CONTRIBUCIÓN CIENTÍFICA

EL FACTOR HUMANO EN LA INTELIGENCIA ARTIFICIAL

Recomendaciones Estratégicas para la Sostenibilidad

ORDEM DOS PSICÓLOGOS

Aportación Científica OPP – El Factor Humano en la Inteligencia Artificial – Recomendaciones estratégicas para la sostenibilidad, publicado por la Ordem dos Psicólogos Portugueses (Colegio de Psicólogos de Portugal).

Sugerencia de cita

Ordem dos Psicólogos Portugueses (2023). Contributo Científico OPP – O Factor Humano na Inteligência Artificial – Recomendações Estratégicas para a Sustentabilidade. Lisboa.

Para más aclaraciones póngase en contacto con Ciência e Prática Psicológicas

andresa.oliveira@ordemdospsicologos.pt

ISBN
978-989-53170-7-3

Ordem dos Psicólogos Portugueses
Av. Fontes Pereira de Melo 19 D
1050-116 — Lisboa

+351 213 400 250

www.ordemdospsicologos.pt

EL FACTOR HUMANO EN LA INTELIGENCIA ARTIFICIAL
RECOMENDACIONES ESTRATÉGICAS PARA LA SOSTENIBILIDAD

El presente documento surge como una aportación de la Ordem dos Psicólogos Portugueses (OPP) al debate en relación con la Inteligencia Artificial, la definición de los principales conceptos asociados a esta, las evidencias científicas que apoyan el área, sus ventajas, riesgos y aplicabilidades.

La OPP es una asociación pública profesional que representa y regula la práctica de los profesionales de la psicología que ejercen la profesión de psicólogo en Portugal (en virtud de la ley nº 57/2008, de 4 de septiembre, con las modificaciones de la ley nº 138/2015, de 7 de septiembre). La misión de la OPP es controlar el ejercicio y acceso a la profesión de Psicólogo, así como elaborar las respectivas normas técnicas y deontológicas y ejercer el poder disciplinario sobre sus miembros. Las atribuciones de la OPP incluyen igualmente defender los intereses generales de la profesión y de los usuarios de los servicios de psicología; prestar servicios a los miembros en relación con la información y formación profesional; colaborar con las restantes entidades de la administración pública para perseguir fines de interés público relacionados con la profesión; participar en la elaboración de la legislación referentes a la profesión y en los procesos oficiales de acreditación y en la evaluación de las carreras que dan acceso a la profesión.

En este sentido, la OPP considera pertinente contribuir al debate sobre el desarrollo de sistemas de Inteligencia Artificial (IA), considerando que reconocer sus singularidades, en lo referente a las ventajas y desafíos, es fundamental para garantizar que estas tecnologías son seguras, transparentes y funcionan en pro del desarrollo y bienestar de las personas y de las comunidades. El uso de IA se da, actualmente, en diferentes contextos, entre ellos el sanitario, el laboral y el judicial, orientando las tomas de decisión que tienen implicaciones prácticas en la vida de los ciudadanos.

Consideramos que la cooperación multidisciplinaria, en la que se incluyen las aportaciones de los especialistas en el comportamiento humano, es una estrategia necesaria para regular el desarrollo, implementación y uso de las tecnologías de IA.

“La OPP considera pertinente contribuir al debate sobre el desarrollo de sistemas de Inteligencia Artificial (IA), considerando que reconocer sus singularidades, en lo referente a las ventajas y desafíos, es fundamental para garantizar que estas tecnologías son seguras, transparentes y funcionan en pro del desarrollo y bienestar de las personas y de las comunidades.”

INTRODUCCIÓN

El séptimo arte ha poblado el imaginario colectivo de ideas de lo que sería una inteligencia diferente de la humana. Las películas 2001: Odisea en el Espacio (Kubrick, 1968), Her (Jonze & Johnson, 2013), Ex Machina (Garland & Macdonald, 2014) y Automata (Ibañez, 2014) son solo algunos ejemplos que han contribuido a una noción de IA como una máquina, un chatbot o un robot humanoide que consigue resolver problemas lógicos, tomar decisiones con base en múltiples fuentes de información e, incluso, influir sobre el comportamiento humano.

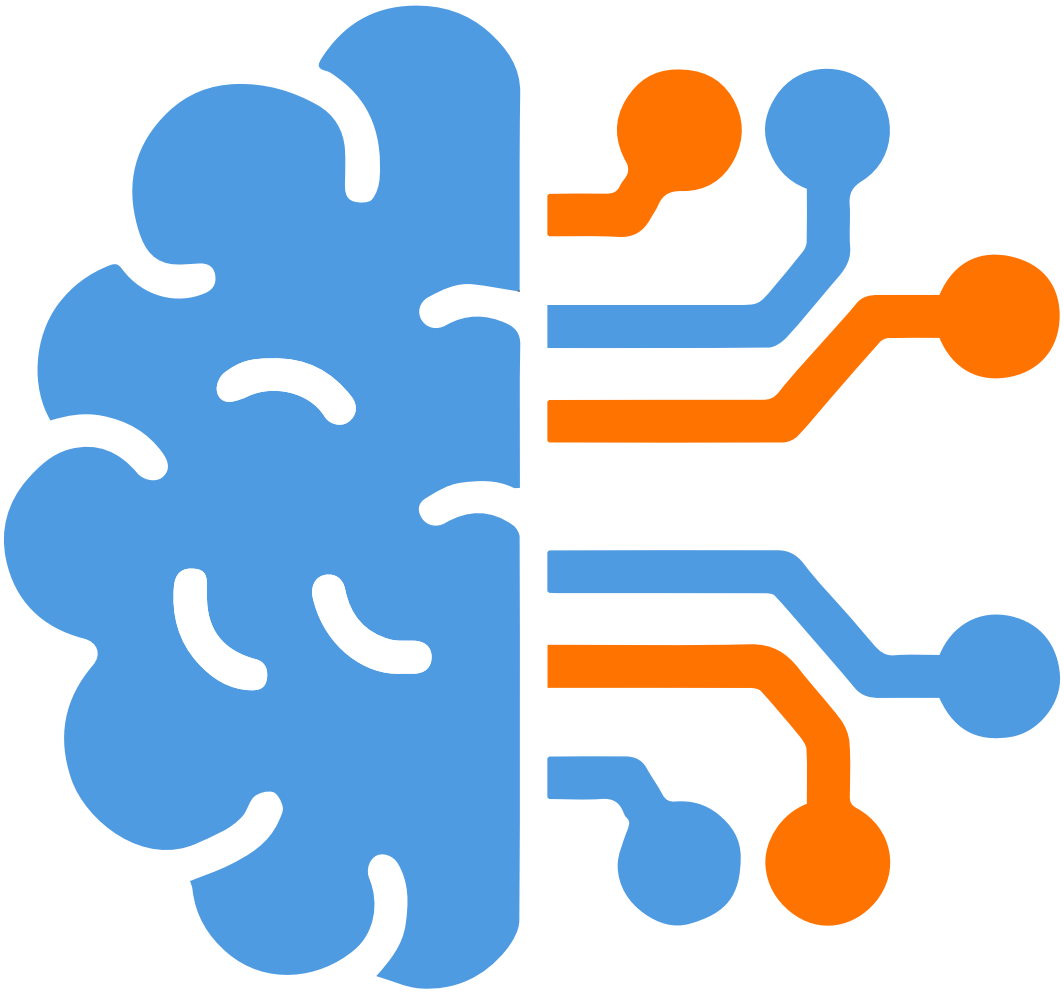
Actualmente, aunque las obras cinematográficas anteriormente referidas presenten realidades distópicas y que los impactos reales a largo plazo de IA sean desconocidos, se reconoce que las **tecnologías con la IA tienen potencial para:**

Resolver problemas humanos que persisten, por ejemplo, la crisis climática, las desigualdades estructurales o problemas de salud como enfermedades oncológicas.

En contrapartida, sin regulación y sin aportaciones técnicas y éticas que orienten su desarrollo, existe la posibilidad de que la IA influya sobre varios ámbitos de la vida humana de forma nefasta e imprevisible.

El desarrollo de sistemas de IA se sitúa en la intersección entre varias áreas científicas, en las que se incluyen la Computación, la Biología, las Matemáticas, la Medicina, la Filosofía e, incluso, la Ciencia Psicológica (Oliveira, 2019). La creación de soluciones de IA justas, seguras e innovadoras implicará el trabajo multidisciplinar entre ingenieros, científicos de la computación, profesionales sanitarios, profesores, educadores, juristas, filósofos, investigadores, decisores, psicólogos, entre otros.

1. 1. INTELIGENCIA ARTIFICIAL: ¿QUÉ ES?



Genéricamente, hablamos de **Inteligencia Artificial (IA)** cuando **un ordenador/máquina logra realizar tareas que imitan, o superan, a la inteligencia humana**. Algunas de las tareas desempeñadas por la IA que involucran diferentes formas de comportamiento inteligente son interpretar el lenguaje, reconocer patrones o procesos de toma de decisión (Parlamento Europeo, 2022). La simulación de la inteligencia humana, en sistemas de IA, incluye **algoritmos**, que se pueden definir como **procesos de cálculo que, por medio de un conjunto de reglas preprogramadas, operacionalizan datos, traduciéndolos en un resultado que responde a la tarea o problema presentado** (Vicente, 2023).

En este sentido, a través de un conjunto de diversos algoritmos y modelos, **ordenadores y máquinas pueden simular procesos cognitivos humanos** –por ejemplo, la percepción y el procesamiento de lenguaje– y, así, conseguir reconocer caras o movimientos en sistemas de videovigilancia o, por ejemplo, identificar el estado emocional de una persona teniendo como base el lenguaje que utiliza (ej. Agbavor & Liang, 2022).

EL TÉRMINO IA TAMBIÉN PUEDE SER CONSIDERADO UN PARAGUAS QUE ABARCA **CUATRO DOMINIOS DISTINTOS** (MALONE ET AL., 2020; RUSSELL & NORVIG, 2021):

APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

01

¿Y si un algoritmo pudiese adaptarse a los datos, reprogramarse a sí mismo y obtener resultados más eficientes? En el marco de la IA, esto es posible por medio del **Aprendizaje Automático** (del inglés, Machine Learning). El trabajo de los programadores es desarrollar algoritmos que, detalladamente y por medio de programación de software, dirigen al ordenador para que haga exactamente aquello que se pretende que haga. A través del Aprendizaje Automático, los programadores no necesitan desarrollar programas para todos los problemas específicos. Por medio de programas generales, con instrucciones que permitan a los ordenadores aprender de sus propias experiencias y del análisis de diferentes datos, estos pueden automatizar sus funciones y producir nuevos resultados por sí mismos.

Además, como subdominio del Aprendizaje Automático, el **Aprendizaje Profundo** (del inglés, Deep Learning), se presenta como la posibilidad de que un sistema artificial aprenda con enormes cantidades de datos, pudiendo identificar patrones imperceptibles al ser humano. El Aprendizaje Profundo sucede por medio de redes neuronales artificiales que simulan la actividad de las neuronas del sistema nervioso central.

ROBÓTICA

02

La robótica es el dominio que **aúna sistemas de IA con interfaces físicas** que permiten a cualquier máquina realizar, de forma independiente, una actividad que antes era realizada por un ser humano, o que, por limitaciones del cuerpo humano, sería imposible de realizar (ej. explorar el interior de un volcán activo). En general, la robótica persigue la automatización de tareas físicas arduas, demoradas, repetitivas y/o peligrosas. El uso de algoritmos que permiten la percepción, planificación, aprendizaje y control, así como la necesidad de un diseño adaptado a las funciones, son necesarios para que estas tecnologías puedan moverse, tomar decisiones y manipular el medio en pro de determinados objetivos.

VISIÓN COMPUTACIONAL

03

Este dominio engloba **sistemas de IA que se dedican a capturar, analizar e interpretar información visual de diferentes fuentes y objetos**, en los que se incluyen diversos elementos del mundo natural, como, por ejemplo, caras humanas, especies de fauna y flora, imágenes de videovigilancia o imágenes de satélite referentes a movimientos, por ejemplo, de buques o tormentas. Estos sistemas no solo consiguen identificar objetos o fenómenos, sino que también consiguen extraer información, clasificarla en múltiples categorías, comprender el contexto en el que se da la información visual y, además, prever acontecimientos –lo que implica algoritmos capaces de establecer múltiples asociaciones y deducir información de naturaleza diferente.

PROCESAMIENTO DEL LENGUAJE NATURAL (PLN)

04

Al ámbito de la IA que se dedica a la simulación del lenguaje natural, por ejemplo, del lenguaje humano, se le llama Procesamiento del Lenguaje Natural (PLN). En lo referente al lenguaje humano, estos sistemas de IA son capaces de **comprender e identificar diferentes lenguas y formas de comunicación**, habladas y escritas, consiguiendo analizar múltiples y amplias fuentes de información, crear productos escritos y responder, con alto grado de precisión, a diferentes preguntas. Los sistemas de PLN más recientes incluyen diferentes sistemas estadísticos, entre otros, de Aprendizaje Automático y Profundo, y funcionan con una lógica predictiva y de aprendizaje por refuerzo (tal como el sistema cognitivo humano).

