



Facultad de Ciencias Económicas y Empresariales
ICADE

Ética empresarial: requisitos en la adopción de la inteligencia artificial

Autora: María Dolores Yuste Sánchez
Director: José Luis Fernández Fernández

MADRID | Marzo 2024

ÍNDICE

1. INTRODUCCIÓN	5
1.1 OBJETIVO	5
1.2 METODOLOGÍA.....	5
1.3 CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TEMA.....	6
1.4 ESTRUCTURA.....	7
2. LA INTELIGENCIA ARTIFICIAL	8
2.1 CONCEPTO	8
2.2 USO EMPRESARIAL DE LA INTELIGENCIA ARTIFICIAL.....	10
3. RIESGOS DEL USO EMPRESARIAL DE LA INTELIGENCIA ARTIFICIAL	12
3.1 PROTECCIÓN DE DATOS	12
3.2 FALTA DE TRANSPARENCIA.....	13
3.3 SESGO ALGORÍTMICO – DISCRIMINACIÓN	14
4. REQUISITOS ÉTICOS A LA HORA DE IMPLEMENTAR LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO EMPRESARIAL	16
4.1 REGULACIÓN ACTUAL (UNIÓN EUROPEA)	16
4.2 PRIVACIDAD	17
4.3 TRANSPARENCIA.....	18
4.4 LIBERTAD Y AUTONOMÍA	19
5. PROPUESTA – PLAN EMPRESARIAL ESTRATÉGICO	20
5.1 ¿CÓMO IMPLEMENTAR LA INTELIGENCIA ARTIFICIAL EN LAS EMPRESAS DE MANERA ÉTICA?	21
5.1.1 MODELO OPERATIVO ADAPTADO A LA INTELIGENCIA ARTIFICIAL	22
5.1.2 EJEMPLOS REALES DE ENFOQUES Y ESTRATEGIAS.....	25
6. CONCLUSIÓN	28
7. BIBLIOGRAFÍA	32

RESUMEN

En los últimos años ha tenido lugar una notable revolución tecnológica a partir del desarrollo de la inteligencia artificial, cuya integración es particularmente considerable en el mundo empresarial debido a los numerosos beneficios que ofrece (como la automatización de procesos, la simplificación en la toma de decisiones o la personalización de servicios).

No obstante, es necesario tener en cuenta que la implementación empresarial de dicha tecnología presenta riesgos éticos significativos que deben ser abordados de manera efectiva (como pueden ser la falta de la protección de los datos utilizados, la opacidad en la forma en la que opera el sistema o la presencia de sesgos en el algoritmo), pues de lo contrario puede acarrear graves consecuencias como la vulneración de derechos de las personas involucradas, la propagación de información errónea o el fomento de la inequidad.

En este contexto, la privacidad, la transparencia, la libertad y la autonomía se configuran como los principales requisitos que las empresas deben garantizar para lograr una implementación ética de la inteligencia artificial. Para lograr este objetivo, se debe formular una estrategia que sea materializada y puesta en práctica a través de un modelo operativo que incluya dichos requisitos y que esté diseñado específicamente para esta tecnología emergente.

Palabras clave: inteligencia artificial, ética, riesgos, requisitos, modelo operativo.

ABSTRACT

In recent years, a remarkable technological revolution has taken place with the development of artificial intelligence, the integration of which is particularly significant in the business world due to the numerous benefits it offers (such as process automation, simplification of decision-making or personalization of services).

However, it is necessary to bear in mind that the business implementation of such technology presents significant ethical risks that must be effectively addressed (such as the lack of protection of the data used, the opacity in the way the system operates or the presence of biases in the algorithm), otherwise it can lead to serious consequences such as the violation of the rights of the people involved, the spread of misinformation or the promotion of inequity.

In this context, privacy, transparency, freedom and autonomy are the main requirements that companies must guarantee in order to achieve an ethical implementation of artificial intelligence. To achieve this goal, a strategy must be formulated, materialized and implemented through an operational model that includes these requirements and is specifically designed for this emerging technology.

Key words: artificial intelligence, ethics, risks, requirements, operating model.

1. INTRODUCCIÓN

1.1 OBJETIVO

El objetivo principal de este estudio es analizar cómo se puede implementar la inteligencia artificial de manera ética en las empresas.

Para ello, en primer lugar analizaré cuáles son los principales riesgos que tiene asociado el uso de esta tecnología en el ámbito empresarial y, a partir de ellos, identificaré los requisitos que garanticen su implementación ética y responsable. Concluiré el estudio planteando un modelo operativo alineado con dichos estándares éticos (e incluiré en un apartado modelos operativos reales a modo de ejemplo) con el fin de proporcionar una base que sirva de guía a posibles compañías que quieran introducir en sus procesos la inteligencia artificial de manera responsable.

Dado que se trata de un tema relativamente nuevo y en constante evolución, el valor que se pretende aportar es ofrecer un marco general sobre el estado actual de la inteligencia artificial en el mundo empresarial y una guía práctica que facilite su implementación de manera ética.

1.2 METODOLOGÍA

El enfoque general de la investigación es inductivo, pues parte de la recopilación de información sobre la adopción empresarial de la inteligencia artificial. Principalmente se tomarán como referencia artículos académicos, estudios especializados en el tema y documentos publicados por empresas en sus páginas web.

Por otra parte, la metodología utilizada es cualitativa. Analizaré la información recopilada para identificar patrones y así poder determinar los riesgos asociados al uso de esta tecnología, establecer los requisitos éticos necesarios para su correcta implementación empresarial y esbozar un modelo operativo pertinente.

1.3 CONTEXTUALIZACIÓN Y JUSTIFICACIÓN DEL TEMA

En los últimos años, ha tenido lugar una notable revolución tecnológica a partir del desarrollo de la inteligencia artificial. Cada vez son más los usos que se le da a esta tecnología en todos los ámbitos de la sociedad, pero su integración es particularmente notable en el mundo empresarial debido a los numerosos beneficios que ofrece (como la automatización de procesos o la simplificación en la toma de decisiones).

El problema identificado radica en que las empresas, atraídas por dichos beneficios, muchas veces incluyen la inteligencia artificial en sus operaciones sin estar plenamente preparadas para ello, dejando en un segundo plano sus riesgos asociados (ya sea porque al ser una tecnología relativamente nueva no hay todavía una comprensión completa de sus implicaciones o porque en el mundo empresarial no se le da la debida prioridad a la ética frente a, por ejemplo, la productividad o la optimización de recursos). Esta falta de preparación y conciencia sobre los riesgos éticos del uso de esta tecnología puede acarrear diversas y graves consecuencias, desde el daño a la reputación de la empresa o la pérdida de la confianza del cliente hasta la vulneración de ciertos derechos como el de la igualdad.

Se trata de un problema que afecta a todas las empresas, independientemente de su tamaño o sector. De hecho, incluso las grandes compañías tecnológicas (que en principio deberían estar mejor preparadas para enfrentarlo) son susceptibles de experimentar complicaciones éticas con el uso de la inteligencia artificial. Como prueba de ello, Google ha tenido recientemente un problema que pone de manifiesto las consecuencias que una inadecuada implementación de esta tecnología puede acarrear. En febrero de 2024, la empresa introdujo en su chatbot llamado “Gemini” la función de generar imágenes realistas, y pronto se dieron cuenta de que el sistema presentaba un sesgo significativo que dio lugar a contenido inexacto y ofensivo (por ejemplo, se abstenía de mostrar personas de raza blanca al solicitar imágenes de figuras históricas como los Padres Fundadores de los Estados Unidos o un soldado alemán durante la Segunda Guerra Mundial). Ante esta situación, el 23 de febrero Prabhakar Raghavan, el vicepresidente senior de la compañía, emitió un comunicado en el que informaba de que la función había sido desactivada y reconocía el error cometido (Raghavan, 2024).

Resulta alarmante destacar que este tipo de fallos no son casos aislados (pues con frecuencia se encuentran noticias de situaciones similares en periódicos) y, aunque puedan parecer menores a primera vista, pueden tener serias consecuencias como son la propagación de información errónea o el refuerzo de sesgos existentes que pueden propiciar situaciones de discriminación.

