UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Elaboración de una guía de Buenas Prácticas para el empleo de la Inteligencia Artificial para profesionales TIC.

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Naharros Pérez, Salvador

Tutor/a: Oltra Gutiérrez, Juan Vicente

CURSO ACADÉMICO: 2023/2024

Resumen

La inteligencia artificial está cada vez más presente en nuestras vidas, tanto desde un prisma ciudadano como profesional. El derecho avanza con menor velocidad que la tecnología, así que para evitar daños a los seres humanos provocados por el código que dan cuerpo a los algoritmos, ya que no tenemos un apoyo suficiente en los códigos legales, tenemos que recurrir a un tercer tipo de códigos: los códigos éticos, en particular los profesionales.

Desde el Consejo de la Unión Europea se ha publicado durante la realización de este trabajo el Reglamento que organiza la necesaria implantación de la IA en nuestras vidas, apoyado en una base ética, según los documentos que han trascendido.

El objetivo del presente TFG es elaborar una guía de Buenas Prácticas en la que poder hacer uso de la inteligencia artificial al profesional TIC de una forma responsable a la vez que eficiente, permitiendo un correcto desarrollo que consiga seguir cubriendo necesidades y mejoras sin amenazar los derechos humanos, apoyándola en un estudio de casos y en documentos de fuentes oficiales (CCN-CERT, Autoridades de Protección de Datos, Grupo de expertos de alto nivel sobre inteligencia artificial de la UE, etc.)

Palabras clave: Inteligencia Artificial, ética.

Resum

La intel·ligència artificial està cada vegada més present en les nostres vides, tant des de la vista d'un ciutadà com d'un professional. El dret avança amb menor velocitat que la tecnologia, així que per a evitar danys als éssers humans provocats pel codi que dona cos als algoritmes, ja que no tenim un suport suficient als codis legals, hem de recórrer a un tercer tipus de codis: els codis ètics, en particular els professionals.

Des del Consell de la Unió Europea s'ha publicat durant la realització d'aquest treball el Reglament que organitza la necessària implantació de la IA en les nostres vides, suportat en una base ètica, segons els documents que han transcendit.

L'objectiu del present TFG es elaborar una guia de Bones Pràctiques en la que poder fer ús de la intel·ligència artificial al professional TIC d'una forma responsable ala vegada que eficient, permetent un correcte desenvolupament que aconseguisca seguir cobrint necessitats i millores sense amenaçar els drets humans, suportant-la en un estudi de casos y documents de fonts oficials (CCN-CERT, Autoritats de Protecció de Dades, Grup d'experts d'alt nivell sobre intel·ligència artificial de la UE, etc.)

Paraules clau: Intel·ligència Artificial, ètica.

Abstract

Artificial intelligence is increasingly present in our lives, both from a citizen and professional perspective. Law advances more slowly than technology, so in order to avoid harm to human beings caused by the code that gives shape to algorithms, since we do not have enough support in legal codes, we have to resort to a third type of codes: ethical codes, particularly professional ones.

From the Council of the European Union, the Regulation that organizes the necessary implementation of AI in our lives has been published during the realization of this work, which is supported on an ethical basis, according to the documents that have emerged.

The objective of this work is to preparate a guide of Good Practices in which the ICT professional can make use of artificial intelligence in a responsible and efficient way, allowing a correct development that continues to cover needs and improvements without threatening human rights, supporting it with a case study and documents from official sources (CCN-CERT, Data Protection Authorities, EU High Level Expert Group on Artificial Intelligence, etc.).

Keywords: Artificial Intelligence, ethics.

Agradecimientos

A mi madre, que siempre me apoyó en todas mis decisiones, siempre velaba por mi felicidad y me enseñó a no rendirme.

A mi padre y a mi hermana, que siempre me han acompañado, especialmente en estos últimos años pese a las dificultades.

Índice de contenidos

Índice de figuras

1. Introducción

1.1 Motivación

Los avances en la inteligencia artificial están ocurriendo a una velocidad vertiginosa. Esta rápida transformación está impactando profundamente la industria de las tecnologías de la información y la comunicación (en adelante, TIC). Como resultado, los profesionales de las TIC requieren orientación específica sobre cómo aplicar la inteligencia artificial de manera ética, eficiente y segura.

Aunque actualmente la inteligencia artificial plantea algunos dilemas éticos, su prohibición total no es concebible, puesto que se ha demostrado que su uso aporta grandes beneficios a la sociedad. Es posible corregir los comportamientos problemáticos, manteniendo o incluso mejorando la eficiencia de la inteligencia artificial tal y como la conocemos en la actualidad.

Desde mi punto de vista, pienso que estos últimos años el crecimiento de la inteligencia artificial está siendo muy interesante, con gran capacidad de conseguir unos objetivos cada vez más cerca y a mayor velocidad. El hecho de que se le haya puesto el foco en este tiempo, principalmente gracias a la rama generativa, me proporciona un sentimiento de esperanza hacia la llegada de una nueva era tecnológica, aunque a partes iguales, todo lo que está consiguiendo ser capaz está asustando a la sociedad, lo que podría ralentizar este proceso.

Tengo un gran interés en ver cómo ha avanzado el marco legal en la inteligencia artificial y con este trabajo conseguiré asentar conocimientos al respecto. Con la elaboración de la guía, se busca marcar los límites legales y éticos de la inteligencia artificial para poder recobrar la confianza de la sociedad en esta tecnología y avanzar con pie firme hacia el futuro que cada vez es más presente.

Profesionalmente, me gustaría poder formarme en ámbitos relacionados con el Big Data y también con el marketing digital, por lo que pienso que con la elaboración de este trabajo voy a poder curtirme en ambos campos, al menos en el aspecto ético, puesto que el uso de la inteligencia artificial en técnicas de marketing cada vez es mayor y las grandes cantidades de datos y la inteligencia artificial tienen una conexión inevitable.

1.2 Objetivos

Para que los profesionales TIC puedan seguir utilizando herramientas de inteligencia artificial de manera responsable, es fundamental establecer directrices que guíen sus acciones y consideraciones durante el proceso. Un medio para poder lograr esto es la creación de una guía de Buenas Prácticas.

En el contexto actual, es crucial identificar los puntos en los que la inteligencia artificial presenta deficiencias o genera consecuencias negativas. De esta manera, podemos corregir esos aspectos sin perder los beneficios que esta tecnología aporta.

Este trabajo va a crear una guía que servirá de gran ayuda para aquellos que vayan a desarrollar sistemas que utilicen inteligencia artificial con el objetivo de cumplir con las leyes que lo definen, basándose especialmente en el Reglamento europeo. Además, se proporcionará un formulario de autoevaluación para que los desarrolladores puedan verificar rápidamente el cumplimiento de las directrices de la guía.

También se realizará un cuestionario a un grupo acotado de personas acerca de sus usos de herramientas de inteligencia artificial, conocimiento de implicaciones éticas en la IA y relacionados.

De esta manera, con la realización de este trabajo pretendo estudiar el espectro legal que afecta a la parte más ética de la inteligencia artificial, consiguiendo así mejorar mi conocimiento al respecto de las leyes que rodean el desarrollo de tecnologías, especialmente las de inteligencia artificial.

1.3 Metodología

Con el objetivo de llevar a cabo este trabajo, se realizará una investigación exhaustiva sobre las principales problemáticas que afectan al uso ético de la inteligencia artificial y sus resultados, ya sea de manera intencionada o no. El propósito es proporcionar una guía que restrinja estos comportamientos y ayude a corregirlos.

Para lograrlo, se analizará el reglamento publicado recientemente por parte de la Unión Europea, en el cual se establecen directrices específicas para el desarrollo y uso adecuado de la inteligencia artificial.

Además, se llevará a cabo una encuesta dirigida a un grupo determinado de personas para comprender sus hábitos de uso de herramientas de inteligencia artificial, así como su conocimiento sobre prácticas irresponsables y no éticas. También se explorará la percepción general sobre la ética en el ámbito de la inteligencia artificial.

1.4 Estructura del documento

El documento estará dividido en siete capítulos y dos anexos:

Capítulo 1:

En el primer capítulo se describen los objetivos del trabajo y se presentan los recursos empleados y la metodología utilizada para alcanzarlos. Además, se aborda la motivación, tanto en términos generales como personales y profesionales.

Capítulo 2:

En este capítulo se expondrá el estado del arte de la inteligencia artificial y la ética. Para ello, se estudiará el marco teórico, con la definición de los 4 principales componentes del trabajo, así como la evolución de la Inteligencia Artificial, desde sus orígenes hasta las implementaciones actuales. Se hace especial hincapié en las mejoras sustanciales que han aportado en el pasado y siguen aportando en el presente a la sociedad, la

economía y otros ámbitos. También se destacan los desafíos éticos a los que la inteligencia artificial se enfrenta actualmente y que explican la importancia de este documento.

Capítulo 3:

Se examina a fondo el marco regulatorio, desde recomendaciones de la UNESCO hasta el análisis del documento del Reglamento Europeo de la Inteligencia Artificial, explorando su alcance y disposiciones clave, poniendo el foco en directrices a seguir por parte de los profesionales TIC. También se detallan otras directrices de otras fuentes.

Capítulo 4:

El núcleo de este trabajo se concentra en este punto, ya que se desarrolla la Guía de Buenas Prácticas para el empleo de la inteligencia artificial para los profesionales TIC, partiendo de la base del documento europeo examinado en el capítulo anterior, así como otras directrices analizadas de otras fuentes. Además, se presenta un formulario de autoevaluación basado en la guía para exploraciones éticas sencillas.

Capítulo 5:

Tras encuestar a un segmento de la población acerca del uso que dan actualmente a herramientas de inteligencia artificial y su conocimiento del ámbito ético en las mismas, se estudian y analizan las respuestas para ver el grado de satisfacción con la tecnología y poner en valor la necesidad de convertir la inteligencia artificial en una tecnología responsable. De esta manera, entenderemos la necesidad de la realización de esta guía.

Capítulo 6:

En el penúltimo capítulo se presentan las conclusiones derivadas del análisis de las respuestas de la encuesta y la guía de Buenas Prácticas y se relacionan los conceptos previamente abordados.

Capítulo 7:

Por último, se incluye la bibliografía utilizada y contrastada durante la realización del trabajo.

Anexos:

Como documentación adicional, en primer lugar, se incluye una reflexión sobre la relación de los Objetivos de Desarrollo Sostenible con la elaboración de este trabajo. En segundo lugar, se proporciona íntegramente el formulario de autoevaluación de sistemas de IA responsables.

2. Estado del arte

2.1 Trabajos previos

Durante una búsqueda de ideas para mi TFG, aquellos que tratasen sobre inteligencia artificial eran los que más me interesaban. Sabiendo que ya está comenzando a implantarse de forma más general el uso de inteligencia artificial en los trabajadores a casi todos los niveles de las empresas tecnológicas, mi interés al respecto es incuestionable. Encontré el trabajo “De Python a Kubeflow” de Sergio Gutiérrez, realizado en la UPV, el cual trata todas las fases desde el análisis y el entrenamiento hasta el uso de herramientas con inteligencia artificial. A raíz de esto, me interesé en el apartado de trabajos futuros de la conclusión de su TFG, y uno de los puntos me llamó la atención: “Abordar desafíos éticos y de privacidad”.

Las mejoras en inteligencia artificial nos brindan nuevas formas de utilizar los datos y de utilizar nuestras vidas también, ¿pero se estaba prestando la suficiente atención a las propias personas y sus derechos? Desde ese momento, despertó en mí el pensamiento filosófico que llevaba varios años aparcado y se vio en la necesidad de profundizar en aquello que me causaba gran interés: la inteligencia artificial; y relacionarla con un aspecto que, por el contrario, parecía haber perdido interés general: la ética. En el momento en el que encontré que la legislación concreta de la IA en la Unión Europea estaba, en aquel momento, cada vez más cerca de publicarse, vi la necesidad de crear una guía que, basándose en ella y en otras directrices, pudiese permitir que el desarrollo de sistemas de inteligencia artificial, como hizo Sergio, sigan prosperando mientras los derechos de las personas se mantienen inalterados.

2.2 Marco teórico

La inteligencia artificial es un término que en los últimos años ha tenido tanto éxito que es difícil encontrar a alguien que no haya oído hablar de ella. Sin embargo, lo cierto es que este campo informático lleva ya varios años a la espalda en los que ha ido evolucionando hasta llegar a ser tal y como hoy la conocemos. Es por eso por lo que en este apartado explicaremos este y otros términos necesarios para comprender el trabajo.

Debido a su larga trayectoria, hemos sido conscientes de no solo sus beneficios, sino también algunos problemas que generan, especialmente en el ámbito ético. Hay ciertos problemas que han sido descubiertos y se han ido arreglando o moderando, otros descubiertos a los que todavía hay que hacerles frente.

2.2.1 Inteligencia artificial

Para el público general es fácil simplificar el significado de “inteligencia artificial” es un chat en el que una máquina te responde a lo que te preguntes. Sin embargo, la inteligencia artificial es más que eso. Y también hay que comprender qué hay por detrás de eso.

La inteligencia artificial es un conjunto de diferentes tecnologías que tiene como objetivo dotar a las máquinas de una inteligencia que se asemeje a la humana. Pero para ello, hay que tener claros los conceptos de la "inteligencia"; la Real Academia Española (RAE) la define como "la capacidad de entender o comprender y de resolver problemas", mientras que algunos psicólogos añaden: "la habilidad de razonar, planificar, pensar en abstracto, comprender ideas complejas, aprender rápido y aprender de la experiencia, que es más que una destreza académica o del aprendizaje por medio de libros" (Gottfredson, 1997) o "Conjunto de habilidades para adaptarse al entorno" (Humphreys, 1979)

El término de "inteligencia artificial", acuñado en 1956 por John McCarthy, ha ido englobando cada vez más algoritmos, tecnologías y herramientas que van dando forma a una especie de cerebro artificial que consiga tener la inteligencia que pueda llegar a tener un humano. Más adelante en este mismo capítulo examinaremos los hitos más importantes que han ido consiguiendo que la inteligencia artificial sea cada vez más inteligente.

Para lograr esa inteligencia, se puede enseñar a cómo hacer las cosas (machine Learning), o bien puede aprenderlas por sí mismo (Deep Learning). Estas son disciplinas que se incluyen dentro de la inteligencia artificial, una dentro de la anterior.

Con el Machine Learning, las máquinas aprenden mediante entrenamiento. Puede realizarse un entrenamiento supervisado, en el que se dan unos datos de entrada etiquetados para que la máquina, tras analizarlos, tenga ejemplos clasificados correctamente y pueda clasificar nuevas entradas en base a estos. Para ello, es común utilizar árboles de decisión que extraen, simplifican y clasifican los datos en base a diferentes criterios.

Por otra parte, está el aprendizaje no supervisado, en el que es la máquina quien asocia patrones entre todas las muestras y las agrupa a partir de estos sin seguir una pauta dada. Para estos casos se hace uso del Deep Learning.

El Deep Learning, conocido como aprendizaje profundo se caracteriza por el uso de redes neuronales, formadas por neuronas artificiales. Estas neuronas artificiales están basadas en las neuronas humanas y su división biológica. Las neuronas biológicas reciben en sus entradas información y después de procesarlas envía su información a cientos de neuronas. A grandes rasgos, ambas tienen dendritas, por donde reciben entradas; el núcleo, en el que procesan la información; y los axiomas, que son las salidas. A partir de eso, se construyen estas redes neuronales con varias capas para poder procesar la información.

Dentro de las redes neuronales, existen diferentes modelos que emplean diferentes técnicas y que funcionan mejor para diferentes tipos de tareas. Por un lado, están las redes neuronales convolucionales, idóneas para tareas de visión como la clasificación de imágenes o detección de objetos. Por otro lado, las recurrentes, que se utilizan en secuencias de datos, como el procesamiento del lenguaje natural y las series.

Pero sin duda, el campo que más demandado está actualmente es el de la inteligencia artificial generativa, incluido dentro del Deep Learning. Se utiliza para generar nuevos datos, como imágenes, música o texto. Estos modelos aprenden a reconocer patrones

y estructuras en los datos y luego utilizan ese conocimiento para crear algo nuevo que siga esos patrones.

2.2.2 Ética

Para poder comprender este trabajo, además de tener claros el significado y las posibilidades de la inteligencia artificial, también hay que definir el concepto de la “ética”. Al igual que las personas, las máquinas en un intento de ser inteligentes, pueden restar en el aspecto ético, lo que hacen cuestionables ciertos avances.

Según la RAE, la ética es el “conjunto de normas morales que rigen la conducta de la persona en cualquier ámbito de la vida”. Lo cierto es que es una rama de la filosofía que aborda el comportamiento humano y su relación con las nociones del bien y del mal, las normas morales, el deber, la felicidad y el bienestar común.

Las mayores influencias de la filosofía han aportado su verdad sobre la ética: desde la Antigua Grecia con Sócrates, Aristóteles y Platón, hasta más recientes como José Antonio Marina

Aristóteles, analizó la ética como el estudio de cómo vivir bien y alcanzar la felicidad y el crecimiento humano. En sus obras hacía referencia a esa felicidad a través del término griego “eudaimonía”, que para él era el bienestar común. Según Aristóteles, la felicidad es el último fin de la vida humana y todas nuestras acciones deben orientarse para conseguirla (Aristóteles, s. IV a. C).

Pone en valor otros conceptos, como la virtud, necesaria para poder conseguir ese bienestar, siendo esta una disposición adquirida que nos permite actuar de acuerdo con la razón; o la sabiduría práctica, con la que indica que no solo se necesita el conocimiento teórico, sino que también es necesario saber tomar decisiones prudentes.

Platón también habló de la sabiduría práctica en sus diferentes obras como “La República” o “Menón”, en las que la defendía como "entendimiento moral”. Sin embargo, aunque consideraba que era el atributo más valioso de los que se podían obtener, él pensaba que no puede ser enseñada, sino que debía ser uno mismo quien descubriese ese conocimiento a través de la práctica y la comprensión moral.

Sócrates por otra parte introdujo por primera vez la idea del intelectualismo moral, planteando que lo más importante en la vida es “obrar justamente” y que es preferible “sufrir una injusticia que cometerla”. Según su filosofía, la bondad está intrínsecamente ligada a la sabiduría, estableciendo una conexión directa entre la verdad, la virtud y la felicidad. Solo quien comprende verdaderamente qué es la justicia puede actuar de manera justa y únicamente aquel que conoce el bien puede obrar en consecuencia.

Pese a la antigüedad del término, la ética sigue explorando nuevos significados. El también filósofo español José Antonio Marina define a la ética como “moral transcultural”: “el conjunto de normas universales que transcienden las peculiaridades culturales” (Marina, 2020). Con ello, busca eliminar los sesgos que los diferentes tipos de morales podrían aplicar entre ellas, poniendo la Declaración de los Derechos Humanos como núcleo de la ética.

En las empresas es común tener un código ético para reflejar sus valores y propósitos que funciona como una brújula moral para que los empleados sepan actuar correctamente. Los principios éticos principales que se declaran suelen relacionarse con garantizar la dignidad de las personas, cumplir compromisos, actuar con integridad o proteger la confidencialidad. En Telefónica, una de las empresas pioneras en inteligencia artificial, actualizó su código ético en relación con la IA para que su uso y desarrollo esté centrado en las personas (Telefónica, 2024), un objetivo que se proyectará en la guía

2.2.3 Guía de Buenas Prácticas

Como objetivo de este trabajo, vamos a crear una Guía de Buenas Prácticas sobre el uso de la inteligencia artificial, que inevitablemente va a estar relacionado con la ética. Para poder hacerlo correctamente, vamos a definir brevemente de lo que debe estar formada esta guía, para qué sirve y qué debería suponer su existencia.

Primero tendremos que definir el concepto de “guía”. Si buscamos este término en el diccionario de la RAE, encontraremos más de 20 acepciones diferentes en los que la mayoría tratan de “encaminar” y “dirigir”. Hay diferentes maneras de señalar el camino: a través de unas instrucciones de una receta de cocina, colocando una vara en un árbol para que sigan una dirección determinada, o consultando la ruta más rápida en Google Maps. En todas estas situaciones existe algo que te da indicaciones de cuál es el camino para conseguir tu propósito.

Al respecto de la parte de “buenas prácticas”, nos encontramos con unas estrategias, métodos o procedimientos que son recomendados para poder conseguir unos resultados óptimos específicos. Estas se basan en la experiencia acumulada, la investigación y las normativas vigentes, y están diseñadas para promover la calidad, la seguridad y la eficacia en diversos ámbitos.

Por lo tanto, si unimos ambos conceptos, obtenemos una Guía de Buenas Prácticas: un documento que proporciona directrices y recomendaciones para realizar actividades o procesos de manera eficiente, segura y ética. Su propósito es estandarizar procedimientos, promover la calidad y garantizar la conformidad a partir de unas normas y principios establecidos. Además, estas guías se elaboran basándose en experiencias previas, investigaciones y normativas vigentes para ayudar a individuos y organizaciones a tomar decisiones informadas y realizar tareas de forma efectiva.

2.2.4 Profesionales TIC

Toda Guía de Buenas Prácticas tiene su razón de ser. En este caso, los profesionales de las Tecnologías de la Información y la Comunicación. Antes de querer guiar al usuario final para utilizar una herramienta desarrollada por estos profesionales, deben ser ellos los que hayan actuado correctamente para poder facilitar el uso de la misma.

Primeramente, hay que saber a qué nos referimos concretamente con “TIC”. Este término engloba una amplia gama de tecnologías y herramientas que permiten la creación, gestión, almacenamiento, transmisión y comunicación de información en formatos digitales. Las TIC abarcan tanto el hardware y los sistemas físicos como el software y las aplicaciones, y desempeñan un papel fundamental en casi todos los aspectos de la vida moderna y profesional. En ello se incluye la recopilación, procesamiento, almacenamiento y transmisión de datos a través de cualquier medio digital para la comunicación, el acceso a información y la automatización de procesos en diversos contextos, como el educativo, el empresarial y el personal.

Por otra parte, un profesional es aquella persona que ha adquirido un alto nivel de competencia y capacitación en una disciplina específica y que ejerce su labor de manera cualificada y con responsabilidad. Los profesionales suelen tienen experiencia práctica y, además, se comprometen a mantener sus habilidades actualizadas y a seguir prácticas y principios que aseguren la calidad y la integridad en su trabajo.

Uniendo ambos conceptos, nos encontramos con un profesional de las Tecnologías de la Información y Comunicación, un campo que está en constante evolución, por lo que es necesario que siga aprendiendo y se adapte a las circunstancias actuales. De igual modo, le servirá para poder actualizar y/o aplicar en futuros trabajos la Guía de Buenas Prácticas que se ha preparado.

2.3 Orígenes de la IA

A pesar de haber cobrado especial relevancia durante los últimos años, los cimientos de la Inteligencia Artificial se remontan a muchos años atrás. Desde el desarrollo de la lógica por Aristóteles, continuando con matemáticos y filósofos de la altura de George Boole y las aportaciones del español Ramon Llull y el alemán Gottfried Willhelm Leibniz, se pudieron observar grandes avances hacia la evolución del pensamiento computacional. El español abrió la puerta a la informática moderna a través de su propuesta de un sistema mecánico que combinaba conceptos fundamentales, algo en lo que el alemán se inspiraría para dar cabida a un lenguaje universal que pudiese procesar una máquina y que alcanzar todo el conocimiento posible (Manrique, 2007).

Con el objetivo de demostrar que el razonamiento teológico y filosófico podía ser representado mediante un dispositivo mecánico, a finales del siglo XIII y principios del XIV, Ramon Llull ideó una máquina conocida como *Ars Magna*. En ella, las teorías, los sujetos y los predicados, tanto filosóficos como teológicos, estaban dispuestos en diversas figuras geométricas que con movimientos podrían ofrecer respuestas. El filósofo estaba convencido de que esta máquina permitiría alcanzar el conocimiento de la Verdad (Baz & Cornelius, 1998).

Llull utilizó un alfabeto de solo nueve letras (BCDEFGHIK) junto con discos de pergamino que funcionaban como memoria, logrando así un sistema mecanizado por primera vez a través de métodos heurísticos, anticipándose en 600 años a Alan Turing, reconocido fundador de la informática moderna. De esta manera, con el artefacto *Ars Magna* en su versión llulliana, se le considera un precursor de esta disciplina.

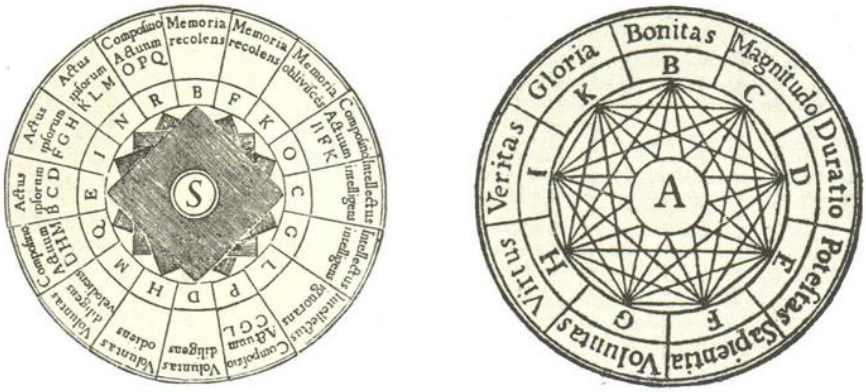

Figura 1: Dos ejemplos de discos de Ars Magna *llulliana (Fuente: enciclopedia.cat)*

Aproximadamente 4 siglos después, entre el siglo XVII y el XVIII, Leibniz consiguió mejorar el *Ars Magna* de Ramon Llull al introducir un sistema más formal y lógico de representación del conocimiento, que se apoyaba en un lenguaje simbólico y en principios matemáticos avanzados. Su enfoque en la lógica, el cálculo y la universalidad del conocimiento estableció las bases para desarrollos posteriores en la lógica y la computación, haciendo que sus contribuciones fueran más aplicables y extensibles en el contexto del pensamiento moderno.

Leibniz propuso un sistema más simplificado y eficiente llamado "lenguaje universal" que consistía en un lenguaje simbólico y algebraico. Este sistema permitía expresar cualquier concepto mediante símbolos y relaciones lógicas, lo que facilitaba su manipulación y análisis respecto al sistema geométrico de Llull. Además, mientras que estaba principalmente para responder cuestiones teológicas o filosóficas, Leibniz puso un enfoque más universal para abarcar todas las disciplinas, proponiendo un sistema que pudiera integrar matemáticas, filosofía y ciencias en un marco común. Se enfocó en la idea de que cualquier proposición podría ser descompuesta en componentes más simples y combinada de diversas maneras, facilitando una exploración más profunda y completa del conocimiento. De esta manera, Leibniz sentaría las bases para la lógica moderna y la computación (Manrique, 2007).

Más recientemente, tenemos nombres que están más directamente relacionados con la informática y la inteligencia artificial. Uno de ellos es el británico George Boole, el que, gracias a sus contribuciones durante el siglo XIX, es conocido como el padre de la lógica. Su creación del álgebra booleana fue y sigue siendo esencial en el diseño de circuitos electrónicos y sistemas de computación. De esta manera, a través de operaciones lógicas utilizando valores binarios se pueden representar las respuestas “verdadero” o “falso”. Estos valores son la base del sistema binario utilizado en la computación modernas. Por ejemplo, en el álgebra booleana, se pueden definir clases o funciones que agrupan elementos con características comunes, lo que es una práctica común en la programación y el procesamiento de datos. Específicamente en la inteligencia artificial, es imprescindible el álgebra booleana puesto que permite a las máquinas tomar decisiones y aprender de los datos.

Karel Čapek aportó a la inteligencia artificial a través de la literatura. Introdujo la palabra “robot” a través de su obra “Rossum's Universal Robots”, publicada en 1920. Este término describía a seres artificiales diseñados para realizar tareas laborales en lugar de los humanos, puesto que deriva de la palabra checa "robota", que significa "trabajo" o "servicio". Esta obra anticipó muchos de los debates éticos que hoy rodean a la inteligencia artificial y la robótica. Por ejemplo, se exploraron temas como la deshumanización del trabajo, la relación entre el ser humano y las máquinas, la rebelión de las máquinas y las consecuencias de crear seres con capacidad de pensamiento y emociones. Estas cuestiones siguen siendo relevantes en la actualidad, ya que los desarrolladores de IA enfrentan desafíos relacionados con la autonomía de las máquinas, los derechos de las inteligencias artificiales y el impacto social de la automatización.

Quien sí influyó indiscutiblemente a la inteligencia artificial fue Alan Turing, matemático, lógico y criptógrafo británico. Es considerado uno de los pioneros más influyentes no solo en el campo de la inteligencia artificial sino también en la computación moderna, con contribuciones fundamentales que sentaron las bases teóricas para el desarrollo de las máquinas inteligentes y el procesamiento de la información.

Una de las aportaciones más significativas de Turing fue la "Máquina de Turing", un concepto teórico que describió en 1936 en su artículo "On Computable Numbers, with an Application to the Entscheidungsproblem". La Máquina de Turing es un modelo abstracto de una máquina capaz de realizar cualquier cálculo que pueda ser formulado matemáticamente, utilizando una cinta infinita y un conjunto de reglas. Aunque esta máquina no fue construida físicamente, el concepto permitió formalizar la noción de algoritmo y computabilidad, lo que se considera un pilar en la teoría de la computación y, por ende, en la inteligencia artificial.

En 1950, Turing publicó su artículo "Computing Machinery and Intelligence", desde el que planteó la pregunta: "¿Pueden las máquinas pensar?". Además, propuso lo que hoy conocemos como el "Test de Turing", un experimento diseñado para determinar si una máquina es capaz de exhibir un comportamiento similar al de un ser humano en una conversación. Este experimento se convirtió en un estándar para evaluar la inteligencia de las máquinas y sigue siendo un punto de referencia en el debate sobre la conciencia y la inteligencia artificial.