TAMBIÉN PODEMOS DIVIDIR LA IA EN **TRES NIVELES** SEGÚN SU NIVEL DE DESARROLLO (GOERTZEL, 2014; RUSSELL & NORVIG, 2021):

INTELIGENCIA ARTIFICIAL RESTRINGIDA *(ARTIFICIAL NARROW INTELLIGENCE)*

TIENE UN CONJUNTO DE COMPETENCIAS LIMITADO.

La Inteligencia Artificial Restringida solo consigue realizar **tareas sencillas**, a veces una única tarea, como jugar al ajedrez, reconocer voces o imágenes. La mayoría de los sistemas de Inteligencia Artificial se encuentra en este nivel (Malone et al., 2020). Este tipo de inteligencia artificial, también conocido como Narrow AI, se restringe a datos específicos y opera de una forma predeterminada y predefinida. A modo de ejemplo, Siri es un sistema de inteligencia artificial restringido porque ha sido entrenada para procesar el lenguaje humano, introduciendo los pedidos de los usuarios en un motor de búsqueda y presentando los resultados (Goertzel, 2014).

INTELIGENCIA ARTIFICIAL GENERAL *(ARTIFICIAL GENERAL INTELLIGENCE)*

TIENE CAPACIDADES SEMEJANTES A LAS CAPACIDADES HUMANAS.

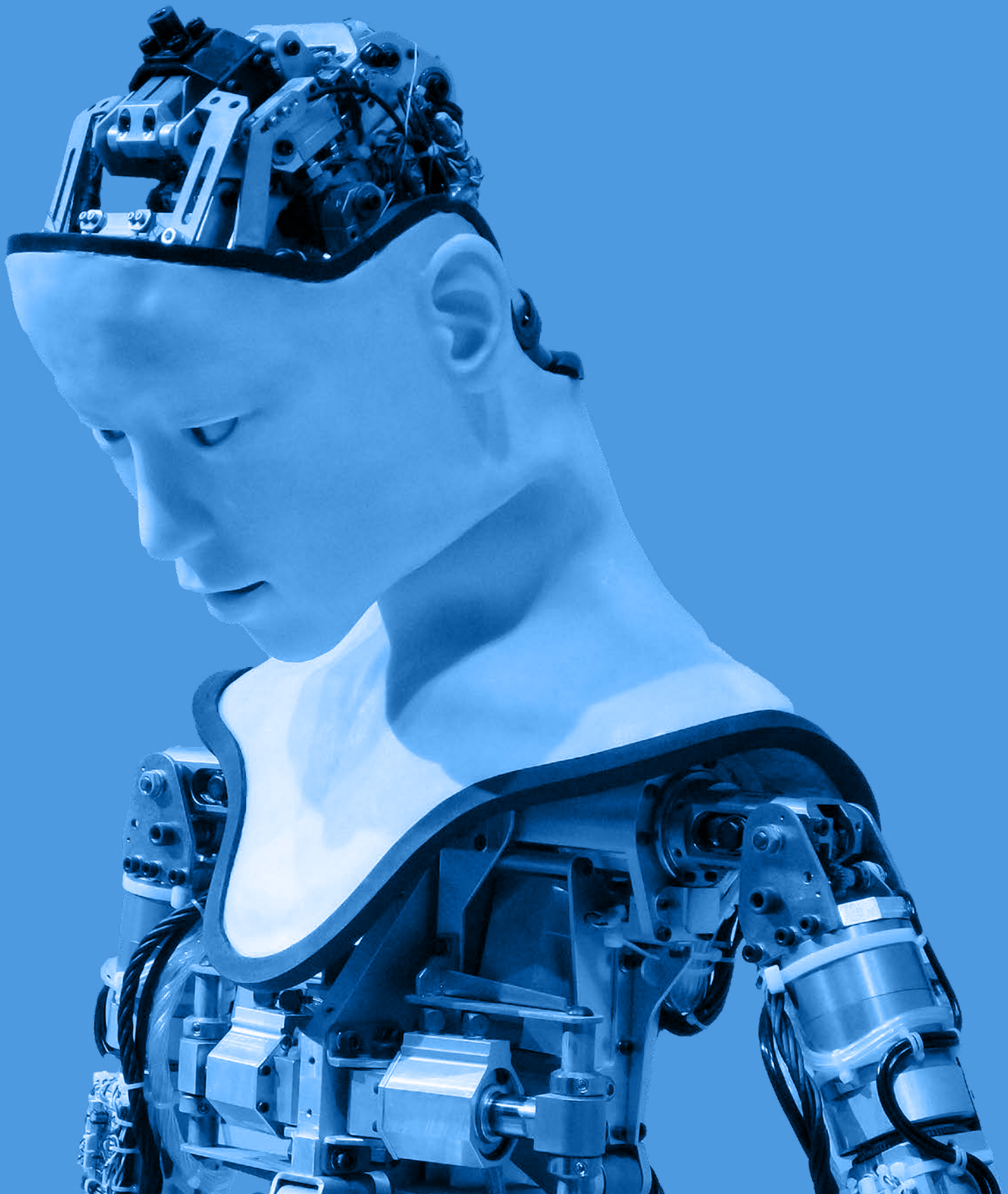
La Inteligencia Artificial General puede exhibir un nivel de **inteligencia semejante a la inteligencia humana**, siendo capaz de desempeñar cualquier tarea intelectual que un ser humano también consiga desempeñar. Un sistema de inteligencia artificial general es capaz de comprender, aprender, adaptarse e implementar información en diferentes áreas del conocimiento. Su competencia no se limita a un campo específico, y consigue realizar desafíos y tareas de naturaleza diferente. Pese a que este nivel de inteligencia no se haya alcanzado totalmente, en 2020, la OpenAI rompió algunas barreras importantes y presentó un modelo de lenguaje natural capaz de desempeñar tareas para las que no fue explícitamente entrenado –el Generative Pre-trained Transformer (chat GPT) (Russell & Norvig, 2021; Malone et al., 2020). Encontramos un ejemplo sorprendente de la AGI en un artículo de Kosinski (2023) en el que large language models, como ChatGPT-4, parecen haber sido capaces de desarrollar o, al menos, replicar la Teoría de la Mente –capacidad de identificar estados mentales no observables.

SUPERINTELIGENCIA ARTIFICIAL *(ARTIFICIAL SUPERINTELLIGENCE)*

SUPERA LA INTELIGENCIA HUMANA.

El último nivel de desarrollo de la Inteligencia Artificial es la categoría de la Superinteligencia Artificial. Esta categoría presupone que el intelecto del sistema **supera al intelecto humano** en las dimensiones de la toma de decisiones, resolución de problemas, creatividad, entre otros (Russell & Norvig, 2021).

2. DESAFÍOS, BENEFICIOS Y RIESGOS DE LA INTELIGENCIA ARTIFICIAL



DESAFÍOS DE LA INTELIGENCIA ARTIFICIAL

Una vez explorada la definición de IA, sus principales áreas, y sus diferentes niveles de desarrollo, es posible empezar a reflexionar sobre los **desafíos que estas tecnologías pueden suponer**. Los avances en el área de la IA describen un progreso de ordenadores y máquinas que realizan tareas específicas y preprogramadas para nuevas interfaces que realizan tareas más amplias, de forma más autónoma y con mayor potencial de analizar una enorme cantidad de datos que se producen diariamente en diferentes formatos (ej. imágenes, historial de búsquedas y publicaciones en redes sociales, historiales de compras, acceso a servicios de salud, finanzas, etc.), en la utilización de diferentes tecnologías y servicios de su día a día. **Esta producción masiva y continua de datos, pasible de ser analizada por sistemas de IA, es denominada Big Data** –concepto que no es consensual entre los diferentes campos de estudio (Favaretto et al., 2020).

BIG DATA

Trabajar con **Big Data**, es decir, con la recogida, procesamiento y análisis de grandes cantidades de información abrió nuevas posibilidades, pero también **aumentó las dificultades para desarrollar sistemas de IA objetivos y neutros, libres de sesgos, y que tomen decisiones justas** (Vicente, 2023).

SESGOS

Uno de los mayores desafíos en el desarrollo de la IA tiene que ver con los **sesgos en la IA**. Los sistemas de IA recurren a algoritmos construidos por **humanos –que tienen valores, creencias y conocimientos limitados–**, y en cualquier momento del proceso de desarrollo de los algoritmos, tanto durante la recogida de datos como del diseño del propio algoritmo, los humanos pueden transmitir sus preferencias, prejuicios y sesgos, resultando en **sistemas de IA que perpetúan y acentúan las desigualdades existentes** (Pedersen & Johansen, 2020).

CAJA NEGRA

Otro desafío importante para el desarrollo de IA es el fenómeno «**caja negra**» (del inglés, black-box). **Este fenómeno remite a la dificultad de comprender de qué forma llegan los algoritmos a sus decisiones**. Al inspeccionar el modelo es posible ver lo que entra en el sistema (datos recogidos) y lo que sale (previsiones o decisiones), **pero la forma en la que el sistema llega estas conclusiones no es totalmente clara. La opacidad de los datos es problemática porque los modelos complejos no toman decisiones con base en las reglas definidas por los humanos, en lugar de eso, actualizan y reprograman los propios algoritmos**, haciendo este proceso cada vez más difícil de **descifrar** (Rahwan et al., 2019).

TRANSPARENCIA DE LOS DATOS

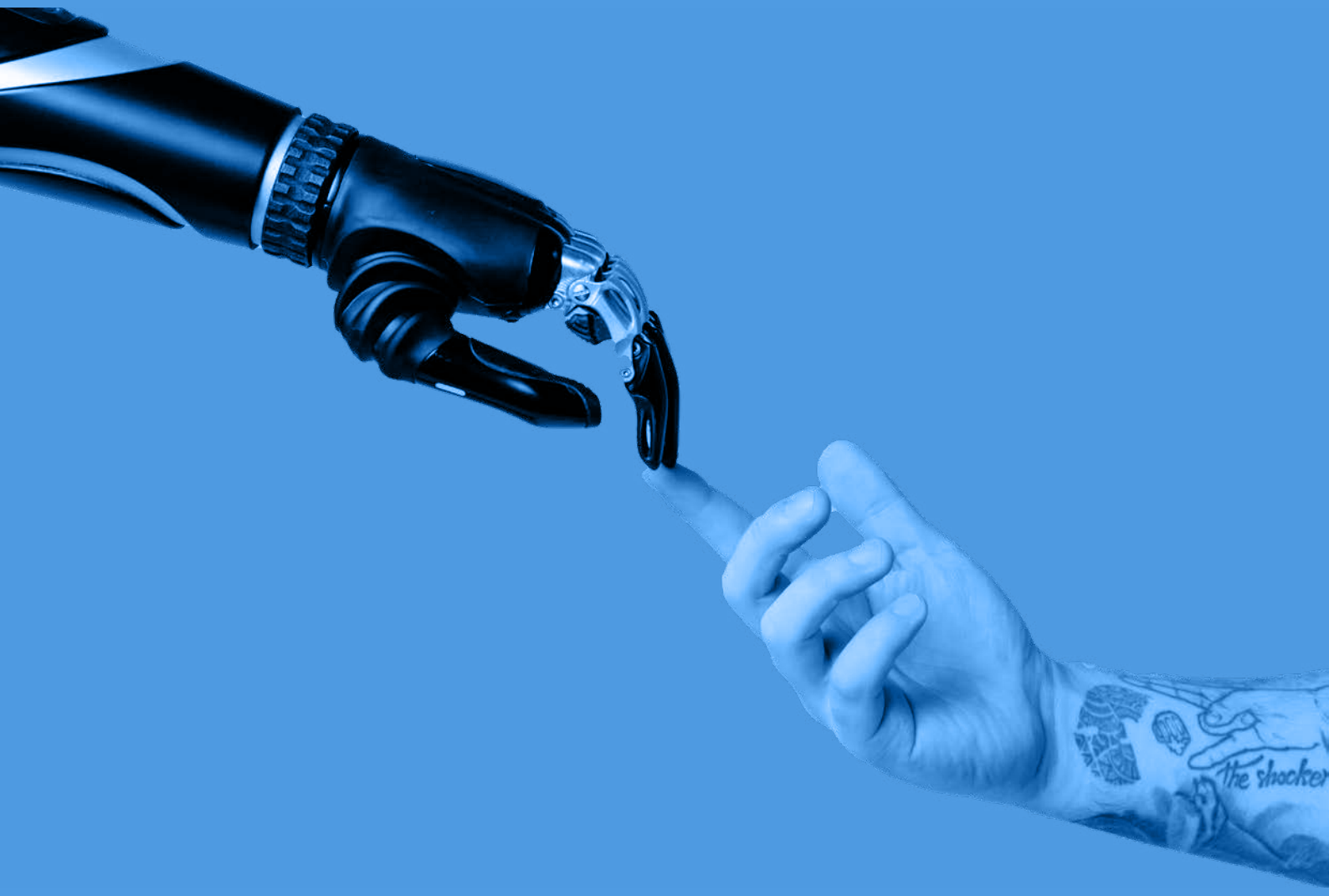
Lidiar con la **falta de transparencia de los datos** es un desafío urgente puesto que, de una forma algo imprevisible y sin monitorización humana, estos sistemas de IA **pueden condicionar masivamente la vida de las personas y de las comunidades**. Es fácil constatar cómo la IA influencia, en diferentes grados, **la vida de las personas** cuando comparamos, por ejemplo, las implicaciones de una IA que sugiere canciones a un usuario de una plataforma de **streaming** con las implicaciones de **una IA que determina si una determinada familia tiene derecho a un apoyo social**. Cuanto mayores las capacidades de los sistemas de IA, mayores las preocupaciones con la opacidad de su funcionamiento.