Es por ello por lo que es necesario llevar a cabo este estudio, no sólo porque se trata de una cuestión de candente actualidad, sino también para tratar de evitar que se den este tipo de situaciones, ofreciendo una visión general del uso empresarial de esta tecnología (sus riesgos, requisitos éticos...) y proporcionando pautas que puedan ayudar a las empresas a llevar a cabo una adecuada implementación.

1.4 ESTRUCTURA

El trabajo comienza con una breve exposición sobre la inteligencia artificial y su aplicación en el ámbito empresarial con el fin de proporcionar contexto y presentar el tema.

A continuación, se analizan los principales riesgos vinculados al uso empresarial de la inteligencia artificial. A partir de este análisis, se identifican los requisitos cuya garantía asegura que dicha implementación en las empresas sea realice de manera ética.

Por último, a modo de propuesta, el estudio concluye con un apartado que incluye las bases de un modelo operativo que las empresas pueden seguir para una adecuada y responsable implementación de la inteligencia artificial (basándose en los requisitos previamente identificados e incluyendo ejemplos reales de empresas que han aplicado estas prácticas).

2. LA INTELIGENCIA ARTIFICIAL

La inteligencia artificial (IA) es una tecnología emergente y disruptiva que ha transformado la forma en la que funciona el mundo. En este apartado abordaré su concepto (qué es, sus distintos tipos y su evolución en el tiempo) y analizaré su uso empresarial (las aplicaciones más comunes, las dificultades que las empresas afrontan en su uso, estrategias para una correcta implementación y expectativas de futuro).

2.1 CONCEPTO

La inteligencia artificial es una tecnológica que tiene como principal actividad la simulación de las actividades intelectuales del ser humano, centrándose para ello en el estudio y análisis de su comportamiento (comprensión, percepción, resolución de problemas y toma de decisiones). En su concepción más contundente, la postura conocida como inteligencia artificial “fuerte”, se defiende que los dispositivos no sólo son inteligentes y tienen mente, sino que pueden exhibir determinadas cualidades mentales a través de un funcionamiento lógico (Hardy, 2001).

Se trata de una tecnología presente en prácticamente todos los ámbitos de nuestra sociedad, siendo aplicada en una amplia variedad de sectores para automatizar procesos y mejorar funcionamientos. De esta manera, está presente en neveras, despertadores, teléfonos, satélites, automóviles, cajeros automáticos, cámaras, aviones, impresoras... Esto se debe en parte a los distintos tipos de inteligencia artificial que podemos identificar, cada uno con sus propias aplicaciones y características. Las principales categorías son las siguientes: la inteligencia artificial analítica, con gran aplicación en el ámbito empresarial y utilizada principalmente para realizar recomendaciones a través de la identificación, interpretación y comunicación de patrones de datos, la inteligencia artificial funcional, con gran aplicación en áreas como la robótica y utilizada para ejecutar acciones, la inteligencia artificial interactiva, con gran aplicación en el ámbito comercial ya que facilita una comunicación eficiente mediante la automatización, la inteligencia artificial textual, utilizada para realizar análisis de texto y procesar el lenguaje (como puede ser la conversión de voz a texto) y la inteligencia artificial visual, con gran aplicación en ámbitos como la visión por ordenador y realidad aumentada, y utilizada para el reconocimiento, clasificación y orden de elementos (Sarker, 2022).

La evolución de esta tecnología emergente ha sido exponencial. Alan Turing fue el primero en explorar la posibilidad matemática de la inteligencia artificial. En su artículo de 1950 “*Maquinaria computacional e inteligencia*” planteó la pregunta “¿Pueden las máquinas pensar?”, y argumentó que, si los humanos utilizaban la información de la que disponían y la razón para resolver problemas y tomar decisiones, las máquinas también podrían hacerlo. No obstante, el nacimiento de la inteligencia artificial se suele marcar con el “*Proyecto de Investigación de Verano sobre Inteligencia Artificial de Dartmouth*” organizado en 1956 por John McCarthy y Marvin Minsky, que se considera como su primera presentación. Por otro lado, una de las primeras demostraciones de algoritmos de aprendizaje automático fue el programa *General Problem Solver (GPS)* desarrollado en 1957 por Allen Newell y Herbert Simon, proyecto que supuso prometedores avances hacia la creación de una máquina con capacidad de resolver problemas (Anyoha, 2017).

En la década de los años 70 tiene lugar el desarrollo de distintos trabajos que sentaron las bases fundamentales para el estudio de esta disciplina. Uno de los logros más destacados fue la creación de *Dendral* en la Universidad de Stanford, el primer sistema experto que tenía la capacidad de realizar tareas propias de un químico, como la reconstrucción de la estructura molecular de un compuesto orgánico. En esta época también se desarrollan importantes programas como *PROLOG*, que marcó los inicios de la programación basada en la lógica de primer orden, y *MYCIN*, un sistema aplicable al ámbito de la medicina diseñado para diagnosticar infecciones bacterianas en la sangre. La década de los años 80 se caracteriza por la aplicación de la inteligencia artificial al ámbito económico y por el proyecto lanzado en Japón de desarrollar computadoras de quinta generación. Por último, es en la década de los años 90 en el que se empieza a aplicar la inteligencia artificial en diversos campos relacionados con la comunicación entre humanos y máquinas (Hardy, 2001). Desde entonces, en los últimos 20 años la inteligencia artificial ha ido evolucionando poco a poco hasta convertirse en el fenómeno que es hoy en día.

En relación con qué se puede esperar en el futuro, es relevante mencionar la Ley de Moore, la cual explica que la velocidad de los ordenadores, medida por la cantidad de transistores que pueden colocarse en un único chip, se duplica aproximadamente cada uno o dos años (Mollick, 2006). Con un crecimiento exponencial de la capacidad de procesamiento de los chips, se prevé que la inteligencia artificial siga desarrollándose y que se amplíe aún más su aplicación en los diversos ámbitos de la sociedad.

2.2 USO EMPRESARIAL DE LA INTELIGENCIA ARTIFICIAL

La implementación de la inteligencia artificial en el mundo empresarial ha transformado significativamente la forma en la que las empresas operan. Se trata de un recurso muy valioso que permite desde la automatización de procesos hasta la toma de decisiones basadas en el análisis de datos, por lo que se configura como una herramienta competitiva que permite mejorar la eficiencia, ahorrar tiempo y reducir costes.

En un estudio llevado a cabo en 2023 por IBM, en el que se entrevistó a 2.342 profesionales trabajando en el área de tecnología de la información (IT) en empresas con más de 1.000 empleados en 20 países diferentes, el 42% de los trabajadores manifestaron que en sus empresas ya se habían implementado la inteligencia artificial, y un 40% adicional reportó que estaban llevando a cabo exploración activa sobre ello. Además, el estudio reveló que las empresas del sector de servicios financieros son las más propensas a utilizar la inteligencia artificial en sus procesos, con casi la mitad de sus profesionales respondiendo afirmativamente respecto a su implementación. Para tener cierta perspectiva del incremento del uso de la inteligencia artificial en las empresas, desde abril de 2023 su aplicación ha incrementado un 13% en Japón, un 14% en Singapur, un 16% en Corea del Sur y un 21% en Reino Unido (IBM, 2023).

Para analizar el uso que las empresas hacen de la inteligencia artificial, una encuesta llevada a cabo por *Forbes Advisor*, en la que se preguntó a 600 propietarios de empresas que ya la utilizaban o planeaban incorporarla en sus negocios, reveló que las aplicaciones empresariales más comunes se daban en las áreas de servicio al cliente (56%), ciberseguridad y gestión del fraude (51%), asistentes personales digitales (47%), gestión de relaciones con el cliente (46%) y gestión del inventario (40%) (Haan, 2023).

No obstante, pese a la ventaja competitiva que supone su uso, la implementación de inteligencia artificial en empresas no es una gestión exenta de dificultades. Austin Ambrozi, cofundador y COO de una empresa de inteligencia artificial llamada Doxci, además de las preocupaciones empresariales más evidentes y notorias como pueden ser la excesiva dependencia en proveedores de servicios externos o la pérdida del elemento humano en el servicio al cliente, recopiló, entre otros, los siguientes desafíos a los que se enfrentan las empresas al empezar a utilizar la inteligencia artificial: la posible falta de

formación en la plantilla que puede causar una brecha en las habilidades requeridas, líderes que no tienen en cuenta y no abordan la incertidumbre que estas tecnologías generan en los empleados (en cuanto a ser reemplazados y su futuro en la compañía), no saber diferenciar qué tareas automatizar y cuales no (aquellas que requieren intervención humana) o no encontrar servicios desarrollados para gestionar dicha implementación en la empresa (Ambrozi, 2023).