Por otra parte, durante la Segunda Guerra Mundial, Turing desempeñó un papel crucial en el criptoanálisis, desarrollando métodos para descifrar los códigos generados por la máquina Enigma utilizada por los nazis. Su trabajo contribuyó al avance en el campo de la computación práctica al idear técnicas que se consideran precursoras del desarrollo de las computadoras digitales modernas (Bejerano, 2014).

Compartiendo la época revolucionaria junto con Turing se encontraba Konrad Zuse, ingeniero alemán considerado uno de los pioneros de la computación moderna,

desarrolló la primera computadora programable y completamente funcional en 1941. Aunque fueron 4 máquinas “Z”, la más destacada fue “Z3” y fue una máquina capaz de realizar cálculos complejos automáticamente, utilizando un sistema de relés y programas almacenados en cinta perforada, lo que la convierte en la primera computadora electromecánica del mundo (Luna, s.f.).

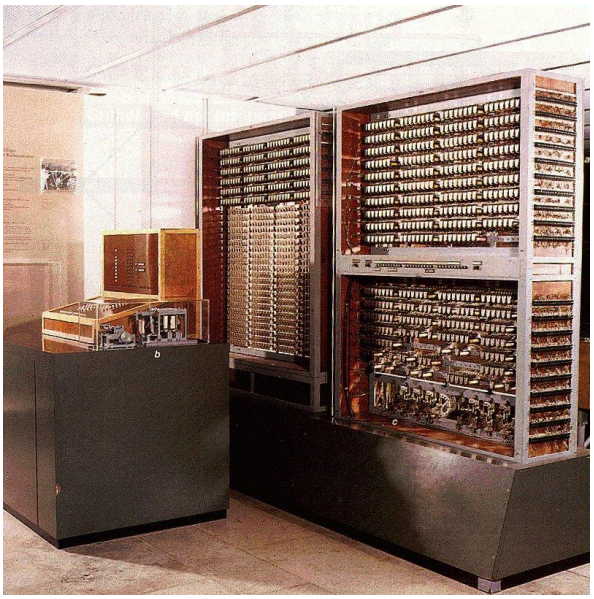

Figura 2: Reconstrucción de la máquina "Z3" (Fuente: angelfire.com)

Posteriormente creó “Z4”, una nueva versión en la que, entre otras mejoras, se admitían entradas en coma flotante de 32 bits y conseguía mejorar tiempos de procesamiento respecto a su antecesora. Por ejemplo, resolvía multiplicaciones unas 30 veces más rápido. Además, creó el primer lenguaje de programación de alto nivel, “Plankalkül”, que, aunque no se implementó durante su tiempo, sentó las bases para futuros lenguajes de programación.

Desde la presentación del “Test de Turing” por parte de Alan Turing en 1950, la inteligencia artificial comenzó a ser reconocida como un campo de estudio independiente, con objetivos claramente definidos y una creciente comunidad de investigadores dedicados a avanzar en esta disciplina. Es por eso que este año se considera un punto de partida fundamental para la evolución de la inteligencia artificial. En los años posteriores, la inteligencia artificial experimentó un desarrollo rápido y continuo, con hitos clave que han dado forma a la tecnología moderna. Por lo tanto, a continuación se tratará la evolución cronológica de la IA desde 1950 para capturar el surgimiento de las ideas y desarrollos más significativos que han llevado a la inteligencia artificial a su estado actual.

2.3.1 1950-1970 Génesis

Para comprender el desarrollo de la inteligencia artificial entre las décadas de 1950 y 1970, es esencial profundizar en el concepto del Test de Turing, ya que marcó un punto de inflexión que podría considerarse el verdadero comienzo del desarrollo de la inteligencia artificial.

El Test de Turing es una prueba diseñada para evaluar la capacidad de una máquina para exhibir un comportamiento inteligente indistinguible del de un ser humano. Esta consiste en un experimento en el que un evaluador humano interactúa con dos participantes ocultos: uno es una máquina y el otro un ser humano. Si el evaluador no puede distinguir consistentemente entre la máquina y el humano basándose únicamente en las respuestas de texto, la máquina se considera que ha pasado el test, demostrando así un nivel de "inteligencia" comparable al humano (Turing, 1950).

En la siguiente figura se puede observar una representación simple de este test:

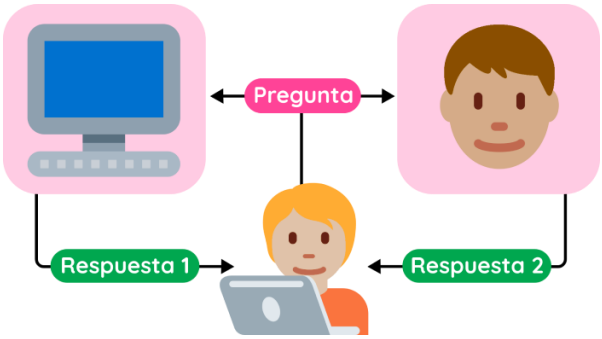

Figura 3: Representación del Test de Turing (Fuente: elaboración propia)

El Test de Turing fue revolucionario porque cambió la perspectiva sobre cómo evaluar la inteligencia en las máquinas. En lugar de intentar definir qué es la inteligencia, Turing propuso un criterio pragmático basado en la interacción humana. Esta idea desencadenó debates filosóficos y científicos que impulsaron a investigadores a explorar las posibilidades de crear máquinas capaces de pensar o, al menos, de simular el pensamiento humano de manera convincente. De hecho, Turing en su publicación admitía que un humano no podría alcanzar la velocidad o la precisión con la que una máquina puede responder en, por ejemplo, cuestiones aritméticas, por lo que le delataría fácilmente.

Por otra parte, dejó en recado para el futuro, puesto que reflejó en su artículo su creencia de que a finales del siglo XX los ordenadores tendrían una programación mucho más elevada que conseguiría que en 7 de cada 10 ocasiones, las máquinas conseguirían hacer creer en el Test de Turing que son humanos.

En marzo de 1950 Claude Shannon, conocido como el "padre de la teoría de la información", publicó un artículo titulado "Programming a Computer for Playing Chess". En este artículo, Shannon presentó uno de los primeros enfoques teóricos para enseñar a una máquina a jugar al ajedrez (Shannon, 1950).

Shannon propuso dos estrategias principales para que una computadora jugara al ajedrez: un método de búsqueda "completa" (evaluar todas las posibles jugadas) y un método de búsqueda "selectiva" (evaluar solo las jugadas más prometedoras). De hecho, calculó que serían 10^{120} jugadas posibles, lo cual se conoce como el “número de Shannon”. Debido al alto número de jugadas posibles, Shannon no consideraba factible que una máquina consiguiera resolver el ajedrez por su alto coste temporal.

Con esto, el ajedrez se convertiría en un campo clave para probar y avanzar en las capacidades de las primeras inteligencias artificiales. Además, es un ejemplo importante de cómo los primeros investigadores comenzaron a explorar la posibilidad de enseñar a las máquinas a realizar tareas complejas que requieren toma de decisiones y planificación estratégica.

En 1956, se celebró un evento que marcaría un antes y un después en la historia de la inteligencia artificial: la Conferencia de Dartmouth. Esta conferencia es considerada el punto de partida oficial del campo de la inteligencia artificial como disciplina científica. Fue en este contexto donde John McCarthy estableció el término "inteligencia artificial" al campo de estudio que buscaba crear máquinas capaces de simular procesos de la inteligencia humana (McCarthy, Minsky, Rochester, & Shannon, 1955).

La Conferencia de Dartmouth no solo introdujo el término "inteligencia artificial", sino que también estableció las bases teóricas y metodológicas que guiarían la investigación en las décadas siguientes. Este evento reunió a un grupo selecto de investigadores que sentaron las bases de lo que sería un campo de estudio independiente, con sus propios objetivos y métodos. Así, se produjo el nacimiento oficial de la inteligencia artificial como un campo de estudio, consolidando su identidad y estableciendo un punto de partida para las investigaciones que transformarían la tecnología en los años venideros.

> *"Todo aspecto del aprendizaje o cualquier otra característica de la inteligencia puede describirse en principio con tanta precisión que una máquina puede ser creada para simularla"*
>
> John McCarthy, Dartmouth College -- 1956

Más tarde en 1957, Alex Bernstein, junto con un equipo de IBM, desarrolló el primer programa de ajedrez para ordenador. Este programa, aunque primitivo en comparación con los estándares actuales, fue un avance significativo porque demostró la capacidad de una máquina para realizar una tarea compleja que requiere planificación, estrategia y toma de decisiones.

El programa de Bernstein fue uno de los primeros intentos de aplicar conceptos de IA a un juego que, históricamente, había sido considerado una demostración del intelecto humano. Funcionaba en una computadora IBM 704 (Compujedrez, 2016)y aunque el programa solo podía calcular hasta cuatro movimientos por jugada, marcó el inicio de un camino que llevaría, décadas después, al desarrollo de programas de ajedrez altamente sofisticados como Deep Blue, que en 1997 derrotó al campeón mundial Garry Kasparov.

Figura 4: Vista general de la computadora IBM 704 (Fuente: Wikimedia)

En 1958, Frank Rosenblatt presentó el perceptrón, un modelo revolucionario de red neuronal artificial que representó un avance importante en el campo de la inteligencia artificial. El perceptrón fue concebido como un sistema capaz de aprender y clasificar datos mediante una estructura de red simple.

El perceptrón se compone de entradas, cada una asociada con un peso, una función de activación y una salida. En este modelo, las entradas se multiplican por sus respectivos pesos y los resultados se suman. Esta suma se pasa a través de una función de activación que determina si la salida será activa (por ejemplo, 1) o inactiva (por ejemplo, 0). Esta estructura permite al perceptrón realizar tareas básicas de clasificación binaria. Utilizando un algoritmo de retroalimentación, el perceptrón aprende y corrige sus errores al modificar estos pesos, mejorando así su precisión en futuras clasificaciones. Esta idea de ajustar parámetros en función del error de predicción fue una de las primeras manifestaciones de aprendizaje automático en las máquinas.

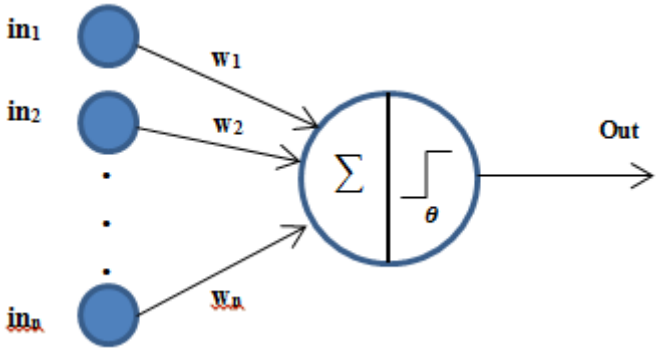

Figura 5: Representación gráfica de un perceptrón (Fuente: Google)

El perceptrón de Rosenblatt estableció las bases para el desarrollo de redes neuronales más complejas con las que se pretendía simular las neuronas de los seres humanos a una mayor escala. Su introducción marcó el inicio de un campo que continuaría evolucionando y llevando al desarrollo de modelos modernos de redes neuronales profundas.

Ese mismo año, John McCarthy volvió a aportar su granito de arena a la informática con el desarrollo del lenguaje de programación LISP (List Processing). LISP fue diseñado específicamente para manipular símbolos y listas, lo que lo hizo particularmente adecuado para las aplicaciones de inteligencia artificial por la representación y manipulación de datos complejos. Este lenguaje destacó por su capacidad de manejar funciones como datos y su soporte para la recursión y la abstracción de datos, características que lo hicieron extremadamente flexible y eficiente. Con esto, se conseguían manejar estructuras de datos más fácilmente, poder llamar a los algoritmos a sí mismos y tratar las funciones de primera clase.

A lo largo de los años, LISP se consolidó como el lenguaje de programación principal en el campo de la inteligencia artificial, siendo utilizado en la creación de programas de procesamiento de lenguaje natural, sistemas expertos y otros sistemas de inteligencia artificial. De hecho, es uno de los lenguajes de programación de alto nivel más antiguos que siguen siendo relevantes, solo por detrás de Fortran.

Con la creación del primer robot industrial del mundo en 1961 por George Devol y Joseph Engelberger, se consiguió automatizar tareas repetitivas y peligrosas en la industria manufacturera. Este robot fue instalado en la planta de General Motors de Nueva Jersey, donde se encargó de manejar piezas de metal caliente en la línea de ensamblaje. De esta manera mejoró la eficiencia, la seguridad y la precisión en la fabricación. Su introducción marcó el comienzo de la era de la robótica industrial, un campo que ha crecido exponencialmente desde entonces y que ha transformado radicalmente la producción en masa.

En 1964, Joseph Weizenbaum desarrolló ELIZA, un programa diseñado para simular una conversación humana, permitiendo a los usuarios interactuar con la computadora a través de un diálogo en lenguaje natural. Este programa se convirtió en un hito en la historia de la inteligencia artificial, ya que demostró cómo las máquinas podían emular la comunicación humana (Signorelli, 2024).

Mediante un conjunto de reglas predefinidas que le permitían analizar el texto ingresado por el usuario, ELIZA respondía de manera que imitaba el lenguaje humano. La versión más famosa de ELIZA fue el "doctor", que simulaba un psicoterapeuta Rogeriano.

Este enfoque permitió que el programa "escuchara" al usuario y devolviera respuestas basadas en las palabras claves identificadas en la entrada, lo que daba la impresión de que la máquina comprendía y participaba en una conversación significativa. Sin embargo, realmente no llegaba a comprender exactamente el contenido de los mensajes humanos.

Figura 6: Demostración del programa ELIZA (Fuente: ResearchGate)

El programa que sí conseguía analizar los datos es DENDRAL, creado por Edward Feigenbaum, Bruce Buchanan y Joshua Lederberg en desarrollo entre 1965 y 1975. Este proyecto fue diseñado para analizar datos químicos y ayudar a los profesionales a deducir la estructura molecular de compuestos orgánicos. DENDRAL es considerado el primer sistema experto ya que emulaba el conocimiento y las habilidades de un especialista humano en el campo específico de la química.

En 1965, Lotfi Zadeh presentó la lógica difusa, una extensión de la lógica clásica introducida como una forma de manejar la incertidumbre y la imprecisión en el razonamiento y la toma de decisiones. Se basa en la idea de que no todos los conceptos son absolutamente ciertos o falsos, sino que pueden tener grados de veracidad.

A diferencia de la lógica clásica, donde una afirmación es estrictamente verdadera o falsa, representados con los valores 1 y 0 respectivamente, la lógica difusa permite que las afirmaciones tengan un valor en un rango entre 0 y 1. Este enfoque es más adecuado para modelar fenómenos que no se ajustan a límites estrictos y definidos, como el lenguaje natural y la percepción humana. Esto pudo aplicarse posteriormente en el control de electrodomésticos o vehículos, procesamiento de imágenes o incluso en la toma de decisiones.

Para finalizar la etapa del Génesis, en 1969, Marvin Minsky y Seymour Papert, dos influyentes investigadores en el campo de la inteligencia artificial, publicaron la obra "Perceptrons", un análisis crítico de las redes neuronales artificiales, centrado principalmente en el modelo de perceptrón desarrollado por Frank Rosenblatt

Minsky y Papert examinaron las limitaciones matemáticas del perceptrón, demostrando que este tipo de red neuronal de una sola capa tenía serias restricciones en cuanto a los problemas que podía resolver. En particular, mostraron que los perceptrones eran incapaces de resolver problemas no lineales, poniendo en entredicho su capacidad para abordar tareas complejas.

La publicación de esta obra destapó los desafíos pendientes de la inteligencia artificial, lo que propició cierta incertidumbre sobre la verdadera viabilidad de la misma. Es por ello que durante los próximos años el avance de la misma sería tibio.

2.3.2 Década de los 70: El invierno

La publicación de "Perceptrons" por parte de Minsky y Papert tuvo un efecto desalentador sobre la investigación en redes neuronales durante este periodo. Muchos investigadores se alejaron del mundo de las redes neuronales y la financiación para este tipo de investigación disminuyó considerablemente. Estos años estuvieron marcados por el escepticismo sobre la viabilidad de las redes neuronales como herramienta para desarrollar inteligencia artificial. Aun así, se produjeron ciertos avances imprescindibles para su evolución.

En 1972, Alain Colmerauer y Robert Kowalski desarrollaron PROLOG, un lenguaje de programación basado en la lógica de predicados. Por ello mismo, se distingue por su capacidad para manejar problemas relacionados con la lógica y la búsqueda automática de soluciones, lo que lo convierte en una herramienta fundamental para la inteligencia artificial, especialmente en áreas como los sistemas expertos y el procesamiento de lenguaje natural.

La peculiaridad de este lenguaje es que los programadores describen lo que quieren lograr en lugar de cómo hacerlo. Para ello, se aprovecha su capacidad para representar y manipular conocimientos basados en reglas lógicas.

Un año más tarde, tras ser encargado por el gobierno británico, el matemático James Lighthill elaboró un informe en 1973 para evaluar el progreso hasta ahora de la inteligencia artificial, así como el posible potencial que podría tener en el futuro. Este informe es conocido como el “Informe Lighthill” y marcó un punto de inflexión en su evolución.

Lighthill fue muy crítico con el estado de la investigación en IA de la época, argumentando que muchos de los avances prometidos no se habían materializado en aplicaciones prácticas. Según el informe, los logros de la IA eran limitados y estaban muy lejos de cumplir las expectativas, especialmente en áreas como la comprensión del lenguaje natural, la visión por computadora y la robótica. Esto propició una reducción de la financiación para la investigación en inteligencia artificial (IAMAI, 2024).

A consecuencia de esta fría noticia, muchos investigadores y proyectos se cancelaron o se redirigieron hacia áreas con un éxito más asegurado. Sin embargo, también forzó a la comunidad de IA a reevaluar sus enfoques y a concentrarse en problemas más específicos, lo que a largo plazo contribuyó a un resurgimiento del campo en el futuro.

Durante estos años, Edward Shortliffe desarrolló el sistema experto MYCIN para ayudar a los médicos en el diagnóstico y tratamiento de infecciones bacterianas en la sangre, como la bacteriemia y la meningitis.

Para hacer uso del sistema, se regía por una serie de reglas if-then que permitían analizar los síntomas y los resultados de laboratorio proporcionados por el usuario para generar un diagnóstico y recomendar un tratamiento antibiótico específico. Además, era capaz de calcular la probabilidad de que un paciente tuviera una determinada infección, abriendo las puertas a la aplicación de sistemas en datos no absolutos.

Sin embargo, aunque en pruebas realizadas en comparación con expertos humanos, MYCIN demostró ser tan efectivo, o incluso más, que los médicos especialistas en ciertas áreas, MYCIN nunca se implementó clínicamente debido a preocupaciones legales y éticas relacionadas con la toma de decisiones médicas automatizadas.

En 1978 surgió la creación del sistema experto XCON (De “configuración experta”) por Digital Equipment Corporation, en adelante DEC. Su tarea era ayudar a configurar las órdenes de venta de sistemas de computación complejos, garantizando que las configuraciones propuestas cumplieran con los requisitos técnicos y las restricciones de compatibilidad. La configuración de sistemas en DEC involucraba la selección de una variedad de componentes y opciones y XCON ayudó a automatizar este proceso, reduciendo errores y optimizando el tiempo de configuración.

El sistema se basaba en un extenso conjunto de reglas y conocimientos específicos del dominio. Utilizaba una base de conocimientos codificada en forma de reglas de producción para realizar inferencias y tomar decisiones sobre las configuraciones adecuadas. Este sistema podía interpretar las especificaciones del cliente y generar configuraciones válidas, lo que permitía manejar un gran volumen de pedidos de manera más eficiente y precisa.

El éxito de XCON influyó en la evolución de los sistemas expertos y en la investigación sobre aplicaciones de la inteligencia artificial en la industria. Este sistema contribuyó a la comprensión de cómo se pueden aplicar las técnicas de IA para resolver problemas específicos del dominio y estableció una base para el desarrollo de sistemas expertos más sofisticados y especializados en diversas áreas.

Para finalizar la década más sombría de la inteligencia artificial, en 1979 se desarrolló el “Stanford Cart”, uno de los primeros vehículos autónomos. Estaba equipado con una cámara de video y una serie de sensores para captar información sobre su entorno y estaba diseñado para seguir un camino específico sin intervención humana. La navegación se realizaba mediante algoritmos de visión por ordenador y procesamiento de imágenes que permitían al coche detectar y evitar obstáculos.

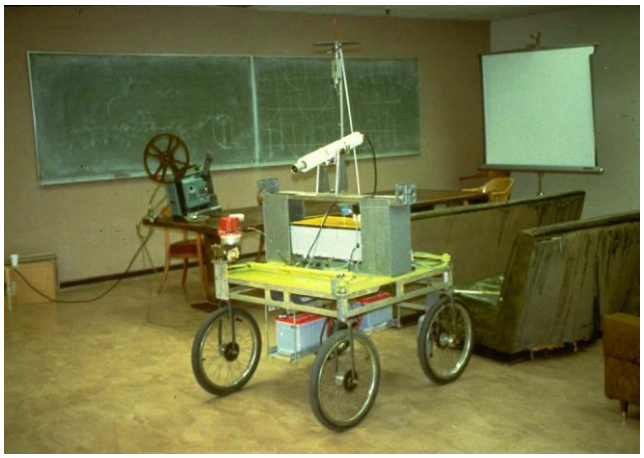

Figura 7: Stanford Cart (Fuente: Rodney Brooks)

Esta década condujo hacia una reestructuración y refinamiento de las ideas y tecnologías existentes. Las lecciones aprendidas durante este período crítico sentaron las bases para el resurgimiento de la inteligencia artificial en la próxima década. En los años siguientes, nuevos avances en hardware, algoritmos y aplicaciones prácticas revitalizarían el campo, llevando a una nueva era de crecimiento e innovación en la inteligencia artificial.

2.3.3 Década de los 80: El renacimiento

Con el comienzo de la década de 1980, el enfoque se trasladaría hacia la resolución de problemas complejos y la integración de técnicas emergentes, marcando el inicio de una era de expansión y desarrollo que impulsaría la IA hacia nuevas fronteras.

En esta década las redes neuronales volverían a coger fuerza gracias al desarrollo del algoritmo de retropropagación, fundamental para el entrenamiento de las redes neuronales multicapa. Esto permite ajustar los pesos de las conexiones entre las neuronas en una red para minimizar el error entre la salida producida por la red y la salida deseada. Este ajuste se realiza mediante la propagación del error desde la capa de salida hacia las capas de entrada.

Aunque el concepto de retropropagación ya existía desde los años 60, este año se popularizó y lo hizo accesible para una amplia gama de aplicaciones. Su trabajo demostró cómo el algoritmo de retropropagación podía ser utilizado eficazmente para entrenar redes neuronales con múltiples capas ocultas, resolviendo problemas complejos de clasificación y regresión.

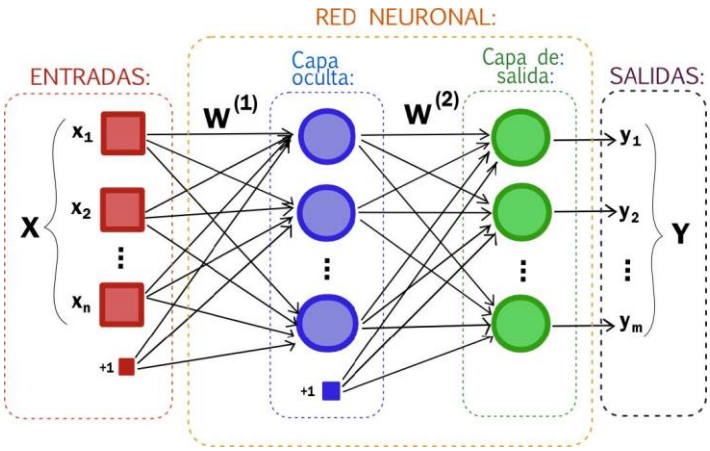

Figura 8: Representación de un perceptrón con capas ocultas (Fuente: IDECOR)

En 1986 terminó de revitalizarse el interés en las redes neuronales con la publicación de “Parallel Distributed Processing: Explorations in the Microstructure of Cognition", una obra de David Rumelhart, James McClelland y Jerome A. Feldman. En él se presentaba una visión detallada de cómo los sistemas cognitivos pueden ser modelados mediante redes neuronales artificiales que operan en paralelo y distribuyen el procesamiento a través de múltiples unidades (Rumelhart, McClelland, & Feldman, 1986).

Además, desarrolló la idea de redes neuronales distribuidas, en las que el conocimiento y el procesamiento se distribuyen a través de numerosas neuronas artificiales interconectadas. Al proporcionar la obra un marco teórico robusto y ejemplos prácticos, el libro ayudó a superar el "invierno de la inteligencia artificial" y a atraer atención y recursos hacia la investigación en redes neuronales. También sentó las bases para el desarrollo del aprendizaje profundo (deep learning) y el uso de redes neuronales profundas en la resolución de problemas complejos.

En 1987, los sistemas expertos comenzaron a expandirse a entornos comerciales y médicos, demostrando su versatilidad y utilidad en una amplia variedad de entornos prácticos. Algunos ámbitos fueron la planificación de la producción, la gestión de inventarios y la configuración de productos. Un ejemplo más concreto es el uso de estos sistemas para la configuración automática de productos complejos, donde el sistema ayudaba a elegir los componentes correctos y a asegurar la compatibilidad.

Los sistemas expertos también se utilizaron para asistir en la toma de decisiones estratégicas y operativas, proporcionando recomendaciones basadas en un análisis detallado de datos y escenarios. Esto permitió a las empresas mejorar la eficiencia y reducir los errores humanos en la toma de decisiones, y a los médicos ayudó a interpretar resultados de pruebas y a seleccionar opciones de tratamiento basadas en una amplia base de conocimiento médico.

En 1988, Rodney Brooks comenzó a desarrollar Genghis, uno de los primeros robots autónomos verdaderamente innovadores, ya que era capaz de operar de manera

independiente en entornos reales, utilizando un enfoque revolucionario de la inteligencia distribuida.

El comportamiento del robot emergía de la interacción de varios módulos simples que funcionaban en paralelo. Esto permitía al robot reaccionar de manera más rápida y efectiva a los cambios en su entorno, en lugar de depender de un único sistema central. Estaba pensado para operar en entornos no estructurados, donde debía enfrentarse a situaciones impredecibles. La simplicidad en el diseño, con el uso de sensores básicos y estrategias de comportamiento elemental, resultó en una navegación autónoma eficiente. Con alrededor de 1 kilogramo de peso, 6 patas y 22 sensores, demostró que un enfoque sencillo podía superar a los modelos más complejos en términos de eficacia en el comportamiento robótico (Robots Guide, s.f.).

Figura 9: Representación digital del robot Genghis (Fuente: robotsguide.com)

Para acabar la década, en 1989 siguió la implementación de las redes neuronales, esta vez en el procesamiento del lenguaje natural (PLN), marcando un cambio significativo desde los enfoques tradicionales basados en reglas y gramáticas formales. Estos modelos neuronales permitieron la creación de sistemas más flexibles y adaptativos para el análisis y la generación de texto.

También se introdujeron y popularizaron métodos estadísticos en el procesamiento del lenguaje, como los modelos de n-gramas, utilizando probabilidades de secuencias de palabras para mejorar la precisión en tareas como la traducción automática y la predicción de palabras.

Los avances en el PLN llevaron a mejoras significativas en la capacidad de los sistemas para interpretar y generar lenguaje humano de manera más precisa y natural. También hubo mejoras en la precisión de los sistemas de traducción automática y el análisis de texto. Esto facilitó el desarrollo de aplicaciones más avanzadas en traducción automática, análisis de texto y sistemas de diálogo.

Este año también se refinó otro concepto ya existente en los 60: los algoritmos genéticos. Estos algoritmos basados en los principios de la teoría evolutiva de la selección natural simulan el proceso de evolución biológica mediante la aplicación de

operadores como selección, cruce y mutación para evolucionar soluciones potenciales a problemas complejos.

En este año, David E. Goldberg presentó una serie de métodos y técnicas avanzadas para mejorar la eficiencia de los algoritmos genéticos y demostrar su aplicabilidad en una amplia gama de problemas de optimización en su libro "Genetic Algorithms in Search, Optimization, and Machine Learning". Estos algoritmos tienen la capacidad de resolver problemas de optimización difíciles de explorar mediante métodos más tradicionales, además de que son fáciles de adaptar a una gran variedad de problemas.

La década de 1980 marcó un período de renovación y avances cruciales en el campo de la inteligencia artificial, destacándose por la reintroducción de redes neuronales a través del algoritmo de retropropagación, la expansión de los sistemas expertos en aplicaciones comerciales y médicas y los avances en procesamiento del lenguaje natural.

2.3.4 Década de los 90: Entra en la sociedad

Esta década fue testigo de una expansión sin precedentes en el poder de computación, la disponibilidad de datos y el refinamiento de los algoritmos, dando lugar a innovaciones que transformaron profundamente la tecnología y la investigación en inteligencia artificial, introduciéndose directamente en la población en varios ámbitos.

En el inicio de esta década, se desarrollaron las máquinas de Soporte Vectorial (SVM), introducidas por Vladimir Vapnik, clave para el progreso de las tareas de clasificación y regresión. Gracias a esto, ahora se podían manejar problemas no lineales mediante el uso de funciones de núcleo, transformando los datos en un hiperplano que puede separarlos linealmente. Además, pueden hacer predicciones más precisas en base a los datos de entrenamiento. Así, se ha podido llevar a campos como la bioinformática, detección de fraudes y el procesamiento de texto.

