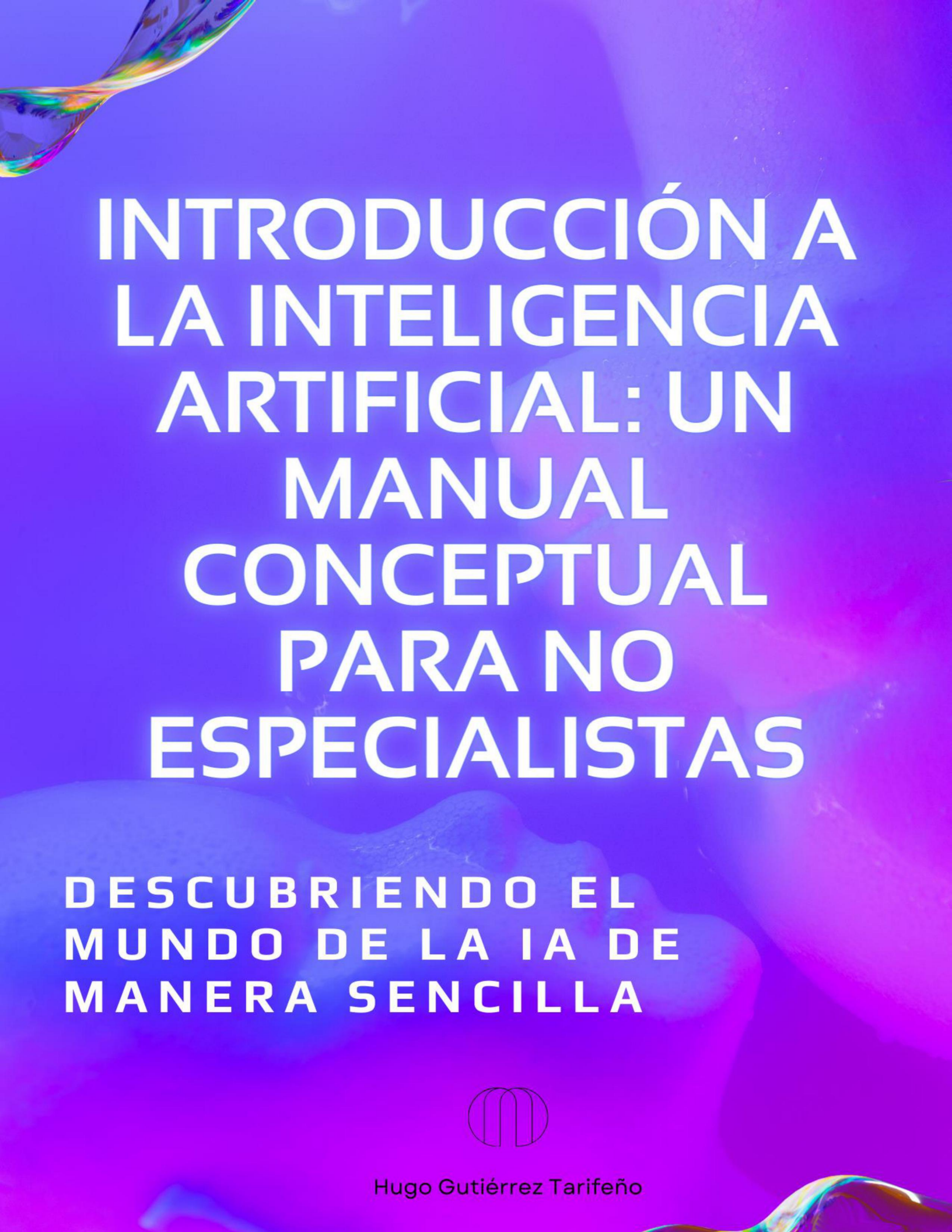




INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL: UN MANUAL CONCEPTUAL PARA NO ESPECIALISTAS

DESCUBRIENDO EL
MUNDO DE LA IA DE
MANERA SENCILLA



Hugo Gutiérrez Tarifeño

INTRODUCCIÓN A LA IA: UN MANUAL CONCEPTUAL PARA NO ESPECIALISTAS

HUGO GUTIÉRREZ TARIFEÑO

INNOVACIENTE
2023



Todos los derechos reservados ©Innovaciente
Berta Silva 10, Jahuel, Valparaíso, Chile
Tel: +56930605070
www.innovaciente.cl

©Hugo Gutiérrez Tarifeño, Autor
ISBN 978-956-416-439-7
Portada e Imágenes generadas con ideogram y canva por Hugo Gutiérrez Tarifeño

Queda prohibida toda reproducción total o parcial de la obra, por cualquier medio, sin la debida autorización previa del autor



Este texto está dedicado a todas y todos los que vivimos esta nueva era, en especial para las personas con neurodiversidades y a quienes vienen avanzados ya desde una era pasada muy diferente a la que experimentamos hoy, vaya mi cariño profundo.

Las ganas de aportar a quienes quieren y creen que aprender, a pesar de las dificultades, es algo que nos hace sentir útiles y conectados, a los que creen en el desarrollo de tecnología a escala humana, a los que sabemos lo importante de la colaboración y compartir conocimiento, pero sobre todo a los que nos despertamos en la mañana con la alegría infinita de seguir vivos.

Un abrazo para mis amigos Fernando Jiménez Motte, Mauricio Garrot, Igor Morales Villena, Christian Garrido Falcon, Sergio Tapia Rodríguez, a mis profesores y guías Ricardo Salazar Chávez de Perú con métodos y John Atkinson Abutridy de Chile con aplicaciones, a mis compañeros Christian, Miguel y Pato, a mis padres Juanita y Aldo por darme la vida, a mi tío Eduardo por su ejemplo, a mi tío Mario por su confianza, y a Loreto, Javi, Tomás, Seba, Fran, August y al Benja por ser porfiados y geniales y mi Lore querida y amada, con ellos la vida es más feliz.

ÍNDICE



6.	Dedicatoria
9.	Bienvenida
10.	Introducción a la IA
16.	Conceptos básicos de la IA
20.	Aplicaciones de la IA en la vida cotidiana
25.	Ética y responsabilidad en la IA
33.	Glosario de términos de IA
36.	Referencias y recursos adicionales
40.	Un adelanto del próximo libro

BIENVENIDA(O)S

A LA ERA DE LA TECNOLOGÍA Y LA IA, DONDE LAS POSIBILIDADES DE INNOVAR SON ILIMITADAS, DONDE LOS SUEÑOS DE DÉCADAS DE CIENCIA FICCIÓN HOY TIENEN UNA REALIDAD QUE IMPACTA FUERTEMENTE EN NUESTRA VIDA COTIDIANA, A LOS QUE ESTAMOS PRESENCIANDO ESTE AVANCE Y A LOS QUE LO ESTAN PROTAGONIZANDO, A UNA REVOLUCIÓN TECNOLÓGICA QUE ESTA REDISEÑANDO EL MUNDO, A LAS SOCIEDADES Y CULTURAS, AL SER HUMANO TAL COMO LO CONOCEMOS, ESTA BIENVENIDA ES PARA TODAS Y TODOS, ES UNA INVITACIÓN A COMPRENDER, COMO Y PORQUE AVANZAMOS EN LA DIRECCIÓN EN QUE LOS HACEMOS.