Por ello, a la hora de implementar la inteligencia artificial, las empresas deben de seguir una estrategia y hacerlo de manera organizada. En primer lugar, es necesario conocer las distintas tecnologías disponibles (cómo funcionan, fortalezas, limitaciones...) a través de la investigación y formación. También es importante, aunque normalmente sea un factor que pase desapercibido, la receptividad de la plantilla al uso de inteligencia artificial, pues es crucial que los trabajadores estén motivados y dispuestos a aprender y adaptarse, y no adopten una postura de rechazo ante innovaciones. Otro factor esencial es que la empresa cuente con la infraestructura necesaria para poder implementar y hacer un uso óptimo de la inteligencia artificial (capacidad interna de ciencia de datos o análisis), para así evitar tener que acudir a proveedores de servicios externos. Por último, es necesario analizar sistemáticamente las necesidades de la empresa (ya sea mediante talleres o mediante contratos de consultoría): identificar las oportunidades, determinar los casos de uso, evaluar la tecnología utilizada y rediseñar si es necesario los procesos empresariales (Davenport & Ronanki, 2018).

En la configuración del futuro uso empresarial, se anticipa que la inteligencia artificial generativa (como ChatGPT o Stable Diffusion) sea la que impulse la siguiente revolución de productividad. En su aplicación en los negocios, se estima que cerca del 75% del valor generado recaerá principalmente en las áreas de operaciones de clientes, marketing y ventas, ingeniería de software e investigación y desarrollo, siendo el sector bancario uno de los más afectados (la inteligencia artificial generativa podría generar en él entre 200.000 y 340.000 billones de dólares anuales). Debido al potencial de la automatización técnica, se estima que aproximadamente la mitad de las actividades laborales actuales puedan ser sustituidas entre 2030 y 2060, teniendo en cuenta que la inteligencia artificial generativa podría llegar a automatizar aquellas diligencias que en la actualidad ocupan entre el 60 y 70% del tiempo de los empleados (McKinsey & Company, 2023).

3. RIESGOS DEL USO EMPRESARIAL DE LA INTELIGENCIA ARTIFICIAL

La implementación de la inteligencia artificial en las empresas, aunque pueda ser muy beneficiosa en términos de productividad, puede resultar peligrosa si no se realiza de manera responsable.

Los principales riesgos identificados en el uso empresarial de esta tecnología son la falta de protección de los datos que utiliza (colocando en una posición vulnerable tanto a los usuarios como a las propias empresas), la falta de transparencia en la forma en la que opera (opacidad que genera incertidumbre sobre su funcionamiento e impide la posible rendición de cuentas) y la presencia de sesgos en el algoritmo (que da lugar a resultados discriminatorios y fomenta la inequidad).

3.1 PROTECCIÓN DE DATOS

La principal preocupación que deriva del uso de la inteligencia artificial es la privacidad de la información de los sistemas que la emplean (que puede verse comprometida ya sea por brechas de seguridad o por aspectos técnicos y organizacionales). Pese a que las empresas que utilizan sistemas tecnológicos reconocen públicamente la importancia de la privacidad y seguridad de sus usuarios, se enfrentan a muchas dificultades a la hora de adaptar sus procesos a esas exigencias de protección de datos ya que la tecnología utilizada todavía es moldeable y está en constante evolución (Brusoni & Vaccaro, 2017).

Parte de la problemática reside en que la inteligencia artificial basada en el aprendizaje automático necesita una gran cantidad de datos para poder desarrollarse y tiene capacidad de detectar patrones (incluso cuando no hay acceso directo a determinada información). Es decir, las tecnologías utilizadas en la inteligencia artificial tienen la capacidad de generar datos que de otra manera serían privados o no estarían disponibles. Un ejemplo claro de esto se puede observar en el estudio llevado a cabo en 2009 por Jernigan y Mistree, en el que se afirma poder identificar la orientación sexual de una persona por sus amistades en la red social Facebook. Además, este tipo de inteligencia artificial puede generar y almacenar datos personales más sensibles, como los de naturaleza emocional, agravando aún más la situación (Carsten, 2021).

En el contexto empresarial, un ejemplo novedoso y conocido de un producto que utiliza innovadores programas de inteligencia artificial (y que plantea una gran inquietud en términos de privacidad) son los asistentes digitales basados en voz como “Alexa” de Amazon o “Siri” de Apple. Se trata de un producto que se encuentra presente en dispositivos como teléfonos móviles, altavoces o televisores, y que recopila, usa y almacena una gran cantidad de información de sus usuarios para así personalizar su experiencia y ofrecer un funcionamiento más individualizado. Este tipo de programas procesan datos en tiempo real y perfeccionan sus respuestas basándose en la retroalimentación de acciones anteriores, generando serias preocupaciones de privacidad en relación con la información personal de sus usuarios (contactos, historial de búsqueda, localización, consultas de voz...) (Vimalkumar et al., 2021).

Esta concentración de datos personales en los servidores de empresas no solo plantea un dilema ético con relación al acceso a información tan personal, sino que también supone el peligro de ser el objetivo de posibles robos de datos por parte de terceros (principalmente para cometer fraudes de identidad). En este contexto hay que tener en cuenta que, pese al énfasis empresarial en la implementación de controles de seguridad de datos, los hackers continúan encontrando formas de vulnerarlos para obtener acceso no autorizado a información sensible. Un ataque de este tipo puede tener serias consecuencias tanto para la empresa (pérdidas financieras, daño reputacional, acciones legales y pérdida de información sensible) como para los usuarios afectados (falta de privacidad y de seguridad) (Barati & Yankson, 2022).

3.2 FALTA DE TRANSPARENCIA

El auge en el uso de la inteligencia artificial y los sistemas de aprendizaje automático han dado lugar a sistemas que, en gran medida, pueden tomar decisiones autónomas. El problema reside en que, por definición, los sistemas de aprendizaje automático que utiliza la inteligencia artificial no son transparentes a la hora de producir y tratar información, y si se trata de sistemas/algoritmos con un propietario determinado, la confidencialidad comercial puede limitar aún más la transparencia (Carsten, 2021).

La “paradoja de la transparencia” (una gran cantidad de información puede generar una mayor opacidad en vez de una mayor explicabilidad) se utiliza para advertir de que el acceso y procesamiento de los datos utilizados por los sistemas de inteligencia artificial en numerosas ocasiones se hace de manera unidireccional y que los propios algoritmos son capaces de generar nuevos algoritmos cada vez más complejos, hasta el punto de llegar a ser incomprensibles. Además, el procesamiento de datos es una tarea compleja, siendo difícil seguir sus distintas fases de manera exhaustiva (Blázquez, 2022). Esta falta de claridad, en ocasiones utilizada por los propios creadores de los algoritmos para mantener la confidencialidad de sus diseños, da lugar a cierta imposibilidad de conocer cómo se toman decisiones o se generan resultados por parte de los sistemas basados en la inteligencia artificial.

Un equipo de investigadores de la Universidad de Stanford, el Instituto de Tecnología de Massachusetts (MIT) y la Universidad de Princeton evaluó la transparencia de 10 de los principales desarrolladores de modelos base de inteligencia artificial (incluyendo gigantes de la industria como Meta, Amazon o Google). Los hallazgos de la investigación indicaron una notable falta de transparencia, con una calificación promedia de 37 puntos sobre 100. Además, se observó una disparidad en la transparencia entre las distintas empresas analizadas (lo que resalta la necesidad de adoptar prácticas estandarizadas para armonizar los distintos sistemas utilizados), y las puntuaciones más bajas se obtuvieron en el ámbito de la creación de datos, el estado de los derechos de autor y la licencia de los datos, lo que refleja una falta de divulgación sobre el proceso de desarrollo de la inteligencia artificial (Bommasani et al., 2023).