En 1995, se publicó el algoritmo AdaBoost (Adaptative Boosting), que representa un avance significativo en el campo del aprendizaje automático. Este es un algoritmo de ensamblado que mejora el rendimiento de los modelos de clasificación al combinar múltiples clasificadores débiles en un clasificador robusto y preciso.

AdaBoost aumenta el peso de las instancias que fueron mal clasificadas durante el entrenamiento, permitiendo que los clasificadores subsiguientes se concentren en los ejemplos más difíciles. Esto se realiza en una serie de iteraciones, donde cada clasificador adicional se enfoca en los errores cometidos por los anteriores. Mejoraba la precisión de los modelos de clasificación y su flexibilidad para adaptarse a diferentes tipos de datos, lo que lo convirtieron en una herramienta valiosa para una amplia gama de aplicaciones. AdaBoost se convirtió en un algoritmo fundamental en el campo de la

IA, influyendo en el desarrollo de métodos de aprendizaje automático más avanzados y en la evolución de técnicas de ensamblado.

Con la llegada de Office 97 en 1996, Microsoft implementó una función en su entorno de ofimática que pretendía ayudar al usuario con sus tareas: el Ayudante de Office. El ayudante predeterminado era Clippit, aunque popularmente era llamado Clippy.

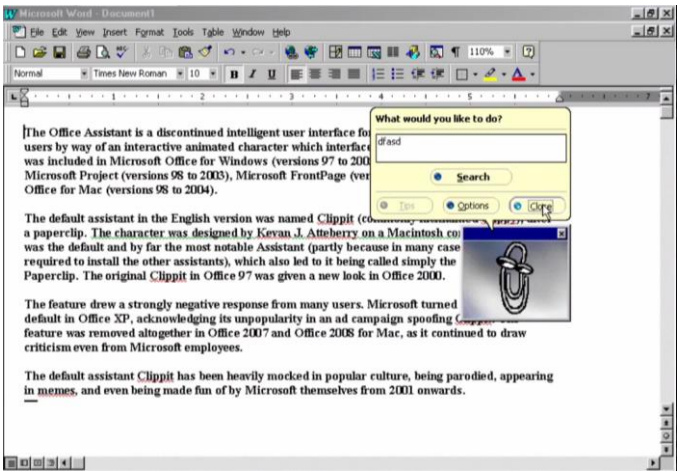

Figura 10: Bocadillo de asistencia de Clippit en Microsoft Word 97 (Fuente: NostalgiaWIndows)

Tenía como objetivo ayudar a los usuarios ofreciendo asistencia contextual y sugerencias mientras trabajaban en documentos. Utilizaba un enfoque de inteligencia artificial para interpretar las acciones del usuario y proporcionar ayuda basada en el contexto de lo que se estaba haciendo. Por ejemplo, si un usuario escribía “estimado:”, Clippit reconocía que estaba escribiendo una carta y ofrecía plantillas y consejos para mejorar el formato del documento.

El diseño y la implementación de Clippy marcaron un esfuerzo temprano en el desarrollo de asistentes virtuales con la intención de hacer que el software fuera más accesible y fácil de usar. Aunque Clippy fue innovador en su tiempo, recibió críticas mixtas debido a su intrusividad y su tendencia a aparecer en momentos inoportunos, lo que llevó a algunas frustraciones entre los usuarios (Swartz, 2003).

También en 1996, el campo de la inteligencia artificial alcanzó un hito significativo con la histórica victoria de Deep Blue, una supercomputadora desarrollada por IBM. Fue una máquina diseñada específicamente para jugar ajedrez a nivel profesional, que contaba con una potencia de procesamiento excepcional para poder evaluar hasta 200 millones de posiciones por segundo. Su estrategia y capacidad para analizar el juego le permitieron superar al entonces campeón mundial, Garry Kasparov, en una de las seis partidas de ajedrez.

No fue hasta un año después, en 1997, cuando jugaron la revancha y la máquina de IBM consiguió llevarse la victoria total en una serie de seis partidas con gran impacto social (Britannica, 2009).

Figura 11: Garry Kasparov jugando contra Deep Blue en mayo de 1997 (Fuente: Britannica)

El triunfo de Deep Blue sobre Garry Kasparov fue un evento simbólico que capturó la atención del público y la prensa, elevando el perfil de la inteligencia artificial y estimulando un mayor interés en sus aplicaciones no solo en el ajedrez, sino también en otras disciplinas estratégicas, dada la capacidad para competir en niveles que antes se consideraban exclusivamente humanos.

A partir de 1998, se empiezan a escuchar nombres más actuales (y comerciales). Google fue lanzado ese año por Larry Page y Sergey Brin mientras eran estudiantes en la Universidad de Stanford. Este motor de búsqueda innovador marcó un cambio radical en la forma en que las personas buscaban información en la web.

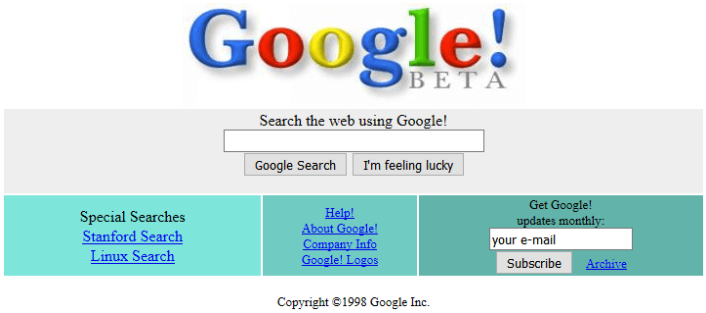

Figura 12: Página de inicio de Google en 1998 (Fuente: webdesignmuseum.org)

El principal avance que Google introdujo fue el algoritmo PageRank, que revolucionó el ranking de páginas web. Este algoritmo innovó al evaluar la importancia de una página web basándose en la cantidad y calidad de los enlaces que recibía de otras páginas.

Google se destacó no solo por su innovador algoritmo, sino también por su interfaz de usuario simple y rápida. La página principal de Google, con su diseño minimalista y su enfoque en la velocidad de búsqueda, contrastaba con las interfaces más complejas y cargadas de otros motores de búsqueda de la época, permitiendo su uso a usuarios menos experimentados y casuales.

En 1998, Furby hizo su debut en el mercado, marcando un hito importante en la industria de los juguetes. Desarrollado por Tiger Electronics, Furby fue uno de los primeros

juguetes interactivos que incorporó tecnología de inteligencia artificial en su diseño, revolucionando la manera en que los consumidores interactuaban con los juguetes.

Figura 13: Classic Furby, de 1998 (Fuente: onourshelf)

Utiliza una combinación de sensores, motores y algoritmos de procesamiento de lenguaje para interactuar con los usuarios. El juguete es capaz de reconocer y responder a estímulos como el tacto, el sonido y el movimiento, ofreciendo respuestas y comportamientos que parecen adaptarse a la interacción del usuario.

Una de las características distintivas de Furby es su habilidad para simular el aprendizaje de nuevas palabras y frases a medida que interactúa con el usuario. Aunque el aprendizaje era limitado en comparación con los estándares actuales, Furby podía cambiar su comportamiento y vocabulario en función de la cantidad y tipo de interacción recibida. Empezaba hablando su propio idioma “furbish” y con el tiempo empezaba a utilizar palabras del idioma programado (CNN, 1998). Esta capacidad de respuesta hizo que el juguete pareciera más "inteligente" y personalizado.

La implementación de esta inteligencia a un juguete hizo que se convirtiera en un fenómeno cultural y en un éxito comercial, evidenciando el potencial de la tecnología interactiva en el entretenimiento y el aprendizaje.

Empezando a cerrar los años 90, en 1999, Sony lanzó Aibo, un innovador perro robot cuyo nombre es una combinación de "AI" (inteligencia artificial en inglés) y "bō" (compañero en japonés). Fue diseñado para simular el comportamiento de un perro real, ofreciendo una experiencia interactiva y emocionalmente enriquecedora para los usuarios (Estrada, 2021).

Aibo estaba capacitado para realizar una variedad de acciones, como caminar, sentarse, girar la cabeza y jugar con juguetes interactivos. Además, Aibo tenía la capacidad de aprender y adaptar su comportamiento en función de las interacciones con sus propietarios, lo que le permitía desarrollar su propia personalidad única (Sony, 1999).

Aibo fue bien recibido por su capacidad para ofrecer una experiencia lúdica y emocional, aunque también se enfrentó a críticas y desafíos relacionados con su coste y la durabilidad de sus componentes. A pesar de estos desafíos, Aibo dejó una marca

duradera en el campo de la robótica, inspirando futuros desarrollos en robots personales y de compañía.

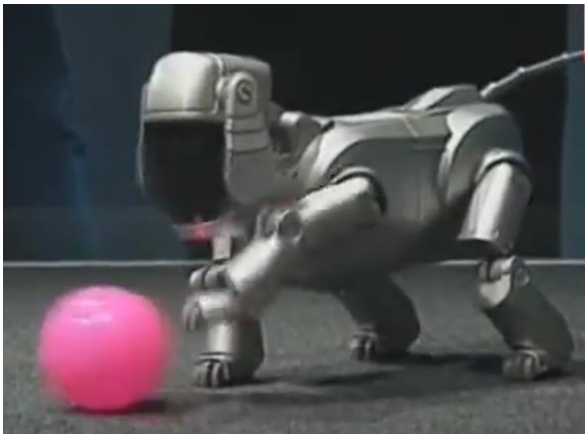

Figura 14: Aibo empujando una pelota en su presentación (Fuente: AP Archive)

A finales de esta década, el comercio electrónico experimentó una rápida expansión, marcando un hito importante en la integración de la inteligencia artificial en el entorno digital. Este crecimiento no solo amplió el alcance de las compras en línea, sino que también facilitó el desarrollo de sistemas de recomendaciones personalizadas.

Estos sistemas, que utilizan algoritmos de IA para analizar el comportamiento de los usuarios y predecir sus preferencias, comenzaron a jugar un papel crucial en la personalización de la experiencia de compra. Amazon y eBay fueron unos de los pioneros en recomendar productos a sus clientes en base a sus compras u otros artículos que hubiesen estado mirando. De esta manera se mejoró significativamente la experiencia del cliente y potenciando las ventas en plataformas de comercio electrónico.

Durante estos años, junto a la mejora del procesamiento del lenguaje natural con la introducción de modelos ocultos de Markov o el algoritmo de Viterbi, la inteligencia artificial seguía su evolución mientras cada vez abarcaba más áreas. Fue la época en la que se empezó a introducir en más productos directos al consumidor y las personas podían hacer uso de productos que ofrecieran tanto asistencia para realizar otros trabajos, como interactuar directamente con productos inteligentes como en el entretenimiento.

2.3.5 Década de los 2000: Consolidación del avance

La primera década del siglo XXI marcó un período de notable evolución en la inteligencia artificial, consolidando avances previos y abriendo nuevas posibilidades. Este periodo estuvo caracterizado por una profundización en tecnologías emergentes y la integración de la inteligencia artificial en aplicaciones cotidianas, reflejando el creciente impacto de la inteligencia artificial en diversos aspectos de la vida moderna.

Uno de los inventos que marcó este nuevo siglo fue ASIMO, un robot humanoide desarrollado por Honda. Podía interactuar con el entorno y las personas de manera natural.

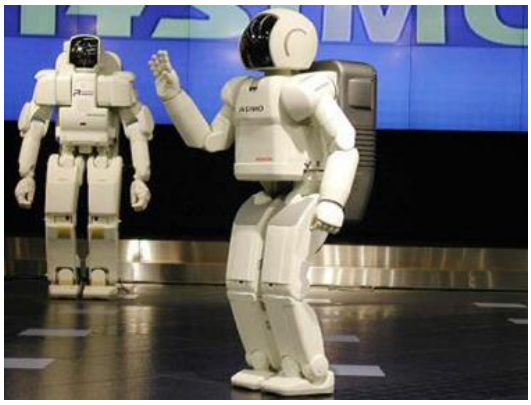

Figura 15: ASIMO en Robot Dream Exhibition, 2000 (Fuente: Kurita Kaku)

Desde su presentación en el 2000 y durante casi dos décadas hasta su descontinuación en 2018, fue recibiendo diversas actualizaciones que iban mejorando sus prestaciones. Era capaz de realizar acciones de movimiento como caminar, correr, saltar o bajar escaleras, pero también podía ocuparse de otras tareas como servir bebidas o abrir puertas.

ASIMO incorporaba una combinación de sensores avanzados, sistemas de equilibrio dinámico y procesamiento en tiempo real, lo que le permitía adaptarse a situaciones cambiantes y realizar movimientos complejos con gran fluidez.

En 2002, iRobot introdujo al mercado Roomba, un robot aspirador que revolucionó la limpieza doméstica al incorporar inteligencia artificial en su diseño. Este dispositivo compacto y autónomo está equipado con sensores y algoritmos que le permiten navegar de manera eficiente por el hogar, evitando obstáculos y adaptándose a diferentes tipos de superficies.

Figura 16: Caja y contenido del primer modelo Roomba de iRobot (Fuente: Vacuum Wars)

Roomba utiliza una combinación de tecnologías, incluyendo sensores de colisión, sensores de caída y un sistema de navegación basado en patrones de movimiento

aleatorios de manera que cubra sistemáticamente las diferentes áreas del suelo. El robot es capaz de ajustar su ruta y detectar áreas que necesitan más atención, mejorando así la eficacia de la limpieza sin intervención humana constante.

El lanzamiento de Roomba representó un avance significativo en la robótica de consumo, demostrando que los robots domésticos podían ser prácticos y accesibles para el público general. Su éxito abrió la puerta a una mayor innovación en la robótica personal y doméstica y reafirmó el potencial de la inteligencia artificial para automatizar tareas cotidianas.

En 2006, se lanzó oficialmente Apache Hadoop, una plataforma de software de código abierto diseñada para el almacenamiento y procesamiento de grandes volúmenes de datos a través de un clúster de computadoras, inspirándose en Google con el MapReduce y el Google File System, Hadoop fue desarrollado para gestionar y procesar datos masivos de manera eficiente y escalable, convirtiéndose rápidamente en una herramienta esencial en el ecosistema de big data (Cantalapiedra & Soler, 2017).

Hadoop permite dividir grandes conjuntos de datos en fragmentos más pequeños, que pueden ser procesados de manera paralela en múltiples nodos de un clúster. Esto no solo mejora la velocidad de procesamiento, sino que también facilita el manejo de datos estructurados y no estructurados en cantidades que antes eran difíciles de gestionar. Esto potenció el desarrollo de la inteligencia artificial y el aprendizaje automático, al proporcionar el entorno adecuado para manejar y procesar los grandes conjuntos de datos necesarios para entrenar modelos avanzados de inteligencia artificial.

En el mismo año, hubo un importante avance para el “Deep Learning”, puesto que Geoffrey Hinton junto con Simon Osindero y Yee-Whye Teh publicaron un trabajo que revitalizó el campo de las redes neuronales y marcó el inicio del Deep Learning moderno.

En este trabajo se mostró cómo las redes neuronales profundas podían superar las dificultades que habían limitado el desarrollo de las redes neuronales en décadas anteriores, particularmente en relación con la convergencia de datos y la capacidad de generalización. Mediante el uso de un algoritmo de preentrenamiento no supervisado, Hinton y compañía lograron mejorar significativamente el rendimiento de las redes profundas, abriendo la puerta a su aplicación en tareas más complejas.

También el 2006, aunque con mayor desarrollo en años posteriores, se presentó Watson, un sistema de inteligencia artificial desarrollado por IBM. Watson emplea una combinación de técnicas avanzadas de procesamiento del lenguaje natural, aprendizaje automático y minería de datos para interpretar preguntas complejas, buscar información relevante en su vasto conjunto de datos y proporcionar respuestas precisas. Este hito fue un claro ejemplo del potencial de la inteligencia artificial para superar las capacidades humanas en tareas específicas de análisis de información y toma de decisiones.

Unos años más tarde, en 2011, Watson se enfrentó a dos de los mejores jugadores de la historia del programa estadounidense ‘Jeopardy!’: Ken Jennings y Brad Rutter. En el concurso, Watson demostró su habilidad para manejar preguntas con ambigüedades, dobles sentidos y referencias culturales. Pese a todo eso, la inteligencia artificial de IBM consiguió vencer a ambos y alzarse con la victoria.

Figura 17: Watson (centro) concursando en 'Jeopardy!' en 2011 (Fuente: IBM Research)

Sin embargo, Watson no se limitó a concursos televisivos. También se aplicó en otros ámbitos, como la medicina, donde ayudó a oncólogos a diagnosticar y proponer tratamientos para el cáncer, analizando rápidamente estudios clínicos y datos de pacientes.

En 2008, Cynthia Mason presentó en su artículo "Giving Robots Compassion" la idea de otorgar a los robots emociones o comportamientos como tener compasión y empatía hacia los humanos. La propuesta incluía la integración de mecanismos que permitieran a los robots interpretar y responder a las señales emocionales humanas de manera que fomentaran una interacción más cercana y humana.

El artículo destacó la importancia de desarrollar sistemas de IA que no solo sean eficientes y funcionales, sino también capaces de construir relaciones más profundas y significativas con las personas. Esto tendría especial relevancia en aplicaciones relacionadas con el cuidado de personas y la asistencia en la vida cotidiana. Sin embargo, a parte de las mejoras, la propuesta de Mason abrió un debate ético acerca de la inteligencia artificial.

Los hitos de esta década reflejan una serie de rápidos avances en el campo de la inteligencia artificial, desde mejoras en la robótica y el aprendizaje profundo hasta la creación de sistemas que revolucionaron el reconocimiento de voz y la comprensión del lenguaje natural. Cada vez se introduce más el uso de la inteligencia artificial en la sociedad y las personas en consecuencia comienzan a beneficiarse de ella.

2.3.6 Del 2010 hasta 2024: el público general

Entrar en la década del 2010 trae consigo una gran cantidad de avances muy recientes pero que rápidamente se han ido quedando obsoletos, puesto que el avance de la

inteligencia artificial parece exponencial (aunque en algunos sentidos pueda ser realmente logístico). Durante este periodo se dan a conocer primeras versiones de sistemas o productos que siguen en desarrollo en la actualidad con una gran mejora desde su nacimiento o que han servido de influencia para otros sistemas.

En 2010, Microsoft lanzó Kinect para la consola Xbox 360, marcando un hito significativo en la interacción de los usuarios con los videojuegos y la tecnología de reconocimiento de movimientos. Kinect es un dispositivo de control por movimiento que utiliza una combinación de sensores de profundidad, cámaras de vídeo y micrófonos para permitir a los jugadores interactuar con los videojuegos mediante gestos y comandos de voz, sin necesidad de un mando físico.

Gracias a la capacidad de rastreo de movimiento y reconocimiento de la postura del jugador permitió una experiencia de juego más inmersiva y natural. Además, influyó en otras áreas fuera del entretenimiento, siendo capaz de traducir el lenguaje de signos, escanear cuerpos y objetos en 3D y mejorar la precisión de los microscopios.

Figura 18: Dispositivo Kinect para Xbox 360 (Fuente: E3)

Sin embargo, pese a seguir apostando inicialmente por Kinect con una nueva versión junto a Xbox One, la sucesora de 360, parece que ni usuarios ni desarrolladores sacaban todo el partido posible de este dispositivo, puesto que en 2017 Microsoft anunció que dejaría de fabricarlo (Pastor, 2017). La realidad aumentada y otros dispositivos como nuestro propio móvil consiguieron que su existencia fuese perdiendo relevancia hasta sentenciarla, lo que no significa que durante su vida haya aportado e influenciado desarrollo de otros sistemas y aplicaciones

En 2011, Mary Lou Maher y Doug Fisher hablaron sobre inteligencia artificial y sostenibilidad. Este programa, celebrado como parte de la conferencia anual de la Asociación para el Avance de la Inteligencia Artificial (AAAI), reunió a investigadores y expertos para investigar y discutir cómo la IA puede ser utilizada para enfrentar desafíos globales como el cambio climático, la gestión de recursos naturales y la reducción de residuos. Los participantes debatieron sobre temas como el diseño de algoritmos eficientes en términos de energía, la optimización de procesos para minimizar el impacto ambiental y el uso de modelos predictivos para apoyar la toma de decisiones sostenibles (Maher & Fisher, 2011).

La organización de este evento reflejó un interés creciente en combinar el avance tecnológico con los objetivos de sostenibilidad. Maher y Fisher facilitaron un espacio para la colaboración interdisciplinaria y la formulación de estrategias que integren la inteligencia artificial en las prácticas que promuevan un futuro más sostenible.

Google Brain fue lanzado en este año también como un proyecto de investigación. Fue fundado por Andrew Ng, Jeff Dean y Greg Corrado. Este proyecto se centró en el uso de redes neuronales profundas para mejorar diversas aplicaciones de inteligencia artificial. El objetivo principal de Google Brain era integrar la inteligencia artificial en productos y servicios cotidianos, aprovechando la gran cantidad de datos y la capacidad computacional de Google para entrenar modelos de aprendizaje profundo.

Los resultados de este proyecto se reflejaron al mejorar los productos de Google, como la precisión en las búsquedas, el reconocimiento de voz y las recomendaciones en YouTube, sino que también han influido significativamente en la comunidad de investigación en IA.

Al año siguiente, Google Brain entrenó el sistema con 10 millones de imágenes extraídas aleatoriamente de videos de YouTube. Durante el proceso, la red neuronal comenzó a identificar gatos sin haber sido específicamente programada para ello (Clark, 2012), demostrando que las máquinas podían aprender a reconocer patrones y objetos complejos por sí solas.

En 2011 también, comienza la aparición de sistemas de asistencia por voz en los dispositivos personales. Aunque al inicio abarcaban un número más reducido de resoluciones, su propósito era responder preguntas, hacer recomendaciones y realizar acciones dictadas por voz por el usuario, integrando progresivamente la inteligencia artificial a la vida cotidiana de las personas de manera más directa.

El primero en implantarlo fue Apple con Siri en 2011. Su introducción marcó un cambio importante en la forma en que los usuarios interactúan con sus teléfonos inteligentes. Por primera vez, podías pedir por voz activar una alarma, indicaciones en un mapa, el pronóstico del tiempo, poner recordatorios en el calendario y otras tareas.

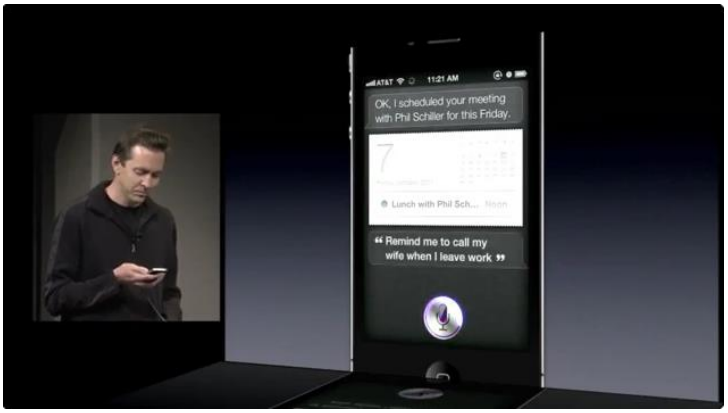

Figura 19: Presentación de Siri en un evento de Apple, 2011 (Fuente: YouTube)

Al año siguiente, en 2012, llegaría Google Now, de Google. Se destacó por su capacidad para proporcionar información contextualizada y proactiva basada en el análisis de datos y el historial del usuario. Google Now ofrecía tarjetas informativas sobre el clima, el tráfico y otros temas relevantes, utilizando la inteligencia artificial para anticipar las

necesidades del usuario antes de que fueran expresadas. Desde 2016 fue sustituido por el Asistente de Google.

Microsoft no se quiso quedar atrás y en 2014, junto con el lanzamiento de Windows 10 se dio a conocer Cortana. Este asistente se enfocó en ofrecer un asistente proactivo y personalizado, con capacidades de recordatorios, búsquedas web y gestión de tareas, así como la integración con otros servicios de Microsoft.

En el mismo año, Amazon lanzó también su propio asistente virtual, pero a diferencia del resto, no era un complemento de un dispositivo personal, sino que era el propio dispositivo. Con el lanzamiento del altavoz inteligente Echo, llegó Alexa. En este caso, en la asistencia primaba la integración con el hogar, dando pie a vincular y controlar otros dispositivos a través de Alexa, además de otras funciones como alarmas, escuchar música y una amplia gama de “skills”.

A raíz del éxito de estos asistentes virtuales, es normal que aparezcan cada vez más asistentes, como Bixby de Samsung, Alice de Yandex o Aura de Movistar. Algunos asistentes pueden centrarse en tareas menores o específicas de sus propias aplicaciones que los asistentes genéricos no estén capacitados de atender.

En este último año, Facebook dio a conocer su desarrollo de DeepFace (Simonite, 2014), aunque el despliegue no comenzaría hasta 2015. Con este sistema, la inteligencia artificial podía identificar y reconocer caras en fotografías con una precisión máxima del 97,3%, mientras que el estimado para los humanos es un 97,5%.

Para poder entrenar DeepFace, Facebook utilizó millones de imágenes de rostros disponibles en su propia red. Aunque fue un gran avance en el campo del reconocimiento facial al poder etiquetar y organizar fotos de manera automática, generó controversia sobre la privacidad y el uso ético de la IA, dado el potencial para el uso indebido de tecnologías de reconocimiento facial en vigilancia y control.

Sin embargo, la rápida evolución de la inteligencia artificial comenzaba a generar preocupaciones. En 2015, científicos e investigadores de renombre, como Stephen Hawking, Elon Musk y Steve Wozniak, junto con expertos en inteligencia artificial y robótica, hicieron un llamado urgente para prohibir el desarrollo y uso de armas autónomas. Esta carta fue presentada en la Conferencia Internacional sobre Inteligencia Artificial (IJCAI) y subrayó las preocupaciones éticas y de seguridad asociadas con el uso de sistemas de IA en armamento militar (Future of Life, 2015).

Los firmantes advirtieron sobre los riesgos potenciales de permitir que máquinas tomen decisiones letales sin intervención humana. Argumentaron que las armas autónomas podrían desencadenar conflictos no deseados, cometer errores fatales y agravar las tensiones internacionales. Además, expresaron su preocupación por la posibilidad de que estas tecnologías sean mal utilizadas por actores no estatales o que caigan en manos equivocadas.

La carta instó a los gobiernos y a la comunidad internacional a establecer regulaciones estrictas para controlar el desarrollo y la implementación de tales tecnologías. Los firmantes enfatizaron la necesidad de un marco legal y ético que garantice que las decisiones sobre el uso de la fuerza en situaciones de conflicto permanezcan bajo control humano directo y no sean delegadas a sistemas autónomos.

Este movimiento reflejó una creciente preocupación en la comunidad científica sobre los límites y las implicaciones morales del avance en la inteligencia artificial y la robótica, subrayando la importancia de abordar los desafíos éticos y de seguridad en la integración de estas tecnologías en áreas tan sensibles como la defensa y la seguridad global.

A partir de 2015 y durante los próximos años, en el mundo del Go, considerado uno de los juegos más complicados, aparecería un nuevo rival: AlphaGo, desarrollado por Google DeepMind. AlphaGo es un programa informático de inteligencia artificial que fue entrenado por millones de movimientos de humanos en GO. En 2015 venció al profesional de 2 dan Fan Hui por 5-0, en 2016 al 9 dan Lee Sedol por 4-1 y en 2017 al campeón durante dos años Ke Jie por 3-0.

Además, este último año desarrollaron una nueva versión que, a diferencia del anterior, no necesitaba aprender de movimientos humanos, sino que ya era capaz de aprender jugando contra sí mismo. Además, su versión AlphaZero consiguió aprender a jugar al ajedrez en cuatro horas y superar al hasta entonces mejor motor de ajedrez en 28 de 100 juegos, de los cuales los 72 restantes resultaron en empate.

A finales de 2015, aterrizaría una empresa que ha adquirido especial relevancia en los últimos años: OpenAI. Fue fundada por rostros reconocidos como Elon Musk o Greg Brockman como una organización de investigación en inteligencia artificial con la misión de asegurar que la inteligencia artificial beneficie a toda la humanidad, con un enfoque en la investigación abierta y colaborativa y con el compromiso de desarrollar IA de manera segura y equitativa.

A diferencia de otros proyectos de inteligencia artificial, OpenAI nació sin ánimo de lucro en contraparte al sector privado en el que los algoritmos desarrollados sí tenían fines lucrativos, a menudo sin considerar plenamente las implicaciones éticas y sociales.

Desde su creación, OpenAI, ha estado en primera línea de los avances de inteligencia artificial, como el desarrollo de la IA generativa con GPT (Generative Pre-trained Transformer). Su objetivo era maximizar los beneficios que se pudiesen obtener con la inteligencia artificial mientras se minimizaban los riesgos.

En 2017, Google vuelve a atraer los focos con el lanzamiento de Google Lens, una innovadora herramienta de reconocimiento de imágenes basada en inteligencia artificial. Google Lens está diseñada para permitir a los usuarios interactuar con el mundo a

través de la cámara de su smartphone, ofreciendo una variedad de funcionalidades que aprovechan el procesamiento avanzado de imágenes y el aprendizaje automático.

Con esta aplicación permite a los usuarios obtener información contextual sobre objetos, texto y lugares simplemente apuntando la cámara de su dispositivo. Entre sus capacidades destacan la identificación de objetos, la traducción de texto en tiempo real, la búsqueda de información sobre productos y la resolución de dudas sobre monumentos y lugares de interés. Además, Google Lens puede escanear y extraer texto de documentos y etiquetas, facilitando su almacenamiento y edición. Esta herramienta refleja el avance significativo en la inteligencia artificial y el procesamiento de imágenes, así como la interacción con el entorno.