Capítulo I

Introducción a la IA

IA y su importancia en la sociedad actual

La Inteligencia Artificial (**IA**) es una rama de la informática que desarrolla algoritmos y sistemas capaces de aprender y realizar tareas que hasta hace poco sólo podían ser realizadas por humanos. Este avance tecnológico ha cambiado por completo diversos aspectos de la vida diaria y de la sociedad en general. La IA extrae grandes cantidades de datos y utiliza redes neuronales complejas para procesar esos datos, lo que permite a las máquinas tomar decisiones, predecir patrones y resolver problemas automáticamente, desde asistentes virtuales en nuestros dispositivos móviles hasta diagnósticos médicos más precisos y vehículos autónomos, la IA se ha vuelto omnipresente en muchas industrias y sectores.

Su importancia radica en su capacidad para aumentar la eficiencia y la productividad en diversos campos, así como resolver problemas complejos que requieren un procesamiento de datos a una escala más allá de las capacidades humanas.

La IA está revolucionando la forma en que interactuamos con la tecnología, haciéndola más intuitiva y accesible, sin embargo, estos avances conllevan la necesidad de tener en cuenta los impactos éticos y sociales de la IA, la transparencia, la justicia y la rendición de cuentas en el desarrollo y uso de estas tecnologías, se están convirtiendo en aspectos críticos para garantizar que la IA beneficie a la sociedad en su conjunto.

En definitiva, la IA representa no sólo un avance tecnológico sino también un cambio fundamental en la forma en que vivimos y trabajamos, su influencia en la sociedad actual es innegable y su continuo desarrollo promete un futuro aún más apasionante y con muchas posibilidades.

Historia de los orígenes de la IA

El desarrollo de la IA es un viaje fascinante a través de avances tecnológicos y conceptuales que han cambiado la forma en que interactuamos con la tecnología y el mundo en general, desde la década de los 50 con la conferencia de Dartmouth de 1956, donde por primera vez se expuso la posibilidad de avanzar a través de los pioneros de la informática y la

cognición, como John McCarthy y Marvin Minsky, reunidos explorando "hacer pensar a las máquinas".

Luego, en la década del 60, dio lugar al desafío de generar lógica y resolución de problemas, desarrollándose los primeros programas de IA basados en lógica, como lo fueron, el programa de resolución de teoremas de Newell y Simon (1) y el programa "ELIZA" de Joseph Weizenbaum (2), que simulaba una conversación terapéutica con un psicólogo.

En los 70 se comenzaron a desarrollar sistemas expertos que utilizaban conocimientos especializados y reglas lógicas para resolver problemas específicos en campos como la medicina y las finanzas. El concepto de "lenguaje Lisp" se introdujo como un lenguaje de programación importante para la IA.

Durante los 80 se generaron nuevos conceptos como conectividad y redes neuronales artificiales, una técnica inspirada en la estructura del cerebro, que generaron y aún generan nuevos temas populares, profundizando la exploración en el campo del aprendizaje automático. La IA se comenzó a aplicar con mayor frecuencia mediante las aplicaciones prácticas como el reconocimiento de voz y la visión por computadora.

La década de los 90 llevó a los algoritmos de aprendizaje automático a un nivel de uso y desarrollo nunca antes visto, así como las máquinas de vectores de soporte (SVM), que ganaron una increíble popularidad, comenzaron a desarrollarse motores de búsqueda avanzados y sistemas de sugerencias.

Para el año 2000 y en adelante los avances en aprendizaje profundo y big data, basados en redes neuronales profundas, hicieron que despegaran muchos ámbitos desconocidos de los desarrollos tecnológicos ligados a la IA, la posibilidad de acceder a fuentes con enormes cantidades de datos (Big Data) (3), estimularon la mejora de algoritmos, apalancando operaciones y procesos que se hicieron más eficientes.

Luego, en la década del 2010, ya la IA se integró a una impresionante cantidad de

1. <https://revistasdex.uchile.cl/index.php/bits/article/download/2767/2700>

2. <https://www.redalyc.org/pdf/5217/521752338014.pdf>

3 <https://www.oracle.com/cl/big-data/what-is-big-data/>

aplicaciones cotidianas, como asistentes virtuales, sistemas de recomendación y reconocimiento facial, los vehículos autónomos y los chatbots basados en IA son cada vez más populares.

Finalmente, desde el 2020 y a futuro, la IA ha seguido creciendo y está desempeñando un papel central en la innovación tecnológica, la medicina, la industria automotriz y muchos otros campos. La historia de la IA es evidencia de un progreso constante en la capacidad de las máquinas para aprender y realizar tareas previamente reservadas a los humanos. A medida que avancemos, la IA promete tener un impacto aún mayor en la forma en que vivimos y en todos los ámbitos de nuestra existencia, abriendo nuevas fronteras en tecnología e innovación.

Objetivos generales y aplicaciones de la IA en diferentes campos

De los objetivos aplicados al desarrollo humano y a mejorar nuestro planeta, claramente el primero es hacer muchísimo más eficaz a la salud y la medicina, a través de mejorar los diagnósticos, los tratamientos de enfermedades y condiciones, en general la administración y gestión de la salud del ser humano, aplicando diagnóstico asistido por IA en imágenes médicas (4), personalización genética de tratamientos y toda una nueva generación de medicamentos, monitoreando a los pacientes en tiempo real.

Luego, por uso y productividad, el desarrollo de vehículos autónomos y mejoras a la seguridad vial, son también prioridades en los objetivos de esta nueva era con la herramienta IA a través de aplicarlos, por ejemplo, en la asistencia a la conducción y sistema de estacionamiento automático, con el aporte en la navegación automática y reconocimiento de señales de tráfico mediante nuevos sistemas de detección y realidad aumentada.

Es claro que uno de los puntos críticos son las finanzas y la banca, a nivel mundial, regional, estatal y local desde el sector público y sobre todo el privado, teniendo como objetivo primordial optimizar las operaciones y servicios financieros, mediante aplicar análisis de riesgo crediticio y detección de fraudes, además de la consultoría financiera

4 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7752970/>

personalizada y la gestión de carteras de inversiones, para cada cliente o usuario del sistema financiero.

Como una derivada no menos importante del proceso financiero y bancario, es parte importante, el consumo, a través de la venta minorista y el comercio electrónico, que representan un sector masivo dentro de los mercados, teniendo como su objetivo más relevante el ofrecer una experiencia de compra más personalizada, rápida, cómoda y eficiente posible, mediante la integración de desarrollos en torno a la recomendación de productos y el análisis de las tendencias de compra, apalancando la optimización de la gestión del inventario y la cadena de suministro en su totalidad.

Si hablamos de importancia suma, no hay otra que la educación, ya que, como objetivo principal de este período, y teniendo que cumplir con el objetivo cercano de personalizar el aprendizaje y mejorar la experiencia educativa, a través de sistemas de autoaprendizaje, de aprendizaje guiado virtualmente y adaptativamente, se hace necesario generar evaluaciones y seguimiento automático del progreso de los estudiantes, pudiendo integrar el aprendizaje, junto a aplicaciones efectivas y útiles de este.

La generación de estas nuevas tecnologías y herramientas han permitido que las industrias vayan mejorando sus procesos y operación a través de centrar el uso de las nuevas herramientas de IA en optimizar y mejora la calidad de la cadena completa de producción, operación, administración y distribución, productos incluidos en esto, por ejemplo, el mantenimiento predictivo de máquinas, la automatización de las líneas de producción y el control de calidad.

No puede estar ausente temas tan críticos como la energía y medio ambiente, esto ya que el uso de tecnologías y herramientas de IA, deben ir alineados con el cuidado de nuestro entorno, por ello como objetivo principal están la promoción de la sostenibilidad y de la eficiencia energética, ante el gasto mayor de ella, aplicando de esta forma una permanente optimización de las redes eléctricas y gestionando e incluyendo en las matrices energéticas las energías renovables no convencionales, y como derivada de este objetivo un subproceso no menos importante, acerca del seguimiento y previsión de fenómenos climáticos extremos, donde el control de riesgos toma un papel relevante a través de la generación de modelos predictivos de desastres naturales y artificiales.