3.3 SESGO ALGORÍTMICO – DISCRIMINACIÓN

Uno de los elementos más notorios para tener en cuenta al tratar la preparación de los algoritmos de la inteligencia artificial es la presencia de sesgos humanos, ya sean sesgos de género o de raza. Es decir, estos sistemas pueden reflejar prejuicios humanos presentes en la sociedad, que incluyen desigualdades sociales históricas y actuales. La razón de ello es que para desarrollar dichos algoritmos y lograr precisión se requiere una gran cantidad de datos, datos que son previamente seleccionados por humanos (que pueden incluir sesgos existentes) y que posteriormente se transfieren al algoritmo (Baker-Brunnbauer, 2021).

La fuente de los posibles sesgos puede variar. En primer lugar, puede tener su origen en los datos de entrenamiento, al incluir grupos excesivamente representados y obviar otros (por ejemplo, si se entrenan los datos para un algoritmo de reconocimiento facial únicamente con personas caucásicas) o al etiquetar dichos datos (por ejemplo, excluyendo ciertas características en las herramientas de contratación). Por otra parte, puede tener su origen en el propio algoritmo, ya sea por el uso de datos de entrenamiento defectuosos o por errores de programación (por ejemplo, estableciendo indicadores excluyentes como puede ser un único idioma). Por último, pueden tener su origen en los prejuicios de las personas que trabajan con el sistema de inteligencia artificial, que incluyen sus preferencias y propios sesgos a la hora de determinar la selección de datos y cómo éstos se ponderan (Equipo de datos e inteligencia artificial de IBM, 2023).

La presencia de dichos sesgos puede tener un impacto muy negativo tanto para los individuos como la sociedad en su conjunto, dando lugar a resultados estereotipados, prejuiciosos o discriminatorios que perpetúan o incluso amplían desigualdades existentes.

Un estudio llevado a cabo sobre el sesgo de género en GPT-3 reveló que las palabras asociadas a puestos de trabajo suelen tener sesgos sociales asociados que resultan en la suposición de que la mayoría de los trabajadores son, por defecto, hombres. De hecho, la investigación reveló que el 83% de las 388 profesiones analizadas tenían una mayor probabilidad de estar asociadas con un identificador masculino. También se realizó otro estudio sobre el sesgo de religión en GPT-3 que reveló que, entre otras, algunas de las palabras asociadas al cristianismo eran “críticos” e “ignorantes”, al budismo eran “vegetarianos” y “sabios”, y al judaísmo eran “racistas” y “listos” (Brown et al., 2020).

En el ámbito empresarial, es necesario tener presente esta posible existencia de sesgos para evitar discriminaciones. Una clara muestra de esto se observa en los procesos de selección que emplean inteligencia artificial para filtrar candidatos, donde pueden darse sesgos de género, raza, color y personalidad que generen oportunidades laborales injustas e inequidad. Si por ejemplo se entrena al sistema con distintos *curriculum vitae* prototipo y éstos son mayoritariamente de empleados masculinos, el algoritmo de reclutamiento percibirá ser hombre como una ventaja para la empresa (empeorando las posibilidades de los demás candidatos que no reúnan esa condición) (Chen, 2023).

4. REQUISITOS ÉTICOS A LA HORA DE IMPLEMENTAR LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO EMPRESARIAL

A medida que la inteligencia artificial está cada vez más presente en el ámbito empresarial, es crucial delimitar un acervo moral (libertad, seguridad, transparencia...) que fije los límites éticos necesarios para que el uso de dicha tecnología salvaguarde el bienestar de todas las partes involucradas.

La existencia de los riesgos previamente analizados (falta en la protección de los datos utilizados, falta de transparencia en el funcionamiento del sistema y posible presencia de sesgos en el algoritmo) es la razón expresa de esta necesidad, pues las personas pueden quedar desprotegidas y puede darse una vulneración de sus derechos fundamentales. Al establecer unos requisitos éticos como límites que las empresas deben respetar, se asegura la seguridad de los derechos de las personas ante el peligro de un uso irresponsable de la inteligencia artificial.

4.1 REGULACIÓN ACTUAL (UNIÓN EUROPEA)

La Unión Europea ha desempeñado un papel fundamental y pionero al regular el uso de la inteligencia artificial, sirviendo como referente para otras regiones y países que buscan abordar los desafíos éticos y legales asociados con esta tecnología emergente.

En primer lugar, el Grupo independiente de expertos de alto nivel sobre inteligencia artificial constituido por la Comisión Europea en junio de 2018 publicó un documento llamado “*Directrices Éticas para una IA fiable*”, que basada la fiabilidad de dicha tecnología en tres componentes fundamentales: que fuese lícita, ética y robusta. Para ello, se identificaron cuatro principios éticos que debían servir de base (respeto de la autonomía humana, prevención del daño, equidad y explicabilidad) y siete requisitos clave que debían cumplirse en dichos sistemas (acción y supervisión humanas, solidez técnica y seguridad, gestión de la privacidad y de los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas) (Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, 2019).

Por otra parte, es importante mencionar una normativa que afecta directamente a las empresas que utilizan inteligencia artificial: el Reglamento General de Protección de Datos (RGPD) de la Unión Europea, norma del ámbito de la privacidad y seguridad que regula la forma en la que se deben tratar los datos personales de las personas físicas. El Reglamento identifica cuáles son los derechos fundamentales de las personas en la era digital, las obligaciones que deben tener aquellos que procesan datos (como implementar medidas de seguridad pertinentes), las formas de garantizar el cumplimiento de las normas establecidas y las sanciones para quienes las infrinjan. Algunos de los aspectos esenciales de la normativa hacen referencia a derechos individuales como el derecho a supresión de datos (conocido como el derecho al olvido), un nuevo derecho a la portabilidad de los datos o el derecho a ser informado en caso de darse una violación en la seguridad de los datos personales.

Por último, en diciembre de 2023 el Consejo de la Unión Europea y el Parlamento Europeo llegaron a un acuerdo provisional sobre la que será la primera Ley de Inteligencia Artificial (“*IA Act*”) del mundo, marcando un importante precedente en la regulación de la inteligencia artificial. El principal objetivo del proyecto es garantizar la seguridad y el respeto de los derechos fundamentales de los ciudadanos al usar sistemas de inteligencia artificial (además de estimular la innovación e inversión en Europa), y se prevé su implementación plena en 2026. En relación con el ámbito empresarial, las multas por violaciones de la Ley se establecerán en base a un porcentaje de los ingresos anuales de la empresa en cuestión o una cantidad preestablecida (la mayor de las dos): el 7% o 35 millones de euros para aplicaciones de IA prohibidas, el 3% o 15 millones de euros para violaciones de las obligaciones establecidas en la ley y el 1,5% o 7,5 millones de euros por el suministro de información errónea (Consejo de la Unión Europea, 2023).

4.2 PRIVACIDAD

Dado que la recopilación de información es un componente fundamental en el funcionamiento de las empresas que utilizan inteligencia artificial, y teniendo en cuenta que dicha información puede incluir datos de clientes, patrones de comportamiento, preferencias individuales y más tipos de información sensible, la privacidad y la protección de los datos (que en el contexto actual son un activo corporativo crítico) deben ser una prioridad empresarial.

El desarrollo de un plan integral de protección de datos empresariales (Todt, 2019) es esencial para garantizar que aquellas empresas que implementen inteligencia artificial lo hagan de manera ética y salvaguardando la privacidad de sus usuarios. Este plan debe abordar elementos clave para no solo proteger la información sensible sino también mitigar los riesgos asociados con su almacenamiento.

En primer lugar, es necesario crear un inventario de todos los datos que la empresa recopila, procesa y almacena. Esto conlleva identificar y marcar el origen de dichos datos y llevar un seguimiento de la forma en la que posteriormente se utilizan. De esta manera, la empresa puede obtener una visión detallada de la información en circulación y se facilita la implementación de medidas de seguridad apropiadas y adaptadas para cada caso concreto.

Por otra parte, la empresa debe divulgar públicamente su política de privacidad para garantizar la transparencia y confianza de los usuarios. Es importante que se informe de manera clara de la forma en la que los datos se recopilan, utilizan y comparten, y que los individuos puedan conocer en profundidad cómo se maneja su información personal y los derechos que tienen en relación con la privacidad de ésta.