La utilización de sistemas inteligentes en diferentes ámbitos de la vida pública, por ejemplo, en la salud, en el trabajo, en la seguridad, en el ocio y, a nivel macro, en la economía, condiciona, hoy en día, los comportamientos de las personas y las decisiones estratégicas de organizaciones y gobiernos –algo que, con la continua automatización de la IA, **puede resultar en consecuencias imprevisibles para la sociedad** (Vicente, 2023).

Por estos motivos, inmensos investigadores han instado a una **moratoria en el desarrollo de la IA** y abogado por una planificación y gestión cuidadosa en su desarrollo (Russell & Norvig, 2021; Future of Life, 2023; Yudkowsky, 2023).

Es fácil comprender que la IA, siendo un arma poderosa, tiene **incontables beneficios, pero también riesgos**, asociados. La tensión entre las ventajas e inconvenientes de la IA se siente en instituciones internacionales como el Foro Económico Mundial y la Comisión Europea. El Foro Económico Mundial (2023a) pone esta tecnología en **el top 10 de tecnologías emergentes** y, simultáneamente, la reporta como una **amenaza futura** en un informe sobre riesgos globales (Foro Económico Mundial, 2023b). En el marco europeo, la Comisión Europea se muestra comprometida con el desarrollo de la IA, y en lo que puede aportar al progreso europeo, a la vez que muestra aprehensión por la forma en la que la utilización de determinados algoritmos puede llevar a la violación de derechos humanos fundamentales (Consejo de Europa, 2019).

Antes de analizar los beneficios y riesgos de la IA, cabe destacar que, aunque sea una área emergente y prometedora, también es un **ámbito que debe ser objetivo de regulación**. Precisamente una señal de eso es el debate que aún se celebra en el Parlamento Europeo a raíz de la **Ley de la Unión Europea sobre Inteligencia Artificial** (Parlamento Europeo, 2023a; Parlamento Europeo, 2023b). De modo a apalancar los beneficios de la IA y mitigar sus peligros, es determinante reflexionar sobre sus riesgos más significativos y establecer políticas y estrategias que regulen este ámbito (Acemoglu, 2021).



BENEFICIOS DE LA INTELIGENCIA ARTIFICIAL

La IA, en todas sus formas, puede aportar beneficios a los ciudadanos, a las organizaciones y a la sociedad en general (Parlamento Europeo, 2023c). De acuerdo con la investigación de Vinuesa y Colaboradores (2020), la **Inteligencia Artificial puede tener un impacto positivo en el 79 % de los Objetivos de Desarrollo Sostenible**.

BENEFICIOS PARA LOS CIUDADANOS

La IA puede tener un impacto en diferentes áreas de la vida de los ciudadanos (Parlamento Europeo, 2023c):

SALUD

La implementación de sistemas de IA podrá aliviar la sobrecarga sobre los sistemas sanitarios, de acuerdo con un informe de Accenture (2018) para el contexto norteamericano, satisfaciendo el 20 % de las necesidades no satisfechas de cuidados sanitarios. Algunos desarrollos recientes, como sistemas de IA aplicados a diagnósticos clínicos que parecen alcanzar una precisión diagnóstica similar a expertos en la materia, muestran también un camino para la IA en el área de los diagnósticos preliminares (De Fauw et al., 2018; Liu et al., 2019).

EDUCACIÓN

Facilitar el acceso a la información, educación y formación

MERCADO DE TRABAJO

Trabajos considerados más peligrosos pueden ser asumidos por máquinas de IA. Las nuevas tecnologías dan origen a nuevos puestos de trabajo en diversas áreas de la economía.

BENEFICIOS PARA ORGANIZACIONES

Las aportaciones de la IA se pueden traducir en mayor productividad, sostenibilidad y creatividad en las organizaciones por medio de (Parlamento Europeo, 2023c):

INNOVACIÓN

Los algoritmos ayudan a identificar nuevas áreas y a crear nuevos productos y servicios en diversas áreas.

ATENCIÓN AL CLIENTE

Pueden desarrollar un conocimiento profundo de los clientes y de sus necesidades, creando servicios eficientes y personalizados. Los chatbots son una herramienta más para mejorar el servicio de atención al cliente, reduciendo los tiempos de espera y personalizando la experiencia de cada cliente.

PRODUCTIVIDAD

Se estima que en 2035 la productividad laboral relacionada con la IA crecerá entre un 11 y un 37%.

EQUIDAD

Fomentar la diversidad y la apertura, mitigando los prejuicios y sesgos en los procesos de reclutamiento, utilizando datos analíticos.

BENEFICIOS SOCIALES

La IA también puede tener efectos transformadores en muchos de los desafíos sociales a los que nos enfrentamos:

CAMBIO CLIMÁTICO

La IA puede traer nuevas posibilidades para: el transporte público, haciéndolo más eficiente y sostenible; la gestión energética de edificios (ej. permite ahorro del 35 % en el consumo de energía) (Lee et al., 2022). Hasta 2030, se espera que la IA pueda ayudar en el esfuerzo para reducir los gases de efecto invernadero entre un 1,5-4 % (Parlamento Europeo, 2023c).

EDUCACIÓN

Las herramientas de IA permiten personalizar las experiencias de aprendizaje de los estudiantes y crean formas innovadoras de evaluación del proceso para beneficio de los profesores (Kulik & Fletcher, 2016).

JUSTICIA

Los sistemas de IA se pueden utilizar para prevenir delitos, riesgo de fuga o para acelerar el ritmo de los procesos judiciales

PROSPERIDAD

Se estima que hasta 2030 la IA contribuya a la economía global con más de 11 trillones de euros (Voss, 2021).

CONOCIMIENTO DE LA INFORMACIÓN

La IA se puede utilizar para reducir y prevenir ciberataques y desinformación, por ejemplo, recurriendo a algoritmos que detectan y eliminan bulos (fake news y deepfakes), y automatizando el proceso de fact-checking (verificación de los hechos) (Voss, 2021).

SEGURIDAD

La IA puede apoyar la construcción y el mantenimiento de máquinas e infraestructuras con mayores niveles de seguridad. Una de las áreas en mayor evolución es la construcción de coches más seguros (ej. con asistencia a la conducción), la construcción de coches de conducción autónoma, entre otros (Parlamento Europeo, 2023c).

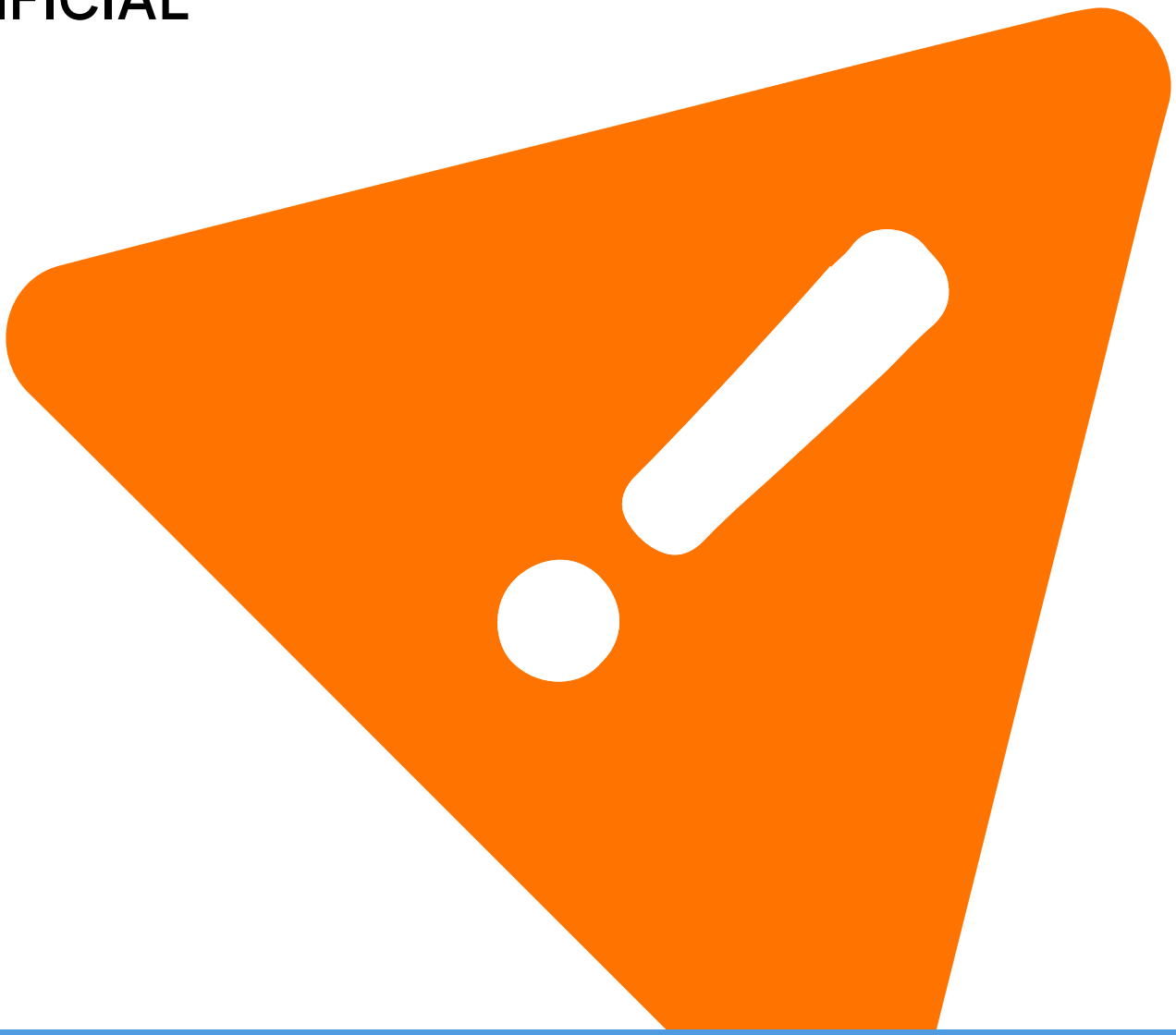
INCLUSIÓN

Las personas con discapacidad pueden recibir asistencia de la IA para ver, oír y desplazarse (Russel & Norvig, 2021).

MERCADO DE TRABAJO

De acuerdo con el Foro Económico Mundial (2020), se espera que la Inteligencia Artificial haga surgir 97 millones de nuevos puestos de trabajo. En Portugal, la IA podrá crear entre 600 000 y 1,1 millones de puestos de trabajo hasta 2030, tal como calcula el estudio de la Universidade NOVA en colaboración con la CIP (Duarte et al., 2019).

RIESGOS DE LA INTELIGENCIA ARTIFICIAL



La IA puede tener un enorme potencial para el progreso humano, aportando, al igual que los avances tecnológicos del pasado, nuevas formas de crear valor, innovación, mejorar la calidad de vida de la población y responder a desafíos sociales (Oliveira, 2019). Sin embargo, la tecnología de IA también puede crear riesgos y agravar riesgos ya existentes, por ejemplo, se estima que pueda tener un impacto negativo en un 35 % de los Objetivos de Desarrollo Sostenible (Vinuesa et al., 2020). Demarcar las áreas en las que la IA puede tener potenciales efectos adversos es un paso fundamental para el desarrollo de políticas de regulación y de medidas de mitigación de los riesgos (Acemoglu, 2021).

MERCADO DE TRABAJO

Existen inmensas estimativas sobre el impacto de la IA en el mercado de trabajo. No se puede negar que tendrá algún tipo de impacto. Por ejemplo, Accenture (2023) prevé que la IA generativa tenga influencia sobre el 40 % de todas las horas de trabajo. En 2023, la OCDE (2023a) propuso que la IA podría afectar a cerca del 30 % de los puestos de trabajo en Portugal (3 p.p. por encima de la media de la OCDE). Duarte y colaboradores (2019), en un informe de la Universidade NOVA en colaboración con la CIP, presentan un riesgo más negativo y consideran que la IA puede influir en cerca del 50 % de los puestos de trabajo, representando una reducción de 1,1 millones de puestos de trabajo – subrayando que, de igual modo, también podrá crear entre 0,6 y 1,1 millones de puestos de trabajo hasta 2030.