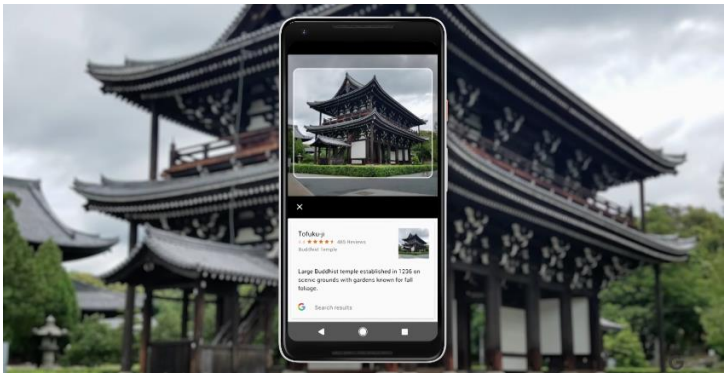

Figura 20: Presentación de Google Lens en el Google Event de 2017 (Fuente: Made by Google)

En 2018, la misma compañía lanzó BERT, un innovador modelo de procesamiento de lenguaje natural. Este se basa en la arquitectura de Transformer (desarrollado también por Google, en 2017) para comprender el texto de manera bidireccional. Es decir, consideraba tanto el contexto que precede a una palabra como el que le sigue.

La llegada de BERT mejoró significativamente el rendimiento en una variedad de tareas de procesamiento del lenguaje, como la búsqueda de información, la clasificación de texto y la traducción automática. En particular, Google integró BERT en su motor de búsqueda, lo que mejoró la precisión de las consultas de los usuarios, al comprender mejor las intenciones detrás de las palabras y frases utilizadas en las búsquedas.

En 2019, OpenAI lanzó GPT-2, el sucesor de GPT-1 como modelo de lenguaje natural Con esta nueva versión, se amplió de 117 millones de parámetros a 1,5 mil millones, permitiendo generar texto de una manera mucho más coherente y detallada para asemejarse a la capacidad de escritura humana.

Aunque es cierto que GPT-1 ya mostraba avances en la generación de texto fluidamente, GPT-2 fue capaz de mantener coherencia en textos más largos y abordar una mayor variedad de temas con precisión. Además, GPT-2 demostró ser más eficaz en tareas multitarea.

Sin embargo, debido a su capacidad para generar texto, se tomaron medidas para prevenir el mal uso de la herramienta ya que podía contener algunos sesgos. De esta

manera, OpenAI optó por una liberación controlada de GPT-2, a diferencia de su antecesor que sí se publicó sin restricciones.

Al año siguiente, fue el turno de GPT-3, que aumentó de nuevo los parámetros sobre su predecesor en un multiplicador de 116 veces más. Este modelo fue capaz de realizar diversas tareas, como traducir idiomas, escribir código, responder preguntas complejas y hasta crear obras de ficción.

Por último, hasta ahora, llegó GPT-4 en 2023 y continuó superando las expectativas. Con una arquitectura aún más compleja y un entrenamiento en una base de datos mucho más amplia, GPT-4 ha demostrado una capacidad sin precedentes para comprender y generar texto, código y otras formas de contenido creativo. Además, este modelo ha mostrado una mayor capacidad para razonar, resolver problemas y adaptarse a diferentes contextos más complejos.

En 2020, DeepMind presentó la segunda versión de AlphaFold, que consiguió predecir estructuras proteicas con un alto nivel de precisión durante la competición CASP14 (Critical Assessment of protein Structure Prediction). Demostró su gran capacidad al alcanzar un rendimiento comparable al de los métodos experimentales de referencia, como la cristalografía de rayos X y la resonancia magnética nuclear.

AlphaFold utiliza técnicas avanzadas de inteligencia artificial, como redes neuronales profundas y algoritmos de aprendizaje automático, para modelar cómo las proteínas se pliegan en su forma tridimensional. Su éxito ha sido un hito en la biología computacional y ha abierto nuevas posibilidades para el descubrimiento de fármacos, la ingeniería de proteínas y el estudio de enfermedades genéticas.

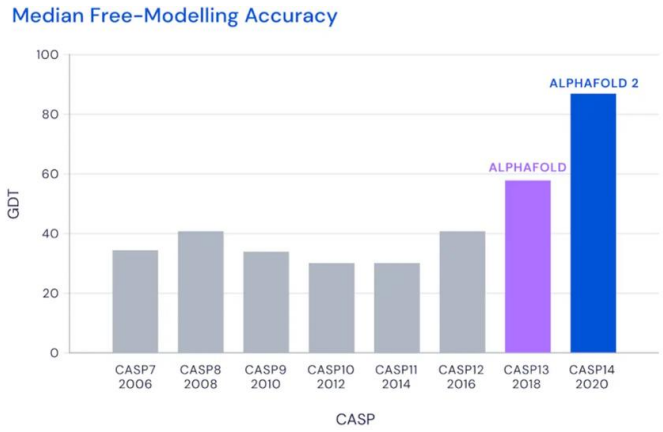

Figura 21: Resultados CASP a lo largo de los años. Se incluyen AlphaFold (2018) y AlphaFold 2 (2020) (Fuente: ResearchGate)

En 2021, DeepMind publicó el código fuente de AlphaFold y las estructuras para casi todas las proteínas conocidas en la ciencia, lo que ha permitido a investigadores de todo el mundo acceder a esta herramienta. Como consecuencia, AlphaFold ha sido reconocido como uno de los mayores logros en la historia de la biología y la inteligencia artificial.

A finales de 2022 tuvo lugar uno de los acontecimientos con mayor impacto social en referencia a la inteligencia artificial: la llegada de ChatGPT. Es una variante especializada de la serie de modelos de lenguaje GPT, diseñada para interactuar de manera conversacional con los usuarios. En sus inicios utilizaba el modelo GPT-3.5, ganando rápidamente popularidad debido a su capacidad para responder preguntas, asistir en tareas de escritura, generar ideas creativas y participar en diálogos complejos sobre una amplia gama de temas. Su lanzamiento representó una evolución significativa en la accesibilidad de la inteligencia artificial, haciéndola más accesible e interactiva para el público general.

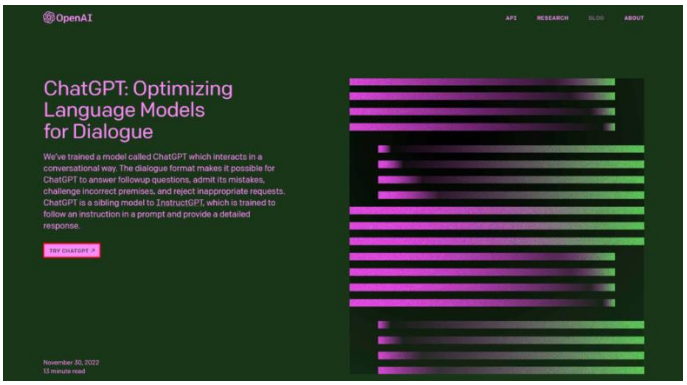

Figura 22: Página principal de ChatGPT en su lanzamiento (Fuente: OpenAI)

No obstante, la aparición de ChatGPT también causó revuelo en su parte ética. La habilidad del modelo para producir respuestas incorrectas o inventadas (lo que se conoce como "alucinaciones") que podían dar pie a la desinformación generaba cierta inquietud. La privacidad y la posibilidad de que se utilice para manipular a gran escala y, directamente, la sencillez con la que los usuarios pueden recurrir a ChatGPT para obtener información subrayó la importancia de supervisar y regular el uso de estos sistemas para prevenir efectos no deseados.

Aun así, el fácil acceso a esta herramienta al público general puso la inteligencia artificial en boca de todos. Dos meses después de su lanzamiento, ChatGPT ya contaba con 100 millones de usuarios, por lo que las grandes empresas empezaron a interesarse por los chatbots.

Microsoft dio el paso primero con Bing Chat, que ha ido evolucionando a lo largo de los años renombrándose a Bing AI y actualmente conocido como Copilot. La ventaja de este chatbot es que se integró en el motor de búsqueda de Bing y también en el propio navegador Microsoft Edge, por lo que todavía facilitaba aún más el acceso a las personas a la herramienta. Este utilizaba la tecnología de GPT-4, mientras que el propio ChatGPT solo permitía acceder gratuitamente a GPT-3.5.

Un mes después, llegó el turno de Google Bard, que a finales de año evolucionó a Google Gemini. Esta, a diferencia de la de Microsoft, no utilizaba GPT, sino que estaba basada en un modelo de lenguaje LaMDA (Language Model for Dialogue Applications) diseñado directamente por Google. Además, en 2024 con su versión Ultra de Gemini consiguió comprender no solo texto plano, sino también otros tipos de información, como imágenes, videos y código.

Desde los primeros desarrollos en los años 50, la inteligencia artificial ha experimentado una evolución impresionante, pasando de las primeras teorías y modelos rudimentarios a tecnologías avanzadas que moldean nuestro presente y empiezan a pintar nuestro futuro. A lo largo de las décadas, hemos visto desde el surgimiento de los primeros sistemas expertos hasta los sofisticados modelos de lenguaje y las aplicaciones prácticas actuales, como los asistentes virtuales y las herramientas de procesamiento avanzado.

Y uno de los ámbitos más importantes en los que la inteligencia artificial ha actuado es en la ciberseguridad (CCN-CERT BP/30, 2023). Gracias a ella, se puede combatir en tiempo real contra amenazas al identificar patrones anómalos o se anticipa a ellos. Además, puede responder de manera orquestada ante los incidentes de seguridad, proporcionando una rápida reacción. Además, ha conseguido que se pueda utilizar la biometría para verificar identidades con mayor precisión. Sin embargo, desde el uso de la IA en la ciberseguridad se han encontrado también con ciertas adversidades, como la manipulación maliciosa para generar resultados incorrectos intencionadamente, causando falsos positivos y falsos negativos. Además, genera un exceso de confianza y la supervisión humana se rebaja. Además, el CCN también advierte de las preocupaciones sobre la privacidad y las implicaciones éticas.

La constante innovación, impulsada por avances en hardware y algoritmos, ha permitido a la inteligencia artificial introducirse e integrarse en diversos aspectos de nuestra vida cotidiana, mejorando la eficiencia y abriendo nuevas posibilidades. Al mirar hacia el futuro, el continuo desarrollo en IA promete transformar aún más la forma en que interactuamos con la tecnología y enfrentamos desafíos globales.

2.4 Desafíos éticos actuales

En la búsqueda de una inteligencia que pueda compararse a la de un ser humano, o que incluso en algunos ámbitos quiera superarla y trascender sobre los límites que estableció la mente humana, no todo son ventajas. Detrás de una inteligencia artificial hay unos algoritmos que pueden haberse programado erróneamente, pero también pueden surgir problemas a raíz del mal uso del diseño final.

De esta manera, pueden aparecer errores que generan problemas por cómo aprenden las máquinas debido a un mal diseño, asociando conceptos erróneamente y, por otra parte, podemos encontrar personas que aprovechen las debilidades de la inteligencia artificial para sacar rédito, vulnerando el bienestar que se pretende conseguir.

A continuación, se analizan varios problemas de la inteligencia artificial que afectan directamente a la ética y que se están sufriendo actualmente.

2.4.1 Sesgo

Uno de los principales problemas a los que se enfrenta la inteligencia artificial día a día es el sesgo algorítmico, con el que los sistemas pueden dar resultados o tomar decisiones subjetivas y que provocan injusticia hacia ciertos grupos o individuos. Además, es difícil de detectar porque puede introducirse en diferentes etapas, como la recopilación de datos inicial o la fase de entrenamiento.

Debido a que los algoritmos de inteligencia artificial se entrenan utilizando grandes cantidades de datos, muchas veces históricos, si estos datos reflejan desigualdades o prejuicios sociales, el sistema de inteligencia artificial puede confundir esos datos como todavía relevantes en la actualidad y actuar y evolucionar en base a esos sesgos. En la última década, con el auge de los algoritmos de reconocimiento facial, este problema se ha visto destapado.

Desde 2017, la investigadora del MIT Joy Buolamwini ha estado trabajando en 'The Coded Gaze', con el que ha demostrado que los algoritmos de reconocimiento facial están mejor entrenados cuanto más blanca es la piel del individuo analizado y también reconoce mejor a los hombres que a las mujeres (Buolamwini, MIT Media Lab, 2017). Esto se debe a que hay ciertos rasgos faciales para los que los sistemas no están tan entrenados. De hecho, hizo una prueba en la que se enfrentaba a un sistema de reconocimiento facial y no reconocía ningún rostro, mientras que una máscara totalmente blanca sí la reconocía.

Figura 23: Buolamwini (izquierda) y Buolamwini con máscara (izquierda) (Fuente: MIT ML)

Buolamwini se juntó con el investigador Timnit Gebru para probar la precisión de algunos de los servicios populares de reconocimiento facial de grandes empresas. En este caso, se estudiaron los de Microsoft, Face++ e IBM. Dieron como entradas una serie de caras a cada servicio y descubrieron que los servicios funcionaban mejor con caras masculinas que con caras femeninas. Además, el tipo de cara que más se les resistía a los sistemas era el de las mujeres más oscuras, con una tasa de error de hasta un 34,7% en el caso de la de IBM frente al 0,3% del hombre blanco (Buolamwini & Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, 2018).

El gran problema se debe en parte a que los datos de entrenamiento utilizados para desarrollar estos sistemas a menudo incluyen una cantidad desproporcionada de imágenes de personas blancas, lo que resulta en un rendimiento deficiente cuando se enfrentan a rostros de otras etnias.

Tal y como ocurre con las caras y las personas, puede ocurrir con otros objetos o sentencias, en las que, al haber sido entrenadas principalmente con datos en base a sus creadores, si no se tiene en consideración otras vivencias o ejemplos de aquellos grupos de personas que no han participado en el entrenamiento de datos, pueden verse afectadas con resultados no aceptables en ciertos contextos.

Así como los sistemas de inteligencia artificial pueden tener sesgos por sus datos de entrenamiento, también pueden tomar decisiones discriminatorias por un mal diseño. De esta manera no son los datos de entrada los culpables de que aprendan de una manera u otra, si no en cómo el propio sistema procesa y pondera la información.

Existen casos en los que estos sistemas evidencian una falta de comprobación de diseño, como el sistema COMPAS, un algoritmo utilizado para evaluar el riesgo de reincidencia de los delincuentes. Este sistema tiene dos modelos de riesgo: reincidencia general y reincidencia violenta, a partir de los cuales cuenta con diferentes escalas basadas en riesgos dinámicos y estáticos. En base a esto, se ha demostrado que COMPAS tiende a sobreestimar el riesgo de reincidencia en personas negras y a subestimar el riesgo en personas blancas. De esta manera, la clasificación tanto de “mayor riesgo, pero no reincidió” como “menor riesgo, pero reincidió” es aproximadamente 20 puntos superior en los afroamericanos que en los blancos (Avella, Sanabria-Moyano, & Dinas-Hurtado, 2022), explicando el error discriminatorio.

También pueden aplicarse sesgos en el ámbito económico, ya que podrían negarse arbitrariamente préstamos a personas que viven en barrios tradicionalmente marginados. Esto se debe a que los algoritmos utilizados para evaluar la solvencia de los solicitantes de crédito pueden discriminar a minorías étnicas o a personas de bajos ingresos si se basan en datos que reflejan disparidades históricas en el acceso al crédito. Por ejemplo, esto ocurre cuando un algoritmo considera la ubicación geográfica como un factor importante sin tener en cuenta la discriminación histórica en ciertas áreas.

Aunque no siempre es culpa de los desarrolladores: a veces el usuario final dirige a la inteligencia artificial hacia una mala conducta, sea consciente o inconscientemente. Por ejemplo, en páginas como YouTube que basan sus recomendaciones en los sesgos de otras personas, replicando esas recomendaciones a otros usuarios para descubrir nuevo contenido, pueden reproducirse recomendaciones sesgadas y polarizadas. Esto también puede causarse de manera intencionada. Uno de los mayores ejemplos fue la inteligencia artificial Simsimi, que en 2016 consiguió gran popularidad en redes puesto que fue uno de los primeros chatbots que respondía a mensajes humanos. Este estaba basado principalmente en respuestas de otros usuarios, por lo que no era de extrañar que cambiase de actitud fácilmente. Y así fue, al poco tiempo de popularizarse, sus conversaciones adoptasen un tono más ofensivo, con mayores connotaciones sexuales y actitudes racistas, lo que causó gran polémica y puso el foco culpabilizador a la inteligencia artificial (Ramíguez, 2016).

En definitiva, no son pocos los ejemplos a lo largo de estos años en los que se excluye o infravalora sistemáticamente a minorías, generando desigualdades e injusticia social. Esto conduce hacia una pérdida de confianza en la tecnología en la que el público se niega a utilizar herramientas que la utilicen, frenando su evolución y capando sus posibilidades.

2.4.2 Alucinaciones

Las alucinaciones en la inteligencia artificial se producen especialmente en los modelos de procesamiento del lenguaje natural como los generadores de texto. Estas "alucinaciones" ocurren cuando un sistema de IA genera información que es incorrecta, incoherente o completamente ficticia, a pesar de que el sistema lo presente con gran confianza y aparente coherencia. Estas salidas erróneas pueden tener serias implicaciones, especialmente cuando los usuarios confían en la precisión y veracidad de la información generada. Es por ello por lo que, normalmente, los generadores de texto te advierten de la posibilidad de contener errores en sus respuestas.

La creación de estas alucinaciones normalmente viene dada cuando los datos contienen información ambigua, incorrecta o contradictoria, de manera que el modelo genere respuestas que mezclen o combinen de manera errónea esa información. También puede ocurrir cuando la cantidad de entradas es tan grande que el contenido sea demasiado complejo y la inteligencia artificial distorsione sus respuestas.

Las alucinaciones también pueden ocurrir cuando el sistema de IA no comprende completamente el contexto en el que se realiza una pregunta o solicitud. La IA puede generar respuestas basadas en patrones previos sin captar los matices necesarios que rodean la pregunta para una respuesta precisa.

Aunque las alucinaciones de la inteligencia artificial son más frecuentes de lo que parece, Google fue señalado en febrero de este 2024 por su generador de imágenes en Gemini. A través de él, se podían crear imágenes históricamente erróneas. Entre ellas, esta inteligencia artificial daba como resultados imágenes como soldados nazis alemanes de raza negra o vikingos asiáticos (Castillo, 2024). Esto provocó furor en las redes, lo que desembocó en la desactivación de la generación de imágenes en la IA de Google, que sigue inactiva a fecha de hoy.

Figura 24: Ejemplos de imágenes generadas por Google Gemini con alucinaciones (Fuente: El Diario)

Por otra parte, también se puede “incitar” a la inteligencia artificial a que de un resultado erróneo. Por ejemplo, preguntas como “¿Quién fue el superviviente del Titanic?” puede dar como respuesta “La superviviente del Titanic fue Millvina Dean. […]”, dando a entender que fue la única superviviente, mientras que fueron más de 700 personas las que lo hicieron.

La generación de información incorrecta por parte de la inteligencia artificial puede tener implicaciones legales y éticas, especialmente si se usa en contextos donde la precisión es crítica, como en la redacción de contratos o la interpretación de leyes. También puede causar daños a la integridad física con algunas afirmaciones como en indicaciones erróneas de uso de medicamentos.

2.4.3 Transparencia

Muchos sistemas de inteligencia artificial, especialmente aquellos basados en redes neuronales profundas, son considerados "cajas negras" porque sus procesos internos son opacos y difíciles de entender incluso para sus propios desarrolladores. Esto significa que, aunque los sistemas puedan ser altamente efectivos, sus decisiones no pueden ser fácilmente explicadas. Sin este acceso, es difícil para los usuarios y reguladores entender las limitaciones y sesgos potenciales del sistema.

Para ello, existe el término de la “explicabilidad”. Un sistema de inteligencia artificial es explicable cuando sus decisiones pueden ser interpretadas y comprendidas por humanos. Esto es particularmente importante en modelos complejos como los de deep learning, donde las decisiones se basan en múltiples capas de cálculos matemáticos que son difíciles de desgranar., De este modo, la transparencia no solo se refiere a la capacidad de "ver" cómo funciona un modelo, sino también a la capacidad de comprender, explicar y auditar sus decisiones y resultados.

En el ámbito críticos, como el médico, un sistema de inteligencia artificial que recomienda tratamientos debe poder explicar por qué ha sugerido un determinado

tratamiento, basándose en los datos o síntomas del paciente, para que los médicos puedan confiar y comprender su decisión para poder actuar en consecuencia. La opacidad puede llevar a desconfianza o a decisiones erróneas si los médicos no consiguen entender el porqué de la decisión tomada por la inteligencia artificial.

Además, gracias a la transparencia se pueden mejorar los propios sistemas de inteligencia artificial gracias a la trazabilidad. Con ella se permite la capacidad de seguir la cadena de decisiones de un algoritmo de inteligencia artificial desde su entrada hasta su salida. Esto es fundamental cuando se producen errores o se presentan problemas, puesto que así los investigadores pueden comprender qué decisiones tomó el sistema en cada momento y por qué, ayudando a determinar si hubo un fallo en el sistema o en su diseño.

2.4.4 Tratamiento de datos

Una de las características de la inteligencia artificial es la capacidad de tratar con grandes cantidades de datos. El problema viene cuando los datos utilizados tienen dueño y se usan tanto explícitamente para entrenar los sistemas como también, a través de información del usuario final, el sistema recopila información para luego utilizarla. Esto precisa de un correcto tratamiento de datos para garantizar una buena gestión, seguridad y privacidad. La negligencia en esta área puede conducir a violaciones de privacidad, brechas de seguridad, pérdida de confianza del público y consecuencias legales significativas.

Uno de los problemas éticos en el tratamiento de datos radica en el consentimiento. Para que el uso de los datos sea ético, los individuos deben dar su consentimiento informado sobre cómo se recopilarán, procesarán y utilizarán sus datos. Sin embargo, en la práctica, el consentimiento informado a menudo es insuficiente debido a la complejidad de los sistemas de IA y la falta de transparencia en las políticas de privacidad.

Por otro lado, está se encuentra el uso de datos para otros fines sin el consentimiento explícito de los individuos. Esto viola el principio de limitación del propósito, que establece que los datos deben ser utilizados solo para los fines para los que fueron recolectados. La práctica más común es vender información a terceros para publicidad dirigida. Este tipo de uso puede llevar a la explotación de datos personales de manera que los usuarios no anticiparon ni autorizaron, por lo que es ética y legalmente cuestionable.

Tal vez el problema no tendría una criticidad tan alta si no fuese porque en muchas ocasiones, esos datos incluyen información sensible, como datos personales o incluso confidenciales. Además, esto ocasiona preocupaciones acerca de la vigilancia y el control social, puesto que se cree que los gobiernos y las empresas pueden utilizar datos para monitorear y controlar el comportamiento de los individuos, lo que puede llevar a la erosión de libertades civiles y derechos humanos. Varios países incluyen el uso de sistemas de reconocimiento facial por parte de gobiernos para vigilar a la población, lo

que puede facilitar la represión de la disidencia y la violación de derechos fundamentales.

Uno de los países que utiliza esta tecnología es Reino Unido, aunque su fin del uso de esta tecnología va enfocado más a la detección de criminales (Metropolitan Police, 2024), proteger a personas vulnerables y a la sociedad en general. Sin embargo, otros países como China están acusados de utilizar esta tecnología para perseguir a los musulmanes de la etnia uigur (CBS News, 2019), o como Rusia, para controlar a los protestantes o periodistas (Salaru, 2022). Además, también utilizan otras tecnologías como datos biométricos para fines similares.

Por otra parte, aunque los datos pueden ser anonimizados para proteger la privacidad, las propias técnicas avanzadas de inteligencia artificial pueden reidentificar a los individuos combinando diferentes fuentes de datos (Romero, 2019). Esto plantea un problema ético grave, ya que los individuos pueden ser identificados a partir de datos que se suponían anónimos. Un estudio de Nature Communications reveló que combinando 15 atributos de datos podrían reidentificar al 99,98% de la población estadounidense de Massachusetts (Rocher, Hendrickx, & Montjoye, 2019).

Otro caso más reciente ha sido el del anuncio de comienzo de uso de datos para entrenar la inteligencia artificial generativa de Meta desde finales de junio de 2024, afectando a Facebook, Instagram, WhatsApp y Threads. Todos los usuarios estarán marcados por predeterminado como aceptados para participar en su entrenamiento de IA. Los usuarios de países de la Unión Europea, gracias la protección de datos en sus leyes, pueden aplicar para no participar en este entrenamiento. Sin embargo, la compañía de Mark Zuckerberg no quiere facilitar las cosas y en vez de existir una opción de marcar y desmarcar, hay que rellenar un formulario en el que expliques por qué no quieres participar en el entrenamiento de su inteligencia artificial y en qué te afecta su procesamiento de datos. Este formulario será revisado por el personal de Meta y decidirán si tu razonamiento es aceptado y dejarán de utilizar tus datos (Heikkilä, 2024). Sin embargo, otros usuarios no residentes de la UE, como por ejemplo los estadounidenses, al no tener leyes tan estrictas que protejan sus datos, no pueden aplicar a este formulario.

Es necesario destacar que no siempre son las inteligencias artificiales las que hacen mal uso de los datos que recopilan. A pesar de las cada vez más sofisticadas medidas de seguridad que se implementan en los sistemas, los ciberdelincuentes encuentran constantemente nuevas vías para explotar las debilidades de los sistemas digitales. La filtración de información sensible, como datos personales o corporativos, puede acarrear consecuencias devastadoras, desde el daño reputacional hasta el chantaje y el robo de identidad. Es importante que tanto los usuarios individuales como las empresas comprendan la gravedad de estos riesgos y adopten medidas constantes para proteger su información de los ciberdelincuentes que progresivamente tienen más puertas de acceso que intentarán abrir.

2.4.5 Influencia en decisiones

Al rastrear nuestra actividad y lo que nos gusta, las grandes empresas pueden analizar cómo reaccionamos ante las cosas que vemos o usamos en internet. Estos perfiles son tan precisos que pueden predecir qué pensamos o sentimos y hasta qué nos va a gustar. Con los datos que recopilan crean perfiles muy detallados de cada persona, teniendo en cuenta actitudes o gustos que nosotros no percibimos de nosotros mismos. Es como si supieran lo que vamos a hacer antes de que nosotros mismos lo hubiésemos pensado siquiera. Esto les da la capacidad de poder anticiparse a nuestros pensamientos, lo que nos hace más fáciles de influenciar.

Esta capacidad de la inteligencia artificial para predecir nuestros comportamientos plantea un escenario inquietante. Al conocer nuestros gustos, miedos y hábitos, la inteligencia artificial puede manipular nuestras decisiones de manera sutil pero efectiva. Pueden llegar a conocer el funcionamiento de nuestra mente casi a la perfección, saben cuáles son nuestros principios y, en consecuencia, cómo pueden neutralizarlos o atacar a puntos débiles. Esto nos hace más vulnerables a la influencia externa, ya que nuestras decisiones, a menudo basadas en emociones y experiencias pasadas, pueden ser fácilmente moldeadas por algoritmos diseñados para ese fin. De esta manera, pueden llegar a establecer sus propios valores en nuestra mente, capando nuestros razonamientos propios.

Uno de los casos más sonados fue el de Cambridge Analytica en 2018. Sin relación directa con la universidad homónima, fue una empresa que se dedicaba a “cambiar el comportamiento de la audiencia”, según indicaba su propia web, a través del análisis de datos. Era principalmente utilizada para campañas publicitarias de marcas y políticos, cuyas campañas más notorias fueron una a favor de la presidencia de Trump en las elecciones de 2016 de Estados Unidos y otra a favor del Brexit con *Leave.EU*. El escándalo que dilapidó a la empresa fue el del continente americano.

Todo comenzó en 2013 con un test de personalidad desarrollado por Aleksandr Kogan a modo de proyecto personal. Al realizar este test, había que conectar tu cuenta de Facebook y aceptar que la aplicación tuviera acceso a varios datos. De hecho, aunque fueron unos 265.000 usuarios los que realizaron el test, los afectados fueron más de 50 millones de usuarios, ya que, en los permisos, además de a la información personal, se daba acceso a la red de amigos. Fue entonces cuando Kogan decidió vender estos datos a la empresa de Analytica, pese a que en las políticas de Facebook eso no estaba permitido.

De esta manera, teniendo información de casi el 15% de la población de Estados Unidos, combinaron la información del test de personalidad junto con la recopilada de los perfiles de Facebook para poder crear mensajes personalizados par la psicología de cada uno de los usuarios, sabiendo tanto el tema, como el contenido y el tono para poder captar a cada usuario y poder inferir en su razonamiento lógico. Este caso, además propiciar el cierre de Cambridge Analytica, supuso un fuerte revés en bolsa para Facebook, al perder confianza de los usuarios por la seguridad y privacidad de su

información, con una caída de casi un 7% y 37.000 millones de dólares (BBC Mundo, 2018).

A medida que los sistemas de inteligencia artificial están más integrados en la toma de decisiones de diversas áreas, es crucial que se desarrollen marcos éticos sólidos para guiar su diseño y uso. Además, es fundamental que los desarrolladores, las empresas y los reguladores trabajen juntos para establecer estas normas y prácticas que aseguren que las decisiones que tomen los sistemas de inteligencia artificial sean justas y responsables. Esto no solo ayudará a mitigar los riesgos éticos, sino que también fomentará la confianza pública en la tecnología, lo que es esencial para su adopción y éxito a largo plazo.

2.4.6 Propiedad intelectual

Con la llegada de la inteligencia artificial generativa, cada vez es más fácil crear una imagen a partir de tres palabras o incluso un vídeo con una simple descripción. Incluso ya pueden modular voces o crear canciones. Cada vez es más complicado saber si el autor de un documento audiovisual tiene nombre y apellidos o si simplemente tiene un nombre artificial. Lo que sí se sabe es que muchos de estos sistemas de inteligencia artificial que crean obras audiovisuales han sido entrenadas con muestras de datos sin el consentimiento de los autores originales, lo que ha generado polémica en los últimos años.