Por último, es crucial desarrollar una respuesta empresarial efectiva y estratégica ante posibles incidentes/brechas de seguridad. Toda empresa que trabaja con datos debe contar con protocolos claros y efectivos para gestionar posibles incidentes y poder responder rápidamente ante cualquier problema de seguridad que pueda surgir.

4.3 TRANSPARENCIA

Dada la capacidad de los sistemas de inteligencia artificial de tomar decisiones de manera automatizada, las empresas se enfrentan a la necesidad de dar a conocer la forma en la que dichos sistemas operan y gestionan la información. La transparencia implica divulgar de manera clara y detallada el proceso de funcionamiento de los algoritmos para así asegurar que éstos sean comprensibles y justificables para cualquier parte interesada (qué datos son utilizados, cómo son utilizados, para qué son utilizados...).

No obstante, no basta únicamente con la publicación del algoritmo. Evert Haasdijk, directivo senior y miembro del Centro de Excelencia de Inteligencia Artificial en Deloitte, explica que para que la inteligencia artificial sea transparente debe ser explicable: la transparencia no se consigue únicamente por la publicación de los algoritmos en internet, también es fundamental que se proporcionen las explicaciones necesarias para poder entender de forma clara cómo los modelos en cuestión llegan a sus decisiones. También declara que el nivel de transparencia exigido debe depender del impacto de la tecnología utilizada (a mayor impacto mayor necesidad de explicabilidad y transparencia), no siendo equiparable la importancia de conocer como un algoritmo utilizado para enviar mensajes comerciales personalizados funciona a como lo hace uno utilizado para otorgar créditos bancarios. Este tipo de modelos que toman decisiones de alto impacto únicamente deberían ser permitidos si cumplen con los más altos estándares de transparencia y responsabilidad (Deloitte, 2019).

La transparencia es el cimiento sobre el cual se construye la confianza en estos algoritmos, ya que tanto los usuarios como la alta dirección empresarial requieren una comprensión clara sobre el funcionamiento del proceso de toma de decisiones para poder aceptar las conclusiones a las que los sistemas llegan. Los usuarios demandan una mayor claridad en el proceso para así poder comprobar que sus derechos no sean vulnerados y la alta dirección busca entender el fundamento detrás de las decisiones técnicas para así poder ajustarlas en caso de ser necesario (García, 2019).

4.4 LIBERTAD Y AUTONOMÍA

La preservación de la libertad y la autonomía humana se rigen como un requisito ético a la hora de configurar y utilizar un sistema de inteligencia artificial, siendo fundamental que las personas mantengan el control con relación a las decisiones que les puedan afectar (aún cuando esas decisiones estén asistidas/tomadas por este tipo de tecnologías).

Los modelos de inteligencia artificial deben poder asegurar la libertad y autonomía plena y efectiva de las personas que los usan, sin de ninguna manera subordinarlas, manipularlas, condicionarlas o dirigir las injustificadamente. Para ello, estos modelos deben de diseñarse con el objetivo de complementar y potenciar sus aptitudes sociales, culturales y cognitivas, permitiéndoles interactuar e intervenir en el proceso de

funcionamiento (e incluso evaluar y modificar el sistema en caso de ser necesario). De esta manera, se consigue salvaguardar la libertad individual en sus diversas manifestaciones (libertad de conciencia, de expresión, de pensamiento...) (OdiseaIA & PwC, 2022).

Al fomentar la libertad y la autonomía en la adopción de la inteligencia artificial, las empresas se comprometen a promover un entorno digital inclusivo en el que todas las personas involucradas puedan prosperar y desarrollarse plenamente, sin que la tecnología suponga un sustituto ni suprima su capacidad de decisión.

5. PROPUESTA – PLAN EMPRESARIAL ESTRATÉGICO

La adopción de sistemas de inteligencia artificial en el entorno empresarial está creciendo de manera exponencial, lo que está transformando profundamente los modelos de negocio convencionales. No obstante, este avance no está exento de desafíos y de riesgos éticos, lo que recalca la necesidad de diseñar un plan empresarial estratégico para poder integrar esta tecnología de manera sostenible y responsable, cumpliendo con los requisitos éticos correspondientes.

El diseño de un plan estratégico basado en la implementación ética de la inteligencia artificial en las empresas es importante a la vez que necesario, pues en el ámbito empresarial, la ética, pese a su importancia, no siempre recibe la debida prioridad. A pesar de que algunas compañías han empezado a tomar iniciativas como la publicación de directrices y principios para el desarrollo y utilización de la inteligencia artificial o la contratación de ingenieros para el desarrollo de soluciones técnicas a problemas relacionados con la ética y justicia en la aplicación de este tipo de tecnología, todavía hay muchas empresas que no reconocen la importancia de esta cuestión. Las organizaciones suelen explorar y discutir desde una perspectiva teórica la ética de la inteligencia artificial, pero lo que no suelen hacer es “recorrer el camino” necesario para ello, emprendiendo acciones concretas, proporcionando los recursos necesarios y capacitando a los equipos involucrados. Es decir, en lo que respecta a la ética de la inteligencia artificial, hay una marcada brecha entre la política empresarial y la práctica de ella (Ali et al., 2023).

Ignorar o minimizar los riesgos éticos asociados con las tecnologías emergentes conlleva serias consecuencias para las empresas. En su segundo informe anual sobre el “Estado de la Ética y la Confianza en la Tecnología”, Deloitte realizó una encuesta a 1.700 profesionales empresariales y técnicos en el que se preguntaba sobre los daños que suponía un problema ético relacionado con este tipo de tecnología y los resultados (ordenados en función de la gravedad percibida) fueron los siguientes: un 38% de daño reputacional, un 27% de daño humano, un 17% de penalizaciones regulatorias, un 9% de daño financiero y un 9% de insatisfacción de los empleados (Ammanath, 2023). Por ello, ignorar el impacto significativo que la inteligencia artificial está teniendo en el mundo empresarial ya no es una opción. Es fundamental que las organizaciones reconozcan este cambio y se configuren de tal manera que puedan adaptarse de manera ética y responsable.

5.1 ¿CÓMO IMPLEMENTAR LA INTELIGENCIA ARTIFICIAL EN LAS EMPRESAS DE MANERA ÉTICA?

A la hora de implementar la inteligencia artificial de manera ética en una empresa, es esencial la elaboración de un plan estratégico que garantice el cumplimiento de ciertos requisitos como la privacidad, la transparencia, la libertad o la autonomía y que aborde de manera integral la gestión de los riesgos asociados como el sesgo algorítmico o la falta de protección de datos.

Configurar un plan estratégico sobre la implementación de la inteligencia artificial supone un cambio empresarial significativo y, a menudo, los esfuerzos de transformación de una compañía se inician con la revisión de su modelo operativo. De hecho, los líderes empresariales consideran que establecer un modelo operativo efectivo es el principal obstáculo para lograr una transformación digital sostenible y duradera. Una organización aumenta significativamente sus probabilidades de éxito cuando su modelo operativo se alinea con su estrategia general, por lo que para que la implementación tecnológica tenga éxito, es necesario que los equipos directivos ajusten y revisen dichos modelos operativos (Kwan et al., 2023)

Un modelo operativo (¿Qué estamos haciendo?) bien estructurado y definido sirve de puente entre la estrategia (¿Por qué lo estamos haciendo?) de la organización y sus operaciones diarias (¿Cómo lo estamos haciendo?), y es un elemento clave para materializar la visión empresarial. No hay un único modelo operativo para todas las empresas, éstos son adaptables y dependerán del tamaño de la organización, su complejidad y los objetivos que pretende cumplir. No obstante, sí hay ciertos elementos “estándar” que todo modelo operativo debería incluir: principios base de diseño (traducen la estrategia en un conjunto de normas a seguir y proporcionan rigor y estructura al proceso), gobernanza (define la toma de decisiones y establece el marco de balanza entre las partes interesadas y la mitigación de riesgos), cultura y valores (establece los puntos clave que deben estar presentes en el funcionamiento de la organización) y procesos (concreta las capacidades de la empresa) (Murphy et al., 2016).