Uno de los temas más complejos por lo multidimensional es la seguridad y defensa, tanto a nivel de estado como a nivel de personas, los objetivos son mejorar la seguridad y responder adecuadamente a las amenazas, precaviendo ataques a pequeña, mediana y gran escala, generando a través de tecnologías de IA aplicaciones para realizar análisis de imágenes de vigilancia y de reconocimiento facial, por supuesto no quedan afuera la detección de actividades sospechosas y la prevención de ataques cibernéticos.

El sector del entretenimiento y medios de comunicación son vitales para el desarrollo del ser humano, por supuesto las nuevas tecnologías y la IA tiene como objetivo la personalización del contenido y mejorar la experiencia del usuario, esto a través de la generación de sistemas de recomendación personalizada de películas, música y contenido en línea, creando efectos especiales y animaciones, llevando a nuevas formas de artes escénicas y audiovisuales.



No son todos los impactos que tiene la IA y las nuevas tecnologías, pero son los ejemplos más claros de cómo se están aplicando en diferentes industrias y sectores culturales, económicos y sociales, la IA está en constante evolución, hoy con IA Generativa, llevándonos a una Super IA, su influencia en la sociedad continúa creciendo a una velocidad abismante y que a muchos nos deja pensando en si estamos preparados para ello, proyectando transformaciones significativas en la forma en que vivimos y nos relacionamos entre nosotros y con nuestro entorno, con la ética como valor humano principal.

Capítulo II

Conceptos Básicos de la IA

Definición sencilla de algoritmos y porque son esenciales en la IA

Un algoritmo (5) es un conjunto de instrucciones paso a paso para completar una tarea o problema específico. En otras palabras, es como una receta que especifica de forma clara y precisa cómo completar una tarea, en el contexto de la IA, los algoritmos son de fundamental importancia ya que forman la base para construir sistemas de IA, estos algoritmos permiten que las máquinas aprendan de los datos, reconozcan patrones, apoyen la toma de decisiones y realicen tareas complejas sin intervención humana directa.

Por ejemplo, en el aprendizaje automático (6), se pueden utilizar algoritmos para permitir que los modelos de IA ajusten sus parámetros para mejorar el rendimiento en tareas específicas cuando se exponen a más datos. De esta forma, los algoritmos están en el corazón de la IA y proporcionan a las máquinas las instrucciones que necesitan para realizar tareas inteligentes y adaptables.

Explicación de las redes neuronales y su papel en la simulación del pensamiento humano

Las redes neuronales son un tipo de modelo matemático inspirado en el funcionamiento del cerebro humano, están diseñados para aprender y reconocer patrones complejos en conjuntos de datos, estas redes constan de capas de nodos interconectados, donde cada capa realiza una operación matemática simple.

El papel de las redes neuronales(7) en la simulación del pensamiento humano radica en su capacidad para aprender y generalizar a partir de datos, al igual que, el cerebro humano, las redes neuronales pueden adaptarse y mejorar su desempeño en tareas específicas cuando se les proporciona más información, esto se logra ajustando los pesos y sesgos de las conexiones entre nodos, sin embargo, es importante enfatizar que las redes neuronales actuales están lejos de imitar completamente el pensamiento humano, aunque puede realizar tareas complejas como el reconocimiento de imágenes

5. <https://sa24377df8873c4a3.jimcontent.com/download/version/1633405791/module/10848800371/name/Principios.pdf>

6. <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419>

7. <https://www.unir.net/ingenieria/revista/redes-neuronales-artificiales/>

y el procesamiento del lenguaje natural, su funcionamiento es esencialmente una simulación de ciertos aspectos del procesamiento cognitivo humano.

El cerebro humano es extremadamente complejo y aún no se comprende del todo. Las redes neuronales en la IA son simplificaciones y abstracciones de los procesos neuronales biológico-cuántico que ocurren en el cerebro. Además, la IA actual carece de conciencia, intencionalidad y emoción, que son elementos esenciales del pensamiento humano. En resumen, las redes neuronales son herramientas poderosas para el aprendizaje automático y la IA, pero el objetivo de simular completamente el pensamiento humano todavía está en el ámbito de la ciencia ficción y es inalcanzable con la tecnología actual.

Aprendizaje automático, cómo las máquinas aprenden de los datos

El aprendizaje automático es un campo de la IA que se centra en la capacidad de las máquinas para aprender de los datos y realizar tareas específicas sin estar programadas explícitamente para cada tarea, aquí explicaremos cómo las máquinas aprenden de los datos.

El primer paso es la recopilación de datos, para ello necesitamos un conjunto de datos relacionados con la tarea que se desea abordar, por ejemplo, si desea crear un modelo que prediga los precios de la vivienda, necesita un conjunto de datos que contenga información sobre las características de la casa (como el tamaño, la ubicación y la cantidad de habitaciones) y sus precios, luego es muy importante preparar los datos, debido a que los datos recopilados a menudo requieren ciertas operaciones para asegurar su consistencia y coherencia, para ello se limpian los datos (eliminar valores atípicos y errores), se normalizan los datos (asegurarse de que estén en un formato coherente) y se dividen en conjuntos de entrenamiento y prueba, el conjunto de entrenamiento se utiliza para entrenar el modelo y el conjunto de prueba se reserva para evaluar el rendimiento.

En adelante se debe gestionar la selección de modelo, de esta forma elegir el mejor tipo de modelo de aprendizaje automático (8), uno apropiado para la tarea proyectada, estos modelos típicamente incluyen algoritmos de regresión para predicción numérica, algoritmos de clasificación para dividir datos en categorías o modelos más complejos

8.- <https://learn.microsoft.com/es-es/windows/ai/windows-ml/what-is-a-machine-learning-model>

como redes neuronales. Para que el modelo sea eficiente debe existir el entrenamiento del modelo, en el cual el modelo se "alimenta" con un conjunto de entrenamiento que consta de datos y respuestas conocidas (como los precios de bienes raíces), luego se ajustan los parámetros internos en función de los datos, intentando minimizar el error entre la respuesta prevista y la real.

Para completar el proceso de uso de la IA se debe validar y optimizar el modelo, esto luego del entrenamiento, para ello, el modelo se evalúa utilizando un conjunto de pruebas

que contiene datos nunca antes vistos, esto representa una medida de la capacidad de generalizar a nuevos datos, si no está satisfecho con el rendimiento, puede ajustar el modelo o los datos, finalmente se produce la magia, el modelo predice y se implementa, es decir, una vez que el modelo está entrenado y validado, se puede utilizar para hacer predicciones sobre nuevos datos, con ello, por ejemplo, se puede utilizar para predecir el



precio de una vivienda en función de sus características. Es importante señalar que el aprendizaje automático es un proceso iterativo, a medida que se recopilan más datos o se identifican mejores algoritmos, mejorar el rendimiento del modelo a menudo requiere volver a los pasos anteriores.

En un buen resumen, el aprendizaje automático permite a las máquinas aprender patrones a partir de datos, lo que les permite realizar tareas sin tener que programarse explícitamente para cada situación específica.