Aunque las nuevas tecnologías creen puestos de trabajo, no todas las personas podrán recibir nuevas cualificaciones y transitar al área tecnológica, en particular aquellas que pierdan sus puestos de trabajo debido a la introducción de la IA. Entre los puestos de trabajo más amenazados están las profesiones administrativas y de la justicia (Hatzius et al., 2023). Sin embargo, los progresos en el área de la IA generativa pueden poner en entredicho puestos de trabajo en áreas que anteriormente estaban salvaguardadas, como finanzas, medicina, ciencias, cultura, ingeniería, entre otras, como refiere la OCDE (2023b). Una consecuencia adicional de la automatización es que los profesionales pasan a tener más competencia para los mismos trabajos, lo que llevará a una reducción de los salarios. Por ejemplo, las tecnologías de IA generativas tienen buenas competencias de escritura y pueden escribir artículos y textos. De esta forma, los y las periodistas podrán tener que competir contra la IA por los mismos trabajos (Acemoglu, 2021; Vallance, 2023).

OCDE (2023) propuso que la IA podría afectar

30%

puestos de trabajo en Portugal

3 p.p. por encima de la media de la OCDE

la IA puede influir en cerca del 50% de los puestos de trabajo, representando una reducción de 1,1 millones de puestos de trabajo –subrayando que, de igual modo, también podrá crear entre 0,6 y 1,1 millones de puestos de trabajo hasta 2030.

MONOPOLIOS DE INFORMACIÓN

Los datos son la materia prima de la IA. Estos se pueden utilizar de forma positiva en la innovación, predicción y toma de decisión. Sin embargo, los datos también se pueden utilizar de forma perjudicial (Acemoglu, 2021).

Se ha expresado una preocupación cada vez mayor con el surgimiento de monopolios –empresas globales de tecnología que concentran la riqueza generada en sus respectivas áreas– responsables del desarrollo de tecnologías de IA. El desarrollo de sistemas de large language models está concentrado en organizaciones como Meta, Google, Amazon y Microsoft (Chatterjee, 2023; Parlamento Europeo, 2023c).

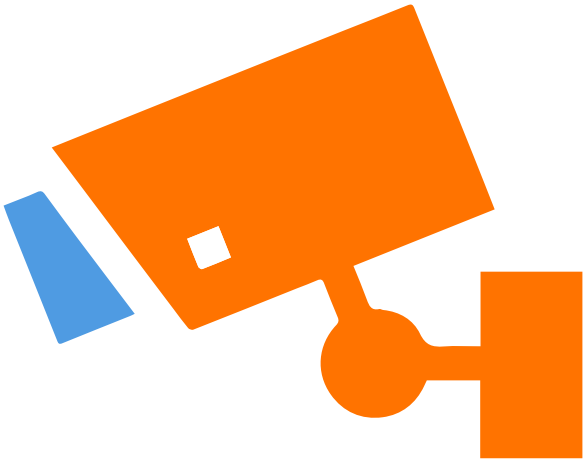
Una de las principales consecuencias de los monopolios es crear situaciones de desventaja para las otras organizaciones. Recurriendo a enormes cantidades de datos de los ciudadanos, estas organizaciones consiguen anticipar el mercado. Un ejemplo de ello, entre muchos otros, es la situación en la que Google aprovechó su posición privilegiada, como principal motor de búsqueda, dando ventaja a la venta de sus propios productos. En vista de esta situación, la Comisión Europea multó a la organización con 4,2 mil millones de euros por considerar que abusó ilegalmente de su posición en el mercado (Comisión Europea, 2017).

El fenómeno de los monopolios puede contribuir a la creación de productos de menor calidad y con menos respeto por la privacidad de los usuarios. Cuando existen muchas organizaciones, la competición por mejorar la privacidad y protección de datos resulta en un mayor respeto por los datos de los usuarios. Ante los monopolios, existe poca o ninguna presión para mejorar los estándares de seguridad y privacidad de los datos. Aunque estos servicios sean muchas veces gratuitos, es importante entender que en estos casos la recopilación de una cantidad excesiva de datos equivale a decir que pagamos un precio excesivo (Stucke, 2018).

Los monopolios de datos se enfrentan a menos presión para mejorar sus políticas de recopilación de datos y privacidad –qué información recopilan y cómo van a utilizarla. Incluso en las situaciones en las que estas organizaciones mejoren los temas relacionados con la privacidad, si no existen alternativas viables competitivas, siempre va a existir una desigualdad en la relación de fuerzas entre los monopolios y sus usuarios (Stucke, 2018).

Las organizaciones con monopolios de datos pueden causar problemas a nivel de la monitorización, seguridad y manipulación comportamental. Que un número reducido de organizaciones acumulen miles de millones de datos personales aumenta el nivel de riesgo de violaciones de la privacidad y de vigilancia autoritaria. Por ejemplo, los monopolios de datos, con políticas débiles de privacidad y seguridad de los datos, pueden poner en riesgo a sus usuarios y terceros involucrados. El escándalo de Cambridge Analytica, empresa que quiso crear un perfil psicográfico de millones de electores, consiguió recopilar, a partir de Facebook, datos no solo de usuarios como de sus amigos que no habían autorizado compartir datos (Cyphers & Doctorow, 2021).

Por otro lado, la centralización de datos de ciudadanos recopilados de múltiples fuentes –redes sociales, websites, cámaras de seguridad– se puede utilizar para crear sistemas de vigilancia de la población, tanto por parte de las organizaciones como por parte de los gobiernos. En el caso de las agencias gubernamentales, durante la pandemia COVID-19 hubo un aumento en el uso de sistemas de vigilancia para apoyar el cumplimiento de las medidas de distanciamiento social y el uso de máscaras. Esta no es una práctica libre de peligros, y es importante que exista una supervisión más estrecha de sus prácticas de vigilancia, para que estas medidas no puedan ser utilizadas para fines de control, represión y establecimiento de regímenes autoritarios (CBS News, 2020; Kerry, 2020).



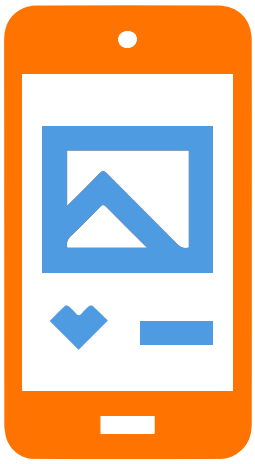
Acumular datos de los consumidores y de las organizaciones proveedoras de servicios facilita la manipulación comportamental. Es decir, las organizaciones pueden utilizar el historial del comportamiento de sus usuarios para vender productos que los consumidores no necesitan. Podemos encontrar una situación de estas cuando una empresa estima los prime vulnerability moments –momentos en los que las personas tienden a comprar impulsivamente– y envía, en esos momentos, anuncios personalizados para que los usuarios compren (Acemoglu, 2021; Stucke, 2018).

DEMOCRACIA

En los últimos años hemos asistido a un debilitamiento de las democracias. Varios países han sustituido sus sistemas democráticos y muchos otros han visto atacadas sus instituciones y normas democráticas. El debilitamiento de las instituciones democráticas se hace acompañar de la polarización de la población en relación con cuestiones políticas y sociales. Entre los factores que han contribuido a la polarización se encuentran la comunicación en línea, las redes sociales y la desinformación. Los sistemas de IA pueden, por medio de echo chambers, deepfakes y tecnología generativa de texto, acelerar el proceso de fragmentación de la democracia (Acemoglu, 2021).

ECHO CHAMBER

A medida que se navega en las redes sociales –haciendo «me gusta», viendo vídeos, compartiendo contenido– los algoritmos del feed empiezan a recoger información sobre los comportamientos y preferencias de cada persona, y a identificar nuevos contenidos que se enmarcan en esas preferencias. El algoritmo del feed puede crear una **echo chamber** –es decir, solo se exhiben contenidos con los que el usuario se identifica– que amplifica el punto de vista del usuario y refuerza sus ideales. La espiral creada por estas echo chambers puede exponer con relativa facilidad al usuario a información errónea, falsa o a rumores y contribuir a su diseminación (Gao et al., 2023).

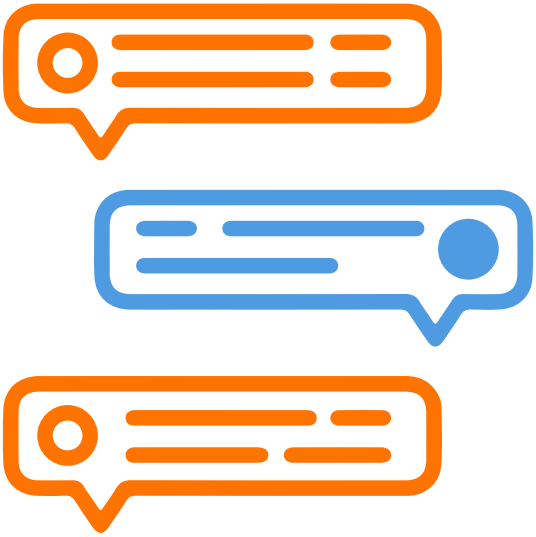


DEEPFAKES

Los deepfakes –contenidos que incluyen vídeos, imágenes, sonidos sintéticamente modificados– pueden causar graves problemas sociales. Los deepfakes pueden causar problemas durante periodos electorales (ej. un video de un candidato haciendo un comentario o acción polémica), pueden reducir la confianza en las instituciones y en las autoridades, aumentar la discordia en la evaluación de hechos y de datos, crear confusión entre opiniones y hechos; generar desconfianza en las fuentes de información respetadas y aumentar el escepticismo de la población en relación con la información (Helmus, 2022).

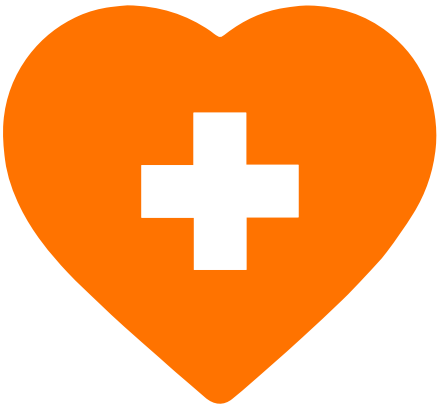
SISTEMAS DE IA GENERATIVOS DE TEXTO

Los estudios existentes sobre mensajes generadas por sistemas de IA generativos de texto, por su capacidad de crear textos de fácil lectura y persuasivos, colocan estas máquinas entre un elevado potencial para fomentar la Alfabetización y un potencial pernicioso para la diseminación de desinformación. En estas investigaciones queda clara la dificultad de los participantes en distinguir los mensajes creados por humanos de los creados por sistemas de IA. De esta forma, esta tecnología se puede usar para crear bots que inunden las redes sociales y páginas web con fake news o comentarios polarizadores, creando desinformación y discordia (Helmus, 2022; Spitale et al., 2023).



DESIGUALDADES

La IA puede tener un impacto determinante en la vida social y económica de las personas. Incluso no siendo los algoritmos los que tomen las decisiones, sus sesgos, intencionales o accidentales, guían la toma de decisión de personas en funciones ejecutivas. Diversas investigaciones defienden que los algoritmos pueden también replicar o agravar prácticas discriminatorias y potenciar las desigualdades estructurales existentes (Acemoglu, 2021; Parlamento Europeo, 2023c).



SALUD

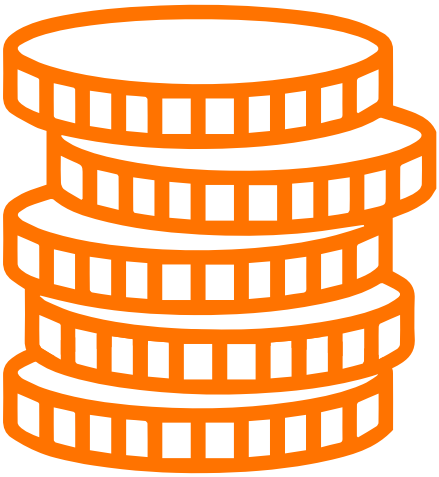
En el contexto sanitario, un algoritmo clínico utilizado en muchos hospitales para apoyar la toma de decisión sobre qué pacientes deben recibir cuidados sanitarios reveló un sesgo racial.

Los pacientes no caucásicos tenían que estar mucho más enfermos que los pacientes caucásicos para ser remitidos a tratamiento. Una de las justificaciones para este sesgo es que, durante el entrenamiento del algoritmo, se recurrió a datos relativos a los gastos en cuidados sanitarios, que reflejaban un historial de menos inversión en los pacientes no caucásicos en comparación con los usuarios caucásicos (Obermeyer et al., 2019).

EMPLEABILIDAD

En el contexto de la empleabilidad, especialmente en la contratación, han surgido varios relatos de reproducción de sesgos por parte de la IA.