Cuando una inteligencia artificial genera una obra, como una pieza musical o un diseño gráfico, surge la pregunta de quién debería ser considerado el autor y quién posee los derechos de propiedad sobre esa obra. Sin embargo, cuando esta genera contenido, a menudo lo hace combinando o reinterpretando trabajos ya existentes. Esto plantea preguntas sobre si las creaciones de la IA son verdaderamente originales y si deberían recibir la misma protección que las obras creadas por humanos.

La protección de la propiedad intelectual a menudo se basa en la originalidad y la creatividad. La generación de creaciones a través de inteligencia artificial plantea la cuestión de si su creación es lo suficientemente original como para ser protegida por derechos de autor. Este problema es éticamente relevante porque afecta la forma en que se valoran y protegen las innovaciones en el mercado.

De hecho, muchas obras están directamente protegidas por estos derechos de autor o *copyright*, por lo que la generación de contenido en base a esos materiales podría considerarse una infracción de derechos de autor, al igual que se han cometido esas infracciones con creaciones de obras por humanos. Esto lo que consigue es lucrarse a costa del trabajo de otros artistas.

Estos problemas son los que suelen tener mayor repercusión en redes sociales, puesto que muchas veces son los usuarios afectados los que denuncian este tipo de prácticas. Precisamente por este motivo, las herramientas de generación de imágenes son las más despreciadas en redes sociales, ya que muchas se nutren de dibujos o creaciones de otros artistas humanos sin consentimiento.

La gota que colmó el bote de pintura de los artistas ocurrió en noviembre de 2022, cuando Deviantart, uno de los portales web con una inmensa colección de arte creado por usuarios de Internet, anunció su propia herramienta de creación de imágenes entrenada con todos los dibujos subidos hasta la plataforma desde su creación en el año 2000, ofreciendo la posibilidad a los artistas de quitar el permiso del uso de sus obras para la herramienta individualmente. El revuelo fue tan grande que en menos de 24 horas la compañía reculó e hizo que ninguna obra estuviese incluida en el uso de la herramienta de IA por defecto, sino que fuese el artista el que tuviese que activar la opción (Whiddington, 2022).

Otro de los sucesos más sonados ocurrió con la red social TikTok a principios de 2024. Esta red social de gran éxito entre los jóvenes está basada en vídeos cortos para una rápida visualización. Aunque cada vez hay más tipos de vídeos en esta red, uno de los tipos de vídeos más subidos y consumidos son los bailes y *challenges* de canciones. De hecho, en los últimos años muchas canciones del mercado actual deben su éxito y fama a su viralización en esta red social y sus *trends*. De hecho, puedes añadir canciones de fondo a tus vídeos aunque no sean de bailes.

Uno de los golpes duros para TikTok llegó en enero de 2024, cuando la discográfica Universal Music Group (UMG) retiró todas las canciones de esta red social que habían sido publicadas bajo su firma (López, 2024). Teniendo en cuenta que la cuota de mercado de UMG en el primer cuarto de 2024 fue de 33,9% (Rys, 2024), siendo líder frente a Sony Music Entertainment (27,33%) y Warner Music Group (18,97%), el catálogo de canciones afectado era muy notorio. Bajo esta discográfica se encuentran artistas internacionales de la tala de Taylor Swift, Drake o Ariana Grande, aunque también afectó a artistas españoles como Rosalía, Aitana o Alejandro Sanz. Además, este bloqueo de canciones no solo limitaba a las nuevas subidas, sino que, además, los vídeos ya subidos a la plataforma que utilizaran esa canción, incluyendo el sonido de fondo, fueron silenciados, lo que afectó frontalmente al catálogo de vídeos de TikTok.

La razón de esta retirada, además de reclamar una mayor compensación económica por el uso de sus canciones, fue que la red social china no quería adoptar mayores medidas de protección contra el uso de la inteligencia artificial pese a las exigencias de Universal. Según indica la compañía discográfica, con la inteligencia artificial se hace uso no autorizado y sin compensación de obras protegidas, como deepfakes, y se suben a esta red social sin ningún tipo de sanción. De hecho, incluso los propios artistas, como Bad Bunny, se han mostrado molestos en alguna ocasión con el éxito de canciones con su voz generada por inteligencia artificial (Zarco, 2023).

Sin embargo, al final tanto UMG como TikTok llegaron a un acuerdo y toda la música bajo su firma volvió a estar disponible en mayo de 2024, así como la reactivación del sonido en aquellos videos que fuesen silenciados por este conflicto. La razón sería que

finalmente se trabajaría en proteger más a los artistas de la inteligencia artificial generativa, así como controlar el contenido generado por la misma para que se desarrolle de manera responsable mientras se protege la creatividad humana (TikTok Newsroom, 2024).

Por otra parte, también han sido aplaudidos otros usos de la inteligencia artificial en obras audiovisuales por ser responsables. Las películas de animación de Spiderman aclamadas por la crítica, “Un nuevo universo” y “Cruzando el multiverso” utilizaron inteligencia artificial para facilitar el trabajo de los animadores. Sin embargo, el productor Chris Miller afirma que no se trata de inteligencia artificial generativa entrenada con obras sin consentimiento (Fuster, 2024), si no que utilizaron técnicas de machine learning desarrolladas por el equipo de efectos especiales para acelerar el dibujado y trazado en algunas animaciones de la película.

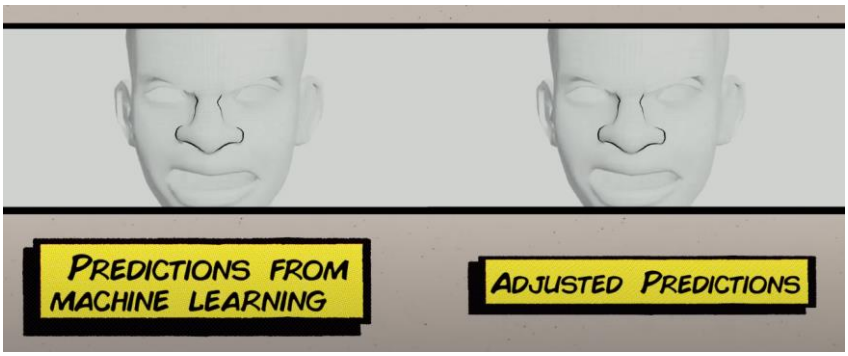

Figura 25: Demostración uso de machine learning en ‘Un nuevo universo' (2018) (Fuente: Sony Pictures)

Además de las consideraciones anteriores, se presenta el dilema de la competencia desleal. Si las creaciones de inteligencia artificial se vuelven indistinguibles de las obras humanas y comienzan a dominar ciertos mercados, los artistas, escritores, diseñadores y otros creadores humanos podrían enfrentarse a esta competencia desleal, puesto que para un usuario actualmente es mucho más fácil, rápido y económico recurrir a una inteligencia artificial antes que a un humano. Esto que podría afectar sus medios de vida y desincentiva la creatividad humana, que a largo plazo podría estancar también el sistema de inteligencia artificial.

3. Marco regulatorio

El empleo de la inteligencia artificial en el ámbito profesional está sujeto a un creciente número de normativas y directrices que buscan garantizar un uso ético, seguro y conforme a la ley de esta tecnología emergente. Tanto a nivel europeo como internacional, la regulación de la inteligencia artificial se encuentra en constante evolución, con organismos globales y regionales desarrollando marcos legales que establecen los principios y obligaciones que deben guiar la implementación de

soluciones basadas en esta tecnología. En este apartado, se analizarán las principales normativas aplicables, incluyendo las recomendaciones de la UNESCO, la recientemente publicada ley 2024/1689 de la Unión Europea y otras directrices que configuran el entorno regulatorio actual.

3.1 UNESCO: "Recomendación sobre la Ética de la Inteligencia Artificial"

Tras varios meses de redacciones de diferentes versiones, en noviembre de 2021 la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) publicó la versión definitiva de un documento en el que incluía recomendaciones éticas sobre la aplicación de la inteligencia artificial. En él, se destaca que las cuestiones éticas de la IA abarcan todas las etapas de su ciclo de vida, desde la investigación y desarrollo hasta su despliegue, uso y eventual desmantelamiento. Además, se indica que los actores implicados en el desarrollo de sistemas de inteligencia artificial pueden ser personas tanto físicas como jurídicas, incluyendo investigadores, ingenieros, empresas y entidades públicas (UNESCO, 2021).

También se reclama una acción internacional y cooperativa, sugiriendo que el desarrollo y uso de la IA debe estar alineado con los valores universales y los principios éticos. Asimismo, se menciona la importancia de abordar las diferencias entre los países desarrollados y en desarrollo en el acceso a la tecnología y la alfabetización digital.

En el documento, se proponen una serie de valores y principios los cuales habrá que considerar y no vulnerar en ningún momento a lo largo del ciclo de vida de los sistemas de inteligencia artificial.

- Dignidad y derechos:
  Se resalta la importancia de que la IA no sea utilizada de manera que viole derechos como la privacidad, la igualdad ante la ley, la libertad de expresión y sin discriminación. Del mismo modo, se debe evitar que la inteligencia artificial beneficie más a un sector o, por el contrario, lo discrimine, evitando las injusticias por sesgos sociales, económicos o políticos.
- Sostenibilidad:
  Promover la prosperidad del medio ambiente y los ecosistemas en los sistemas de IA. Se subraya la necesidad de que todos los actores involucrados en la IA respeten las leyes internacionales y nacionales para la protección ambiental y el desarrollo sostenible. Además, se insta a reducir el impacto ambiental de la IA, incluyendo la huella de carbono, para minimizar el cambio climático y prevenir la explotación insostenible de los recursos naturales.
- Seguridad, privacidad y protección de datos
  Debe ser respetada y protegida la privacidad en todo en todo momento en los sistemas de IA. Es imprescindible que la gestión de datos para la IA, desde su recolección hasta su eliminación, se realice en conformidad con el derecho internacional y los marcos legales pertinentes. Deben establecerse marcos de protección de datos y mecanismos de gobernanza que sigan principios internacionales y que aseguren el consentimiento informado y la protección

adecuada de los datos personales. Además, los sistemas de IA deben someterse a evaluaciones de impacto en la privacidad, considerando también aspectos sociales y éticos, para garantizar la protección de la información personal.

- Transparencia y responsabilidad
- Se debe informar a las personas cuando las decisiones que afectan su seguridad o derechos son influenciadas por IA, permitiéndoles solicitar explicaciones y revisar esas decisiones. De esta manera, se podrá comprender cómo y por qué se toman ciertas decisiones para poder hacer los algoritmos más justos. De igual modo habrá que equilibrar la transparencia con la privacidad. A raíz de estos ámbitos, habrá que tener mecanismos de supervisión y evaluación, pudiendo recaer responsabilidades en los actores involucrados.

El documento también aborda el impacto de la IA en diversas áreas específicas, como la educación, la cultura, el medio ambiente, la economía, el trabajo y la salud.

- En educación se destaca la necesidad de preparar a las personas para la era de la IA, no solo en términos de habilidades técnicas, sino también en ética y pensamiento crítico. Además, se explica la necesidad de evitar la brecha digital incentivando a aumentar la accesibilidad de la IA.
- En cultura se enfatiza la importancia de proteger y promover la diversidad cultural, evitando que la IA contribuya a la homogeneización cultural. Además, se debe apoyar la preservación del patrimonio cultural y la creatividad, evitando la homogeneización cultural.
- En el ámbito ambiental, se reconoce el papel crucial de la IA en la protección del medio ambiente, aunque también se reconocen los riesgos asociados con el aumento del consumo energético al utilizar estas tecnologías, cuyos impactos deben ser evaluados y mitigados.
- En la economía y el trabajo, se aborda la transformación del mundo laboral debido a la IA, destacando la necesidad de políticas que aseguren una transición justa para los trabajadores afectados. Además, se declara la necesidad de regular el uso de la IA en el lugar de trabajo para proteger los derechos de los trabajadores, garantizar condiciones laborales justas y prevenir la discriminación.
- En el sector de la salud, la IA tiene el potencial de mejorar la atención médica, pero se requiere garantizar que sea segura, efectiva y respetuosa de los derechos de los pacientes. Los sistemas deben poder comprenderse por parte del profesional y el paciente y se debe asegurar que estas tecnologías no contribuyan a la desigualdad en el acceso a la atención médica.

Por otra parte, en el documento se dan recomendaciones específicas para los Estados Miembros de la UNESCO. Se insta a estos Estados a desarrollar marcos normativos y políticas nacionales que reflejen los principios éticos delineados, asegurando que todas las partes interesadas sean incluidas en el proceso. Además, se recomienda invertir en educación y formación en temas relacionados con la IA, promover la investigación ética en este campo y fortalecer la cooperación internacional para gestionar colectivamente los riesgos y desafíos que plantea la IA. También se establecen las responsabilidades de los Estados Miembros en la implementación de la Recomendación y se invita a informar sobre las medidas adoptadas y los desafíos enfrentados. La UNESCO destaca

la importancia del intercambio de experiencias y la cooperación internacional para garantizar que los beneficios de la IA se distribuyan equitativamente y que los riesgos sean abordados de manera colectiva.

3.2 Norma ISO/IEC 27001

La norma ISO/IEC 27001, que se actualizó recientemente en 2022 con modificaciones adicionales en 2024, es un estándar internacional reconocido para la gestión de la seguridad de la información, proporcionando un marco sistemático para proteger datos sensibles y garantizar la integridad, confidencialidad y disponibilidad de la información, pudiendo certificar las empresas su cumplimiento. Aunque esta norma no se refiere específicamente a la inteligencia artificia, es especialmente relevante, ya que los sistemas de IA manejan grandes volúmenes de datos, normalmente sensibles, y son susceptibles a diversos riesgos de seguridad.

El objetivo principal de la norma es proteger la confidencialidad, la integridad y la disponibilidad de la información en una organización mediante un enfoque sistemático de gestión de riesgos. Para lograrlo, la norma establece un marco que permite a las organizaciones identificar, evaluar y gestionar riesgos relacionados con la seguridad de la información, lo cual ayuda a prevenir incidentes como ciberataques, accesos no autorizados, pérdida de datos y otros tipos de amenazas.

El hecho de que se pueda certificar el cumplimiento de esta norma ayudar a las organizaciones a demostrar el cumplimiento de otras regulaciones, como el RGPD de la Unión Europea, evitando sanciones y fortaleciendo su posición frente a las autoridades reguladoras. Además, facilita la armonización con otras normativas y estándares de gestión de seguridad, traduciéndose en una ventaja competitiva significativa, atrayendo nuevos negocios y manteniendo relaciones comerciales sólidas y de largo plazo.

Además, no solo se trata de añadir controles de seguridad, sino también de optimizar procesos internos relacionados con la gestión de la información. La norma fomenta la eficiencia mediante la identificación de áreas de mejora, la reducción de redundancias y la implementación de mejores prácticas en la gestión de la seguridad. Como resultado, las organizaciones pueden aumentar su productividad general y reducir los costes asociados a incidentes de seguridad, como la pérdida de datos o la interrupción de servicios. La calidad de los datos de entrada de un sistema de IA y la integridad del proceso de entrenamiento del modelo son esenciales para evitar resultados erróneos o sesgados.

La versión 2022 de la norma incluye actualizaciones que reflejan las nuevas amenazas y desafíos del entorno digital, como el aumento del trabajo remoto, la adopción de tecnologías en la nube y la creciente sofisticación de los ciberataques. Al adaptar sus controles a estos cambios, las directrices de la norma indican a las organizaciones a ser más resilientes y preparadas para enfrentar riesgos emergentes, asegurando que sus sistemas de seguridad estén alineados con las mejores prácticas globales actuales. Con la mejora continua se busca responder a cambios en el entorno de las amenazas o en

los requisitos del negocio. La adopción de esta mentalidad de mejora continua ayuda a las organizaciones a mantenerse a la vanguardia en seguridad.

3.3 Reglamento (UE) 2024/1689

Por su parte, la Unión Europea ha estado trabajando en un Reglamento sobre la Inteligencia Artificial desde 2021, culminando su publicación en el Diario Oficial de la Unión Europea en julio de 2024, y en consecuencia en el Boletín Oficial del Estado (Agencia Estatal, 2024), y su entrada en vigor en agosto del mismo año, aunque su aplicación se extenderá paulatinamente hasta agosto de 2027 (UE 2024/1689, art 111). Esta legislación establece un marco normativo que equilibre la innovación tecnológica con la protección de los derechos fundamentales y la seguridad pública.

El propósito de esta ley es regular el uso de sistemas de inteligencia artificial en diversos contextos, desde aplicaciones comerciales hasta su impacto en derechos humanos y medio ambiente. La normativa introduce requisitos específicos para garantizar que los sistemas de IA sean desarrollados y utilizados de manera que respeten los principios de transparencia, responsabilidad y sostenibilidad, a la vez que acompaña las normas ya existentes.

Con la creación de esta ley se pretende evitar que se creen diferentes normativas por territorios que ralenticen la evolución o aplicación de la inteligencia artificial, además de conducir hacia una competencia justa y un mercado único. Este reglamento se aplica a los sistemas de inteligencia artificial que sean utilizados, comercializados o desarrollados en la Unión Europea. Esto evita la entrada de sistemas desarrollados bajo unos estándares más permisivos que podrían vulnerar aquellos derechos que se pretenden proteger.

En el documento se clarifican prácticas que quedan terminantemente prohibidas en relación con la inteligencia artificial (UE 2024/1689, art. 5). Entre ellas, la manipulación subliminal, en la que se quiera influir a las personas en comportamientos o decisiones sin que sean conscientes de ello, llevando a actuar a estas personas de una manera que no harían normalmente, pudiendo incluso causar daño. Esta práctica atenta frontalmente contra el individuo, porque la manipulación a la que sucumbe puede ser difícil de detectar o también difícil de resistir. Es por ello por lo que está prohibida por comprometer la autonomía y el libre albedrío de los individuos, violando su derecho a tomar decisiones informadas y voluntarias.

También se señala a la explotación de vulnerabilidades, que tiene gran relación con la anterior práctica, especialmente a aquellas personas o grupos que, por su edad, discapacidad, situación económica o psicológica, puedan distorsionar su comportamiento, pudiéndoles perjudicar. De esta manera, se impedirá el abuso de poder que se ejercería injustamente sobre estas personas al ser más vulnerables y, en consecuencia, más fáciles de manipular.

Paralelamente, se prohíbe que los sistemas de IA puntúen a individuos en base a características sensibles para clasificarlos, como raza, opiniones políticas, religión, circunstancias económicas o similares para evitar posibles casos de discriminación o marginación, que violaría el principio de igualdad y los derechos humanos básicos. Esto se realiza especialmente para evitar que se realicen evaluaciones de riesgos de personas físicas de cometer un delito basándose solo en características personales, exceptuando aquellos casos en los que la persona involucrada sí esté objetivamente relacionada con un pasado delictivo. Tampoco podrá hacerse uso de datos biométricos para detectar el estado emocional de las personas en contextos laborales y educativos, ni podrá establecerse una puntuación social en base a los datos o comportamientos recopilados, así como vigilancia masiva en espacios públicos, o la utilización de imágenes de vigilancia para el entrenamiento de sistemas de reconocimiento facial, sin una base legal que lo permita.

Para estas prácticas se aplican excepciones, principalmente cuando se quieran utilizar sistemas de inteligencia artificial en ámbitos policiales o legales, como casos de secuestro, trata de personas, sospechas de atentado terrorista o cualquier amenaza inminente que ponga en peligro la vida o la seguridad física de las personas físicas.

El reglamento además clasifica los sistemas de IA en función de su riesgo para la seguridad, la salud y los derechos fundamentales (UE 2024/1689, art 6-49). Esto asegura que las regulaciones sean proporcionales, evitando una regulación excesiva de sistemas de bajo riesgo que podría decelerar la innovación. Los sistemas de alto riesgo incluyen aquellos que afectan de manera significativa las decisiones individuales, como la IA utilizada en procesos judiciales o en el acceso a servicios esenciales.

Los sistemas de IA de alto riesgo deberán estar sujetos a estrictos requisitos de seguridad, transparencia y gobernanza de datos. De la misma manera, deberán implementar un sistema de gestión de riesgos, siendo identificados y mitigados. Obligatoriamente serán evaluados antes de su comercialización y su funcionamiento será continuamente monitoreado.

Las características principales que etiquetan de alto riesgo a un sistema de inteligencia artificial son:

- Formar parte de un componente de seguridad implicado en actos legislativos de la Unión Europea.
- Necesitar evaluación de conformidad de terceros para poder ser puesta en servicio.
- Elaboración de perfiles de personas físicas.
- Gestión y funcionamiento de infraestructuras digitales críticas del tráfico, suministro de agua, gas, calefacción y electricidad.
- Evaluación de educación de asignaturas e influencia en el nivel de educación o formación al que podrán acceder.
- Relación con la empleabilidad: contratación, gestión de trabajadores, selección de personal…

- Concesión de préstamos y servicios de carácter público y privado y la calificación de solvencia de las personas.

Existirán excepciones que se podrán aplicar en los casos de uso detallados en el reglamento cuando el sistema de IA contemple alguna de estas acciones:

- Realizar una tarea delimitada como estructuración de datos no estructurados
- Mejorar una actividad ya realizada por un humano.
- Detección de patrones de toma de decisiones con respecto a otros patrones de decisiones anteriores.
- Realización de una evaluación a un sistema de IA de alto riesgo.

Además, un sistema de IA se clasificará como riesgo sistémico sin tiene capacidades de gran impacto tras su evaluación con metodologías técnicas adecuadas o si la Comisión Europea determina que su impacto es equivalente (UE 2024/1689, art. 51-56). Este gran impacto vendrá dado por la capacidad de cálculo de operaciones superior a 1025 en coma flotante. Aquellos desarrolladores de un sistema de riesgo sistémico deberán notificar a la Comisión, aunque esta clasificación podrá revisarse si se demuestra sólidamente que pese a cumplir los requisitos, no se presenta realmente ningún riesgo sistémico.

Al respecto de la transparencia, se hace hincapié en aquellos sistemas que interactúen con las personas, categoricen y generen o manipulen contenido. De esta manera, el contenido audiovisual que se haya creado mediante inteligencia artificial y que pueda generar confusión sobre su veracidad o parezcan reales, deberán indicar que han sido creados a partir de esta tecnología (UE 2024/1689, art. 50).

Por otra parte, se ponen sobre la mesa medidas que puedan favorecer la innovación y no paralicen los avances que se realicen (UE 2024/1689, art. 57-62). Debe existir un equilibrio entre la regulación y la innovación y para eso se podrá realizar pruebas de los sistemas de inteligencia artificial dentro de un entorno especial en el que las regulaciones sean más flexibles, pero enfrentándose a un contexto igualmente realista. De esta manera, los desarrolladores podrán perfilar los sistemas sin miedo previo de incumplir alguna restricción.

Del mismo modo, se da la posibilidad de lanzar pruebas piloto, en el que toda su aplicación en el mundo real es controlada. Las pruebas piloto estarán dirigidas especialmente a los objetivos de la Unión Europea, como la salud o la energía. Estas pruebas, además de acelerar los avances respecto a pruebas en entornos cerrados, podrán aportar información a la industria para mejorar los sistemas y la legislación.

También existirá por parte de la Unión Europea un apoyo para los proyectos de investigación y desarrollo de inteligencia artificial, incluyendo formación, acceso a redes de innovación y colaboración entre sectores públicos y privados. Estos programas están diseñados para ayudar a las empresas, especialmente a las pymes, a superar las barreras iniciales al desarrollo de nuevas tecnologías.

Para poder articular la cooperación de autoridades, tras la creación de la Oficina Europea de Inteligencia Artificial en mayo de 2024 (Comisión Europea, 2024) se crean autoridades competentes en cada estado miembro de la UE, formando el Consejo de IA. De esta manera, mientras que las autoridades se encargarán de monitorizar los sistemas, aplicar el reglamento y mostrar colaboración con la UE, la Oficina se encargará de coordinar las supervisiones nacionales junto con las regulaciones que se proponen en el documento europeo. Además, también será el encargado de orientar a todas las autoridades competentes al respecto de conflictos que puedan surgir, así como procurar que estas apliquen correctamente el reglamento europeo y proponer mejoras si lo ven necesario.

Para los sistemas que no sean considerados de alto riesgo, se anima a que cada uno de los desarrolladores de su sistema siga un código de conducta de manera que se pongan los requisitos que se imponen a los sistemas de alto riesgo como referencia, pero en este caso voluntarios.

Sin embargo, cuando sea un sistema de inteligencia artificial de alto riesgo el que comience a comercializarse, se activará un protocolo de seguimiento activo al mismo a través de varias fuentes (UE 2024/1689, art. 72), además de comprobar que se cumplen los requisitos de riesgo. Esto incluye la realización de inspecciones, pruebas de conformidad y el monitoreo de los sistemas de IA una vez que estén en el mercado. De igual manera, se notificarán los incidentes que ocurran con estos sistemas a las autoridades de vigilancia (UE 2024/1689, art. 73).

En todo momento, las actuaciones tendrán carácter confidencial, incluyendo el tratamiento de datos personales, secretos comerciales y otro tipo de información que pueda considerarse sensible. Deberá protegerse la información de cualquier acceso no autorizado, así como su divulgación, aunque se reserva el derecho legítimo como excepción.

Por otra parte, el reglamento establece que las sanciones por incumplimiento del propio reglamento deben ser “efectivas, proporcionales y disuasorias”. Esto puede incluir multas significativas, que se calcularán en función de la gravedad de la infracción y del tamaño de la empresa, pudiendo llegar hasta los 35 millones de euros en las infracciones graves (UE 2024/1689, art. 99). Las autoridades pueden exigir a las empresas que tomen medidas correctivas, como la modificación de un sistema de IA o la retirada del mercado de un sistema.

4. Guía de Buenas Prácticas de la inteligencia artificial

Tras haber analizado los diversos documentos que pretenden encaminar el buen hacer respecto a la inteligencia artificial y haber visto cómo las tecnologías que la componen

han ido cogiendo fuerza y relevancia en la sociedad, siendo cada vez más influyentes, es necesario establecer una serie de pautas para que los profesionales TIC que desarrollen sistemas de IA o los utilicen, puedan hacerlo de forma responsable.

Para poder marcar esas pautas a seguir, procedemos a continuación a la elaboración de la Guía de Buenas Prácticas.

4.1 Introducción y objetivos

Aunque la IA ofrece oportunidades significativas para el progreso económico y social, también plantea desafíos en términos de ética, seguridad, privacidad e igualdad. Los sistemas de IA mal diseñados pueden causar daños a los usuarios, tales como responder en base a sesgos o comprometer la seguridad de los datos. Por lo tanto, es esencial guiar a los profesionales TIC para que desarrollen y desplieguen la IA de manera responsable.

La creciente preocupación por los riesgos asociados con la IA ha llevado a la implementación de regulaciones, como las recomendaciones de la UNESCO o el reglamento de la Unión Europea, que buscan establecer un marco legal claro e incluyente. Esta Guía de Buenas Prácticas se alinea con esos marcos regulatorios y proporciona una orientación práctica para que los profesionales TIC ejerzan su responsabilidad y cumplan con estas normas. Al mismo tiempo no deberá verse afectada la innovación en una búsqueda del equilibrio entre la protección de los derechos fundamentales y el desarrollo tecnológico.

Si la inteligencia artificial consigue demostrar que es segura a la vez que cada vez es más llamativa, se construirá un sentimiento de confianza en la IA entre usuarios, mejorando el bienestar de la sociedad. Es trabajo de los profesionales que desarrollan los algoritmos y aplicaciones con estas tecnologías conseguir que sus sistemas no solo resuelvan problemas, sino que no generen otros. Es por eso por lo que es necesario tener claros cuáles son los objetivos principales del proyecto y para qué se quiere crear un sistema que involucre la inteligencia artificial.

4.2 Principios éticos de los sistemas de IA

El desarrollo y uso de la inteligencia artificial deben estar guiados por un conjunto de principios éticos fundamentales que aseguren que estas tecnologías se implementen de manera justa, segura y en consonancia con los valores humanos y los derechos fundamentales. Esta sección de la guía presenta los principios clave que los profesionales TIC deben considerar y aplicar en cada etapa del ciclo de vida de los sistemas de IA.

Todo desarrollo de un sistema de IA debe tener una visión antropológica, es decir, debe tener el foco centrado en el ser humano. Para poder lograrlo debe diseñarse y desarrollarse con el objetivo de mejorar el bienestar humano y respetar la dignidad, los derechos y las libertades fundamentales de las personas. La IA debe ser una

herramienta para potenciar las capacidades humanas, apoyar la toma de decisiones y facilitar la vida cotidiana, sin sustituir el juicio humano en situaciones críticas. Este enfoque centrado en el ser humano implica también que los sistemas de IA deben ser accesibles y comprensibles para los usuarios, asegurando que puedan interactuar con la tecnología de manera segura y efectiva.

A la hora de utilizar la inteligencia artificial para crear un producto puede llegar a imponerse los intereses comerciales frente a los humanos. Con tal de maximizar las ganancias, pueden generarse influencias contrarias para manipular pensamientos y tomar el control de la sociedad en beneficio propio. Por ello, deben diseñarse unos fundamentos que se apliquen durante todas las etapas del desarrollo. Los principales, sin ser limitantes, deberían ser:

- Beneficencia universal: Debe contribuir al bienestar humano y tener un impacto positivo en la sociedad.
- Evitar malas intenciones: No debe buscar ocasionar daños intencionadamente y evitar el daño no intencional que se pueda causar a personas individuales o a la sociedad.
- Consentimiento: Respetar la autonomía de las personas, permitiéndoles tener control sobre sus propias decisiones y sus propios datos.