Capítulo III

Aplicaciones de la IA en la Vida Cotidiana

Ejemplos prácticos de cómo la IA está presente en nuestras vidas diarias, con asistentes virtuales, recomendaciones de contenido, y mucho más

La IA está cada vez más presente voluntaria o involuntariamente en nuestra vida diaria, facilitando diversas tareas, mejorando la experiencia de las personas en muchos aspectos y trayendo cierta comodidad, que, de no ser atendida, podría provocar trastornos personales y sociales de desadaptación, al no contar con estas nuevas tecnologías.

En concreto los asistentes virtuales como Siri de Apple, Google Assistant y Amazon Alexa utilizan IA para comprender y responder a comandos de voz, pudiendo realizar tareas como configurar recordatorios, buscar información en Internet, enviar mensajes y controlar los dispositivos conectados en su hogar. Todo esto va de la mano del desarrollo de recomendadores de contenido, plataformas como Netflix, Spotify y YouTube utilizan algoritmos de IA para analizar tus preferencias, hábitos de gasto y recomendar películas, series, música y vídeos que te puedan gustar.

Servicios como los de traducción automática del tipo Google Translate utilizan IA para traducir texto entre diferentes idiomas, la IA analiza patrones lingüísticos en grandes conjuntos de datos para crear traducciones más precisas, junto con ello en el mundo virtual se han desarrollado filtros de spam de correo electrónico, utilizando tecnología de IA para identificar y clasificar correos electrónicos no deseados y evitar que lleguen a nuestra bandeja de entrada.

Los nuevos sistemas de recomendación de compras, integrados en plataformas como Amazon y otros sitios web de comercio electrónico utilizan IA para analizar el historial de compras y el comportamiento de navegación para brindar recomendaciones de productos personalizadas para cada persona y situación, automatizando el servicio al cliente, de hecho son muchas empresas que están utilizando chatbots basados en IA para proporcionar respuestas automatizadas a preguntas frecuentes o brindar soporte en tiempo real a los clientes en sus sitios web.

Como lo mencionamos anteriormente, un punto principal es la conducción autónoma, donde la IA es esencial en los vehículos con esta tecnología de autoconducción, ya que procesa

información de los sensores para tomar decisiones en tiempo real, sobre como en qué dirección ir, que velocidad alcanzar y las urgencias de frenado del vehículo.

De las áreas donde más se siente la integración de las nuevas tecnologías y la IA es en la del diagnóstico médico asistido por dispositivos digitales, hoy mismo algunos sistemas de IA están capacitados para analizar imágenes médicas, como rayos X e imágenes de resonancia magnética, para ayudar a los profesionales médicos como radiólogos a identificar y diagnosticar afecciones médicas.

Avanzamos en técnicas de reconocimiento facial y de voz, esto ya que la IA se utiliza en sistemas de reconocimiento facial para desbloquear dispositivos, etiquetar personas en fotografías y mejorar la seguridad en lugares públicos, se está utilizando fuertemente para el reconocimiento de voz en servicios como Siri y el asistente de voz de tu coche con tecnología en pantalla, a su vez y dentro del mismo ámbito de desarrollo hoy la optimización de rutas y tráfico, a través de aplicaciones de navegación como Google Maps, Waze y otras utilizando algoritmos de IA para analizar datos de tráfico en tiempo real y proporcionar rutas optimizadas para llegar a su destino.

Estos son ejemplos de la irrupción profunda en nuestras vidas de las nuevas tecnologías emergentes y la IA, hoy con avances notables en casi todos los campos de la vida humana, la IA se está implementando en una variedad de aplicaciones, desde salud y educación hasta entretenimiento y seguridad, lo que tiene un impacto en nuestra vida diaria, pero debemos estar alertas porque los impactos pueden positivos o negativos, dependiendo de su facilidad de acceso, el uso y los resultados luego de su uso.

Casos de estudio sobre cómo la IA ha revolucionado industrias como la salud, el transporte y el comercio

Las tecnologías emergentes y la IA han revolucionado una variedad de industrias, incluidas la atención médica, el transporte y el comercio, son muchos los casos de estudio donde se muestra claramente el tremendo y vasto impacto en cada una de las áreas que vamos a analizar, que además son las áreas prioritarias para el desarrollo de tecnología a escala humana.

Siempre prioritario es el área Salud, lo mencionamos en el apartado anterior, por ejemplo, en el diagnóstico médico, a través de IBM Watson for Oncology ⁽⁹⁾, que es un sistema de IA que ayuda a los médicos a determinar opciones de tratamiento para pacientes con cáncer basándose en la literatura médica y que contempla una serie de análisis y resultados y junto a IBM, Aidoc y Zebra Medical Vision, que apoyan y soportan a los radiólogos para detectar anomalías en imágenes médicas, como radiografías y tomografías computarizadas.

En el sector transporte, ya también lo mencionamos antes, la conducción autónoma, con desarrollos como el de Waymo (que es una subsidiaria de Google) ⁽¹⁰⁾ y Tesla han incluido tecnología de conducción autónoma que utiliza algoritmos de aprendizaje automático para interpretar datos de sensores y tomar decisiones en tiempo real. Por supuesto también y en base a los largos viajes de los vehículos desde las fábricas y líneas de ensamblaje hasta su destino final, la logística y rutas de las empresas de transporte y logística, también están utilizando algoritmos de IA para optimizar estas rutas y generar alternativas, gestionar flotas y predecir tiempos de entrega.

El comercio ha sido uno de los más impactados por las nuevas tecnologías e IA por ejemplo en el comercio electrónico, esto a través de recomendadores de productos y servicios como lo tienen Amazon y otras plataformas de comercio electrónico, que están utilizando algoritmos de IA para analizar el historial de compras y el comportamiento del usuario para brindar recomendaciones personalizadas, en el mismo sector y ámbito la gestión de inventario, donde tenemos sistemas predictivos con IA integrada, para prever la demanda futura y optimizar los niveles de inventario para reducir costos y evitar pérdidas de ventas debido a falta de existencias y exceso de producto.

Para las finanzas mundiales, el sector de la gestión y análisis de riesgos, hoy utilizan algoritmos de aprendizaje automático para analizar las enormes cantidades de datos y predecir también riesgos financieros para ayudar a tomar decisiones informadas sobre inversiones y préstamos, con ello también en el sector del mercado de valores, que hoy usan de forma constante las nuevas tecnologías y la inteligencia artificial para reconocer patrones en los datos del mercado y ejecutar operaciones de compra y venta en una fracción de segundo.

9. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7105853/>

10. <https://waymo.com/intl/es/>

Sin duda uno de los sectores más beneficiados con la integración de IA es la educación, donde la personalización del aprendizaje de plataformas educativas como Khan Academy y Duolingo, utilizando algoritmos de IA para adaptar el contenido a las necesidades y habilidades individuales de los estudiantes y junto con ellos las plataformas de calificaciones automáticas son herramientas útiles, Turnitin (11) utiliza IA para analizar y calificar automáticamente ensayos y trabajos escritos.

Importante es el sector del entretenimiento, donde los recomendadores de contenido, dentro de las plataformas de streaming como Netflix y Spotify utilizan IA para analizar el comportamiento del usuario y ofrecer recomendaciones de películas, series y música, por supuesto la creación de contenido, con algunas empresas que están hoy experimentando con nuevas



tecnologías e IA para generar contenido, como escribir noticias o componer música.

Estos casos de estudio son sólo algunos sobre el uso de las tecnologías emergentes y la IA, la lista continúa creciendo rápidamente a medida que avanza el desarrollo de la tecnología. Todo este desarrollo sobre todo la IA está transformando una variedad cada vez mayor de industrias al aumentar la eficiencia, la personalización y la precisión de la toma de decisiones, que deben tener un componente ético.