En 2018, en un reportaje de Reuters, Dastin (2018) informó de que un programa de IA de evaluación de CV fue usado por Amazon en 2014 –evaluando a los candidatos de 0 a 5 estrellas– e identificando el top 5 de candidatos. Después de un año de utilización, el programa fue desechado por comprobarse que penalizaba CV con la palabra «mujer», entre otras preferencias por los candidatos del género masculino. Este **sesgo con relación al género dejaba a las mujeres en situación de desventaja en el momento de solicitar un empleo**. En el mismo sentido, resultó, de una auditoría independiente al algoritmo que controla los anuncios de empleo del Facebook, la **conclusión de que este no presenta los mismos anuncios de empleo a hombres y mujeres, incluso cuando ambos tienen las mismas cualificaciones** (Hao, 2021).



FINANCIERO

En el sector financiero, concretamente en aquello que tiene que ver con préstamos, se muestran evidencias de que los algoritmos utilizados derivan en discriminación y disparidades en los precios pagados por diferentes propietarios.

Bartlett y colaboradores (2022) descubrieron que **cuando una persona no caucásica/latina solicitaba un préstamo, el importe que pagaba era superior al de una persona caucásica en las mismas condiciones**. Anualmente, **esta discriminación racial supone cerca de 500 millones de euros pagados de más por minorías raciales/étnicas**.

VIVIENDA

En esta área de la vivienda, la utilización de algoritmos sesgados por cuestiones raciales discrimina a las personas, colocándolas en situaciones de desventaja económica, menores oportunidades y segregación. Esta discriminación deriva de sistemas de IA que evalúan la capacidad de una persona de conseguir o no pagar una vivienda. Sin embargo, este análisis, con base en el nombre, la raza y el código postal, discrimina a personas no caucásicas/latinas, incluso si tienen crédito suficiente, aumentando el valor de sus alquileres o presentando casas en barrios más violentos (Akselrod, 2023; Johnson, 2023).

Un caso semejante de exclusión fue identificado en Facebook –los anuncios para comprar casa eran exhibidos con mayor frecuencia a usuarios caucásicos y los anuncios para alquiler surgían más veces a minorías raciales/étnicas (Hao, 2019).

JUSTICIA

En el área de la justicia, el riesgo de reincidencia ha sido objeto de estudio por su potencial discriminatorio y elevado impacto en la vida de los acusados. En un estudio sobre el algoritmo COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) –programa que genera una previsión del riesgo de reincidencia y riesgo de reincidencia violenta– se descubrió que el algoritmo no calculaba este riesgo de una forma justa.

Los acusados no caucásicos que no reincidían durante dos años tenían una probabilidad dos veces mayor de ser clasificados como reincidentes de alto riesgo que los acusados caucásicos. Y en el sentido inverso, los acusados caucásicos que reincidieron en los dos años siguientes tenían una probabilidad dos veces mayor de ser clasificados como bajo riesgo en comparación con los no caucásicos (Larson et al., 2016). En un estudio posterior, Dressel y Farid (2018) compararon la precisión del mismo programa COMPAS con la precisión de personas con poca o sin ninguna experiencia a nivel de la justicia a la hora de predecir el riesgo de reincidencia –ambos alcanzaron un 65 % de precisión.

La vigilancia predictiva –es decir, la IA analiza datos para prevenir potenciales delitos– ha sido criticada por su falta de transparencia y por mantener injusticias estructurales. Una de las situaciones más graves, la de la vigilancia predictiva basada en la localización –es decir, el sistema considera el historial de registros de antecedentes e identifica los lugares con mayor riesgo de delito– perpetúa el racismo sistémico (ej. exceso de vigilancia policial de barrios con minorías raciales/étnicas está correlacionado con el registro de más incidentes) y no reúne pruebas de su capacidad para reducir el crimen (Heaven, 2020; Rotaru et al., 2022).

OTROS RIESGOS

Las tecnologías de IA pueden conducir a otros problemas como (Acemoglu, 2021; Russel & Norvig, 2021):

ARMAS BASADAS EN IA

Las armas basadas en la tecnología de IA funcionan de forma autónoma y pueden ser más baratas, rápidas y tener un alcance mayor que las armas controladas por humanos. Junto con estas armas autónomas surgen problemas éticos y sociales que deben ser regulados. Sin regulación, estas armas podrán decidir autónomamente matar, atacar indiscriminadamente a civiles o combatientes y no preocuparse de ningún daño colateral. En otras circunstancias, el recurso a estas armas podrá ser dirigido contra ciudadanos, grupos en protesta o grupos de la oposición, con vistas a fortalecer a un gobierno.

IMPACTOS SOBRE LA SALUD MENTAL.

Yam y colaboradores (2023) concluyeron que la exposición a la idea de la IA en los lugares de trabajo genera una sensación de inseguridad con relación al trabajo. Además de eso, ingenieros que trabajaban diariamente con IA, en una relación mediada por la mayor inseguridad en el trabajo, mostraban mayores niveles de burnout y más comportamientos antisociales. En otro estudio, Tang y colaboradores (2023), la interacción con IA en el trabajo estaba correlacionada con mayor necesidad de interacción y más sentimientos de soledad, lo que desencadenaba comportamientos adaptativos (ej. ayudar a otros en el trabajo), pero también desadaptativos (ej. consumo de alcohol e insomnio después del trabajo). Siendo tan solo estudios preliminares y breves sobre el tema, es necesario y recomendable investigar más sobre los impactos de la IA en la salud mental.

DESALINEAMIENTO DE OBJETIVOS

El problema del desalineamiento de los objetivos tiene una dimensión más futurista, pero, aun así, debe ser ponderado, monitorizado y preparado seriamente. Asociada al tema del surgimiento de la Superinteligencia Artificial, resta la duda de cómo se comportará una tecnología que, en ese momento, superará todas las capacidades humanas. De una forma preventiva, es importante intentar garantizar que los objetivos de las máquinas de IA y los objetivos de la humanidad estén alineados.

3. APLICACIONES DE LA IA Y RECOMENDACIONES ESTRATÉGICAS

CONTEXTOS DE SANITARIOS
CONTEXTOS DE TRABAJO
CONTEXTOS EDUCATIVOS

APLICACIONES DE LA IA Y RECOMENDACIONES ESTRATÉGICAS

El desarrollo de sistemas de IA más autónomos y con mayor potencial de influencia sobre el comportamiento humano hace necesario comprender el funcionamiento de estas tecnologías y, al mismo tiempo, que **los especialistas de diferentes áreas sean conscientes y comprendan las implicaciones psicológicas y sociales de su implementación en la vida cotidiana de las personas** (Abrams, 2019; 2023).

A continuación, presentamos algunas aplicaciones de los sistemas de IA en la salud, el trabajo y la educación, así como consecuentes recomendaciones estratégicas. El desarrollo de una **IA centrada en lo humano** (del inglés Human-Centered AI) se presenta como un camino posible que permita **conciliar la eficiencia técnica de estas tecnologías con las necesidades, las competencias, los valores y los derechos de las personas** (Parker & Grote, 2019).

EN LOS CONTEXTOS SANITARIOS

En el marco de la salud, es posible distinguir entre tecnologías de IA que no implican una interacción relacional con las personas (ej. herramientas de evaluación y diagnóstico asistidas por IA) y tecnologías que cuentan con un componente relacional, bien sea con un asistente virtual (como chatbots) o con una interfaz robótica (ej. robots). Dwyer e colegas (2018) identifican que, en el área de la Salud Mental, los sistemas de IA, concretamente de Aprendizaje Automático y de Procesamiento de Lenguaje Natural (PLN), pueden ser bastante útiles como herramientas que ayudan al diagnóstico y que aclaran pronósticos teniendo en consideración la selección del tratamiento.

Los sistemas de PLN tienen una amplia área de aplicación, con importantes aportaciones para los profesionales sanitarios que trabajan con dificultades y problemas de salud psicológica, como el suicidio, la demencia y las perturbaciones psicóticas. Según estudios realizados, los sistemas de PLN parecen bastante precisos en la identificación de patrones de comunicación que puedan revelar riesgo de suicidio (Lejeune et al., 2022), pudiendo contribuir a **la prevención de comportamientos autolesivos y del suicidio** y, además, a la identificación de patrones de comunicación que sugieren **declive cognitivo** (Agbavor & Liang, 2022), pudiendo ayudar a prevenir o diagnosticar precozmente la **demencia**. Los sistemas de PLN también han sido utilizados para identificar trastornos psicóticos, y se espera que a través de su integración con tecnologías de visión computacional sea posible prever, con meses de antelación –y, así, prevenir– el desarrollo de trastornos como la esquizofrenia (Corcoran & Cecchi, 2020).

En el **campo de la robótica** se ha investigado la **utilización de robots** (ej. robot NAO) para realizar **intervenciones con niños que viven con algún trastorno del espectro del autismo** (TEA), con el objetivo de fomentar el aprendizaje de competencias sociales (Pennisi et al., 2016). Los resultados de estas investigaciones, con diferentes prototipos, **parecen demostrar alguna efectividad** (Salimi et al., 2021). Igualmente, en estudios comparativos, las intervenciones con robots parecen tan efectivas como las intervenciones realizadas por humanos en el reconocimiento de gestos y en el reconocimiento de competencias sociales (So et al., 2019). Aunque los niños con TEA parezcan relacionarse más fácilmente con los robots, una mayor participación no parece reflejarse, necesariamente, en mejores resultados (So et al., 2019).

Otros prototipos de robots (ej. PARO; AIBO; NeCoRo; Bandit) han **sido utilizados en la intervención con ancianos**, con el objetivo de reducir el estrés y el aislamiento y de fomentar las conexiones sociales (Bemelmans et al., 2012). Aunque algunos estudios indiquen que la interacción con estos robots puede ser beneficiosa para fomentar la calidad de vida y reducir el uso de medicación en personas con demencia, **no existen pruebas consolidadas que apoyen su efectividad** (Rashid et al., 2023; Wang et al., 2022).

En el **campo de la salud psicológica**, es posible destacar el desarrollo de «**psicoterapeutas virtuales**» (ej. Woebot) que consiguen **identificar, en la comunicación con las personas, dificultades emocionales**, que facilitan educación psicológica deliberada (ej. «¿qué son distorsiones cognitivas?») y, además, que **potencian el aprendizaje de técnicas y competencias** que persiguen, por ejemplo, la reducción de la ansiedad (Fiske et al., 2019). Algunos estudios sugieren que estos chatbots, desarrollados con base en modelos psicoterapéuticos validados, pueden ser eficaces para reducir síntomas depresivos (Fitzpatrick et al., 2019) e, incluso, que algunas personas establecen con ellos un vínculo semejante al de una relación humana (Darcy et al., 2021).

Sin embargo, **estas evidencias no se encuentran consolidadas y temas relacionados con la seguridad** (de la persona y de los datos), **con la eventual dependencia de estas tecnologías y, además, con la receptividad del apoyo** prestado frente a diferentes dificultades psicológicas, **no se encuentran aclaradas** (Abd-Alrazaq et al., 2020; Lim et al., 2022).

Es urgente, pues, la regulación y la definición de orientaciones técnicas y éticas, que incluyan el conocimiento sobre los comportamientos humanos, en el desarrollo de estas tecnologías de IA, sobre todo porque, ya hoy, su proliferación y accesibilidad pueden constituir riesgos para la salud pública (Floridi et al., 2018).

EN ESTE SENTIDO, PROPONEMOS LAS SIGUIENTES **RECOMENDACIONES ESTRATÉGICAS:**

SEGURIDAD, SALUD Y BIENESTAR

Desarrollar algoritmos, utilizados en modelos de Aprendizaje Automático, que consideren variables sociales y psicológicas en un sentido que estos puedan ser más seguros, menos propensos a decisiones discriminatorias (Parker & Grote, 2019) y que, por ejemplo, en la predicción de la ideación suicida, minimicen los falsos negativos (Doorn et al., 2020).

MODELOS BASADOS EN EVIDENCIAS

Desarrollar robots y chatbots seguros, basados en modelos científicamente validados y que, idealmente, sean testados en ensayos clínicos controlados con diferentes poblaciones y en diferentes condiciones de salud mental (Lim et al., 2022) –regulando estas tecnologías como instrumentos basados en evidencias que los profesionales tienen disponibles (Doorn et al., 2020).

RELACIÓN CON USUARIOS

Investigar la interacción entre personas y chatbots, comprendiendo fenómenos como la relación establecida y los efectos a largo plazo para los usuarios y para la salud de las comunidades, concretamente cómo su utilización puede modificar las percepciones de riesgo frente a situaciones clínicas graves, así como la eventual dependencia de estas tecnologías (Fiske et al., 2019).

MONITORIZACIÓN

Desarrollar herramientas de IA que, por medio del procesamiento de la comunicación registrada en consultas, identifican, por ejemplo, diferentes tipos de sintomatología, características de la receptividad del profesional y, además, ayudan a prever la adhesión a determinada intervención o modelo terapéutico y el respectivo pronóstico (Miner et al., 2022).

ACCESIBILIDAD

Investigar de qué forma las tecnologías de IA pueden permitir una mayor accesibilidad y una mejor gestión de recursos de salud, en el sentido de un apoyo más precoz, oportuno y que responda a dificultades de ligeras a moderadas, facilitando mayor disponibilidad de profesionales que se dediquen a seguimientos prioritarios y de mayor gravedad (Fiske et al., 2019).