5.1.1 MODELO OPERATIVO ADAPTADO A LA INTELIGENCIA ARTIFICIAL

Antes de diseñar un modelo operativo, los líderes deben de tener claro el tipo de empresas que quieren liderar (en este caso, una empresa que hace uso de la inteligencia artificial de manera ética y responsable) y el conjunto de capacidades de las que dispone. De esta manera, se consigue sentar las bases principales para un diseño que se alinee con la estrategia empresarial.

Por otra parte, a la hora de diseñar un modelo operativo que pueda ayudar a guiar una transformación digital sostenible y exitosa, es conveniente abordar los siguientes seis puntos que guiarán el proceso. Se trata de medidas que tienen como objetivo lograr la alineación efectiva entre la estrategia organizacional, la estructura y las operaciones del día a día, fomentado de esta manera un entorno favorable para la implementación de la inteligencia artificial. Los puntos son los siguientes (Kwan et al., 2023):

1. Empoderar a los líderes funcionales y empresariales para promover la transformación en la organización
2. Establecer de manera clara los roles y responsabilidades entre las distintas unidades empresariales
3. Implementar sistemas de incentivos para fomentar la colaboración

4. Establecer sistemas de informes multifuncionales para mantener a todas las partes pertinentes informadas
5. Implementar un modelo de gobierno que incluya indicadores clave de rendimiento (KPI) para cada equipo de liderazgo
6. Normalizar e impulsar el intercambio de recursos y conocimientos dentro de la empresa.

En cuanto a los principios base de diseño del modelo, para definirlos es necesario preguntarse: ¿Cómo deberá trabajar la organización para alcanzar los objetivos estratégicos? (Murphy et al., 2016).

En el marco de un modelo operativo adaptado al uso de la inteligencia artificial de manera ética, la organización deberá trabajar cumpliendo con los requisitos éticos identificados: transparencia, privacidad, autonomía y libertad. Esto implica que la empresa debe establecer normas claras que guíen el desarrollo sus políticas y prácticas, y que éstas estén alineadas con estos principios: que promuevan la explicabilidad en el funcionamiento de los sistemas de inteligencia artificial, prioricen la protección de datos personales y respeten la autonomía y libertad de los individuos.

Como ejemplo, desde 2019 Google publica cada año un documento que incluye los principios que inspiran su uso de inteligencia artificial. Los principios de la empresa para el último año eran los siguientes: que el uso de IA sea socialmente beneficiosa, evitar crear o reforzar prejuicios injustos, que el sistema sea construido y probado para así ofrecer una mayor seguridad, tomar responsabilidad ante las personas, incorporar principios de privacidad en los diseños, mantener altos estándares de excelencia y que los usos de esta tecnología sean acordes con todos los demás principios (Google, 2023).

La gobernanza es el segundo aspecto clave que tratar ya que, a pesar de que el 79% de los ejecutivos reconocen que la ética de la inteligencia artificial es importante, menos del 25% ha implementado principios de gobernanza ética (Montgomery & Rossi, 2023).

A la hora de proporcionar directrices y marcos, es fundamental que las empresas consideren la gobernanza como un aspecto clave para implementar la inteligencia artificial, ya que de esta manera se puede lograr no sólo confianza y eficacia en el proceso

sino también garantizar el cumplimiento de los requisitos éticos establecidos. Hay que tener en cuenta que la responsabilidad de la gobernanza de la inteligencia artificial no se limita a un único individuo o departamento, sino que debe ser una responsabilidad compartida en toda la organización. Por ejemplo, el director ejecutivo y la alta dirección se encargan de establecer el tono y la cultura general de la organización (un uso ético de la tecnología), los asesores legales se encargan de evaluar y mitigar los posibles riesgos jurídicos (asegurando que la aplicación de la inteligencia artificial cumple con las leyes y regulaciones pertinentes) y los equipos de auditoría se encargan de verificar la integridad de los datos utilizados y de que funcionen según lo previsto (sin sesgos ni errores) (Mucci & Stryker, 2023).

Una recomendación es establecer un Comité de la Ética de la Inteligencia Artificial interdisciplinar formado por especialistas en ética, abogados, técnicos y estrategias de negocios. La función principal de dicho Comité debe ser la de velar por una correcta implementación de la inteligencia artificial en la empresa e identificar y mitigar posibles riesgos en su uso. Además, con relación a la autoridad que debe tener, teniendo en cuenta la importancia de ser ético en el uso de estas tecnologías, es recomendable que las consultas del Comité no solo sean recomendaciones, sino que sean de obligado cumplimiento y que se le otorgue el poder de vetar determinadas propuestas que no cumplan los requisitos establecidos (de esta manera se asegura un impacto real en la empresa) (Blackman, 2022).

En relación con la cultura y los valores, se trata de un aspecto importante del modelo operativo pues éstos proporcionan una base sólida y orientativa que sirve de guía para todos los empleados (desde directores de marketing hasta ingenieros de datos) para la implementación adecuada de la inteligencia artificial.

La empresa debe elegir los valores corporativos que quiere que guíen todo el proceso de implementación de esta tecnología (ya sean nuevos o existentes). Los principios incluidos en el documento “*Directrices Éticas para una IA Fiable*” desarrollado por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial creado por la Comisión Europea pueden servir como buena referencia. Entre ellos, se incluyen la acción y supervisión humana, la solidez técnica y seguridad, la diversidad, no discriminación y

equidad, el bienestar social y ambiental y la rendición de cuentas (Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, 2019).

Por último, en cuanto a la parte de procesos, que concreta las capacidades de la empresa, es fundamental que los equipos posean las habilidades técnicas necesarias para implementar de manera adecuada la inteligencia artificial.

Las capacidades funcionales esenciales en este tipo de modelo operativo incluyen contar con proveedores de datos, operaciones de la Tecnología de la Información (“IT”), recursos especializados en ciencia de los datos y de la inteligencia artificial, gobierno de datos, consumidores de datos y gestión de entrega de los servicios de inteligencia artificial (IBM, 2020).

5.1.2 EJEMPLOS REALES DE ENFOQUES Y ESTRATEGIAS

IBM es una de las empresas pioneras en el desarrollo e implementación de estrategias éticas para el uso empresarial de la inteligencia artificial, estableciendo los principios de confianza y transparencia como sus principios base.

Analizando su gobernanza, está se basa en 4 pilares fundamentales. En primer lugar, cuentan con una Junta de Ética de Inteligencia Artificial, un órgano central e interdisciplinario que tiene como misión respaldar un proceso unificado de gestión, dirección y evaluación de las prácticas de IBM. También cuentan con un Comité Asesor de Políticas, que está formado por líderes de alto nivel que se encargan de supervisar y ayudar a la Junta Ética de Inteligencia Artificial en el establecimiento de las estrategias pertinentes. Por otra parte, cuentan con representantes capacitados en la ética de la inteligencia artificial de distintas unidades de negocio que actúan como los primeros puntos de contacto encargados de identificar cualquier problema ético que pueda surgir, y con una “red de defensa” formada por empleados de distintos niveles cuya función es promover los principios éticos de IBM dentro de sus equipos y operaciones. (Montgomery & Rossi, 2023).

Además de estos cuatro puntos focales de gobernanza, IBM cuenta con una Oficina del Proyecto de Ética del Jefe de Privacidad (Chief Privacy Office), que actúa como intermediario con los roles de gobierno para asegurar que se cumplen las prioridades éticas establecidas (Montgomery & Rossi, 2023).

Otra empresa que destacar en este ámbito es Telefónica, que ha establecido como principios base para guiar su implementación ética de inteligencia artificial la justicia, transparencia y explicabilidad, privacidad y seguridad desde el diseño, el enfoque en las personas como prioridad y su aplicación a la relación con socios y terceras partes (Telefónica, 2018).

Telefónica desde hace años funciona con un Gobierno de la Privacidad, un modelo de gobernanza basado en el establecimiento de responsabilidades en las distintas áreas de negocio, la orientación a los posibles riesgos y la aplicación de las medidas correctivas adecuadas en cada caso. La estructura incluye un Delegado de Protección de datos (DPO) y una Oficina del DPO, además de una serie de foros e interacciones que facilitaban la comunicación entre las diferentes partes interesadas (relaciones con el Consejo de Administración, con los órganos de dirección de la Compañía, un Comité de seguimiento, Comités de negocios y un Foro DPO).