Para poder cumplir con estos puntos, sería de gran soporte involucrar a los usuarios en el proceso de desarrollo para entender sus necesidades, preocupaciones y contextos. También podrían realizarse estudios de impacto social y medioambiental y pruebas de usabilidad con grupos diversos para garantizar que la IA atienda a las necesidades reales de las personas. De esta manera, los profesionales obtienen un feedback continuo acerca de cómo está avanzando su trabajo y puede prevenir posibles futuras correcciones en fases posteriores que serían más difíciles de solucionar.

El objetivo de la visión antropológica es el más importante de todos, puesto que si un sistema de IA se crea procurando que los intereses de un sistema de inteligencia artificial sea el de poder beneficiar a las personas, significa que hay otros principios éticos por detrás que también se están cumpliendo. Es por ello por lo que es necesario recalcar aquellos principios éticos que deben aplicarse a raíz de este para que pueda cumplirse.

4.2.1 Justicia e igualdad

Para poder poner al centro de los intereses al ser humano, no solo puede considerarse un sector de la población o un grupo determinado de personas. Un error que suele cometerse en el desarrollo de los sistemas de inteligencia artificial es en la proyección del beneficio social desde la visión de un grupo limitado de personas que conforman la sociedad. Lo peor de estos casos es que son errores que suelen cometerse inintencionadamente, por lo que son difíciles de detectar si todos los profesionales involucrados en el desarrollo se ven identificados con la visión del proyecto. Esto puede generar resultados imparciales, discriminando a la parte de la población que no fue contemplada por los desarrolladores.

Los sistemas de IA a menudo se entrenan utilizando grandes conjuntos de datos, que pueden contener sesgos implícitos debido a patrones históricos de desigualdad o a la

falta de representación de ciertos grupos. Si estos sesgos no se abordan adecuadamente, los sistemas de IA pueden perpetuar o incluso amplificar las desigualdades existentes, lo que podría llevar a decisiones injustas o discriminatorias en áreas críticas como el empleo, la educación, la justicia penal, el acceso a créditos y los servicios públicos.

Para poder contener la aparición de estos sesgos, deben aplicarse algunas acciones:

- Documentación de principales sesgos: El historial de discriminaciones e injusticias tanto en la sociedad como en otros sistemas de IA puede servir de ayuda para poner el foco en ellos y así poder impedir que nuestro sistema siga los mismos pasos erróneos.
- Revisión de datos: Tras estar documentado de a qué posibles sesgos puede enfrentarse nuestro sistema, se debe realizar una evaluación del conjunto de datos utilizado para entrenar nuestro modelo de inteligencia artificial. De esta manera, podremos ver si nuestros datos consiguen incluir a toda la población sin discriminación. También se puede comprobar si existe cierta falta de datos para algunos grupos sociales que pueda provocar un mal aprendizaje de nuestra IA, pudiendo balancear su existencia en proporciones ajustadas a la realidad. Esta revisión ayuda a detectar el problema de manera temprana, ahorrando correcciones más difíciles en fases posteriores.
- Evaluación y mitigación: Tras poner en marcha nuestro sistema, debemos ir evaluando su comportamiento. De esta manera podemos fijarnos en si el sistema actúa de una manera justa o si ha desarrollado o amplificado algún sesgo. Para poder contrarrestar esto, se puede hacer uso de técnicas de debiasing, como una modificación de los pesos de instancias para que sea la ponderación o el ajuste del umbral el que pueda corregir alguna desigualdad identificada (Lemmens, s.f.).

También es muy interesante que exista diversidad en los grupos de desarrollo de los sistemas de IA, puesto que facilitará la detección de sesgos de una manera más directa y realista. De esta manera, el objetivo del beneficio para la población amplía su campo de visión y más se contemplan más consideraciones durante el diseño y la implementación que ayudan a enriquecer el funcionamiento del sistema final y reducir las injusticias en la toma de decisiones.

4.2.2 Transparencia y explicabilidad

En un contexto donde las decisiones automatizadas pueden tener un impacto significativo en la vida de las personas, garantizar que las decisiones de la IA sean transparentes y explicables es crucial para fomentar la confianza, la responsabilidad y la equidad en el uso de estas tecnologías. Este principio busca asegurar que los sistemas de IA sean comprensibles para sus usuarios, para las partes interesadas y para los reguladores, permitiendo una supervisión efectiva y un razonamiento de sus acciones.

Teniendo en cuenta que la inteligencia artificial tiene muchos algoritmos por detrás del resultado que proporciona, es necesario saber cuál es el razonamiento de ese resultado.

Muchos de los resultados obtenidos por una inteligencia artificial pueden formar parte de decisiones críticas, por lo que ofrecer transparencia es crucial para que se pueda comprobar que la decisión tomada por el sistema es coherente y no ha sido afectada por alguno de sus problemas más frecuentes, como podría ser la invención mediante alucinaciones. Por otra parte, la transparencia es vital para que los reguladores y las autoridades puedan supervisar adecuadamente los sistemas de IA y asegurar que cumplen con las normativas aplicables.

Para poder dotar de mayor transparencia nuestro sistema de inteligencia artificial, será necesario cumplir con los siguientes puntos:

- Documentación: Crear y mantener actualizada una documentación detallada del desarrollo del modelo junto con su diseño ayuda a garantizar la transparencia y entender las decisiones tomadas y además facilita las auditorías. En la documentación se puede exponer los datos y algoritmos que se utilizan junto con su metodología de entrenamiento, cuáles son sus limitaciones y otras métricas de evaluación. Esta documentación, además de facilitar la trazabilidad, proporciona un recurso educativo para los usuarios, otros desarrolladores y otras partes interesadas.
- Modelos explicables: Diseñar modelos que puedan explicar sus decisiones de manera comprensible. Se pueden combinar modelos complejos con otros más sencillos a la par que interpretables para tener consciencia de cómo se está tratando la información. Sin embargo, la búsqueda de mayor explicabilidad puede implicar comprometer la precisión del modelo. Elegir el equilibrio adecuado depende del contexto y del riesgo asociado con la aplicación del modelo.
- Advertir de su uso: Además de poder seguir la trazabilidad del sistema, también es necesario advertir directamente al usuario que está interactuando con un sistema de IA

También es importante dotar a nuestro sistema de IA de una buena transparencia para que sea posible cumplir con diferentes funciones con relación al tratamiento de datos de acuerdo con el Reglamento General de Protección de Datos (en adelante, RGPD). De hecho, desde la Agencia Española de Protección de datos se recuerda al usuario que el término de transparencia se utiliza con diferentes significados en el RGPD y en el Reglamento de la UE, en el que actúan actores diferentes, pero se complementan entre ellas (AEPD, 2023). Este tema se tratará en a continuación en el principio ético de la Protección de Datos, privacidad y seguridad.

4.2.3 Protección de Datos, privacidad y seguridad

Una de las grandes preocupaciones, no solo en la inteligencia artificial, sino en el ámbito tecnológico general, es el tratamiento que se le da a nuestros datos. Los profesionales TIC tienen la responsabilidad de garantizar que los sistemas de IA cumplan con las normativas de protección de datos, como el RGPD de la Unión Europea y de implementar prácticas que respeten la privacidad de los usuarios. Si estos datos no se gestionan de manera adecuada, pueden ser mal utilizados, comprometidos en brechas de seguridad o utilizados para fines distintos a los previstos, afectando negativamente a los individuos y consiguiendo el objetivo contrario al principal.

Al utilizar grandes cantidades de datos en el entrenamiento de los sistemas de IA y en otras fases, Sin embargo, el problema del tratamiento de datos, al no ser exclusivo de la inteligencia artificial, está más regulado y controlado, por lo que existen directrices más claras de cómo actuar ante ellos, aunque también se pueden utilizar recomendaciones adicionales.

- Anonimizar datos: Los datos que se vayan a tratar deben ser anonimizados, eliminando todo rastro de datos personales, de manera que no haya manera de identificar a un usuario. También pueden sustituirse los datos personales por pseudónimos o directamente recopilar solo los datos necesarios para el propósito específico de nuestro sistema de IA.
- Consentimiento: En el caso de que sea necesario tratar con datos personales, se deberá informar al usuario del tratamiento que se les dará a sus datos. Para esto, deberán dar su consentimiento explícito, es decir, no vale solo con informar, sino que el usuario debe aceptar las condiciones. De igual manera, se puede revocar el consentimiento en cualquier momento y solicitar la eliminación de sus datos.
- Seguridad: Para proteger los datos que se traten deberán implantarse medidas de seguridad como la encriptación de datos y controles de acceso para proteger los datos tanto de acceso no autorizado como de modificaciones, filtraciones o eliminaciones. Es importante que todo el sistema esté protegido, y no únicamente los propios datos, para no obtener respuestas alteradas. Los desarrolladores deberán tener formación en seguridad y actualizar constantemente sus sistemas de protección para defenderse de nuevas formas de ataque. Para poder protegerse de ataques, es aconsejable realizar ataques controlados a nuestro propio sistema para saber cómo respondería y poder solucionar brechas de seguridad, como exploits que puedan aprovechar los hackers en su beneficio, que se descubran en el proceso. En caso de sufrir una brecha de seguridad, se deberá informar a los usuarios y a las autoridades regulatorias.

También es importante tener en cuenta la normativa vigente al respecto de la protección de datos de cada territorio. A nivel europeo, a través del RGPD de la UE se establecen estándares rigurosos, incluyendo el procesamiento legal, finalidad de la recopilación, limitación del almacenamiento de los datos y la confidencialidad.

En el RGPD también se trata la transparencia, que como hemos mencionado anteriormente tiene un significado y proyección diferente al utilizado en el Reglamento de IA de la UE. En este último, la transparencia se enfoca directamente al entendimiento del funcionamiento de los sistemas de IA, mientras que la del RGPD es al entendimiento de cómo se tratan los datos de los usuarios. Hay que tener en cuenta que los responsables de cada transparencia son diferentes, puesto que en la del Reglamento de IA afecta a cualquier usuario que forme parte del despliegue del sistema de IA, como desarrolladores, diseñadores o proveedores, mientras que la del RGPD son los responsables del tratamiento. En los sistemas de IA que manejen datos personales, ambas transparencias pueden entrar en acción, haciendo que los actores del sistema de IA puedan convertirse también en responsables del tratamiento, ofreciendo transparencia de funcionamiento y de tratamiento de datos de acuerdo con ambos reglamentos.

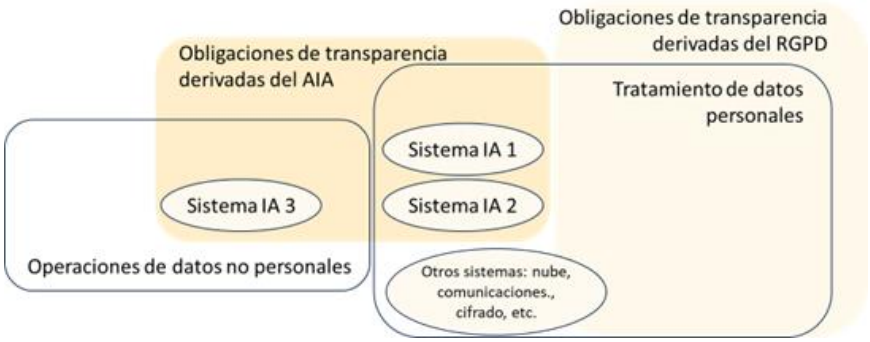

Figura 26: Combinaciones transparencias Reglamento IA y RGPD (Fuente: AEPD)

Por otra parte, se aconseja también evaluar cuál puede ser el impacto en la privacidad si se quieren tratar datos sensibles en un sistema de IA. Este proceso debe ir actualizándose a la par que se actualice el sistema o las normativas que les afecten.

4.2.4 Sostenibilidad

Tener una visión de futuro de lo que puede acarrear el sistema de IA desarrollado, pensando en cómo afectará al medio ambiente y a las futuras generaciones es importante a la hora de desarrollarlo. No es únicamente el objetivo y si cumple con los derechos humanos, sino también controlar el impacto medioambiental que pueda tener nuestro sistema, considerando el consumo de energía y otros recursos para entrenar modelos complejos. En un momento en que la preocupación por los efectos ambientales y sociales de las tecnologías está en aumento, los profesionales TIC deberán trabajar para que estos impactos sean mínimos.

- Eficiencia energética: El uso de la tecnología consume energía, pero en el caso de la IA, especialmente en la fase de entrenamiento debido al uso de grandes cantidades de datos, es muy significativo. Para poder procesar tales cantidades de datos se realiza en servidores, que, a su vez al desprender calor, necesitan ser refrigerados, consumiendo más energía (Peris, 2024). Es importante que se estudie la eficiencia de nuestro sistema de IA para que el trabajo que realice sea menor e implique un menor procesamiento, consumiendo menos energía. También se podría optar por fuentes de energía renovables para menguar el impacto. Por otra parte, en los centros de datos y servidores podrían utilizarse la propia IA para tener un mantenimiento predictivo o gestionar los sistemas de climatización según la necesidad.
- Evaluaciones de impactos: Comprobar con el tiempo cuál está siendo el impacto energético de nuestro sistema de IA, puesto que nuestro sistema seguramente se vaya actualizando y mejorando, lo que podría derivar un mayor consumo de energía. También se debería hacer una gestión del ciclo de vida de los componentes hardware y optar por los dispositivos modulares para sustituciones más sencillas sin necesidad de reemplazo total de la máquina.

Además de que los efectos que se puedan producir en el medio ambiente afectarían a generaciones futuras, los propios sistemas de IA pueden revolucionar los modelos de negocio, afectando a puestos de trabajo y reorganizaciones en empresas. De hecho, en el estudio “Future of Jobs Report 2023” se afirma que entre 2023 y 2028 se producirán reorganizaciones en el 23% de las empresas (World Economic Forum, 2023). Además,

un 44% de las competencias actuales de los trabajadores se verán afectadas en consecuencia (Gómez Cardosa, 2023).

Para poder advertir las consecuencias medioambientales y sociales de nuestro sistema de IA habría que informar de manera transparente acerca de los impactos potenciales y reales del sistema a las partes interesadas. Esto incluye informes sobre el consumo de recursos, las emisiones de carbono y los efectos sociales, tanto positivos como negativos.

4.2.5 Supervisiones y responsabilidades

Estos principios aseguran que los sistemas de IA operen de manera responsable, con ética y en línea con los derechos humanos y las normativas aplicables. El hecho de que exista una responsabilidad clara sobre las decisiones y acciones de los sistemas de IA. Implica la aparición de la figura de rendición de cuentas. Sin embargo, para poder evitar que haya que rendir cuentas y cargar responsabilidades sobre malas acciones cometidas, es necesario que también exista la figura de supervisión con el objetivo de monitorear, intervenir y corregir el comportamiento del sistema de IA cuando sea necesario.

Los sistemas de IA, especialmente aquellos que puedan tomar decisiones críticas o afectar significativamente a las personas, no deben operar de manera autónoma sin la posibilidad de supervisión humana. Esto es determinante porque pese a que los sistemas de IA pueden procesar grandes cantidades de datos y realizar tareas complejas, es más complicado que acierten en la comprensión contextual, la empatía y el juicio moral que los seres humanos pueden aportar. Para ello, se tomarán diferentes medidas:

- Revisión: Se deberán implementar procedimientos en los que las decisiones críticas tomadas por los sistemas de IA sean revisadas y aprobadas por un humano antes de ser ejecutadas. Esto deberá aplicarse en consonancia con la transparencia, con la que será posible revisar cómo funcionan internamente los diferentes algoritmos.
- Reglamento: También habrá que supervisar que no se incumplan ningún punto de las normativas vigentes tanto nacionales como europeas. En los sistemas de alto riesgo deberán supervisarse los sistemas para que cuando ocurra algún crítico se informe a las autoridades de vigilancia.

Para que haya una correcta supervisión y, si llega el momento, rendición de cuentas, habrá que asignar responsabilidades dentro del proyecto de nuestro sistema de IA:

- Roles: La asignación de roles para establecer quién es responsable de cada aspecto del desarrollo y operación del sistema de IA es muy importante. Esto ayudará a tener un control mayor sobre qué aspecto controla más cada persona, impidiendo que varias personas se superpongan en un mismo rol y queden aspectos descubiertos. Los roles deberían asignarse a todos los implicados en el proyecto, desde los desarrolladores y los ingenieros hasta los gerentes y los responsables de cumplimiento. Esto incluye la responsabilidad sobre los datos utilizados, las decisiones de diseño del modelo y la gestión del sistema en producción.

- Gobernanza: Debería existir un control de acceso para asegurar que solo el personal autorizado tenga acceso para modificar o intervenir en los sistemas de IA. Esto protege la integridad del sistema y asegura que cualquier intervención humana esté debidamente registrada y justificada asumiendo responsabilidades. También se deberían revisar las decisiones y políticas relacionadas con la IA, asegurando que se alineen con los valores de la empresa y las expectativas éticas y sociales.

Habiendo asignado las responsabilidades, será más fácil dar respuesta a los incidentes que ocurran durante el desarrollo o la posterior comercialización de un sistema de IA. Desarrollar y mantener planes para gestionar errores, fallos o impactos negativos de los sistemas de IA. Esto incluye protocolos claros para la evaluación de los impactos, la implementación de medidas correctivas y la notificación de incidentes. Como se ha comentado anteriormente, en el caso de los sistemas de alto riesgo, intervendrán las autoridades de vigilancia de mercado de las regiones pertinentes cuando el sistema ya se haya lanzado y se deberá preparar el terreno para esta, tal y como se indica en el Reglamento de la IA: “La supervisión ex post debe garantizar que, una vez que un sistema de IA esté en el mercado, las autoridades públicas tengan las competencias y los recursos necesarios para intervenir en caso de que este genere riesgos inesperados, lo que justificaría una rápida actuación.”.

4.3 Cumplimiento normativo

Si bien con la inteligencia artificial nuestro principal objetivo es otorgar un servicio que sea beneficioso para la sociedad, no puede conseguirse mediante el libre albedrío: es necesario seguir las normas vigentes, que pueden variar en cada territorio. Aunque puedan considerarse al principio como obstáculos en el camino, las diferentes normativas a aplicar son realmente las señales que nos marcan el camino correcto, puesto que estas normativas velan principalmente por el cumplimiento del objetivo principal de beneficiar a la sociedad, pero mantiene a raya posibles acciones que, precisamente sin normas y directrices, podrían obstaculizar alcanzar ese objetivo.

Este apartado se centra en las obligaciones legales y cuáles pueden ser las mejores prácticas para asegurar que los sistemas de IA no solo sean técnicamente eficientes y éticos, sino también conformes con las leyes aplicables, como el RGPD, el Reglamento de IA de la Unión Europea y otras normativas aplicables relevantes. Además, estará alineado con el apartado 4.2.5, ya que para la aplicación de la normativa se deben realizar acciones para tener una mejor supervisión de nuestro sistema y sus influencias.

4.3.1 Identificación de riesgos

La identificación y clasificación de riesgos es un proceso fundamental en el cumplimiento normativo de los sistemas de IA. Este proceso implica identificar, evaluar y priorizar los posibles riesgos asociados con el desarrollo y uso de sistemas de IA, para luego implementar estrategias adecuadas de mitigación. Este procedimiento debe realizarse en las primeras etapas del diseño y desarrollo de nuestro sistema para poder adelantarnos a posibles incidentes y saber reaccionar rápidamente ante ellos, en vez de

pensar cómo actuar cuando ocurra. Esto, además, permitirá ahorrar costes y evitar daños mayores.

Para poder identificar los riesgos potenciales se puede hacer uso de una matriz de riesgos, clasificándolos según su probabilidad de ocurrencia y el impacto que supondría. Aunque la ponderación de cada riesgo es gradual, se clasifican en cuatro grupos de riesgos: los críticos, los significativos, los moderados y los menores. De esta manera, se puede priorizar la actuación ante riesgos más críticos que requieran atención inmediata.

Para poder realizar la matriz de riesgos, primeramente desde las fases iniciales del proyecto de nuestro sistema se deben evaluar los riesgos potenciales a través de los requisitos del sistema, los datos involucrados y el contexto en el que se utilizará nuestro sistema. Posteriormente, a cada riesgo se le deberá asignar una puntuación tanto en términos de probabilidad como de impacto. Para ello, habrá que comprobar qué medidas se podrían aplicar para evitar que ocurran los riesgos, afectando a la puntuación de probabilidad o valorando en apostar por conductas más seguras para minimizar el impacto. Asignadas ya las puntuaciones, se colocarían en la matriz en función de estas y aquellas más cercanas al máximo tanto de impacto como de probabilidad, deberían ser los riesgos prioritarios. Sin embargo, respecto a los riesgos que quedan más dispersos por la matriz, podrían llegar a asignarse prioridades subjetivas.

Sin embargo, aunque la elaboración de la matriz debe hacerse al principio del desarrollo del sistema de IA, no debe considerarse una referencia permanente, puesto que nuestro sistema irá evolucionando. Es por eso por lo que a medida que el sistema evoluciona o se despliega en nuevos contextos, la matriz de riesgos debe revisarse y actualizarse para reflejar cualquier cambio en la probabilidad o impacto de los riesgos. Esto ayudará a focalizar la monitorización y supervisión de los riesgos más importantes.

Hay que tener en cuenta que, si nuestro sistema de IA cumple con las condiciones establecidas en el Reglamento de IA de la Unión Europea de sistema de alto riesgo por su impacto significativo en los derechos fundamentales, la salud, la seguridad o el bienestar de las personas, se deberá cumplir con requisitos adicionales de evaluación de conformidad, transparencia y supervisión humana. Esto incluye la realización de evaluaciones de impacto específicas y la implementación de controles adicionales para mitigar los riesgos identificados.

Como se ha comentado previamente, para poder analizar los riesgos, debemos tener claro cuál va a ser el enfoque de nuestro sistema de IA, ya que en los contextos en los que vaya a funcionar y los datos que vaya a analizar depende de a qué campos van a influir los riesgos. De esta manera, las ponderaciones de algunos riesgos pueden ser diferentes para diferentes contextos, puesto que, para ámbitos más críticos como la salud, los riesgos pueden tener consecuencias más graves que si son para otros ámbitos como podría ser el entretenimiento.

4.3.2 Datos personales

Al respecto del tratamiento de los datos personales, la Unión Europea tiene una normativa bastante rígida a través del RGPD. Para el cumplimiento de este reglamento, puede utilizarse la Evaluación de Impacto en la Protección de Datos (EIPD). Este

proceso evalúa los riesgos que un sistema de IA o sus procesos pueden representar para la privacidad o el consentimiento de las personas cuyos datos serán procesados y de esta manera conseguir mitigar los posibles riesgos relacionados (AEPD, 2021).

Aunque esta evaluación solo es necesaria cuando el impacto de los datos tratados constituye un riesgo alto para los derechos y libertades de los individuos, como procesamientos a gran escala, uso de datos sensibles o la monitorización en tiempo real. Sin embargo, para el caso de tratamientos automatizados como la inteligencia artificial se deben seguir igualmente las indicaciones para todos los sistemas independientemente de su calificación de riesgo:

- Descripción: Es necesario detallar el propósito del procesamiento de los datos, qué tipo de datos se recopilan, de qué manera son utilizados durante el proceso y qué tecnologías son utilizadas
- Identificación y evaluación de riesgos: al igual que hemos hecho en el apartado anterior, se deberán identificar y evaluar los riesgos potenciales
- Mitigación: Se deberán tomar medidas preventivas para disminuir el nivel de riesgo que afecte a los derechos y libertades, así como medidas reactivas que reduzcan el impacto.
- Brechas de seguridad: Se deberán contemplar las posibles brechas de seguridad, puesto que es uno de los riesgos potenciales que siempre deben considerarse.

Figura 27: Medidas de seguridad en la gestión del riesgo para los derechos y libertades (Fuente: AEPD)

Además, si un usuario desea revocar el acceso a sus datos, se debe cumplir con las normativas de protección de datos, tal y como se explica en el RGPD, gestionar la solicitud de revocación de datos. En el caso de que haya sido utilizada para entrenar los modelos, habría que volver a entrenar los modelos sin la muestra eliminada.

4.3.3 Propiedad intelectual

Los sistemas de IA, particularmente los modelos de aprendizaje automático, a menudo requieren grandes volúmenes de datos para entrenarse. Estos datos pueden incluir textos, imágenes, música, videos y otros tipos de contenidos que están protegidos por derechos de autor. El uso no autorizado de estos materiales puede constituir una infracción de los derechos de autor.

Para poder evitar esto, una de las soluciones más simples sería eliminar los datos que contienen derecho protegido por derechos de autor y utilizar en su caso contenido que sea de dominio público. Sin embargo, si es necesario tratar con esos datos, se puede optar por la obtención de licencias necesarias que permitan utilizar el contenido para nuestro fin.

También existe la posibilidad de hacer uso del "fair use" del contenido, aunque actualmente, la legislación al respecto no permite específicamente el tratamiento del contenido de este modo (Gonzalo, 2023). Sin embargo, hay reflexiones al respecto: "Si se copia para poder analizarla, pero no para hacer un uso económico directo de la copia, estamos en un margen complicado de apreciar ya que, en principio, se asemeja al modelo de aprendizaje humano en el que se tienen datos de la obra, pero no está como tal" (Maeztu, 2023).

Para poder tener clara la posición de nuestro sistema de IA ante la propiedad intelectual y los derechos de autor, se deberán establecer políticas claras sobre el uso de contenido de terceros, verificación de derechos y restricciones de uso acorde a las normativas locales, puesto que dependiendo del territorio en el que se desarrolle y se comercialice, se pueden aplicar unas leyes u otras.

4.3.4 Prácticas prohibidas

En el cumplimiento normativo en el desarrollo y uso de sistemas de inteligencia artificial no siempre será posible adherirse a las reglas para poder sacar adelante nuestro sistema. SI este presenta grandes amenazas sobre los principios éticos y derechos fundamentales puede llegar a ser explícitamente prohibidas según las regulaciones vigentes en cada territorio. Estas prohibiciones buscan proteger los derechos fundamentales de las personas y prevenir los daños potenciales que podrían surgir del uso indebido o irresponsable de la IA. Es de vital importancia atender a la lista de prácticas prohibidas y otras restricciones específicas del Reglamento de la Unión Europea sobre IA para asegurar que los sistemas de IA operen de manera segura, justa y en beneficio de la sociedad y, en conceptos de intereses de nuestro proyecto, sea legal.

Siguiendo las indicaciones que se han dado en los anteriores subapartados, como definir políticas internas acorde a la legislación vigente, supervisar el desarrollo y cumplir con los principios éticos, será más sencillo evitar estas prácticas específicamente prohibidas. Habrá que descartar los desarrollos con ese objetivo, pero no reducirse a ello, si no también estar atento a las posibles actualizaciones de las normativas, adición de otras prácticas específicas y adaptar las prácticas internas de conformidad en consecuencia. Para ello, sería conveniente que también existiera un puesto responsable que se asigne a un experto legal y se informe de avances en la legislación del territorio y de los boletines regulatorios.

En cualquier caso, siempre se debería priorizar el desarrollo modular de los sistemas de IA que permitirían tener procesos para la rápida adaptación de los sistemas y políticas internas a los cambios en las normativas, minimizando los riesgos de incumplimiento. De igual manera, estas actualizaciones deberían ser informadas continuamente a los desarrolladores, gerentes de proyecto y otros profesionales TIC.

4.4 Diseño

El diseño responsable y seguro de los sistemas de inteligencia artificial es un principio clave que asegura que estas tecnologías se desarrollen y operen de manera que prioricen la seguridad, la ética, la privacidad y los derechos de las personas. Este principio se enfoca en integrar la seguridad y la responsabilidad en todas las etapas del ciclo de vida de un sistema de IA, desde la conceptualización hasta la implementación y la gestión continua. Un diseño responsable y seguro no solo protege a los usuarios y a la sociedad de posibles daños, sino que también fomenta la confianza y la aceptación de la IA en diversas aplicaciones.

La seguridad debe ser una consideración integral desde las primeras etapas del desarrollo del sistema de IA y no un añadido posterior. Al incorporar medidas de seguridad desde el diseño del proyecto, se puede minimizar la exposición a riesgos potenciales y asegurar que el sistema funcione de manera segura y conforme a los estándares legales y éticos. Esto incluye la identificación, modelado y mitigación de vulnerabilidades y amenazas y el cifrado de datos.

También es conveniente que nuestro sistema de IA tenga un diseño resiliente, de modo que nuestro sistema de IA sea capaz de continuar funcionando de manera efectiva ante fallos técnicos, ataques cibernéticos o condiciones operativas adversas. Un diseño resiliente minimiza las interrupciones y asegura que el sistema pueda recuperarse rápidamente en caso de incidentes. Para ello, habría que implementar mecanismos que permitan al sistema recuperarse, como los sistemas de respaldo o la creación de protocolos de recuperación automática, permitiendo volver a un estado seguro en caso de que su desarrollo o utilización se tuerza. También podrían configurarse sistemas de respuesta que puedan tomar medidas correctivas inmediatas ante la detección de anomalías, como el aislamiento de componentes comprometidos o la activación de protocolos de contingencia.

Nuestro sistema, precisamente por la rápida evolución de la IA, tendrá cambios durante el tiempo, incluso en el proceso de desarrollo previos a su comercialización. Es por eso que para llevar un control del diseño actual, y poder volver a versiones anteriores si fuese necesario, es necesario un sistema de gestión de cambios y versiones para documentar todas las modificaciones realizadas en el sistema de IA a lo largo del tiempo. Para ello, la documentación del proyecto deberá ser detallada y completa, incluyendo decisiones clave, evaluaciones de riesgos, cambios implementados y justificaciones para las elecciones de diseño. Esto no solo facilita la conformidad con normativas, sino que también sirve como base para auditorías y revisiones futuras, además de asegurar la trazabilidad y permitir rastrear cómo y por qué se hicieron ciertos ajustes, lo cual es esencial para la mejora continua.