11. <https://www.turnitin.com/es>

Capítulo IV

Ética y Responsabilidad en la IA

Consideraciones éticas al desarrollar y utilizar sistemas de IA

El desarrollo y uso de nuevos sistemas con tecnologías emergentes y de IA, requiere importantes y críticas consideraciones éticas para garantizar que esta tecnología se implemente de manera responsable y beneficiosa para la sociedad en su conjunto y no sólo para unos pocos que es la regla en nuestra cultura mundial.

Las consideraciones éticas tienen prioridades y dentro de ellas, una de las principales es que se incluyan transparencia y explicabilidad, esto porque es importantísimo ser lo más transparente sobre cómo los sistemas de IA toman decisiones. Todos los que trabajamos a diario como desarrolladores de nuevas tecnologías asociadas a la IA, junto con la transparencia nos deberíamos esforzar en explicar cómo funcionan los algoritmos diseñados y cómo llegan a determinadas conclusiones o resultados, para entender el porqué de ellas.

No podemos sino hablar sobre equidad y sesgo en los sistemas de IA, ya que en los nuevos desarrollos se pueden reflejar sesgos presentes en los paquetes de datos fuentes utilizados para la capacitación o entrenamiento (12), es importante reconocer y reducir estos sesgos para evitar una discriminación injusta basada en características como raza, género y edad. Junto con la equidad y sesgo también es crítico la privacidad y seguridad, para ello se deben implementar medidas estrictas para proteger la privacidad de los datos utilizados para entrenar y operar sistemas de IA, siempre es muy importante obtener el consentimiento informado de los usuarios al momento de recopilar datos personales.

Una de las áreas críticas en el desarrollo de las nuevas tecnologías y la IA es la responsabilidad y rendición de cuentas, todas y todos los desarrolladores y propietarios de sistemas de IA debemos ser responsables de su funcionamiento, prioridad es estar preparados para tomar medidas que solucionen algún fallo o problemática y reparar posibles daños causados por su sistema, esto va de la mano con la benevolencia y no maldad, donde los sistemas de IA deben diseñarse para beneficiar a la sociedad en su totalidad y no causar daño intencional o involuntario, y claramente la consideración sobre

12.<https://blogs.iadb.org/igualdad/es/como-aprovechar-la-inteligencia-artificial-sin-perpetuar-sesgos-de-genero/>

posibles efectos negativos, deben permitir tomar medidas proactivas y no reactivas para minimizarlos.

La importancia de la ciudadanía es vital, la participación y colaboración de todos los sectores protagonistas de las tecnologías emergentes y de IA, siendo lo más crítico el involucramiento de diversas partes interesadas, incluidos especialistas en ética (13), representantes de la comunidad y usuarios finales, todo esto dentro del proceso pre y post desarrollo y resultados de la toma de decisiones de los sistemas de IA. Con ello también la transparencia en el uso, donde usuarios, beneficiados y afectados por los sistemas de IA, deben ser informados y capacitados sobre cómo utilizar e interactuar con los sistemas de IA, siempre con la inclusión de la mayor cantidad y calidad de información sobre las capacidades y limitaciones de los sistemas desarrollados y en desarrollo.

Parte de las consideraciones éticas dan cuenta del desarrollo sostenible, los sistemas de IA deben contribuir al desarrollo sostenible y no deben generar efectos externos negativos que dañen el medio ambiente, la economía y principalmente a la sociedad, donde también es una tarea relevante el cumplimiento de regulaciones y leyes, ya que los sistemas de IA deben cumplir con todas las leyes y regulaciones relevantes, incluidas aquellas relacionadas con la protección de datos, la seguridad y los derechos humanos.

La formación continua y la capacitación deben comprender la conciencia ética, todas y todos los profesionales y técnicos que están involucrados en el desarrollo y uso de sistemas de IA deben recibir capacitación ética y ser conscientes de las implicaciones éticas de su trabajo y el impacto de los resultados en los seres humanos.

Estas consideraciones éticas y de cómo interactuamos entre diferentes sectores y actores son esenciales para garantizar que las tecnologías emergentes y la IA, sean utilizados de manera responsable y se maximice el excedente social con beneficios para el conjunto más amplio de la sociedad, en territorios y comunidades y donde los avances en tecnología deben ir acompañados de un enfoque ético, reflexivo y a escala humana.

13.- https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S2215-34032020000100093

Discusión sobre sesgos y prejuicios en los algoritmos de IA

El sesgo en los algoritmos de IA es una cuestión importante que debe abordarse cuidadosa y sistemáticamente, con responsabilidad y se debe discutir sobre este tema, ¿qué son los sesgos y prejuicios en la IA?

Los sesgos y prejuicios en la IA son errores en las fuentes de los algoritmos, que los hacen inexactos o injustos, debido a la influencia del sesgo en los datos de entrenamiento o el diseño del algoritmo, se refiere a la tendencia a producir resultados que son o pueden conducir a resultados discriminatorios o sesgados para ciertos grupos de personas. Para ello se deben analizar las fuentes de los sesgos en la IA, los sesgos en los datos de entrenamiento, en los datos utilizados para entrenar modelos de IA pueden reflejar y reforzar los sesgos sociales y culturales existentes, por ejemplo, si hay desigualdades en el conjunto de datos históricos, los algoritmos podrían y pueden aprender y mantener estas desigualdades.

Existen otros sesgos ,por ejemplo, los sesgos de características seleccionadas, que se generan cuando la selección de parámetros puede introducir sesgo si las características no se seleccionan por igual o si ciertas características se sobreestiman, así como los sesgos en el diseño de los algoritmos, de esta forma los propios algoritmos pueden diseñarse introduciendo sesgos, por ejemplo, un algoritmo de clasificación particular puede ser más propenso a errores en ciertas clases que en otras, y este tipo de fallas se deben prevenir y precaver desde el inicio del desarrollo.

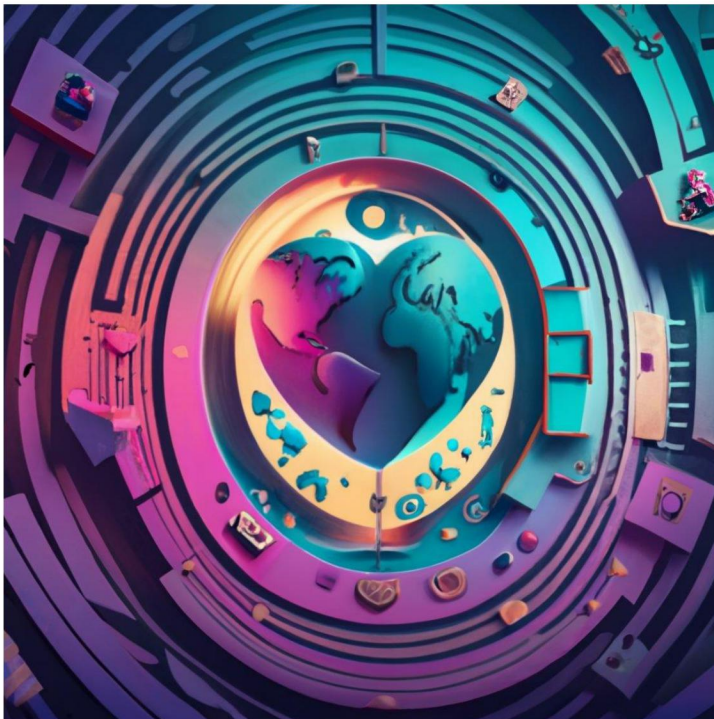
Claramente toda esta tecnología emergente e IA está diseñada y generada por seres humanos, por lo que un peligro latente es la falta de diversidad en los equipos de desarrollo⁽¹⁴⁾, por lo cual los equipos de desarrollo que crean algoritmos pueden no ser lo suficientemente diversos en términos de experiencia y perspectivas, lo que resulta en que no se consideren la existencia de ciertos grupos y sus necesidades, existe una posibilidad de aquello, que se debe prevenir desde el inicio de los proyectos y pudiera esto tener un impacto de sesgo y discriminación. Los nuevos sistemas de IA sesgados pueden conducir a resultados discriminatorios, por ejemplo, en el ámbito de la justicia, donde los sistemas

14.-<https://ts2.space/es/la-importancia-de-la-diversidad-en-el-desarrollo-de-la-inteligencia-artificial/>

de inteligencia artificial que recomiendan sentencias pueden castigar injustamente a ciertos grupos.