PARTICIPACIÓN HUMANA

Garantizar que el desarrollo de tecnologías de IA aplicadas a la salud sigue teniendo profesionales sanitarios involucrados en el proceso de concepción, estudio, implementación y evaluación del sistema (Parlamento Europeo, 2022).

TRANSPARENCIA

En Europa, el RGPD obliga al desarrollo de sistemas de IA comprensibles. De esta forma, el sistema de IA debe ser un explainable AI (sistema de IA explicable) que pueda ser: comprensible para los usuarios, que refleje de una forma razonable su raciocinio y debe permitir entender por qué diferentes usuarios reciben resultados diferentes (Russel & Norvig, 2021; Navas, 2023).

FORMACIÓN

Los profesionales de salud implicados en el desarrollo de IA deben recibir formación adecuada para garantizar que los algoritmos de IA aplicados a la salud son usados de forma correcta (Parlamento Europeo, 2022).

DIVERSIDAD

Para evitar discriminación y sesgos con base en el género, edad, raza/etnia, nivel socioeconómico, los sistemas de IA deben ser entrenados de forma sistemática con muestras representativas. Los equipos de desarrollo de IA deben ser también equipos inclusivos con profesional Parlamento Europeo, 2022).

PRIVACIDAD

El trabajo en el área de la salud siempre debe incluir el consentimiento informado para que los usuarios puedan tomar decisiones informadas en relación con compartir su información personal. Los usuarios deben ser adecuadamente informados sobre la privacidad, anonimato, utilidad y ciberseguridad de sus datos (Parlamento Europeo, 2022).

RESPONSABILIDAD

Deben implementarse procesos que identifiquen el papel y las responsabilidades de los profesionales y organizaciones involucrados en el desarrollo de la tecnología de IA, en concreto, en los casos en los que estos provocan algún tipo de daño (Parlamento Europeo, 2022).

EN LOS CONTEXTOS DE TRABAJO

La aplicación de la IA en el mundo del trabajo ha proporcionado **un cambio en la forma en la que las organizaciones toman decisiones relativas a los trabajadores** y a las **estrategias que utilizan para definir y alcanzar sus objetivos**. En la actuación específica de los profesionales de Recursos Humanos, la IA puede ser una herramienta que permite tomar decisiones más precisas en el reclutamiento, selección y gestión de personas (Oswald et al., 2020).

Aplicado a la realidad de las organizaciones, un sistema de IA, con datos de diversas fuentes, puede, por ejemplo, **prever el desempeño futuro de diferentes candidatos, ayudando en la toma de decisión en los procesos de contratación** –algo que implica que la IA identifique y evalúe características psicológicas, actitudes, comportamientos y otras dimensiones referentes a los candidatos por medio de la comunicación verbal, escrita y/o visual (Landers & Behrend, 2023). Actualmente, en los reclutamientos asistidos por IA, el candidato responde a cuestiones establecidas por asistentes virtuales, siendo los datos analizados posteriormente y cruzados con resultados de testes psicométricos y con otros datos disponibles en línea, definiendo un perfil o identificando características y competencias relevantes.

Otras aplicaciones de la IA en los lugares de trabajo tienen que ver con la **posibilidad de que estas tecnologías complementen y mejoren el desempeño de los trabajadores** (Marikyan et al., 2022) y **permitan una mejor gestión de su tiempo de trabajo** (Bryant et al., 2020). La utilización de sistemas de PLN (ej. Chat GPT), como asistentes de trabajo para organizar tareas, buscar información o, incluso, para desarrollar soluciones para problemas concretos y complejos, demuestra que humanos y sistemas de IA pueden trabajar en un mismo sentido y en pro de los objetivos de la organización. Sin embargo, **las implicaciones en la productividad de las personas y en las relaciones laborales aún están siendo estudiadas**.

Si, por un lado, los sistemas de IA pueden ayudar a tomar decisiones más concretas en la gestión de personas y pueden contribuir, como herramientas de trabajo, a un mejor desempeño de los trabajadores, por otro lado, **destacan los riesgos a los que determinados grupos de trabajadores se enfrentan con la implementación de estas tecnologías**.

La mejora constante de los sistemas de IA exige que miles de personas contribuyan, entre bastidores, a gestionar millones de datos recogidos por grandes empresas tecnológicas (Vicente, 2023). Estos trabajadores, seleccionados por crowdsourcing, realizan varias microtareas de computación asociadas al funcionamiento de algoritmos –por ejemplo, anotación y categorización de datos, transcripción de audio o texto, extracción de información de recibos, interacción experimental con chatbots y/o filtrado de contenido (Newman, 2019). **Imperceptibles para los usuarios de las plataformas y servicios digitales, muchos trabajan en condiciones precarias de trabajo temporal y casi informal, encontrándose expuestos a riesgos psicosociales específicos** (Vicente, 2023). Entre estos riesgos, destaca la **exposición a material gráfico explícito de violencia y de muerte en tareas de filtrado de contenido, con consecuencias graves para su salud mental** (Steiger et al., 2021).

Igualmente, los trabajadores de plataformas digitales que ejercen funciones en **servicios de transporte de pasajeros y en servicios de entregas tienen su trabajo condicionado por sistemas autónomos e inteligentes**. En estos casos, sus tareas laborales, las compensaciones financieras y los horarios y ritmo de trabajo son gestionados directamente por sistemas algorítmicos de Aprendizaje Automático que utilizan datos relativos a su desempeño y cuyos parámetros son definidos por las plataformas y por los usuarios de los servicios (Shestakofsky, 2020). La utilización de estos sistemas de gestión automática puede tener impacto **en la salud psicológica de los trabajadores** (Sun, 2023). La inestabilidad financiera, el **aislamiento, largas e imprevisibles jornadas de trabajo y la ausencia de canales de comunicación con jefes y colegas**, son algunos de los factores que contribuyen al agotamiento emocional y a la sintomatología depresiva y ansiosa que algunos de estos trabajadores experimentan, sobre todo cuando esta actividad profesional es la única fuente de ingresos (Sun, 2023).

La falta de control sobre las condiciones laborales muestra de qué forma el uso de sistemas algorítmicos y modelos de IA, sin considerar implicaciones psicosociales, **puede dejar a los trabajadores a merced de un sistema sesgado por parámetros de productividad en detrimento de parámetros de seguridad y salud mental** (Vignola et al., 2023).

También es previsible que, con los avances e integración de la IA con interfaces robóticas, que trabajadores de determinados sectores, concretamente de la manufactura y el comercio, vean a máquinas sustituirles en sus funciones laborales (Duarte et al., 2019). En fábricas, **la robótica permitirá una mayor eficiencia del trabajo mecanizado** (ej. sector automovilístico), en oficinas, **asistentes virtuales podrán asumir tareas de atención al cliente** (ej. telecomunicaciones). La automatización de tareas repetitivas, morosas y peligrosas tendrá impacto en los empleos disponibles. Los resultados de un estudio comparativo entre países de la Unión Europea indican que los trabajadores portugueses se sienten inseguros en relación con los avances tecnológicos e, igualmente, son aquellos que, por la predominancia de determinadas industrias, tienen mayor probabilidad de ver sus funciones laborales sustituidas (Kozak et al., 2020).

La introducción de la IA en el mundo del trabajo implica prevenir que se perpetúen o agraven desigualdades en el acceso al trabajo, influyan sobre el mantenimiento del empleo o tengan consecuencias nefastas para la salud mental de los trabajadores cuyas funciones son gestionadas o sustituidas por sistemas automáticos e inteligentes (Parker & Grote, 2019).

EN ESTE SENTIDO, PROPONEMOS LAS SIGUIENTES **RECOMENDACIONES ESTRATÉGICAS:**

AUDITORÍAS

Garantizar auditorías en las organizaciones, fomentando la transparencia en los procesos de selección y gestión de personas asistidos por IA y comprobando que los parámetros definidos en los algoritmos resultan en predicciones válidas y sin sesgos de sexo, género, raza, condición socioeconómica u otros factores sociales (SIOP, 2022).

IA CENTRADA EN HUMANOS

Desarrollar contextos de trabajo de IA centrada en humanos (Parker & Grote, 2019), en un sentido en que estas tecnologías son utilizadas como complementarias a las decisiones o a las tareas de trabajo y que todas las partes implicadas –gestores, técnicos y trabajadores– pueden participar en el proceso de construcción y afinamiento de los algoritmos.

RESPONSABILIDAD

Garantizar la responsabilidad y la participación de los profesionales de Recursos Humanos en la comunicación de decisiones de contratación, gestión o despido de trabajadores, fomentando políticas organizacionales que velen por la transparencia de los sistemas cuyas decisiones afectan directamente a la vida de las personas (Maasland & Weißmüller, 2022)

EVALUACIÓN DEL IMPACTO EN LOS TRABAJADORES

Comprender el impacto de riesgos psicosociales, así como las percepciones y actitudes de los trabajadores, en diferentes sectores, acerca de la implementación de sistemas de IA en los lugares de trabajo (Gnambs & Appel, 2019), identificando necesidades de recualificación laboral y nuevos modelos de trabajo necesarios en las organizaciones (Damian et al., 2017). A este efecto, puede ser importante consultar a los trabajadores o a sus representantes sobre, entre otros asuntos: a) ¿quién en la organización está implementando los sistemas de IA?, b) ¿cuál el objetivo de la implementación de estos sistemas?, c) ¿para quién va a ser útil el sistema de IA?, d) ¿para quién puede ser perjudicial el sistema de IA y a quién podrá poner en situación de desigualdad?, e) ¿quién queda excluido de la posibilidad de utilizar IA?, f) ¿cuáles son las potenciales amenazas para los derechos sociales de los trabajadores?, g) ¿quién es responsable de los datos y quién tiene derecho a acceder a los datos de los trabajadores?, y h) ¿los trabajadores serán partícipes del proceso de diseño del sistema de AI? (Madinier, 2021).

SEGURIDAD, SALUD Y BIENESTAR

Crear mecanismos de protección y de asistencia que velen por la seguridad, la salud (física y psicológica) y el bienestar de los trabajadores «invisibles» que, en la actualidad, garantizan operaciones necesarias para el desarrollo de mejores tecnologías de IA (Steiger et al., 2021).

EVALUAR LA EFICACIA DE LAS HERRAMIENTAS DE IA

Investigar nuevos modelos de trabajo que presuponen la complementariedad entre humanos y las herramientas de IA, explorando de qué forma la interacción con asistentes virtuales se refleja en la productividad, satisfacción con el trabajo y bienestar de los trabajadores (Bryant et al., 2020).

IMPLEMENTAR UN SISTEMA DE QUEJAS

Teniendo en cuenta el potencial impacto de los algoritmos en la vida de los trabajadores, las organizaciones deben crear sistemas de quejas para que los trabajadores puedan informar de una forma segura de sus preocupaciones en relación con los sesgos de los algoritmos (Obermeyer et al., 2021).

SUPERVISIÓN

Las organizaciones deben crear equipos de supervisión de los sistemas de IA. Estos equipos reúnen a un grupo diverso de profesionales que trabajan de forma continua para evaluar el impacto de los sesgos de los algoritmos y para garantizar la implementación de soluciones basadas en IA equitativas. Las principales tareas de este equipo son dar respuesta a quejas, registrar toda la información relativa al proceso de toma de decisión de estos modelos y coordinar los procesos de auditoría (Obermeyer et al., 2021).

INCLUSIÓN DE PSICÓLOGOS

Los modelos de gobernanza de las organizaciones deben ser adaptados a la IA. La adaptación de los modelos de gobernanza debe pasar por incluir en las Administraciones de las organizaciones a profesionales, con carácter ejecutivo o no ejecutivo, como los psicólogos, que aporten una visión informada, sensible y justa de la aplicación de la IA. Esta acción prepara a las organizaciones para utilizar estas nuevas tecnologías de una forma ética, transparente y responsable (IMDA & PDPC, 2020).

EN LOS CONTEXTOS EDUCATIVOS

También el área de la educación se encuentra en transformación debido a las tecnologías de IA. El **surgimiento de sistemas inteligentes**, concretamente de sistemas de PLN, permite que **chatbots** funcionen como **herramientas de tutoría** y, además, por ejemplo, que se faciliten **recomendaciones personalizadas de contenidos y métodos de aprendizaje** teniendo en cuenta las características y competencias de los alumnos (Singer, 2023).

Los sistemas de IA, cuando se utilizan bajo un paradigma de aprendizaje que incluye el papel de diferentes agentes (es decir, alumnos, profesores y sistemas de IA), pueden potenciar cambios en la forma en la que los **profesores imparten y desarrollan los contenidos programáticos, en la forma como los alumnos gestionan su aprendizaje** y como los **contextos educativos evalúan el desempeño de los estudiantes** (Ouyan & Jiao, 2021).

En plataformas online, el uso de **asistentes virtuales** puede **fomentar competencias de autorregulación** que mejoran la implicación de los alumnos en cursos en línea en los que, generalmente, existe poco apoyo de terceros. Estos asistentes virtuales, por medio de diversos datos sobre el usuario pueden, por ejemplo, definir objetivos diarios e, incluso, sugerir estrategias de regulación de la atención, siendo que su efectividad depende de su adaptación a las características de personalidad del usuario (Pogorskiy & Beckmann, 2023).