4.5 Formulario de autoevaluación

Con el fin de asegurar el cumplimiento de las buenas prácticas y directrices establecidas en la guía, se ha desarrollado un cuestionario de autoevaluación destinado a los desarrolladores de inteligencia artificial. Este cuestionario proporciona un método sencillo y estructurado para verificar el cumplimiento de los principios mencionados en los apartados anteriores en el desarrollo de sistemas de IA. El objetivo principal es facilitar la identificación de áreas de conformidad y destacar aquellas que requieren mejoras. Dado que la IA puede tener un impacto significativo en la sociedad, es fundamental que los desarrolladores dispongan de mecanismos para revisar continuamente sus prácticas y asegurar su alineación con los estándares establecidos.

El cuestionario está diseñado en un formato de tabla que organiza las 63 afirmaciones en varias secciones clave correspondientes a los diferentes principios abordados en la guía. Cada afirmación contiene dos casillas de respuesta, una que corresponde a "Sí cumple" y la otra a "No cumple", lo que permite a los desarrolladores identificar rápidamente las áreas que cumplen con las directrices y aquellas que necesitan especial atención. Para observaciones adicionales, como cumplimientos parciales, se podrá utilizar la casilla de observaciones disponible al final de cada sección.

Debido a la extensión del formulario, se encuentra disponible al completo al final del documento en el anexo 2.

4.6 Ejemplo: Sistema de IA para selección de personal

Para poner en práctica la guía que hemos creado y comprobar su efectividad, vamos a ejemplificarla con la creación de un sistema que utilizará inteligencia artificial para una empresa que pretende automatizar parte del proceso de selección de personal para ayudar al departamento de recursos humanos. El sistema utilizará algoritmos de aprendizaje automático para analizar currículums, cartas de presentación y realizar evaluaciones iniciales de los candidatos, ayudando a los reclutadores a identificar a los mejores postulantes para los diferentes puestos. Durante el proceso, surgirán problemas en los que habrá que aplicar las medidas de la guía de buenas prácticas para abordarlos.

Por órdenes superiores, el sistema de IA tenía como objetivo automatizar al máximo el proceso, eliminando en la medida de lo posible la actuación de personal de recursos humanos. Sin embargo, los desarrolladores del sistema indicaron que si se eliminaba la revisión humana, el riesgo de obtener candidatos que no fuesen los mejores para el puesto por errores o discriminación en otros candidatos era demasiado alto, por lo que se optará por no eliminar esa revisión de la selección. Además, se establecerán controles para asegurar que los reclutadores puedan cuestionar y ajustar las recomendaciones del sistema.

Al haber leído la guía, como buenos profesionales, decidieron revisar los datos con los que el sistema sería entrenado y se descubrió que la cantidad de muestras de personas asiáticas era bastante baja en comparación con otras, lo cual podría llevar a un sesgo

en la evaluación de candidatos de origen o rasgos asiáticos. Para darle solución, optaron por el sobremuestreo de esta minoría y posteriormente balancearon los datos. Respecto a los propios datos con los que ha sido entrenado el sistema, han sido anonimizados y se han eliminado aquellos datos que no son trascendentes, de manera que las muestras de por sí no son identificables.

Aunque con el tema de los datos el riesgo era bajo al haber sido anonimizado, se pusieron sobre la mesa los diferentes riesgos que podría ocasionar tanto el desarrollo como la comercialización, antes y durante. Para poder organizarse, realizaron una evaluación exhaustiva de estos riesgos en un diagrama, incluyendo escenarios de pruebas, y se establecieron controles adicionales durante todas las fases del desarrollo.

En las primeras pruebas iniciales del sistema, se comparó la elección de candidatos para ser jefe del departamento de Preventa entre varios reclutadores y el sistema de IA. Aunque no todos los reclutadores eligieron al mismo candidato como jefe, gracias a que el sistema se desarrolló de manera transparente se pudo ver en qué cualidades se basó en su elección, señalando algunas que los propios reclutadores no habían considerado.

Durante el desarrollo del sistema de selección se observó un gran incremento de uso de energía en la planta en la que se estaba desarrollando, entrenando y haciendo pruebas con la aplicación. Para poder cumplir con las promesas de consumo de energía de la empresa, se instó a reducir el consumo energético, pese a ya hacerlo con energía renovable en un gran porcentaje. Para ello, los desarrolladores optimizaron los algoritmos para reducir el uso de recursos. Además, también se redujo la frecuencia de entrenamiento del sistema.

Además, se involucró a personal de ciberseguridad que no había participado en el proyecto para que buscase cualquier brecha de seguridad que pudiese tener el sistema que pudiese corromperlo o dar como óptimos a candidatos de manera fraudulenta. De esta manera, se comprobó que se podía enviar currículums manipulados con datos especialmente diseñados para engañar al algoritmo, logrando que un candidato específico reciba una calificación más alta descartando a otros más competentes. De esta manera, se pudieron implementar validaciones más estrictas en los datos de entrada que comprobaban correctamente que, además, al detectar el fraude, descartaba directamente al candidato. Sin embargo, los candidatos se mostrarían al usuario final para hacer una revisión por un humano que verifique el fraude.

El hecho de que el sistema no estuviese preparado para detectar currículums fraudulentos agitó a los superiores y exigió responsabilidades, pero lo cierto es que no estaban estrictamente clasificados los roles. En aquel momento se establecieron los roles claros dentro del equipo para la supervisión de las decisiones del sistema, incluyendo un responsable que garantice la revisión y validación de las decisiones automatizadas y que esté capacitado para responder ante cualquier error o incidencia.

5. Encuesta sobre el uso de la IA y percepción de la ética

Para poder reafirmar la importancia de un buen uso de la ética en la inteligencia artificial, se ha realizado una encuesta en la que se realizan cuestiones sobre el uso de la inteligencia artificial actual, el impacto en el día a día y ámbitos en los que se utiliza. Después de eso, se han introducido otras preguntas al respecto de los problemas que está teniendo el uso de sistemas con estas tecnologías al respecto de la ética y la privacidad, así como el punto de vista y opinión acerca de esta. Por otra parte, se ha puesto a prueba la capacidad de la sociedad de distinguir imágenes generadas por inteligencia artificial de imágenes reales.

La encuesta se ha realizado digitalmente mediante la herramienta Google Forms y se ha publicitado en diferentes redes sociales, como YouTube, X (anteriormente Twitter), Instagram y WhatsApp, para poder cubrir el mayor espectro social. En total, se han conseguido 249 participaciones. Aunque la encuesta no trata con datos personales sensibles, se ha requerido la aceptación de los participantes para el tratamiento de sus respuestas y han sido informados de ello. Para el análisis de las respuestas se han realizado combinaciones de respuestas con la ayuda de macros en Microsoft Excel.

5.1 Finalidad de la encuesta

La presente encuesta tiene como finalidad recopilar y analizar datos que permitan entender mejor el panorama actual de la inteligencia artificial en la sociedad. Con el rápido avance de la IA, es fundamental comprender cómo esta tecnología impacta a las personas en su vida diaria, cuáles son sus percepciones y preocupaciones y cómo podemos orientar su desarrollo hacia un uso ético y beneficioso para todos.

Un primer objetivo es evaluar el nivel de conocimiento que tiene la sociedad sobre la inteligencia artificial. Esto incluye identificar qué tanto saben las personas sobre las aplicaciones de la IA, si están familiarizadas con los conceptos básicos, y qué grado de interés tienen en aprender más sobre el tema. Este objetivo permitirá detectar brechas de conocimiento y áreas donde es necesario fomentar la educación y divulgación sobre la IA. También se busca investigar los diferentes casos de uso de la IA que son relevantes para la sociedad. Esto abarca desde aplicaciones cotidianas, como los asistentes virtuales y el traductor de texto, hasta usos más enfocados en ámbitos concretos, como el académico o el profesional. Se buscará entender cómo la IA está cambiando la forma en que vivimos, trabajamos y nos relacionamos, tanto en términos positivos como negativos. Esto permitirá no solo medir el nivel de adopción de la IA en distintas áreas, sino también evaluar las implicaciones sociales y económicas de estos cambios.

Por otra parte, la ética en la inteligencia artificial es un tema de creciente importancia. Por ello, la encuesta busca conocer cómo percibe la sociedad los aspectos éticos

relacionados con la IA, tales como la privacidad, la transparencia, la igualdad y la responsabilidad en el uso de estos sistemas. Este objetivo ayudará a identificar las principales preocupaciones éticas y a guiar el desarrollo de políticas y prácticas que aborden estos desafíos. Además, se desafiará a los participantes a detectar si un contenido multimedia ha sido creado por un humano o por inteligencia artificial.

5.2 Descripción de participantes

Para poder analizar y comprender posteriormente las respuestas a la encuesta, se han realizado en primero lugar preguntas de clasificación por género, rango de edad, nivel de estudios y hobbies, además de cuestionar el grado de conocimiento sobre IA que consideran que tienen. Los resultados de la encuesta estarán protagonizados por las mujeres al representar al 53,4% de los encuestados, frente al 44,5% de la participación de los hombres y el 2,1% que se identifica con otro género o no ha querido especificarlo.

Respecto a la edad de los encuestados, se ha optado por dividir a los encuestados en 6 rangos de edad. En este caso los tres grupos más jóvenes (Entre 0 y 34 años) han dominado la encuesta liderada por el rango 18-24 años, pese a ser el grupo que menos años comprende, con un 37,3% de respuestas. Le sigue de cerca el grupo de los menores de 18 años (30,1%) y un poco más lejos cierra el pódium en rango de entre 25 y 34 años con un 15,7%. Los participantes de más de 34 años representan al 16,9%, dividiéndose bastante igualados en entre 35 y 44 años (6,4%), entre 45 y 54 años (6%) y mayor de 54 años (4,4%).

A nivel de estudios, el 36,2% posee ya estudios universitarios y el 16,3% tiene ya título de Formación Profesional. Por otra parte, aquellos con Educación secundaria obligatoria representan al 26,4%, mientras que aquellos con Bachiller forman un 21,1%.

Por último, para poder relacionar usos cotidianos de inteligencia artificial, se preguntó por cuáles eran los hobbies favoritos, de los cuáles se podía realizar una selección múltiple. De los participantes de la encuesta, el 82,7% sienten devoción por los videojuegos, siendo la afición más elegida. En cuanto a entretenimiento, el 68,7% se decantan por obras audiovisuales como películas o series, mientras que los libros o novelas son elegidos por el 46,6%. Otras opciones de selección de hobbies más movidos, como el deporte (28,5%) o la cocina (30,5%), mientras que los idiomas identifican al 24,1%. El último hobby, quizá el más relacionado con la inteligencia artificial, es la tecnología, de la cual un 37,8% están apasionados.

Al respecto del conocimiento autoevaluado de inteligencia artificial, el 28,9% de los encuestados admite tener un nivel bajo de conocimiento, con el que pueden llegar a utilizar las herramientas, pero sin saber en absoluto cómo funcionan. Por otra parte, el grueso de los participantes opina tener un grado medio de conocimiento, con el que conocen ligeramente cómo funciona, representando un 54,2% de las respuestas. Por último, con un 16,9% de representación están aquellos que dicen tener un conocimiento alto, saben bastante al respecto y además siguen las actualizaciones asiduamente.

5.3 Análisis de resultados

Para comenzar a analizar los resultados de la encuesta, vamos a comenzar con el uso que da la sociedad a la inteligencia artificial. Primeramente, se preguntó cuáles de las herramientas que utilizan inteligencia artificial entre una selección habían utilizado en los últimos meses. Entre las opciones más elegidas se encuentran los generadores de texto, como ChatGPT o Copilot, con un 71,9%, y un poco más alejadas, pero casi empatadas, fueron elegidos los asistentes virtuales (58,2%), la traducción automática (55,4%) y el procesamiento de imágenes (51,4%). A partir de esto, podemos observar que el poder de las herramientas de IA generativa ha llegado a muchos hogares, mientras que esta tecnología sigue adentrándose en la vida cotidiana mediante asistentes como Alexa o Siri. Por otra parte, sigue con fuerza el uso de traductores, una herramienta más veterana que el resto pero que sigue estando presente.

También es interesante combinar estos datos junto con el conocimiento de inteligencia artificial de los participantes de la encuesta, tal y como podemos observar en la siguiente figura con datos relativos a cada grado de conocimiento:

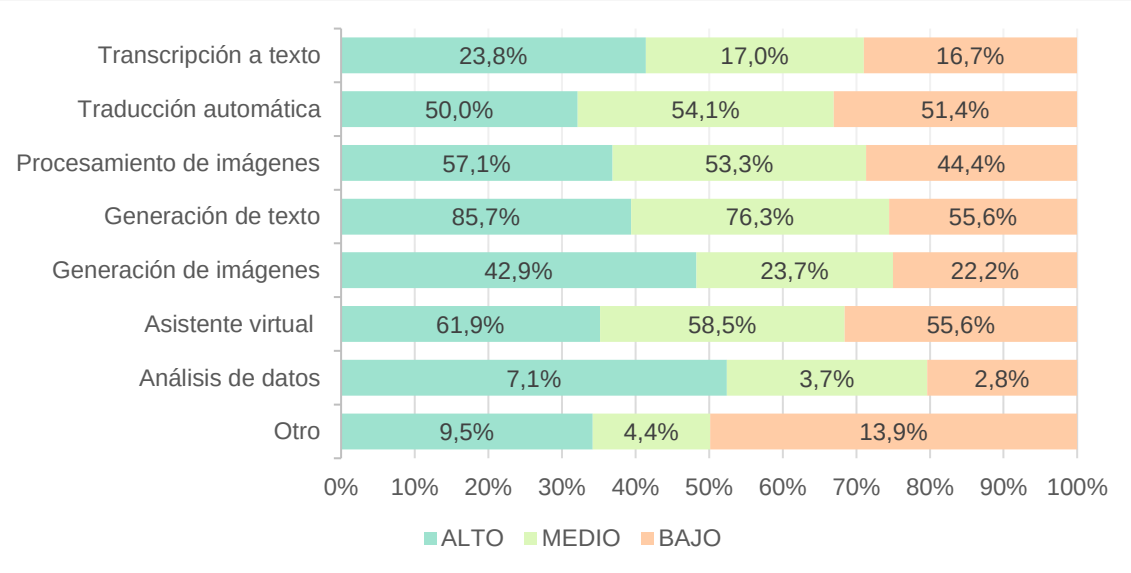

Figura 28: Gráfica de uso de herramientas por conocimiento de IA (Fuente: elaboración propia)

De este modo, podemos observar cómo los generadores de texto siguen siendo los más utilizados, especialmente por los que tienen un conocimiento alto ya que el 85,7% de ellos los utilizaron en los últimos 12 meses. Por otra parte, podemos ver que los generadores de imágenes son más utilizados por aquellos con un conocimiento de IA alto que aquellos con conocimientos inferiores. Precisamente esta herramienta es de las que ha recibido más críticas mediáticamente por sus entrenamientos de datos controversiales. Este grado de conocimiento también destaca en el uso para análisis de datos, probablemente debido a que su complejidad de uso es algo superior. Por otra parte, podemos observar también que cuanto más conocimiento se tiene sobre inteligencia artificial, más herramientas se utilizan: de media, con conocimiento alto han seleccionado el 42,3% de las opciones de la pregunta, con medio un 36,4% y con bajo un 32,8%.

El uso de herramientas de inteligencia artificial siempre tiene algún fin: académico, profesional o incluso por simple diversión. Sin embargo, pese a los diferentes ámbitos en los que se pueda utilizar la inteligencia artificial, no parece ser influenciado directamente por el nivel de estudios de la persona, manteniendo unas proporciones similares y una clara dominancia por los estudios universitarios.

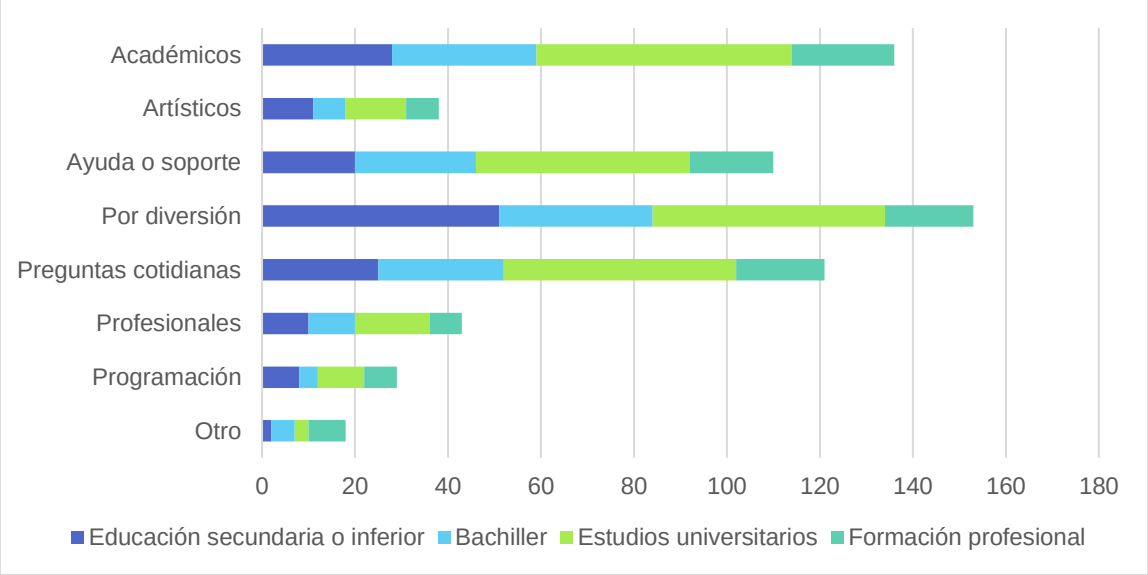

Figura 29: Uso por ámbitos según nivel de estudios (Fuente: Elaboración propia)

Un indicador que sí varía según los ámbitos en los que se utilizan las herramientas de IA es la frecuencia de uso. Se instó a los encuestados a puntuar su uso semanal de estas herramientas del 1 al 5 y la media se sitúa en un 2,55. Podemos ver el desglose por ámbito de uso en la siguiente gráfica:

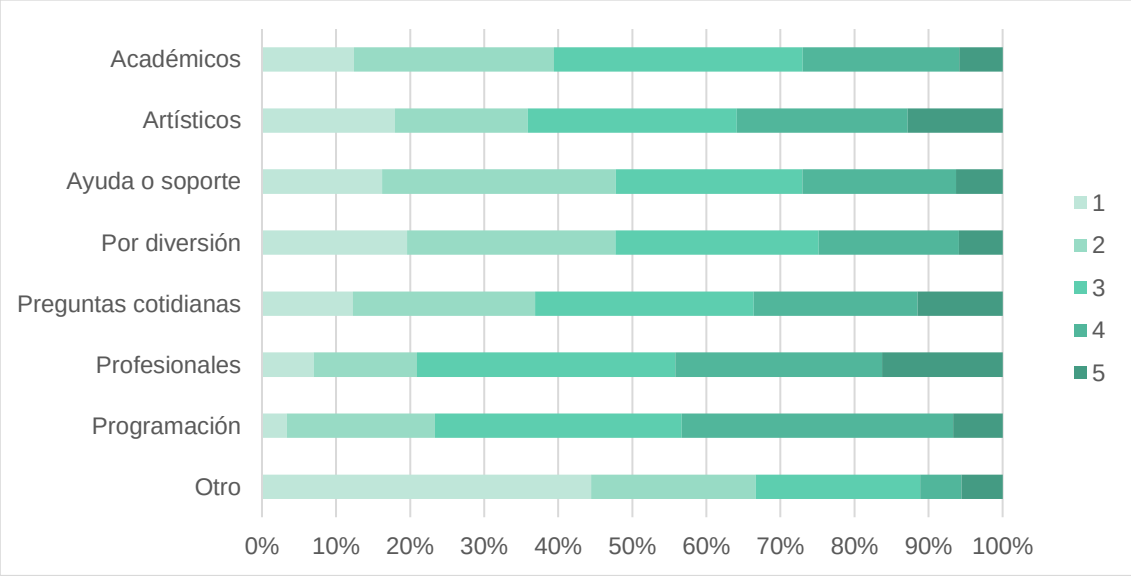

Figura 30: Frecuencia de uso de inteligencia artificial por ámbito (Fuente: Elaboración propia)

Desde la gráfica podemos observar que el ámbito profesional es el que más frecuenta el uso de herramientas de inteligencia artificial aumentando su media hasta el 3,33, pese a haber sido seleccionado por solo el 17,3% de los encuestados. En contraparte, el

ámbito más utilizado es el lúdico (“Por diversión”) con un 61,4%, pero muestra una frecuencia de uso menor que la profesional. Esto puede deberse a que en el ámbito profesional se utilizan herramientas específicas, lo que significa que son utilizadas por personas más concretas, pero con más asiduidad, mientras que el uso lúdico es más esporádico, pero comprende un público más genérico. Otros usos que destacan son el académico (55%), el de preguntas cotidianas (49%) y el de ayuda o soporte (44,6%), reafirmando la importancia que está despertando la inteligencia artificial en el día a día de la sociedad. Por otra parte, la explicación de la baja frecuencia de “Otro” es por la inclusión de “ningún ámbito” además de otros ámbitos no comprendidos, seleccionado por el 7,2% de los encuestados.

Ante la clara integración de los sistemas de inteligencia artificial, se ha preguntado también si han sustituido alguna herramienta tradicional por otra herramienta de inteligencia artificial que ayudase a simplificar el proceso. Los resultados están bastante divididos, aunque la mayoría simple se la lleva el “No” con un 40,6%, mientras que un 33,7% admiten sí haberla sustituido. El 25,7% restante pertenece a aquellos que no saben o no han sido conscientes de si se ha realizado esa sustitución. Si comparamos estos datos con las herramientas que han utilizados nuestros participantes, podemos comprobar que independientemente de que hayan sustituido o no a las tradicionales, las herramientas se usan de proporciones similares, pero con mayor frecuencia en aquellos que sí han sustituido, tal y como se muestra en la siguiente figura:

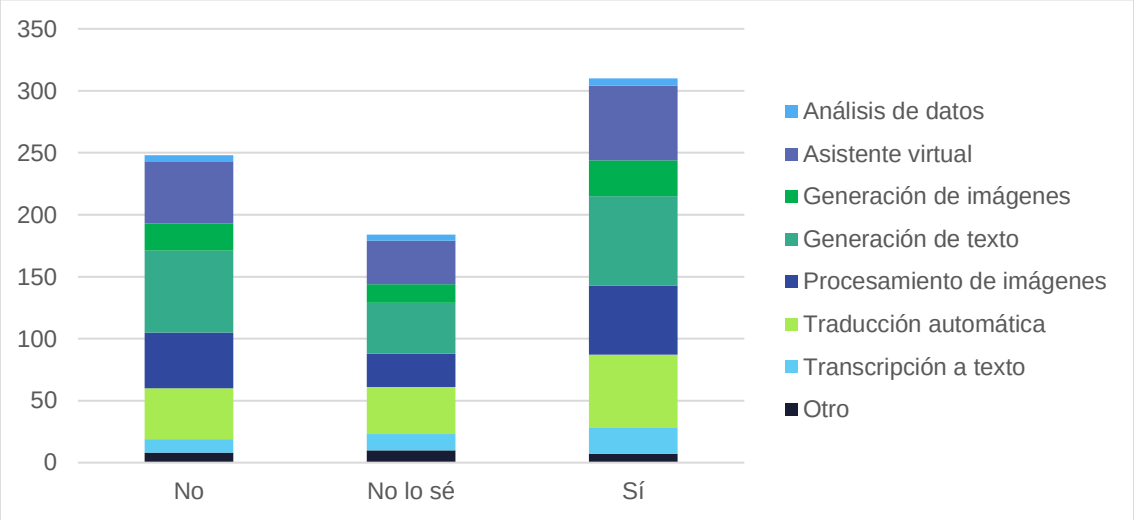

Figura 31: Desglose sustituciones de herramientas tradicionales por IA (Fuente: Elaboración propia)

La incorporación de la inteligencia artificial ha despertado el interés de las personas en ella, y es que el 73,1% de los encuestados estaría interesado en conocer más acerca de la IA, siendo un 45,8% aquellos que simplemente quieren conocimientos más superficiales y un 27,3% los que estarían dispuestos a entrar en detalle y ponerse al día con los avances actuales. Sin embargo, existe cierta resistencia a la tecnología: el 15,3% se conforman con utilizar herramientas de IA sin saber cómo es su funcionamiento, mientras que el 11,6% restante reivindica evitar el uso de inteligencia artificial y, en consecuencia, no tiene interés en conocer más.

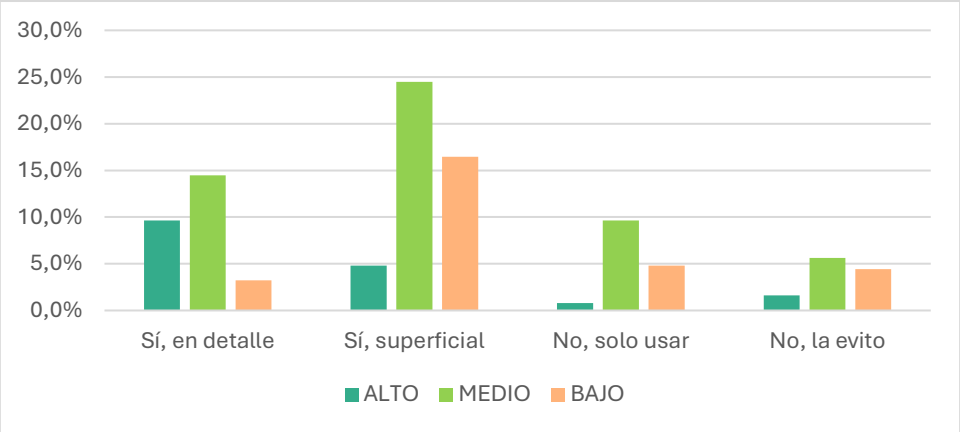

Figura 32: Interés en IA según conocimiento de IA actual (Fuente: Elaboración propia)

SI combinamos los datos sobre el conocimiento actual con los del conocimiento que quieren adquirir en un futuro, se puede observar que aquellos que tienen poco nivel de conocimiento sobre estas tecnologías están muy interesados en conocer más sobre ella, aunque a un nivel más básico, mientras que los que ya tienen un conocimiento alto, no se despreocupan de ella y quieren mantenerse informados sobre los avances actuales.

Tras las cuestiones de conocimientos, se ha preguntado acerca de cuál es su opinión al respecto de la inteligencia artificial. Vamos a empezar por los puntos positivos:

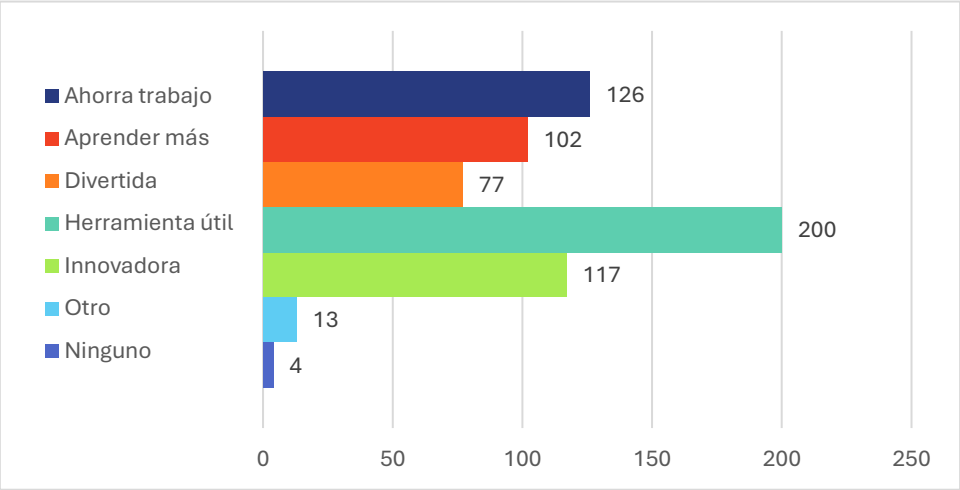

Figura 33: Puntos positivos de la inteligencia artificial (Fuente: elaboración propia)

El 80,3% de los encuestados piensa que la IA es, al menos, una herramienta útil, valorando su indudable aporte a la sociedad. Por otra parte, además de su incuestionable innovación, destacan los puntos más académicos y profesionales: el 50,6% ve positivo poder ahorrarse mucho trabajo gracias a ella y el 41% piensa que les facilita el camino para expandir su conocimiento. Igualmente sigue presente la función lúdica con un 30,9% de votos. Además, se ha ofrecido dar respuestas abiertas a través de la opción “Otros”, en las que destacan sus avances en medicina y también el soporte que pueden hacer a los humanos, remarcando que no puede utilizarse como herramienta sustitutoria de los mismos.