No obstante, en 2022 se creó un modelo de prueba de gobernanza de una inteligencia artificial ética. Dicho modelo incluía un Grupo de Expertos en ética de la Inteligencia artificial, los “campeones” de la inteligencia artificial responsable y una Oficina de Inteligencia Artificial. Las lecciones aprendidas con dicho modelo de prueba (de cara al futuro modelo de gobernanza ética) fueron las siguientes:

1. Necesidad de una extensa difusión de gobernanza en las áreas de negocio en las que se asuma la responsabilidad de establecer los requisitos de los sistemas de IA
2. Necesidad de eficientes sistemas de coordinación
3. Firme enfoque en la evaluación y gestión de riesgos
4. Ampliación en el proceso de toma de decisiones, incluyendo a especialistas en ética
5. La gestión del gobierno debe recaer sobre las áreas de cumplimiento de la compañía

En base a esto, el 1 de diciembre de 2023 se aprobó un reglamento interno que define el Modelo de Gobierno de la Inteligencia Artificial de Telefónica, que se irá implementando de manera gradual (Muñoz et al, 2023).

Por último, cabe destacar que Telefónica se encuentra entre las ocho empresas tecnológicas (junto con GSMA, INNIT, Lenovo Group, LG AI Research, Mastercard, Microsoft y Salesforce) que se han comprometido a aplicar la Recomendación de la UNESCO sobre la ética de la inteligencia artificial. El acuerdo se firmó en febrero de 2024 en Kranj (Eslovenia) en el segundo Foro Mundial de la UNESCO sobre Inteligencia Artificial, y marca la primera colaboración entre empresas y las Naciones Unidas en este ámbito.

La Recomendación aborda temas como la gobernanza y administración ética, política de datos o evaluación del impacto ético. Al firmar el acuerdo, las empresas se comprometen a cumplir plenamente su función en la protección de los derechos humanos en todas las etapas de diseño, desarrollo, adquisición, comercialización y uso de la Inteligencia Artificial (UNESCO, 2024).

Este acuerdo es un claro ejemplo de la creciente tendencia de las empresas a reconocer la necesidad de regular la implementación de la inteligencia artificial en sus operaciones de manera ética y responsable (cada vez más demandada tanto por los consumidores como por los propios trabajadores). A pesar de ser un campo relativamente nuevo en el mundo empresarial, cada vez es mayor la conciencia sobre los riesgos éticos de esta tecnología, y se puede esperar un progreso significativo en este ámbito en los próximos años (tanto por la entrada en vigor de nuevas leyes como por la propia iniciativa empresarial).

6. CONCLUSIÓN

La inteligencia artificial ha transformado significativamente el entorno empresarial, y se prevé que su adopción y aplicación continúe expandiéndose considerablemente en los próximos años. El problema que plantea esta nueva realidad tiene su origen en la relativa novedad de esta tecnología, pues todavía no existe una regulación estandarizada y completa en este ámbito, no se ha desarrollado un modelo empresarial generalizado que pueda servir como guía para llevar a cabo una adecuada implementación de la inteligencia artificial y tampoco hay plena conciencia de los riesgos potenciales asociados con su uso.

Hasta ahora, la estrategia empresarial predominante ha sido abordar la incorporación de esta tecnología mediante un proceso de prueba y error (lo que ha dado lugar a polémicos y graves errores por parte de empresas como la introducción de chatbots con sesgos raciales). No obstante, dada la acumulación de información disponible, se ha llegado a un punto en el que es factible establecer unas bases sólidas sobre las que diseñar una estrategia dirigida a mitigar riesgos y asegurar una implementación ética.

Para que una empresa pueda implementar la inteligencia artificial de manera ética lo primero que se debe hacer es tomar conciencia de los riesgos que lleva asociado su uso. Los principales riesgos identificados son la falta de protección de los datos utilizados (colocando en una posición vulnerable tanto a los usuarios como a las propias empresas), la falta de transparencia en la forma en la que opera el sistema (opacidad que genera incertidumbre sobre su funcionamiento e impide la posible rendición de cuentas) y la presencia de sesgos en el algoritmo (que da lugar a resultados discriminatorios y fomenta la inequidad).

En segundo lugar, es necesario identificar los principales requisitos que la empresa debe garantizar para asegurar una implantación ética y la salvaguarda de los derechos de todas las personas involucradas. Partiendo de los riesgos analizados, los requisitos necesarios son la privacidad de la información utilizada, la transparencia del funcionamiento del sistema y la libertad y la autonomía de los usuarios.

A partir de esto, con esta información se puede esbozar una estrategia basada en un uso ético de la inteligencia artificial, que puede ser materializada y puesta en práctica a través de un modelo operativo diseñado específicamente para esta tecnología emergente. Los puntos clave de este modelo residen en:

1. El establecimiento de unos claros principios éticos base que guíen su diseño
2. La adaptación y actualización de la gobernanza de la compañía (incluyendo, por ejemplo, un Comité de la Ética de la Inteligencia Artificial multidisciplinar)
3. La alineación de la cultura y los valores de la empresa con la estrategia
4. La preparación de los procesos y capacidades para que la implementación sea factible.

Por último, teniendo en cuenta que la integración de la inteligencia artificial ya es una realidad y está redefiniendo el panorama empresarial, es importante destacar que ya no es viable ignorar o subestimar los riesgos éticos asociados con su implementación. Las empresas deben adoptar una postura proactiva y abordar este desafío de manera urgente.

“¿Sugiere mi trabajo una vía para continuar profundizando el tema?”

Mi trabajo proporciona los cimientos básicos sobre los que construir una estrategia empresarial orientada a la implementación ética de la inteligencia artificial. De esta manera, se abre la posibilidad de una exploración más detallada de cada una de las pautas establecidas y su adaptación a las necesidades específicas de cada sector (dado que mi enfoque es general).

“¿Qué debilidades presenta mi trabajo?”

La principal debilidad que presenta mi trabajo es que, dado que trata un tema relativamente nuevo (pues las empresas han empezado a tomar conciencia de la importancia de la implementación ética de la inteligencia artificial en los últimos años), aún no hay muchos ejemplos de empresas reales que aborden este tema ni datos concretos sobre el impacto a largo plazo de las estrategias propuestas. Al ser un tema que se encuentra en las primeras fases de desarrollo, hay dificultad a la hora de realizar proyecciones precisas sobre la efectividad de las pautas propuestas.

“¿Hay maneras con las que mi trabajo pueda ser perfeccionado por investigadores en el futuro?”

En el futuro, cuando ya se cuente con más información disponible y las empresas adquieran más experiencia tratando con la inteligencia artificial, mi trabajo podrá ser perfeccionado llevando a cabo estudios que evalúen la efectividad de las estrategias sobre su implementación ética. Por otra parte, se podrán realizar análisis sobre las posturas más convenientes teniendo en cuenta la diversidad de sectores industriales.

“¿Cuáles son las implicaciones prácticas de mi trabajo?”

Mi trabajo proporciona un enfoque general sobre el estado actual de la inteligencia artificial en el mundo empresarial y una guía básica para facilitar su implementación de manera ética.

DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA EN TRABAJOS FIN DE GRADO

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, María Dolores Yuste Sánchez, estudiante del doble grado de Derecho y Administración y Dirección de Empresas de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " Ética empresarial: requisitos en la adopción de la inteligencia artificial", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
2. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 13 de marzo de 2024

Firma: 

7. BIBLIOGRAFÍA

Ali, S., Christin, A., Smart, A., Katila, R. (2023) “Walking the Walk of AI Ethics in Technology Companies” *Stanford University – Human-Centered Artificial Intelligence* (Disponible en https://hai.stanford.edu/sites/default/files/2023-12/Policy-Brief-AI-Ethics_0.pdf; última consulta 29/02/2024).

Ambrozi, A. (2023). “11 Challenges of Adopting AI In Business (And How to Address Them Head-On),” *Forbes* (Disponible en <https://www.forbes.com/sites/forbesbusinesscouncil/2023/10/24/11-challenges-of-adopting-ai-in-business-and-how-to-address-them-head-on/>; última consulta 08/02/2024).