Junto con lo anterior es también esperable que, si no se toman las medidas correctivas desde el inicio, se reproduzcan desigualdades sociales, ello porque las fuentes de datos de entrenamiento reflejan o podrían reflejar desigualdades históricas, y el/los algoritmos pueden reproducir y mantener estas desigualdades, lo que derivaría en una falta de equidad y justicia, y claro, estos sesgos o prejuicios de la IA ⁽¹⁵⁾ pueden afectar la asignación de recursos y oportunidades, violando así principios fundamentales humanos. Para reducir y eliminar los sesgos y prejuicios se deben generar estrategias para reducirlos, recopilando datos diversos y representativos, a ello la importancia de utilizar datos que sean representativos de la diversidad de la población a la que se aplicará el algoritmo.

Para realizar las tareas de generación de los nuevos sistemas de IA, se deben realizar pruebas rigurosas y evaluación continua, las tecnologías emergentes y los sistemas de IA



deben evaluarse rigurosamente para detectar y corregir sesgos, esto incluye prioridades como las pruebas de conjuntos de datos y métricas de rendimiento, que junto con las funciones y algoritmos de optimización, permiten implementar técnicas que optimicen algoritmos y funciones que reduzcan los sesgos, involucrando a expertos en ética, diversidad y representantes de diversos grupos en el proceso de diseño y desarrollo pre y post, que generan

soporte de ayuda a identificar y abordar posibles sesgos, destacando en este proceso la transparencia y

15.-<https://www.bbvaopenmind.com/tecnologia/inteligencia-artificial/como-crear-inteligencia-artificial-que-no-exhiba-prejuicios-humanos/>

lenguaje claro, donde los sistemas de IA deben ser transparentes al tomar decisiones para que los sesgos puedan identificarse y corregirse más fácilmente.

Finalmente, el sesgo y los prejuicios en el desarrollo presente y futuro de la IA es un punto crítico de fallas, lo que requiere una atención cuidadosa y un enfoque ético para el diseño, el entrenamiento y la implementación de algoritmos. Con ello y en un proceso paralelo la formación, capacitación y concientización de las personas, con las respectivas acciones proactivas son esenciales para que la IA beneficie a la sociedad en su conjunto de manera justa y equitativa.

La importancia de la transparencia y la toma de decisiones éticas en el diseño de sistemas de IA

La transparencia y la toma de decisiones éticas son esenciales en el diseño de sistemas de IA por muchas y profundas razones asociadas a lo humano, creativo y de a ratos irracional de nuestro comportamiento, por ello su importancia en el desarrollo de nuevas tecnologías y la IA.

La prioridad hoy y mañana es la responsabilidad y la rendición de cuentas, la transparencia en el diseño de sistemas de IA permite a desarrolladores y usuarios comprender cómo trabajan y cómo se toman decisiones, de esta forma, facilita la identificación de sesgos, errores o malentendidos, aumentando así la responsabilidad de los creadores, y generando incluso, desde las etapas de diseño conceptual, formas de disminuir y eliminar los prejuicios y la discriminación.

La IA aprende de datos históricos y podría reflejar los sesgos existentes en esos datos, por lo cual es importante limpiar de todo sesgo los datos de origen a través de revisión de transparencia, aportando a identificar y corregir estos sesgos, evitando así la discriminación y promoviendo la equidad.

Respecto a uno de los temas menos valorados y más importantes, debemos hablar sobre la confianza del usuario, los usuarios de los sistemas de IA confían en que las respuestas que muestran los sistemas IA son justas y se basan en criterios éticos, para ellos esa idea preestablecida sobre la transparencia en la toma de decisiones puede fomentar esta

confianza y de paso albergar en los desarrollos presentes y futuros de tecnologías emergentes el cumplimiento de normativas y leyes, ya que muchos países y jurisdicciones establecen regulaciones específicas para el uso de la IA, lo que a menudo requiere decisiones transparentes y firmemente éticas, en el cumplimiento de estas regulaciones, va la esencia para evitar posibles sanciones legales y mantener la reputación de una organización de cualquier sector.

Esta confianza y cumplimiento de normas le da soporte a otro punto crítico, que es como los desarrollos de IA, se alinean con los valores organizacionales y humanos, transparencia

y toma de decisiones éticas garantizan que los sistemas de IA estén alineados con la visión y la misión de la organización privada, pública, social o académica que los desarrolla, siendo esto necesario para mantener la coherencia y la integridad de la implementación de las tecnologías desarrolladas, en desarrollo y por desarrollar, para ello



además debemos establecer normas sobre la protección de la privacidad, ya que la transparencia en la recopilación y el uso de datos en los sistemas de IA es importante para garantizar que se respeten las normas de privacidad y se protejan los derechos de las personas, dentro y fuera de las organizaciones que desarrollan tecnologías emergentes e IA.

Punto prioritario también es la innovación sostenible, fomentando la transparencia y la ética en el diseño de la IA, de forma de impulsa la innovación sostenible a largo plazo, donde los nuevos sistemas que respetan los valores éticos tienen más probabilidades de ser

aceptados y adoptados por la sociedad en general, en territorios y comunidades, sobre todo las más aisladas.

Finalmente, y como conclusión de este texto, decimos que la transparencia y la toma de decisiones éticas en el diseño de sistemas de IA no solo son moralmente importantes, son esenciales para construir sistemas confiables, justos y equitativos y desarrollo de tecnología a escala humana basada en ETAS (<https://www.innovaciente.cl/etas/>), así como para respetar los derechos y valores humanos, a los individuos y la sociedad en su totalidad, desde la educación en cada hogar, en la formación dentro de las instituciones educacionales, de la sustentabilidad y adaptación de los seres humanos a los cambios y procesos naturales y el respeto por ser habitantes de un mismo planeta , nuestra nave Tierra que nos lleva por el Universo, en este camino respetando y cuidando nuestro entorno, así como en la confianza respetarnos entre seres humanos.

Glosario de Términos de IA

**Lista de términos clave con definiciones
simples: algoritmo, aprendizaje automático,
red neuronal y otros**

Algoritmo Definición: Un conjunto de pasos o reglas definidas que una computadora sigue para resolver un problema o realizar una tarea específica.

Aprendizaje Automático (Machine Learning) Definición: Un tipo de IA que permite a las computadoras aprender de datos y mejorar su rendimiento en tareas específicas sin ser explícitamente programadas para cada una.

Red Neuronal (Neural Network) Definición: Un modelo de IA inspirado en la estructura y el funcionamiento del cerebro humano, que consiste en capas de nodos interconectados que procesan y transmiten información.

Aprendizaje Profundo (Deep Learning) Definición: Una rama del aprendizaje automático que utiliza redes neuronales con múltiples capas para aprender patrones complejos en grandes cantidades de datos.