En la implementación de modelos educativos inclusivos, el **uso de diferentes tecnologías de asistencia**, que utilizan sistemas de IA, puede permitir que **niños y jóvenes con discapacidad comuniquen con más eficiencia** y se beneficien de mejores oportunidades de aprendizaje (Zdravkova et al., 2022). Estas tecnologías de Comunicación Aumentada y Alternativa (CAA) permiten, por ejemplo, que un joven con dificultades auditivas graves pueda, gracias a un sistema de PLN instalado en un smartphone (ej. HearMeOUT app), traducir, casi al instante, comunicación verbal en lengua gestual y, en sentido contrario, por medio de visión computacional, traducir comunicación gestual en frases (Gopavaram et al., 2020).

Aunque las tecnologías de IA demuestren potencial en el marco de la Educación, sobre todo en el apoyo al aprendizaje y en la facilitación de la comunicación por medio de tecnologías de asistencia, **las desigualdades en el acceso a estas tecnologías pueden perpetuar, o agravar, desigualdades entre alumnos procedentes de condiciones socioeconómicas distintas** (Dieterle et al., 2022). También los **parámetros de algoritmos**, utilizados en sistemas de Aprendizaje Automático, **pueden perpetuar estas desigualdades en la atribución de resultados escolares o en la previsión de éxito académico condicionado**, por ejemplo, **el acceso a establecimientos de enseñanza superior**.

Desde una perspectiva más amplia de educación, y considerando los riesgos que los sistemas de IA tienen para la democracia en un mundo cada vez más digitalizado, es urgente que la mayoría de los ciudadanos se encuentren informados y capacitados para la identificación de noticas falsas y otros tipos de desinformación, incluyendo deepfakes – **Conocimiento de información y Conocimiento digital** (Ahmed, 2023).

Transformar la educación, fomentándola a lo largo del ciclo de vida, implica que se enmarque el uso de tecnologías de IA en perspectivas pedagógicas y científicas sobre el aprendizaje, considerándose el camino posible (y éticamente deseable; UNESCO, 2019) de contextos educativos que utilizan la IA en pro del desarrollo de las personas.

EN ESTE SENTIDO, PROPONEMOS LAS SIGUIENTES **RECOMENDACIONES ESTRATÉGICAS:**

AUDITORÍAS

Garantizar auditorías, en contextos educativos, a los sistemas de IA utilizados, velando por la transparencia de los parámetros algorítmicos y por la inclusión de variables psicológicas y sociales que influyen sobre la implicación y el desempeño de los alumnos –en el sentido de prevenir sesgos que perpetúen desigualdades (Dieterle et al., 2022).

APORTACIONES DE LA CIENCIA PSICOLÓGICA

Integrar conocimiento de la ciencia psicológica, concretamente sobre las relaciones entre características de la personalidad y competencias de autorregulación del aprendizaje, en el desarrollo de chatbots utilizados en plataformas educativas en línea (ej. Moodle) y que se destinan a fomentar la implicación y desempeño de los alumnos.

EQUIDAD

Garantizar la equidad en el acceso a herramientas de IA y a oportunidades de aprendizaje digitales (hardware, software, Internet), integradas en nuevos modelos educativos, para la diversidad de niños y jóvenes, incluyendo aquellos que vivan con algún tipo de discapacidad y/o con menores recursos socioeconómicos (Dieterle et al., 2022).

FORMACIÓN

Los profesores y educadores deben recibir la formación adecuada, desarrollando competencias de interacción con sistemas de IA, de evaluación de los aprendizajes e implementación de estrategias con base en IA (UNESCO, 2021)

PARTICIPACIÓN HUMANA

Desarrollar modelos educativos y prácticas pedagógicas que centren la relación estudiante-profesor en el proceso de aprendizaje y que incluyan tecnologías de IA en una perspectiva asistencial (UNESCO, 2019), es decir, sistemas de IA diseñados para apoyar a humanos en la realización de tareas o toma de decisiones –investigando, en pruebas piloto, su efectividad con muestras heterogéneas de estudiantes (ej. Chen et al., 2022).

CONOCIMIENTO DE LA INFORMACIÓN Y DIGITAL

Capacitar a las personas contra la desinformación, reconociendo que la creencia y divulgación de noticias falsas es un fenómeno social y definiendo estrategias de fact-checking y de inoculación (ej. Bad News Game; Roozenbeek & van der Linden, 2019) que consideran el efecto de heurísticas relacionadas con las dimensiones cognitiva y emocional del comportamiento humano (Pennycook & Rand, 2021).

APRENDIZAJE A LO LARGO DE LA VIDA

La IA en la educación debe ser puesta al servicio de todas las personas, a lo largo de todo el ciclo de vida, en contextos formales, no formales e informales (UNESCO, 2021).

RECOMENDACIONES ESTRATÉGICAS GENERALES

01
SEGURIDAD, SALUD Y BIENESTAR

Desarrollar sistemas de IA seguros exige que se consideren diferentes variables, incluyendo psicológicas y sociales, en el desarrollo de los algoritmos que rigen su funcionamiento. Los sistemas de IA serán tanto más seguros y generadores de bienestar cuanto mayor sea la participación humana, sobre todo de especialistas en el área en el que la tecnología será aplicada, así como cuanto más se basen en las evidencias científicas disponibles (ej. en las evidencias procedentes de la ciencia psicológica).

02
ÉTICA DE TRANSPARENCIA

Siendo la opacidad de los sistemas de IA una preocupación central, sobre todo por la imprevisibilidad y por la dificultad en escrutar las decisiones de estos sistemas, es especialmente importante una ética de transparencia por parte de las organizaciones y equipos que los desarrollan. Informar y aclarar –haciendo visible elementos de los algoritmos utilizados en los sistemas de IA es una responsabilidad ética.

03
EQUIDAD, DIVERSIDAD E INCLUSIÓN

La constitución de sistemas de IA justos está relacionada con las garantías de que todas las personas se podrán beneficiar de las ventajas de la IA y de que ningún grupo en especial deberá ser perjudicado por ellas. La IA debe ser entrenada para la diversidad y la inclusión, por equipos diversos e inclusivos, recorriendo a muestras representativas, de forma sistemática, para evitar discriminación por género, edad, raza/etnia, estatuto socioeconómico, discapacidad u otro.

04
PARTICIPACIÓN HUMANA

Se debe estudiar y monitorizar la interacción entre humanos y los sistemas de IA implementados en los diferentes contextos. Las percepciones y experiencias de los usuarios con relación a la implementación y uso de las tecnologías de IA permiten tanto comprender las relaciones establecidas entre personas y tecnologías (ej. chatbots, robots) como aclarar metodologías de implementación seguras. El desarrollo de sistemas de IA centrada en humanos debe ser una prioridad, intentando que las tecnologías autónomas e inteligentes complementen el trabajo de los profesionales, mejorando la productividad y la calidad de los servicios.

05
PRIVACIDAD Y REGULACIÓN

Implementar sistemas de IA de forma responsable pasa por prevenir cuestiones relacionadas con la privacidad y confidencialidad de los datos. Es necesario promover la regulación, ayudando a los usuarios a tomar decisiones informadas con relación a compartir su información personal, por medio de un consentimiento informado, considerando la confidencialidad, anonimato, utilidad y ciberseguridad en relación con los datos. También debe existir regulación rigurosa en cuanto al uso de todos los datos por parte de las organizaciones.

06
FORMACIÓN Y PROMOCIÓN DEL CONOCIMIENTO

El compromiso con la formación de profesionales de diferentes áreas y de todos los ciudadanos es fundamental para garantizar que las personas utilizarán y aplicarán sistemas de IA de modo productivo, responsable y ético, potenciando sus beneficios y mitigando los riesgos.

07
APORTACIONES DE LA CIENCIA PSICOLÓGICA

Para el desarrollo de las potencialidades de la IA es determinante que las aportaciones de la ciencia psicológica estén codificadas en los sistemas y modelos de IA desde su diseño hasta su implementación

08
FINANCIACIÓN

Se considera esencial priorizar la financiación de investigaciones que articulen la aportación de la psicología en el desarrollo y aplicación de sistemas de IA.

CONCLUSIÓN

El debate en relación con el desarrollo e implementación de tecnologías de IA es un imperativo de la actualidad, siendo tanto mayor la urgencia cuanto mayor el impacto que estas tienen en el día a día de las personas. En virtud del reciente impacto de sistemas como el Chat GPT en diferentes sectores, por ejemplo, en la industria del entretenimiento, en la educación y en el espíritu empresarial, es difícil negar el potencial de estas tecnologías que, entre otras, marcan el inicio de la anticipada 4ª Revolución Industrial.

Por primera vez en la historia se desarrolla una tecnología inteligente, que puede operar de forma independiente del control humano y tomar decisiones que afectan directamente a su comportamiento y a la vida en sociedad. El potencial prometedor de estas tecnologías es real, al igual que sus riesgos. Estamos dando los primeros pasos en la adopción de la IA y una forma de garantizar que este camino es seguro es desarrollar una IA centrada en humanos, que permita a las personas beneficiarse de sus potencialidades y, al mismo tiempo, salvaguardar sus derechos, su seguridad, salud y bienestar.

La ética en la IA, compuesta por algoritmos transparentes, justos y en línea con los objetivos de desarrollo sociales, solo será posible cuando la sociedad civil, especialistas en diferentes áreas (ej. salud, educación, ciencias de la computación), líderes organizativos y decisores políticos, definan estrategias y actúen en conjunto para regular de una manera informada y responsable esta área emergente.

La Ordem dos Psicólogos Portugueses se asocia a este esfuerzo conjunto, trayendo las aportaciones de la ciencia psicológica y de todos los psicólogos, como especialistas en el comportamiento humano y de los procesos mentales, para garantizar que la IA se desarrolla y utiliza de una forma que promueva la salud, el bienestar, la prosperidad, la sostenibilidad y el progreso.

REFERENCIAS BIBLIOGRÁFICAS

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. JMIR Mental Health, 22(7), 1-17.

Abrams, Z. (2019). The promise and challenges of AI. Monitor on Psychology, 52(8). Retirado de: https://www.apa.org/monitor/2021/11/cover-artificial-intelligence.

Abrams, Z. (2023). AI is changing every aspect of psychology. Here's what to watch for. Monitor on Psychology, 54(5). Retirado de: https://www.apa.org/monitor/2023/07/psychology-embracing-ai.

Accenture (2018). Artificial Intelligence: Healthcare's New Nervous System. EUA: Accenture.

Accenture (2023). A new era of generative AI for everyone – The technology underpinning ChatGPT will transform work and reinvent business. EUA: Accenture.

Acemoglu, D. (2021). Harms of AI. NBER Working Paper Series, 29247, 1-57.

Agbavor, F., & Liang, H. (2022). Predicting dementia from spontaneous speech using large language models. PLOS Digital Health, 1(12), 1- 14.

Akselrod, O. (2021). How Artificial Intelligence Can Deepen Racial and Economic Inequities. Retirado de https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities.

Ahmed, S. (2023). Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. New Media & Society, 25(5), 1-22

Bartlett, R., Morse, A., Stanton, R. & Wallace, N. (2022). Consumer-lending discrimination in the FinTech Era. Journal of Financial Economics, 143(1), 30-56.

Bemelmans, R., Gelderblom, G. J., Jonker, P., & de Witte, L. (2012). Socially assistive robots in elderly care: A systematic review into effects and effectiveness. Journal of the American Medical Directors Association, 13(2), 114-120.

Bowles, J. (2014). 54% of EU jobs at risk of computerisation. Retirado de https://www.bruegel.org/blog-post/chart-week-54-eu-jobs-risk-computerisation.

Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2020). How artificial intelligence will impact K–12 teachers. McKinsey & Company. Retirado de: https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers#/.

CBS News (2020). Yuval Harari warns humans will be "hacked" if artificial intelligence is not globally regulated. Retirado de https://www.cbsnews.com/news/yuval-harari-sapiens-60-minutes-2021-10-29/.

Chatterjee, M. (2023). AI might have already set the stage for the next tech monopoly. Retirado de https://www.politico.com/newsletters/digital-future-daily/2023/03/22/ai-might-have-already-set-the-stage-for-the-next-tech-monopoly-00088382.

Chen, S-Y., Lin, P., & Chien, W-C. (2022). Children's digital art ability training system based on AI-assisted learning: a case study of drawing color perception. Frontiers in Psychology, 13: 823078.

Comissão Europeia (2017). Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison-shopping service. Retirado de https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1784.

Corcoran, C. M., & Cecchi, G. A. (2020). Using language processing and speech analysis for the identification of psychosis and other disorders. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5(8), 770-779.

Council of Europe (2019). Algorithms and human rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Genebra: Council of Europe.

Cyphers & Doctorow (2021). Privacy without monopoly: Data protection and interoperability. EUA. Electronic Frontier Foundation.