Ammanath, B. (2023) “State of Ethics and Trust in Technology – Annual report” *Deloitte – Technology Trust Ethics* (Disponible en <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/us-tte-annual-report-2023.pdf>; última consulta 29/02/2024)

Anyoha, R. (2017). “The History of Artificial Intelligence”, *Special Edition on Artificial Intelligence – The Graduate School of Arts and Sciences of Harvard University* (Disponible en <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>; última consulta 04/02/2024).

Baker-Brunnbauer, Josef. (2021). “Management perspective of ethics in artificial intelligence”. *AI and Ethics*: pp.173-181.

Barati, M., Yankson, B. (2022). “Predicting the Occurrence of a Data Breach” *International Journal of Information Management Data Insights*. vº2 (2).

Blackman, R. (2022). “Why You Need an AI Ethics Committee”, *Harvard Business Review* (Disponible en <https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee>; última consulta 1/03/2024)

Blázquez Ruiz, F. J. (2022). “La paradoja de la transparencia en la IA: Opacidad y explicabilidad. Atribución de responsabilidad.” *Revista Internacional De Pensamiento Político*, 17(1), pp. 261–272.

Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D., Liang, P. (2023). “The Foundation Model Transparency Index” *Stanford Institute for Human-Centered Artificial Intelligence* (Disponible en <https://crfm.stanford.edu/fmti/fmti.pdf>)

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D. (2020). “Language Models are Few-Shot Learners” *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* <https://arxiv.org/pdf/2005.14165.pdf>

Brusoni, S. & Vaccaro, A. (2017). “Ethics, Technology and Organizational Innovation”. *Journal of Business Ethics*. n°143. pp. 223-226.

Carsten, B. (2021). “Ethical Issues of AI” *Artificial Intelligence for a Better Future, An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*. pp. 35-53.

Chen, Z. (2023). “Ethics and discrimination in artificial intelligence-enabled recruitment practices” *Humanit Soc Sci Commun* (Disponible en <https://doi.org/10.1057/s41599-023-02079-x>; última consulta 19/02/2024).

Consejo de la Unión Europea. (2023, 9 de diciembre). *Ley sobre inteligencia artificial: el Consejo y el Parlamento llegan a un acuerdo sobre las primeras normas para la IA en el mundo* [Comunicado de prensa] (Disponible en <https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>)

Davenport, A. & Ronanki, R. (2018). “IA And Machine Learning: Artificial Intelligence for the Real World”, *Harvard Business Review* (Disponible en <https://hbr.org/2018/01/artificial-intelligence-for-the-real-world>; última consulta 05/02/2024)

Deloitte. (2019). “A call for Transparency and Responsibility in Artificial Intelligence” (Disponible en <https://www2.deloitte.com/nl/nl/pages/innovatie/artikelen/a-call-for-transparency-and-responsibility-in-artificial-intelligence.html>, última consulta 28/02/2024)

Equipo de datos e inteligencia de IBM. (2023). “Arrojando luz sobre el sesgo de la IA con ejemplos del mundo real” *IBM* (Disponible en <https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/>; última consulta 03/02/2024).

García, E. (2019). “Unboxing de Inteligencia Artificial: confianza a través de transparencia y explicabilidad”, *KPMG Tendencias – Inteligencia Artificial y Data Analytics* (Disponible en <https://www.tendencias.kpmg.es/2019/07/inteligencia-artificial-confianza-transparencia/>)

Google. (2023). “AI Principles Progress Update 2023” *Google* (Disponible en <https://ai.google/static/documents/ai-principles-2023-progress-update.pdf>; última consulta 02/03/2024).

Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, (2019). “Directrices éticas para una IA fiable” *Oficina de Publicaciones* (Disponible en <https://data.europa.eu/doi/10.2759/14078>)

Haan, K. (2023). “How Businesses Are Using Artificial Intelligence in 2024”, *Forbes – Advisor* (Disponible en <https://www.forbes.com/advisor/business/software/ai-in-business/>; última consulta 04/02/2024)

Hardy, T. (2001). “IA: Inteligencia Artificial”. *POLIS, Revista Latinoamericana*, vol.1. n°2. pp 1-24.

IBM. (2020). “An operating model for AI at scale – How to maximise the expected business value from AI” *IBM* (Disponible en <https://www.ibm.com/downloads/cas/EQA04AW8>; última consulta 02/03/2024).

IBM. (2023). “IBM Global AI Adoption Index – Enterprise Report” (Disponible en <https://ibm.box.com/shared/static/xzfbugrtjvc45cotgd4yyng3vb2pprp0.pdf>; última consulta 08/02/2024).

Kwan, A., Schroeck, M., Kawamura, J. (2019). “Architecting an operating model – A platform for accelerating digital transformation” *Deloitte Insights* (Disponible en https://www2.deloitte.com/content/dam/insights/us/articles/5078_architecting-an-operating-model/DI_architecting-an-operating-model.pdf; última consulta 29/02/2024).

McKinsey & Company. (2023). “The economic potential of generative AI: The next productivity frontier” (Disponible en <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction>; última consulta 10/02/2024)

Mollick, E. (2006). “Establishing Moore's Law”. *Annals of the History of Computing, IEEE*. n°28. pp 62 - 75.

Montgomery, C., Rossi, F. (2023). “A look into IBM’s AI ethics governance framework” *IBM* (Disponible en <https://www.ibm.com/blog/a-look-into-ibms-ai-ethics-governance-framework/>; última consulta 1/03/2024).

Mucci, T., Stryker, C. (2023). “What is AI governance?” *IBM* (Disponible en [https://www.ibm.com/topics/ai-governance#:~:text=Artificial%20intelligence%20\(AI\)%20governance%20refers,and%20respect%20for%20human%20rights](https://www.ibm.com/topics/ai-governance#:~:text=Artificial%20intelligence%20(AI)%20governance%20refers,and%20respect%20for%20human%20rights); última consulta 2/03/2024).

Muñoz, A., Salado, J., Benjamins, R., (2023). “Un Marco de Gobierno de la Inteligencia Artificial en Telefónica” *Telefónica* (Disponible en <https://www.telefonica.com/es/sala-comunicacion/blog/un-marco-de-gobierno-de-la-inteligencia-artificial-en-telefonica/>; última consulta 02/03/2024).

Murphy, A., Kirwin, J., Abdul, K. (2016). “Operating models. Delivering on strategy and optimizing processes” *Ernst & Young* (Disponible en https://assets.ey.com/content/dam/ey-sites/ey-com/en_ca/topics/strategy/ey-operating-models-design-and-implementation.pdf?download; última consulta 29/02/2024).

OdiseaIA & PwC. (2022). “GuIA de buenas prácticas en el uso de la inteligencia artificial ética” *OdiseaIA. Observatorio del Impacto Social y Ético de la Inteligencia Artificial* (Disponible en <https://www.pwc.es/es/publicaciones/tecnologia/assets/guia-buenas-practicas-uso-inteligencia-artificial-pwc-odiseia.pdf>; última consulta 28/02/2024)

Raghavan, P. (2024). “La generación de imágenes de Géminis se equivocó. Lo haremos mejor.” *Geminis – Google* (Disponible en <https://blog.google/products/gemini/gemini-image-generation-issue/>).

Sarker, I. (2022). “AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems.” *SN Computer Science* 3(2), 158.

Telefónica. (2018). “Principios de IA de Telefónica” Disponible en <https://www.telefonica.com/es/wp-content/uploads/sites/4/2021/08/principios-ai-esp-2018.pdf>; última consulta 02/03/2024).

Todt, K. E. (2019). “Data Privacy and Protection: What Businesses Should Do” *The Cyber Defense Review*, 4(2), pp.39–46.

UNESCO. (2024, 5 de febrero). *AI Ethics: 8 global tech companies commit to apply UNESCO’s Recommendation* [Comunicado de prensa] (Disponible en <https://www.unesco.org/en/articles/ai-ethics-8-global-tech-companies-commit-apply-unescos-recommendation#:~:text=Today%2C%20GSMA%2C%20INNIT%2C%20Lenovo,designing%20and%20deploying%20AI%20systems.>)

Vimalkumar, M. Kumar, S. Bahadur, J. & Dwivedi, Y. (2021). ““Okay google, what about my privacy?”: User's privacy perceptions and acceptance of voice based digital assistants” *Computers in Human Behavior*. vol 120.