Sistema de Recomendación (Recomendadores) Definición: Un tipo de sistema de IA que sugiere productos, servicios o contenido a los usuarios basados en sus preferencias y comportamientos anteriores.

Procesamiento del Lenguaje Natural (Natural Language Processing - NLP)

Definición: La capacidad de una máquina para comprender, analizar e interpretar el lenguaje humano en forma escrita o hablada.

Visión por Computadora (Computer Vision) Definición: La habilidad de una máquina para interpretar y analizar imágenes y videos, permitiéndole "ver" y comprender su contenido.

Chatbot Definición: Un programa de computadora diseñado para interactuar con humanos a través de conversaciones en lenguaje natural, generalmente a través de texto.

Prompt Definición: Un prompt se refiere a una indicación, solicitud o estímulo específico que se proporciona a un programa de computadora o a un sistema, con el fin de obtener una respuesta o realizar una acción determinada, no es más que una instrucción dada para iniciar una interacción o proceso.

Automatización Robótica de Procesos (Robotic Process Automation - RPA)

Definición: La automatización de tareas y procesos empresariales mediante software o "bots" que imitan las acciones humanas.

Supervisado vs. No Supervisado (Supervised vs. Unsupervised)

Definición: En el aprendizaje automático supervisado, los algoritmos se entrenan

con datos etiquetados, mientras que, en el no supervisado, se utilizan datos sin etiquetar.

Ética de la IA Definición: El conjunto de principios y estándares que guían el desarrollo y uso de la IA para asegurar que sea ética, justa y beneficiosa para la sociedad.

Sesgo en la IA Definición: La tendencia de los sistemas de IA a tomar decisiones o hacer predicciones injustas o desiguales debido a la influencia de datos de origen sesgados o desequilibrados.

Interfaz de Programación de Aplicaciones (Application Programming Interface - API) Definición: Un conjunto de reglas y protocolos que permite a diferente software comunicarse entre sí.

Ciberseguridad Definición: La práctica de proteger sistemas, redes y datos de ataques, daños o accesos no autorizado.

Cuarta Revolución Industrial Definición: La Cuarta Revolución Industrial es a partir de la integración de tecnologías emergentes y avanzadas en los procesos de producción y gestión de las empresas, siendo estas nuevas tecnologías el ingreso de la IA, la internet de las cosas (IoT), la impresión 3D, la robótica avanzada, el big data, la realidad aumentada, la analítica de datos, entre otras nuevas formas de relación humanidad-máquina-mercado.

Internet de las Cosas (IOT) Definición: El Internet de las cosas (IoT), o Internet of Things en inglés, se refiere a la interconexión digital de dispositivos, objetos y sistemas a través de la red de internet. Estos dispositivos están equipados con sensores, actuadores y tecnología de comunicación que les permiten recopilar y transmitir datos, así como interactuar entre sí y con el entorno.

Alucinación IA Definición: La "alucinación" de la IA se refiere a la capacidad de un modelo de lenguaje o red neuronal generativa para crear contenido que parece auténtico y coherente, pero que en realidad es completamente inventado por la máquina y no tiene base en datos reales. Esto puede incluir textos, imágenes, música, entre otros.

Falso Profundo Definición: Un "deepfake" es una técnica de manipulación digital que utiliza IA y aprendizaje automático para crear contenido multimedia falso, como videos, imágenes o audio, que parece auténtico, pero en realidad es generado de manera sintética, todos estos contenidos pueden hacer que una persona aparezca o suene como si estuviera diciendo o haciendo algo que nunca ocurrió en realidad.

Phishing Definición: El "phishing" se refiere a una técnica de fraude en línea que consiste en engañar a una persona para obtener información confidencial, como

contraseñas, números de tarjeta de crédito o información personal. Los estafadores utilizan mensajes de correo electrónico, mensajes de texto o páginas web falsificadas que parecen ser de fuentes legítimas, como bancos, empresas o servicios en línea, para llevar a las víctimas a proporcionar sus datos.

Referencias y Recursos Adicionales

**Libros, artículos y recursos en línea
recomendados para aprender más sobre
IA y sitios web y plataformas para
practicar y experimentar con IA de manera
interactiva.**

Libros

Deep Learning de Ian Goodfellow, Yoshua Bengio, y Aaron Courville.

Este libro es una referencia clave para entender el aprendizaje profundo, con un enfoque en las redes neuronales.

Python Machine Learning de Sebastian Raschka y Vahid Mirjalili.

Ideal para aprender a aplicar algoritmos de aprendizaje automático utilizando Python.

Artificial Intelligence: A Modern Approach de Stuart Russell y Peter Norvig.

Un libro introductorio ampliamente utilizado que cubre una variedad de temas en IA.

Una introducción a la IA: de Stuart Russell y Peter Norvig es un clásico de la IA que proporciona una introducción completa al campo.

El libro de jugadas de la IA: de Melanie Mitchell ofrece una visión general de la IA y sus aplicaciones en diferentes campos.

The Hundred-Page Machine Learning Book: de Andriy Burkov es una introducción compacta al aprendizaje automático.

IA para todos: Este libro de Pedro Domingos es una introducción a la IA que se centra en los algoritmos y las aplicaciones prácticas.

El futuro de la IA Este libro de Nick Bostrom explora los posibles impactos de la IA en la sociedad.

IA y ética Este libro de John Danaher explora los desafíos éticos que plantea la IA.

El poder de la IA Este libro de Kai-Fu Lee explora el impacto de la IA en la economía y la sociedad.

El mundo de la IA Este libro de Gary Marcus explora los diferentes tipos de IA y sus aplicaciones.

La IA explicada Este libro de Nick Bostrom y Eliezer Yudkowsky explica los fundamentos de la IA de una manera accesible.

La IA a la vista Este libro de Stuart Russell y Peter Norvig explora las aplicaciones de la IA en diferentes campos.

El futuro de la IA Este libro de Max Tegmark explora los posibles impactos de la IA en el universo.

Cursos en línea

Aprende IA con Python Este curso en línea de Coursera te enseña los fundamentos de la IA utilizando el lenguaje de programación Python.

IA para la ciencia Este curso en línea de Coursera te enseña a utilizar la IA para resolver problemas científicos.

Introducción a la IA Este curso en línea de edX te enseña los fundamentos de la IA utilizando la plataforma de aprendizaje automático de Google, TensorFlow.

IA para el arte Este curso en línea de edX te enseña a utilizar la IA para crear arte.

IA para la vida cotidiana Este curso en línea de Audacity te enseña a aplicar la IA a problemas del mundo real.

IA para principiantes Este curso en línea de Udemy te enseña los fundamentos de la IA utilizando el lenguaje de programación Python.

IA para el entretenimiento Este curso en línea de Udemy te enseña a utilizar la IA para crear juegos y otras formas de entretenimiento

Plataformas de Práctica y Experimentación con IA

TensorFlow Playground (playground.tensorflow.org)

Una plataforma interactiva que permite experimentar con redes neuronales utilizando TensorFlow.

Kaggle (kaggle.com)

Ofrece competiciones de aprendizaje automático, conjuntos de datos y kernels (notebooks) para aprender y practicar.

Google Colab (colab.research.google.com)

Una plataforma en la nube basada en Jupyter Notebooks que permite ejecutar código de Python y experimentar con IA sin necesidad de configurar un entorno local.

Artículos y Blogs:

Towards Data Science (towardsdatascience.com)

Una publicación en Medium con numerosos artículos sobre aprendizaje automático y IA.

OpenAI Blog (openai.com/blog)

Ofrece información sobre los avances en IA y temas de interés en el campo.