Damian, R., Spengler, M., & Roberts, B. W. (2017). Whose job will be taken over by a computer? The role of personality in predicting job computerizability over the lifespan. European Journal of Personality, 31(3), 291–310.

Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. JMIR Formative Research, 5(5), 1-7.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Retirado de https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

De Fauw, Ledsam, J., Romera-Paredes, B., … & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine, 24, 1342-1350.

Dieterle, E., Dede, C., & Walker, M. (2022). The cyclical ethical effects of using artificial intelligence in education. AI & Society, 1-11

Doorn, K., Kamsteeg, C., Bate, J., & Arjfes, M. (2020). A scoping review of machine learning in psychotherapy research. Psychotherapy Research, 31(1), 92–116.

Dressel, J. & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4, 1-6.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. Annual Review of Clinical Psychology, 14, 91-118.

Duarte, J., Brinca, P., Gouveia-de-Oliveira, J. & Ferreira, A. (2019). O Futuro do Trabalho em Portugal: O Imperativo da Requalificação – Relatório final. Lisboa: NOVA & CIP.

European Parliament (2022). Artificial intelligence in healthcare: Applications, risks and ethical and societal impacts. European Parliament.

European Parliament (2023a). Lei da UE sobre IA: primeira regulamentação de inteligência artificial. Retirado de https://www.europarl.europa.eu/news/pt/headlines/society/20230601STO93804/lei-da-ue-sobre-ia-primeira-regulamentacao-de-inteligencia-artificial.

European Parliament (2023b). Parlamento negoceia primeiras regras para inteligência artificial mais segura. Retirado de https://www.europarl.europa.eu/news/pt/press-room/20230609IPR96212/parlamento-negoceia-primeiras-regras-para-inteligencia-artificial-mais-segura.

European Parliament (2023c). Artificial intelligence: threats and opportunities. Retirado de https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities.

European Parliamentary Research Service (2022). Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. Bruxelas: European Parliament.

Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS One, 15(2), 1-20.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. JMIR Mental Health, 21(5), 1-12.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 4(2), 1-11.

Floridi, L., Cowls, J., Beltrametti, M., et al…, & Vayena, E. (2018). AI4People - An ethical framework for a good AI Society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.

Future of Life Institute (2023). Pause Giant AI Experiments: An Open Letter. EUA: Future of Life Institute.

Gao, Y., Liu, F. & Gao, L. (2023). Echo chamber effects on short video platforms. Scientific Reports, 13(6282), 1-17.

Garland, A., & Macdonald, A. (2014). Ex Machina [Motion picture]. United Kingdom: Film4.

Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. Journal of Artificial General Intelligence, 5(1), 1-46.

Gopavaram, S. R., Bhide, O., & Camp, L. J. (2020). Can you hear me now? Audio and visual interactions that change app choices. Frontiers in Psychology, 11: 2227.

Gnambs, T., & Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. Computers in Human Behavior, 93, 53-61.

Hao, K. (2019). Facebook's ad-serving algorithm discriminates by gender and race. MIT Technology Review, 1-8.

Hao, K. (2021). Facebook's ad algorithms are still excluding women from seeing jobs. MIT Technology Review, 1-8.

Hatzius, J., Briggs, J., Kodnani, D. & Pierdomenico, G. (2023). Global Economics Analyst: The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani). EUA: Goldman Sachs Economics Research.

Heaven, W. (2020). Predictive policing algorithms are racist. They need to be dismantled. MIT Technology Review, 1-14.

Helmus, T. (2023). Artificial Intelligence, Deepfakes, and Disinformation – A Primer. EUA: Rand Corporation.

Ibañez, G. (2014). Automata. [Motion picture]. Sofia, Bulgaria: Nu Boyana Film Studios.

Info-communication Media Development Authority (IMDA) & Personal Data Protection Comission (PDPC) (2020). Model Artificial Intelligence Governance Framework (2nd Edition). Singapura: IMDA & PDPC.

Johnson, K. (2023). Algorithms Allegedly Penalized BlackRenters. The US Government Is Watching. Retirado de https://www.wired.com/story/algorithms-allegedly-penalized-black-renters-the-us-government-is-watching/.

Jonze, S., & Johnson, M. (2013). Her [Motion picture]. United States: Warner Bros.

Kosinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. arXiv, 1-17. Doi: https://doi.org/10.48550/arXiv.2302.02083.

Kerry, C. (2020). Protecting privacy in an AI-driven world. Retirado de https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/.

Kozak, M., Kozak, S., Kozakova, A., & Martinak, D. (2020). Is fear of robots stealing jobs haunting european workers? A multilevel study of automation insecurity in the EU. IFAC-PapersOnLine, 53(2), 17493-17498.

Kubrick, S. (1968). 2001: A Space Odyssey [Motion picture]. United States: Metro-Goldwyn-Mayer.

Kulik, J. & Fletcher, J. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. Review of Educational Research, 86(1), 42-78.

Landers, R. N., & Behrend, T. S. (2023). Auditing the AI Auditors: a framework for evaluating fairness and bias in high stakes ai predictive models. American Psychologist, 78(1), 36-49.

Larson, J., Mattu, S., Kirchner, L. & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. Retirado de https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Lee, D., Chen, Y. & Chao, S. (2022). Universal workflow of artificial intelligence for energy saving. Energy Reports, 8, 1-32.

Lejeune, A., Le Glaz, A., Perron, P. A., Sebti, J., et al..., & Berrouiguet, S. (2022). Artificial intelligence and suicide prevention: A systematic review. European Psychiatry, 65(1), 1-22.

Lim, S. M., Shiau, C. W., Cheng, L., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. Behavioral Therapy, 53(2), 334-347.

Liu, X., Kale, A., Wagner, S., ... & Denniston, A. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health, 1, 271-297.

Maasland, C., & Weißmüller, K. S. (2022). Blame the machine? Insights from an experiment on algorithm aversion and blame avoidance in computer-aided human resource management. Frontiers in Psychology, 13:779028.

Madinier, F. (2021). A guide to Artificial Intelligence at the Workplace. European Economic and Social Committee.

Malone, T., Rus, D. & Laubacher, R. (2020). Artificial Intelligence and the future of work. Research Brief, 17, 1-39.

Marikyan, D., Papagiannidis, S., Rana, O. F., Ranjan, R., & Morgan, G. (2022). "Alexa, let's talk about my productivity": The impact of digital assistants on work productivity. Journal of Business Research, 142, 572-584.

Miner, A. S., Fleming, S. L., Haque, A., et al..., Shah, N. H. (2022). A computational approach to measure the linguistic characteristics of psychotherapy timing, responsiveness, and consistency. npj Mental Health Research, 1(19), 1-12.

Navas, L. (2023). Explainable Artificial Intelligence needs Human Intelligence. Retirado de https://edps.europa.eu/press-publications/press-news/blog/explainable-artificial-intelligence-needs-human-intelligence_en.

Newman, A. (2019). I Found Work on an Amazon Website. I Made 97 Cents an Hour. The New York Times. Retirado de: https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html.

Obermeyer, Z., Power, B., Vogell, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366, 447-453.

Obermeyer, Z., Nissan, S., Eaneff, S., ... & Mullainathan, S. (2021). Algorithmic Bias Playbook. EUA: Chicago Booth – The Center for Applied Artificial Intelligence.

Oliveira, A. (2019). Inteligência Artificial. Lisboa: Fundação Francisco Manuel dos Santos.

Oliveira, M., Gamito, P., Conde, A., ... & Marina, S. (2021). Ethical decision-making training goes virtual. Technology, Mind & Society, 1-5.

Organização para a Cooperação e Desenvolvimento Económico (OCDE) (2023a). OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market. Genebra: OCDE.

Organização para a Cooperação e Desenvolvimento Económico (OCDE) (2023b). OECD Employment Outlook 2023: Artificial Intelligence and jobs – An urgent need to act. Genebra: OCDE.

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: forward progress for organizational research and practice. Annual Review of Organizational Psychology and Organizational Behavior, 7, 505-533.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, 1-6.

Parker, S., & Grote, G. (2019). Automation, algorithms, and beyond: why work design matters more than ever in a digital world. Applied Psychology, 71(4), 1171-1204.

Pedersen, T., & Johansen, C. (2020). Behavioural artifcial intelligence: an agenda for systematic empirical studies of artifcial inference. AI and Society, 35, 519–532.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, et al..., & Pioggia, G. (2016). Autism and social robotics: A systematic review. Autism Research, 9(2), 165-183.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. Trends in Cognitive Sciences, 25(5), 288-402.

Pogorskiy, E., & Beckmann, J. F. (2023). From procrastination to engagement? An experimental exploration of the effects of an adaptive virtual assistant on self-regulation in online learning. Computers and Education: Artificial Intelligence, 4.

Rahwan, I., Cebrian, M., Obradovich, N., et al..., Wellman, M. (2019). Machine behaviour. Nature, 568, 477–486.

Rashid, L. A., Leow, Y., Klainin-Yobas, P., Itoh, S., & Wu, V. X. (2023). The effectiveness of a therapeutic robot, 'Paro', on behavioural and psychological symptoms, medication use, total sleep time and sociability in older adults with dementia: A systematic review and meta-analysis. International Journal of Nursing Studies, 145.

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. Palgrave Communications, 5(65), 1-10.

Rotaru, V., Huang, Y., Li, T., Evans, J. & Chattopadhyay, I. (2022). Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. Nature human behaviour, 6, 1056-1068.

Russell, S. & Norvig, P. (2021). Artificial Intelligence: A modern approach (4th edition). EUA: Pearson Education, Inc.

Salimi, Z., Jenabi, E., & Bashirian, S. (2021). Are social robots ready yet to be used in care and therapy of autism spectrum disorder: A systematic review of randomized controlled trials. Neuroscience & Biobehavioral Reviews, 129, 1-16.

Shestakofsky, B. (2020). Uberland: how algorithms are rewriting the rules of work. Contemporary Sociology: A Journal of Reviews, 49(3), 294–296.

Singer, N. (2023). New A.I. Chatbot Tutors Could Upend Student Learning. The New York Times. Retirado de: https://www.nytimes.com/2023/06/08/business/khan-ai-gpt-tutoring-bot.html.

So, W., Wong, M. K., Lam, W., Cheng, C., et al..., & Wong, W. (2019). Who is a better teacher for children with autism? Comparison of learning outcomes between robot-based and human-based interventions in gestural production and recognition. Research in Developmental Disabilities, 86(1), 62–75.

Society for Industrial and Organizational (SIOP) (2022). SIOP Statement on the use of Artificial Intelligence (AI) for Hiring: Guidance on the Effective use of AI-Based Assessments. Disponível em: https://www.siop.org/Portals/84/docs/SIOP%20Statement%20on%20the%20Use%20of%20Artificial%20Intelligence.pdf?ver=mSGVRY-z_wR5iluE2NWQPQ%3d%3d.

Spitale, G., Biller-Andorno, N. & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. Sci. Adv., 9, 1-9.

Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M. (2021). The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In CHI Conference on Human Factors in Computing Systems. (pp. 1-14). Yokohama: ACM.

Sun, G. (2023). Quantitative analysis of online labor platforms' algorithmic management influence on psychological health of workers. International Journal of Environmental Research and Public Health, 20(5), 1-17.

Stucke, M. (2018). Here Are All the Reasons It's a Bad Idea to Let a Few Tech Companies Monopolize Our Data. Retirado de https://hbr.org/2018/03/here-are-all-the-reasons-its-a-bad-idea-to-let-a-few-tech-companies-monopolize-our-data.

UNESCO (2019). Beijing Consensus on Artificial Intelligence and Education: Planning education in the AI era: Lead the leap. Beijing: UNESCO.

UNESCO (2021). AI and Education – Guidance for policy-makers. Paris: UNESCO.

Vallance, C. (2023). AI could replace equivalent of 300 million jobs - report. Retirado de https://www.bbc.com/news/technology-65102150.

Vicente, P. N. (2023). Os Algoritmos e Nós. Lisboa: Fundação Francisco Manuel dos Santos.

Vignola, E., Barron, S., Plasencia, E. A., Hussein, M., & Cohen, N. (2023). Workers' health under algorithmic management: emerging findings and urgent research questions. International Journal of Environmental Research and Public Health, 20(2), 1-14.

Voss, A. (2021). Report on Artificial Intelligence in a digital age – Special Committee on Artificial Intelligence in a Digital Age. European Parliament.

World Economic Forum (2020). The Future of Jobs Report 2020. Geneva: World Economic Forum.

World Economic Forum (2023a). Top 10 Emerging Technologies of 2023. Geneva: World Economic Forum.

World Economic Forum (2023b). The Global Risks Report 2023 (18th Edition). Geneva: World Economic Forum.

Yudkowsky, E. (2023). Pausing AI Developments Isn't Enough. We Need to Shut it All Down. Retirado de https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/.

Zdravkova, K., Krasniqi, V., Dalipi, F., & Ferati, M. (2022). Cutting-edge communication and learning assistive technologies for disabled children: An artificial intelligence perspective. Frontiers in Artificial Intelligence, 5, 970430.

ORDEM DOS PSICÓLOGOS

ordemdospsicologos.pt