabilidad de la persona física o jurídica con el enfoque de la seguridad sistémica. Como dice Brent Mittelstadt (2019), “puede ser necesario establecer el desarrollo de la IA como una profesión con una categoría equivalente a otras profesiones de alto riesgo. Es una rareza regulatoria que otorguemos licencias a profesiones que brindan un servicio público, pero no a la profesión responsable de desarrollar sistemas técnicos para aumentar o reemplazar la experiencia humana y la toma de decisiones dentro de ellos. Los riesgos de las profesiones autorizadas no se han disipado, sino que han sido desplazados a la IA” (2019).

De manera afín, Ariel Vercelli sugiere “superar los comités de ética y ofrecer a la sociedad una discusión más amplia, abierta y democrática sobre IA”. Desde su perspectiva, se requieren procesos de co-construcción entre tecnologías y regulaciones; esto es, “desarrollar tecnologías (incluso para fines regulativos: control y gestión del tiempo, el espacio y las conductas)” (2023).

De allí que para afrontar los tipos emergentes de IA desde las ciencias sociales y humanas, para que su gobernanza sea eficaz, es preciso complementar la ética con un enfoque sistémico, que reponga, allí donde hay alto riesgo, las mediaciones institucionales que la aceleración se saltea, que establezca procedimientos de revisión transparentes, estructuras de rendición de cuentas vinculantes, documentación de modelos y conjuntos de datos y auditoría independiente. En suma: defensas en profundidad para que convivir e incluso coevolucionar con las IA sea más seguro y confiable.

## “Conocer es organizar”: una agenda de investigación para la Argentina y la región

Poder intervenir en el ecosistema digital y de las IA requiere cartografiar un territorio complejo y multicapas. Por un lado, esta perspectiva convoca a las ciencias sociales y humanas del siglo XXI a ingresar de lleno en el “giro materialista”: dejar de lado la supuesta “desmaterialización” asociada con las tecnologías digitales y formularse la pregunta por las materialidades (cables submarinos, infraestructura de nube, servidores, satélites, minería de tierras raras) y enfocar la dinámica geopolítica de las luchas por los diferentes extractivismos que la componen (desde el de minerales raros hasta el de datos), así como cartografiar la economía política del ecosistema de datos, algoritmos y plataformas.

El segundo tema de esta agenda es el reverso del primero: la pregunta por los restos y las huellas del desarrollo. Pensemos, por ejemplo, en las cinco “islas de la basura” que recorren los océanos del mundo; la del Pacífico Norte es hoy tan extensa como la superficie del Perú. O en las “zonas de sacrificio” del capitalismo informacional transnacional. Regiones geográficas enteras están permanentemente sujetas a daño medioambiental y falta de inversión. La tarea aquí se extiende desde la medición de esas huellas, la planificación sobre el modo de minimizar el daño, y hasta la decisión de impulsar, parafraseando a Benjamin Bratton, una “terraformación de la Tierra” (mucho antes que de la Luna o de Marte), buscando el modo de intervenir para hacer, de las ruinas, un hogar.

El tercero reconduce la pregunta por las huellas hacia la trazabilidad y el gobierno de los vivientes. Desde hace poco más de una década convivimos con la asombrosa posibilidad de cruzar las huellas comportamentales y las huellas biométricas de los habitantes del mundo digital, inscriptos en una verdadera “cultura de la vigilancia” no elegible, ubicua y distribuida. Aquí se abren preguntas de investigación relevantes, como cuánto dicen de nosotros estos datos, y qué grilla de inteligibilidad se nos propone acerca de nosotros mismos: ¿somos un conjunto de datos?, ¿somos nuestros comportamientos?

El cuarto tema son los desafíos de la aceleración digital. En un país como la Argentina, cuyos habitantes de entre 16 y 64 años pasan en promedio 9 horas diarias en internet, según datos de enero de 2023 de la agencia internacional We Are Social (y lo mismo ocurre en los otros países de la región analizados: 9 horas en Colombia; 8.30 horas en Chile; 8.01 horas en México), es una

cuestión crucial. ¿Cómo impacta en nuestras vidas –en el trabajo, la educación, en la salud– esa aceleración digital?

El quinto tema es el de los accidentes propios de estas tecnologías de alta complejidad con las que convivimos. Los nuevos accidentes sistémicos del Tecnoceno son generales por su extensión geográfica; multiescalares –afectan desde la escala doméstica hasta la planetaria–; transversales –requieren experticias muy diferentes para hacerles frente–; y, en algunos casos, afectan a víctimas de cuarta instancia –no nacidas en el momento del accidente–. Los potenciales incidentes de IA que afectan la calidad de la vida democrática son múltiples: desde la vigilancia masiva hasta el *deep fake*, desde la multiplicación de los sesgos y la infrarrepresentación de las minorías hasta la manipulación maliciosa del comportamiento, la calificación social a través de métodos algorítmicos y las crecientes desigualdades epistémicas. ¿Cómo recabamos e identificamos los incidentes de IA a tiempo? ¿Cómo los estudiamos? ¿Cómo los anticipamos y “capturamos el error” para minimizar su impacto? No es una tarea solo para los expertos en informática, sino para equipos interdisciplinarios expertos en Sociedad Artificial.

## Palabras finales

No es impensable que en el futuro, en retrospectiva, recordemos estos años como un tiempo transcurrido “entre accidentes”: entre la pandemia provocada por el Covid-19 y las controversias sobre los desafíos –incluido el “riesgo existencial”– que conllevan los nuevos tipos emergentes de IA. El primero de estos accidentes, debido a las limitaciones de movimiento que implicó, ubicó a buena parte de la población del mundo en la necesidad de ingresar a un “shock de virtualización”. E instaló en la conversación pública la necesidad de interceptarlo a través de un “nuevo acuerdo tecnológico”: un *tech new deal* para unas nuevas democracias digitales. El segundo, por el momento, ha disparado una enorme cantidad de interrogantes, debido a la fuerza disruptiva con la que parece atravesar las diversas escalas de la experiencia social, y porque la aceleración que imprime a los procesos que automatiza se manifiesta, al menos en un primer momento, como un impulso de “destitución” de muchas de las mediaciones que constituían hasta hace poco lo social. Para las ciencias sociales esta apertura temporal es un momento de gran potencial heurístico. ¿Serán nuestros saberes sobre lo social capaces de afrontar la *sociedad artificial* del siglo XXI?



## Referencias bibliográficas

- Costa, F. (2021). *Tecnoceno*. Buenos Aires, Taurus.
- Costa, F. *et al.* (2023). “Desafíos de la Inteligencia Artificial Generativa”, en *Question*, vol. 3, N° 76, dossier “La Inteligencia Artificial en debate: reflexiones para generar”, La Plata.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nat Mach Intell* 1, 501-507. <https://doi.org/10.1038/s42256-019-0114-4>.
- Vercelli Ariel (2023). “Las inteligencias artificiales y sus regulaciones: Pasos iniciales en Argentina, aspectos analíticos y defensa de los intereses nacionales”, en *Revista de la Escuela del Cuerpo de Abogados y Abogadas del Estado*, mayo 2023, Año 7 N° 9, Buenos Aires, Argentina (ISSN 2796-8642), pp. 195-21. En internet: [revistaecae/article/download/232/213/548](http://revistaecae.ptn.gov.ar/index.php/revistaecae/article/download/232/213/548)
- Zuboff, S. (2020). *La era del capitalismo de la vigilancia*, Barcelona, Paidós.

<https://vocesenelfenix.economicas.uba.ar>