Inteligencia Artificial en español(<https://www.inteligencia-artificial.es/>)

Este blog se enfoca en noticias y desarrollos relacionados con la IA en el mundo hispanohablante.

Tecnología e Inteligencia Artificial

(<https://conecta.tec.mx/es?keys=inteligencia%20artificial>)

El Tecnológico de Monterrey ofrece una sección de noticias sobre IA en su sitio web.

AI Time Journal (<https://www.aitimejournal.com/>)

Aunque no está completamente en español, este sitio proporciona contenido sobre IA y tecnología en múltiples idiomas, incluido el español

Comunidades en línea

Reddit – r/MachineLearning (reddit.com/r/MachineLearning)

Una comunidad activa donde se discuten temas de aprendizaje automático e IA.

Stack Overflow (stackoverflow.com)

Un sitio donde puedes hacer preguntas y encontrar respuestas relacionadas con la programación en IA y aprendizaje automático.

Un adelanto de nuestro próximo libro

**Cuántas IA existen, en profundidad la IA
Generativa y la IA Explicativa**

IA Generativa IAGen

La IA generativa (IAGen) es un campo de la IA que se centra en el desarrollo de sistemas que pueden generar nuevos datos o contenidos, la IAGen tiene como objetivo automatizar tareas tradicionalmente realizadas por humanos, como la creación de arte, música, texto y código de programación, por ejemplo.

Su Concepto se basa en varias características claves, que incluyen: la generación como proceso de creación de nuevos datos o contenidos, el aprendizaje automático, que se centra en el desarrollo de sistemas que pueden aprender de los datos sin estar programados explícitamente, en redes neuronales, que son un tipo de algoritmo de aprendizaje automático inspirado en la forma en que funciona el cerebro humano.

La IAGen se puede utilizar para personalizar el contenido para cada usuario, por ejemplo, descubrir nuevos patrones y tendencias en los datos, sus desafíos de control y mejora contemplan autenticidad, ya que la IAGen puede crear contenido falso o engañoso inclinando la balanza de forma poco ética, se debe asegurar la integridad, ya que la IAGen puede usarse para crear contenido dañino o disruptivo.

La AIGen es un campo emergente que tiene el potencial de automatizar tareas tradicionalmente realizadas por humanos, plantea una serie de desafíos, pero también tiene el potencial de resolver algunos de los mayores problemas de la IA, como la falta de creatividad e innovación.

Referencias

- Goodfellow, I., et al. (2016). Generative adversarial networks. arXiv preprint arXiv:1406.2661.
- Radford, A., et al. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
- Isola, P., et al. (2017). Image-to-image translation with conditional adversarial networks. arXiv preprint arXiv:1701.07875.

Finalmente, además de los conceptos y desafíos mencionados anteriormente, la IAGen también presenta desafíos éticos, pudiendo usarse para crear contenido disruptivo o dañino, como violencia gráfica o discurso de odio, por lo cual es crítico considerar estos desafíos éticos al desarrollar y utilizar el sistema IA.

IA Explicativa IAX

Una buena definición para la IA Explicativa (IAX) es que es un campo de la IA que se centra en el desarrollo de sistemas de IA que puedan explicar o interpretar sus decisiones de una manera que los humanos puedan entender, con la IAX podemos aumentar la transparencia y la confianza en los sistemas de IA, lo que puede ayudar a evitar sesgos y errores.

Como concepto la IAX se basa en diversas claves como, por ejemplo, la interpretación dentro del proceso de proporcionar información sobre cómo tomar una decisión, en este contexto la IAX las explicaciones e interpretaciones pueden ser verbales, visuales o digitales.

Aumentamos así la transparencia, logrando un grado de comprensión mayor de cómo funcionan los sistemas IA, pudiendo ser garantizada mediante documentación, código abierto o explicaciones.

Por otro lado, tenemos la confianza, que en contexto IAX es la creencia de que un sistema es capaz de realizar sus tareas de forma correcta y segura, y es en este contexto donde la podemos establecer, mediante la transparencia, la explicación y la demostración del desempeño.

La IAX ofrece muchos casos de uso, que incluyen, sistema de sugerencias como Netflix y Amazon para recomendar productos y servicios a los usuarios, la IAX puede ayudar a explicar por qué un sistema de recomendación recomienda un determinado producto o servicio, lo que puede ayudar a los usuarios a tomar decisiones informadas.

En sistemas de diagnóstico, donde los médicos utilizan IAX como sistemas para diagnosticar enfermedades, y allí es donde la IAX puede ayudar a explicar cómo el sistema

de diagnóstico llegó a un diagnóstico determinado, lo que puede ayudar a los médicos a tomar decisiones precisas.

Para las organizaciones el utilizar IAX en los sistemas de toma de decisiones para tomar decisiones importantes, es un aporte de esta nueva tecnología que permite explicar cómo un sistema de toma de decisiones llegó a una determinada decisión, lo que puede ayudar a las organizaciones a tomar decisiones informadas y mucho más certeras.

Se establecen junto con la IAX una serie de desafíos, que incluyen, desafíos técnicos para comprender cómo funciona el sistema de IA, cual es el valor y cuan costoso resulta el desarrollo e implementación de estas nuevas técnicas y herramientas.

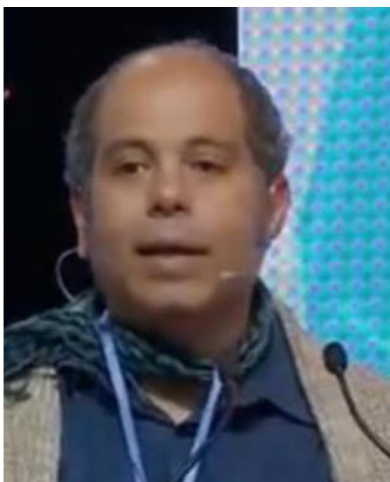
Concluimos que la IAX es un campo emergente que tiene el potencial de aumentar la transparencia y la confianza en los sistemas de IA, también la IAX plantea una serie de desafíos, y tiene el potencial de resolver algunos de los mayores problemas de la IA, como el sesgo y los errores.

Referencias

- Gunning, D., et al. (2022). Explanatory AI: A survey. arXiv preprint arXiv:2201.00175.
- Rudin, C. (2019). The dangers of Stochastic Parrots: Can machine learning be too good to be true? arXiv preprint arXiv:1907.11692.
- Samek, W., et al. (2017). Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges. arXiv preprint arXiv:1708.07228.



**¡Buena suerte en tu viaje
de aprendizaje en IA!**



Hugo Gutierrez Tarifeño, Tarik para los amigos y conocidos, nació en 1974, hijo de Juanita y Aldo, dio sus primeros pasos en la casa de sus abuelos maternos, Berta y Fernando en la ciudad de Viña del Mar en Chile, fue durante su infancia, que se interesó en variadas temáticas de los más diversos orígenes y sectores, leyendo y disfrutando de la playa y el campo, practicando deportes y escribiendo.

Luego en su juventud fue llevado por caminos de distintas culturas, religiones, trabajos y personas de los más diversos orígenes sociales y económicos, pasando por occidente y oriente en ese conocimiento, lo que le valió tener una mente abierta y plena de dudas, haciéndose padre en el inicio de su juventud.

Hoy día ad portas de los 50 años de vida, junto a Lorena, sus hijos, padres y nieto, tíos y primos y amigos que constituyen la base de una carrera plagada de equilibrio, lealtad y compromiso por un país y planeta más educado, justo, abierto, y sostenible, entrega este primer texto a modo de aporte para el nuevo mundo.