Por otra parte, se ha preguntado acerca de cuáles son los puntos negativos. Aunque parece tener ligeramente menos puntos negativos que puntos positivos, puesto que se han seleccionado 503 opciones frente a las 635 de los positivos, se demuestra

igualmente cierta disconformidad en la sociedad con las tecnologías de inteligencia artificial. En la siguiente gráfica podemos ver cómo se han repartido las respuestas:

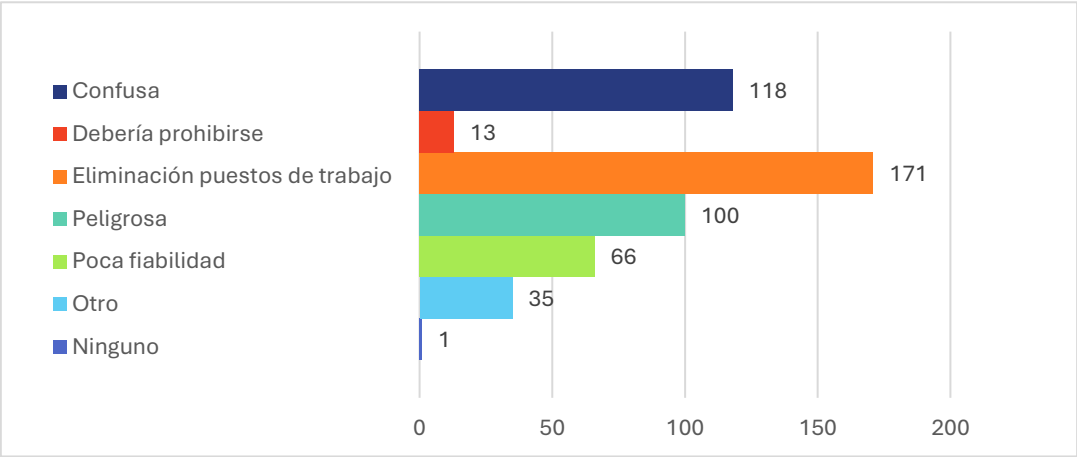

Figura 34: Puntos negativos de la inteligencia artificial (Fuente: Elaboración propia)

El principal punto que inquieta a los encuestados es la posibilidad de que vaya a acabar con muchos puestos de trabajo por su sustitución por sistemas de IA, suponiendo un 68,7% de las respuestas. La siguiente preocupación es la alteración de la realidad, ya que logra confundir a la sociedad según el 47,4%. Por otra parte, aunque el 40,2% piensa que es peligrosa, solo el 5,2% considera que debería ser prohibida, exponiendo que, aunque actualmente pueda haber contextos controversiales en la inteligencia artificial, confían en que se vayan a proponer acciones regulatorias para solucionarlo. Como punto a destacar, la casilla de respuesta abierta de "Otro" ha casi triplicado a la casilla homónima de los puntos positivos. Esto indica que, además de los puntos principales, hay otros problemas de la IA más específicos que les preocupan. Entre ellos, la creación de un sentimiento de dependencia hacia ella, el uso malintencionado para desinformar, el restringido acceso según situación económica, el consumo insostenible de energía o la creación de desinterés o poco esfuerzo en realizar acciones cotidianas.

A continuación, se ha profundizado más acerca de la ética y la privacidad en la inteligencia artificial preguntando si eran conscientes de algún uso de la inteligencia artificial que no fuese ético, de los cuales el 74,3% respondió afirmativamente, y si habían pensado acerca del pensamiento de cómo afecta a la privacidad de las personas la IA, obteniendo en este caso un 71,9%. Además, un 59,8% respondió afirmativamente ambas preguntas, demostrando la afección de ambos campos a las personas. En estas preguntas también se ha otorgado un campo de respuesta abierta opcional para justificar sus elecciones, de los cuales el 33,5% de los participantes ha utilizado para concretar su visión. Las mayores acciones no éticas se componen por el entrenamiento de dibujos artísticos sin consentimiento de los autores, el excesivo uso que se le puede dar en el ámbito académico, creación de imágenes fraudulentas de personas en situaciones falsas y el ahorro de costes con consecuentes despidos humanos. Además, varios participantes apuntan que los problemas éticos que puedan surgir de la IA los ocasiona el mal uso por parte de los humanos. Al respecto de la privacidad, la preocupación por esta ha sido creciente principalmente debido al uso de entrenamientos según la actividad en redes sociales sin preaviso y el uso de datos sin consentimiento. Sin embargo, también hay quien defiende a la IA haciendo un símil con la revolución

industrial en la que muchos puestos de trabajo dejaron de tener cabida y fueron sustituidos por las máquinas.

Para poder focalizar más cuáles eran los principales problemas éticos y de privacidad que conciernen a la inteligencia artificial, se dio una selección de aspectos y se preguntó cuáles les afectaban principalmente:

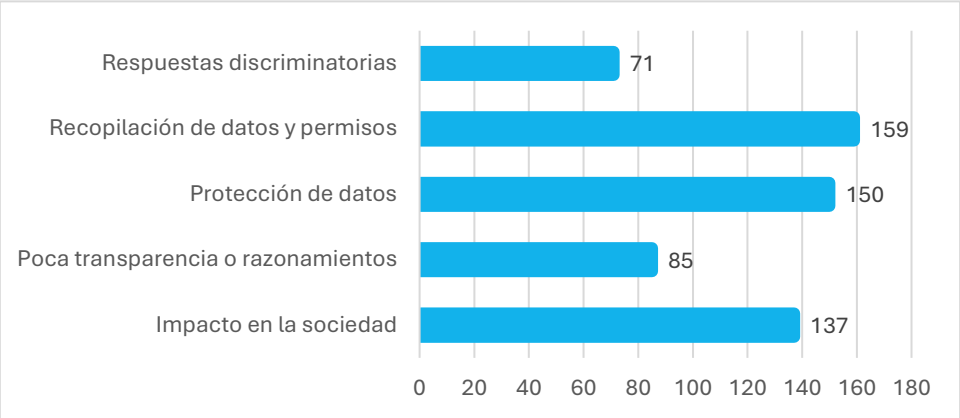

Figura 35: Aspectos de la inteligencia artificial que afectan a la sociedad (Fuente: elaboración propia)

De nuevo, vuelve a destacar la preocupación por la recopilación de datos y permisos y la protección de estos ante el 63,9% y el 60,2% de los encuestados respectivamente. Además, el 55% asegura que va a crear un impacto en la sociedad. Un poco menos preocupante, pero igualmente relevantes, es en relación con las respuestas y los resultados. La falta de razonamiento o transparencia en el proceso afecta a un 34,1% de los encuestados y la connotación discriminatoria de los resultados a un 28,5%. Desde esta perspectiva podemos ver que existe cierta desconfianza tanto en cómo se usa la inteligencia artificial como los resultados que proporciona la misma.

Para hacer la encuesta más dinámica, se propuso una pequeña mecánica que enfrentó la inteligencia humana y la inteligencia artificial. Vamos a analizar los resultados de una pequeña prueba que se les hizo a los encuestados en la que a partir de 4 imágenes tenían que conseguir detectar cuáles habían sido generadas por inteligencia artificial. Para ello, se utilizó la herramienta de Artguru y se solicitó crear una imagen del monumento del Arco del Triunfo, una cueva sobre el mar, un conejo en el césped y unos estudiantes mirando unos apuntes. Efectivamente, las 4 imágenes fueron creadas mediante IA, pero solo 16 de los 249 encuestados (el 6,4%) consiguió detectarlo.

Figura 36: Izquierda: Monumento, derecha: Conejo (Fuente: Elaborado con Artguru)

En primer lugar, las dos imágenes que más han chirriado a los encuestados han sido la del Arco del Triunfo y la del conejo en el césped. El monumento arquitectónico no ha

conseguido engañar al 50,6% de los participantes. Probablemente algunos detalles en las posiciones de las personas o los dibujos de la estructura han provocado que algo más de la mitad de los encuestados detecten el uso de IA. Sin embargo, el concepto general es bastante convincente para el 49,4%, por lo que se queda bastante igualado. Sin embargo, con el conejo la inteligencia artificial no ha sido tan hábil, ya que el 92% de los encuestados se han percatado de que no era una imagen real.

Figura 37: Izquierda: Cueva acuática, derecha: Estudiantes (Fuente: Elaborado con Artguru)

Sin embargo, con estas imágenes los participantes no han podido detectar con tanta certeza el uso de inteligencia artificial. En la de los estudiantes, pese a la dificultad de la inteligencia artificial de dibujar algunas características humanas, el 38,2% ha detectado el uso de inteligencia artificial. Pero la imagen que mejor le ha funcionado a la inteligencia artificial ha sido la de la cueva marina, la cual solo ha sido detectada por el 14,1%. Es decir, el 85,9% de los encuestados pensaba que esta imagen era real. El hecho de no incluir elementos como humanos o animales demuestra el poder de convicción en el frente de los paisajes de la inteligencia artificial.

Para finalizar, se preguntó a los encuestados si tras la realización de la encuesta, había descubierto algunos aspectos éticos o de privacidad que hubiesen cambiado su percepción de la inteligencia artificial. El 69,9% respondieron negativamente, por lo que ya eran conscientes de los problemas a los que se enfrenta esta tecnología.

5.4 Conclusiones de la encuesta

Para cerrar este punto sobre la encuesta, podemos sacar varias conclusiones. La primera, y la más importante, es que la llegada de la inteligencia artificial no es el futuro, es el presente. Esta tecnología se ha adentrado ya en muchos hogares y forma parte de tareas cotidianas. Sin embargo, el ansia de muchas empresas de querer crear sus propios sistemas de inteligencia artificial está generando malestar en la sociedad por el uso poco ético que se está dando, principalmente al uso de datos y la privacidad del usuario.

La inteligencia artificial ha llegado a la sociedad y la sociedad la ha aceptado. Que muestren cierto descontento hacia la misma no significa que quieran erradicarla, sino que confían en que se tomen medidas para que continúe su desarrollo de manera más ética, responsable y sostenible para la sociedad.

6. Conclusión

La inteligencia artificial lleva en nuestras vidas mucho tiempo, pero en los últimos años, los avances que se han producido al respecto han ido calando en la sociedad. Es por ello por lo que el interés en tecnologías que involucren inteligencia artificial es cada vez mayor y cada vez puede aplicarse a más ámbitos. Sin embargo, la evolución actual de esta tecnología, al estar vinculada directamente con la sociedad en muchos campos, está generando conflictos éticos que pueden impactar negativamente tanto en la propia sociedad como en el avance de la inteligencia artificial. Sin embargo, la sociedad no está dispuesta a dejar de utilizar herramientas de inteligencia artificial, por lo que es necesario comenzar a formar un marco regulatorio que permita que el crecimiento de la inteligencia artificial de lugar a un decrecimiento en los derechos fundamentales de las personas. La inteligencia artificial no es solo un avance tecnológico, sino también un reflejo de nuestras decisiones y valores como sociedad. A medida que los sistemas de IA se integran más profundamente en nuestras vidas, los profesionales TIC enfrentan la responsabilidad de asegurarse de que estos sistemas no solo sean técnicamente eficientes, sino también éticamente sólidos y seguros para la humanidad.

La Unión Europea ya se estaba preparando desde 2021 para poder regular el uso y el desarrollo de sistemas de inteligencia artificial, proponiendo varias versiones hasta que finalmente en julio de 2024 se publicó oficialmente el Reglamento 2024/1689 y entró en vigor en agosto del mismo año, con el cual se establecen normas para armonizar el uso y desarrollo de inteligencia artificial. Hasta entonces, el uso y desarrollo de sistemas de inteligencia artificial se regían por otras normativas que les afectaban colateralmente, como el Reglamento General de Protección de Datos. Desde la entrada en vigor del 2024/1689, los territorios pertenecientes a la Unión y aquellos que no pertenezcan, pero quieran introducir sus sistemas de inteligencia artificial, deberán seguir las directrices que se establecen en el mismo.

6.1 Objetivos cumplidos

La guía de buenas prácticas de IA para profesionales TIC se posiciona en el cruce entre la tecnología y la ética, proporcionando un marco de directrices que va más allá de la mera funcionalidad técnica para incluir consideraciones fundamentales sobre la transparencia, la justicia y la protección de los derechos humanos. La ética no es un accesorio opcional en el desarrollo de la IA, sino una necesidad imperativa. La guía subraya la importancia de desarrollar sistemas de IA centrados en el ser humano, una idea que se traduce en diseñar tecnologías que prioricen el bienestar, respeten la dignidad y promuevan los derechos de todos los individuos afectados por sus decisiones. Además, se pone un fuerte énfasis en la necesidad de capacitar y sensibilizar a los profesionales TIC sobre los desafíos éticos y normativos asociados con la IA. En un campo que evoluciona rápidamente por lo que la formación continua no es solo favorable, sino esencial. Los profesionales deben estar preparados para atravesar las cuestiones éticas y legales que emergen en el uso de la IA mientras la innovación en ella sigue su curso. Esto no solo mejorará la calidad de los sistemas

desarrollados, sino que también fortalecerá la confianza pública en la IA y su potencial para ser una fuerza positiva en la sociedad.

6.2 Aportación personal

En la realización de este trabajo he podido comprobar que la sociedad que acepta con los ojos cerrados los Términos y Condiciones en una pulsación táctil es cada vez más consciente de los derechos que la tecnología, en este caso la inteligencia artificial, está neutralizando para el beneficio de unos pocos y el detrimento del resto. Ciertamente tenía mis dudas al respecto, puesto que solo había visto descontentos en algunos ámbitos en los que la irrupción de la inteligencia artificial está causando estragos. Pero saber que como sociedad somos conscientes de que hay algo que no termina de funcionar bien y que hay que arreglarlo, no destruirlo y retroceder al pasado, me ha motivado aún más en la realización del trabajo y escribiendo estas últimas líneas, siento satisfacción por lo realizado.

Por el camino he podido aprender más a fondo tanto de la inteligencia artificial como de la manera en la que se aprovechaban de nuestros derechos en cuanto tenían la oportunidad. Ha sido un recorrido agitado, porque durante la realización del trabajo se producían nuevos avances en la IA y nuevos avances en el reglamento y era necesario ir actualizando y modificando lo ya realizado, pero al mismo tiempo tratar un tema de actualidad al pie del cañón y poder haber visto “de cerca” el proceso de cómo se cerraba la propuesta del Reglamento de la UE tras tantos años de preparación es reconfortante.

6.3 Perspectivas futuras

La realización de esta guía de buenas prácticas tenía el enfoque generalizado para poder abarcar todos los sistemas de inteligencia artificial y todos pudieran beneficiarse de ella. Pero, de todos modos, se podrían crear diferentes variaciones de guías según el ámbito de aplicación para que se ajusten más concretamente a su propósito y objetivos, lo que aumentaría todavía más la efectividad de la guía.

Aunque la legislación se haya aderezado y por fin sea posible que actuar en consecuencia específicamente para los sistemas de inteligencia artificial y la sociedad esté más protegida frente a intereses que vulneran el beneficio común como sociedad, el camino todavía no ha acabado. Tal y como la inteligencia artificial siga avanzando, aparecerán nuevas amenazas para los derechos fundamentales y es posible que esta guía no consiga cubrir todos los aspectos de un futuro tal vez más próximo de lo que esperemos. Es por ello por lo que animo a todo aquel que haya leído mi trabajo a tomar esta guía no solo para desarrollar un sistema de inteligencia artificial, sino para desarrollar una nueva guía que incluya las nuevas directrices que se incluyan en actualizaciones de Reglamentos y/o en otras Leyes. La inteligencia artificial ha llegado para quedarse, pero para quedarse junto con las personas. No permitamos que un mal uso de esta tecnología merme el más mínimo detalle de nuestros derechos fundamentales y utilicémosla correctamente para poder continuar transformando la sociedad indicando nosotros el camino a seguir.

7. Bibliografía

AEPD. (Junio de 2021). *Gestión del riesgo y evaluación de impacto en tratamiento de datos personales*. Obtenido de AEPD: https://www.aepd.es/documento/guia-evaluaciones-de-impacto-rgpd.pdf

AEPD. (23 de septiembre de 2023). *Inteligencia artificial: Transparencia*. Obtenido de AEPD: https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia

Agencia Estatal. (12 de julio de 2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº*. Obtenido de Boletín Oficial del Estado: https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079

Aristóteles. (s. IV a. C). Ética Nicomáquea. En Aristóteles, *Ética Nicomáquea* (págs. 1097b20-21).

Avella, M. d., Sanabria-Moyano, J. E., & Dinas-Hurtado, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal, vol. 8, n. 1,*, 275-310.

Baz, L., & Cornelius, W. K.-H. (1998). EL ARS GENERALIS ULTIMA DE RAMON LLULL. *Revista Española de Filosofía Medieval, 5*, 89-107.

BBC Mundo. (20 de marzo de 2018). *5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera US$37.000 millones en un día*. Obtenido de BBC News: https://www.bbc.com/mundo/noticias-43472797

Bejerano, P. G. (6 de febrero de 2014). *Código Enigma, descifrado: el papel de Turing en la Segunda Guerra Mundial*. Obtenido de elDIario.es: https://www.eldiario.es/turing/criptografia/alan-turing-enigma-codigo_1_5038272.html

Britannica. (2009). *Deep Blue - computer chess-playing system*. Obtenido de Britannica: https://www.britannica.com/topic/Deep-Blue

Buolamwini, J. (2017). *MIT Media Lab*. Obtenido de MIT Media Lab: https://www.media.mit.edu/posts/how-i-m-fighting-bias-in-algorithms/

Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Cambridge.

Cantalapiedra, C. G., & Soler, E. V. (2017). *Evolución de los entornos Big Data y los retos para el arquitecto de datos*.

Castillo, C. d. (2024). *Pizza con pegamento y nazis negros: Google choca con las alucinaciones de la inteligencia artificial*. Obtenido de elDIario.es: https://www.eldiario.es/tecnologia/pizza-pegamento-nazis-negros-google-choca-alucinaciones-inteligencia-artificial_1_11417313.html

CBS News. (3 de julio de 2019). *Reporter on China's treatment of Uighur Muslims: "This is absolute Orwellian style surveillance"*. Obtenido de CBS News: https://www.cbsnews.com/news/china-puts-uighurs-uyghyrs-muslim-children-in-prison-re-education-internment-camps-vice-news/

CCN-CERT BP/30. (2023). *Aproximación a la Inteligencia Artificial y la ciberseguridad*.

Clark, L. (26 de junio de 2012). *Google's Artificial Brain Learns to Find Cat Videos*. Obtenido de Wired: https://www.wired.com/2012/06/google-x-neural-network/

CNN. (5 de octubre de 1998). *New toy an interactive fur ball*. Obtenido de CNN: http://edition.cnn.com/US/9810/05/furby/index.html

Comisión Europea. (29 de mayo de 2024). *La Comisión crea una Oficina de IA para reforzar el liderazgo de la UE en materia de inteligencia artificial segura y fiable*. Obtenido de Comisión Europea: https://ec.europa.eu/commission/presscorner/detail/es/IP_24_2982

Compujedrez. (2016). *Programas de ajedrez. Primeros pasos: Alex Bernstein*. Obtenido de Compujedrez - Ajedrez y computadoras: https://compujedrez.wordpress.com/programas-de-ajedrez-primeros-pasos-alex-bernstein/

Estrada, R. (12 de mayo de 2021). *¿Se acuerdan de AIBO, el perro robot de Sony?* Obtenido de Digital Trends: https://es.digitaltrends.com/tendencias/aibo-robot-perro-mascota-sony/

Fuster, J. (2 de junio de 2024). *'Spider-Man: Beyond the Spider-Verse' Writer-Producer Chris Miller Shuts Down Rumors of Generative AI Use*. Obtenido de The Wrap: https://www.thewrap.com/spider-man-beyond-the-spider-verse-no-ai-chris-miller/

Future of Life. (28 de julio de 2015). *Autonomous Weapons Open Letter: AI & Robotics Researchers*. Obtenido de Future of Life: https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/

Gómez Cardosa, D. (Julio de 2023). *Àmbits d'afectació de la IA en el mercat laboral i les habilitats*. Obtenido de eLearning Innovation Center - Universitat Oberta de Catalunya: http://hdl.handle.net/10609/148606

Gonzalo, M. (28 de agosto de 2023). *Las IA generativas ¿copian o leen? El «uso legítimo» ('fair use') es la clave*. Obtenido de Newtral: https://www.newtral.es/ia-fair-use-scraping-openai-copia-web/20230828/

Gottfredson, L. S. (1997). *Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography*. Elsevier Inc.

Heikkilä, M. (14 de junio de 2024). *How to opt out of Meta's AI training*. Obtenido de MIT Technology Review: https://www.technologyreview.com/2024/06/14/1093789/how-to-opt-out-of-meta-ai-training/

Humphreys, L. G. (1979). *The construct of general intelligence*. Elsevier Inc.

IAMAI. (9 de enero de 2024). *El desarrollo de la IA antes y después del informe Lighthill*. Obtenido de IAMAI: https://www.iamai.es/Blog/El-informe-de-James-Lighthill-sobre-la-inteligencia-artificial/

Lemmens, A. (s.f.). *Debiasing Algorithms: Fair Machine Learning*. Obtenido de Aurélie Lemmens: https://www.aurelielemmens.com/debiasing-algorithms-fair-machine-learning/

López, M. (Enero de 2024). *Universal retira sus canciones de Tik Tok, ¿cuáles son las implicaciones para el negocio musical?* Obtenido de Sympathy for the Lawyer: https://sympathyforthelawyer.com/blog/universal-retirada-catalogo-tiktok

Luna, J. (s.f.). *Máquinas de Zuse*. Obtenido de Jaime Luna: https://jaimeluna.angelfire.com/zuse.html

Maeztu, D. (2023). *¿Plagio masivo, fair use o reutilización del conocimiento?* . Obtenido de Newtral: https://www.newtral.es/ia-fair-use-scraping-openai-copia-web/20230828/

Maher, M. L., & Fisher, D. (2011). AAAI 2011 Spring Symposium Artificial Intelligence and Sustainable Design. California. Obtenido de https://web.archive.org/web/20190729063022/http://dts-web1.it.vanderbilt.edu/~fisherdh//AI-Design-Sustainability.html

Manrique, J. (2007). *La lengua universal de Leibniz*. Universidad Nacional de Colombia.

Marina, J. A. (2 de febrero de 2020). La moral y la ética. *EL MUNDO*.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. (1955). *A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE*. Hanover.

Metropolitan Police. (2024). *Facial Recognition Technology*. Recuperado el agosto de 2024, de Metropolitan Police.

Pastor, J. (26 de octubre de 2017). *Microsoft abandona la fabricación de Kinect, el periférico que podría haberlo cambiado todo*. Obtenido de Xataka: https://www.xataka.com/videojuegos/microsoft-abandona-la-fabricacion-de-kinect-el-periferico-que-podria-haberlo-cambiado-todo

Peris, G. (26 de junio de 2024). *Ingeniería e IA: ¿cómo pueden ayudar a solucionar los retos de la refrigeración en centros de datos?* Obtenido de Sener: https://www.group.sener/insights/ingenieria-e-ia-como-pueden-ayudar-a-solucionar-los-retos-de-la-refrigeracion-en-centros-de-datos/?doing_wp_cron=1725167137.6688230037689208984375

Ramíguez, A. (2016). *SimSimi: la nueva app de inteligencia artificial genera controversia*. Obtenido de Tecnetico: https://www.tecnetico.com/internet/simsimi-la-nueva-app-de-inteligencia-artificial-genera-controversia/64220

Robots Guide. (s.f.). *Genghis*. Obtenido de robotsguide.com: https://robotsguide.com/robots/genghis

Rocher, L., Hendrickx, J. M., & Montjoye, Y.-A. d. (2019). *Estimating the success of re-identifications in incomplete datasets using generative models.* Obtenido de Nat Commun 10, 3069: https://doi.org/10.1038/s41467-019-10933-3

Romero, P. (23 de julio de 2019). *Demuestran cómo los datos personales anonimizados no garantizan la privacidad*. Obtenido de Público: https://www.publico.es/sociedad/tecnologia-demuestran-datos-personales-anonimizados-no-garantizan-privacidad.html

Rumelhart, D., McClelland, J., & Feldman, J. A. (1986). *Parallel distributed processing: explorations in the microstructure of cognition.* MIT Press.

Rys, D. (12 de abril de 2024). *Billboard*. Obtenido de Record Label Market Share Q1 2024: Warner Records Posts Huge Gains While Universal Enters a New Era: https://www.billboard.com/business/record-labels/record-label-market-share-q1-2024-universal-warner-1235655068/

Salaru, D. (23 de junio de 2022). *Russia: Facial recognition software used to target journalists*. Obtenido de ipi.media: https://ipi.media/russia-facial-recognition-software-used-to-target-journalists/

Shannon, C. E. (1950). Programming a Computer for Playing Chess. *Philosophical Magazine, ser. 7, vol. 41, núm. 314*.

Signorelli, A. D. (20 de abril de 2024). *La irónica historia del primer chatbot y su creador, Joseph Weizenbaum*. Obtenido de WIred: https://es.wired.com/articulos/el-primer-chatbot-de-la-historia-y-su-creador-joseph-weizenbaum

Simonite, T. (17 de marzo de 2014). *Facebook Creates Software That Matches Faces Almost as Well as You Do*. Obtenido de MIT Technology Review: https://www.technologyreview.com/2014/03/17/13822/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/

Sony. (1999). *Sony Launches Four-Legged Entertainment Robot*. Obtenido de Sony: https://www.sony.com/en/SonyInfo/News/Press/199905/99-046/

Swartz, L. (2003). *WHY PEOPLE HATE THE PAPERCLIP: LABELS, APPEARANCE, BEHAVIOR AND SOCIAL RESPONSES TO USER INTERFACE AGENT*.

Telefónica. (10 de julio de 2024). *Actualizamos nuestros principios de inteligencia artificial*. Obtenido de Telefónica: https://www.telefonica.com/es/sala-comunicacion/blog/actualizamos-nuestros-principios-de-inteligencia-artificial/

TikTok Newsroom. (2 de mayo de 2024). *Universal Music Group y TikTok anuncian un nuevo acuerdo de licencia*. Obtenido de TikTok Newsroom: https://newsroom.tiktok.com/es-latam/tiktok-universal-music-group-acuerdo-musica-plataforma

Turing, A. M. (1 de octubre de 1950). COMPUTING MACHINERY AND INTELLIGENCE. *Mind, Volume LIX, Issue 236,*, págs. 433-460. Obtenido de https://doi.org/10.1093/mind/LIX.236.433

UE 2024/1689. (art 111). *Sistemas de IA ya introducidos en el mercado o puestos en servicio y modelos de IA de uso general ya introducidos en el mercado*.

UE 2024/1689. (art 6-49). *Sistemas de IA de alto riesgo*.

UE 2024/1689. (art. 5). *Prácticas de IA prohibidas*.

UE 2024/1689. (art. 50). *Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA*.

UE 2024/1689. (art. 51-56). *Modelos de IA de uso general*.

UE 2024/1689. (art. 57-62). *Medidas de apoyo a la innovación*.

UE 2024/1689. (art. 72). *Vigilancia poscomercialización*.

UE 2024/1689. (art. 73). *Notificación de incidentes graves*.

UE 2024/1689. (art. 99). *Sanciones*.

UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*.

Whiddington, R. (15 de noviembre de 2022). *DeviantArt's New A.I. Generator Angers Artists for Promising—But Failing—to Protect Creator's Rights*. Obtenido de Artnet: https://news.artnet.com/art-world/deviantart-dreamup-ai-generator-creators-rights-ip-controversy-2210607

World Economic Forum. (2023). *Future of Jobs Report 2023.* Obtenido de https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

Zarco, J. (8 de noviembre de 2023). *El enorme enfado de Bad Bunny después de que se viralice una canción suya creada por la IA*. Obtenido de Las Provincias: https://www.lasprovincias.es/culturas/musica/enorme-enfado-bad-bunny-despues-viralice-cancion-20231108124224-nt.html

ANEXO 1: Relación del TFG con los Objetivos de Desarrollo Sostenible

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS):

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. **Fin de la pobreza.**				X
ODS 2. **Hambre cero.**				X
ODS 3. **Salud y bienestar.**	X			
ODS 4. **Educación de calidad.**		X		
ODS 5. **Igualdad de género.**	X			
ODS 6. **Agua limpia y saneamiento.**				X
ODS 7. **Energía asequible y no contaminante.**		X		
ODS 8. **Trabajo decente y crecimiento económico.**		X		
ODS 9. **Industria, innovación e infraestructuras.**		X		
ODS 10. **Reducción de las desigualdades**.	X			
ODS 11. **Ciudades y comunidades sostenibles**.		X		
ODS 12. **Producción y consumo responsables.**		X		
ODS 13. **Acción por el clima.**		X		
ODS 14. **Vida submarina.**				X
ODS 15. **Vida de ecosistemas terrestres.**				X
ODS 16. **Paz, justicia e instituciones sólidas.**		X		
ODS 17. **Alianzas para lograr objetivos.**				X

Figura 38: Objetivos Desarrollo Sostenible relacionados con el TFG

Gracias a la guía de buenas prácticas, a los objetivos que actualmente tiene la inteligencia artificial podemos añadir otros más con la incorporación de un uso más ético y sostenible. Es por eso por lo que este trabajo guarda gran relación con el ***ODS 5. Igualdad de género*** y el ***ODS 10. Reducción de las desigualdades***, ya que con la guía se pretende eliminar el actual problema que tiene la inteligencia artificial con los sesgos discriminatorios. Según cómo se haya desarrollado la inteligencia artificial y cómo vaya aprendiendo en el proceso, siguiendo los pasos de este trabajo se debería eliminar las posibles actitudes discriminatorias hacia grupos sociales poco representados por las muestras. Con la eliminación de estos sesgos se promueve la igualdad de oportunidades y una tecnología justa para todos, en relación con el ***ODS 16. Paz, justicia e instituciones sólidas***.

Otro aspecto relacionado es el ***ODS 4. Educación de calidad***. La IA puede personalizar la educación, proporcionar acceso a recursos educativos a personas en áreas remotas

y ayudar a personalizar las necesidades de aprendizaje individuales. Con la guía de buenas prácticas se pretende hacer que la tecnología sea más accesible y se tenga un mayor control para evitar respuestas erróneas que frenarían el campo de la educación. Además, con sistemas más transparentes, se podría comprender el funcionamiento de estos, pudiendo ser este un motivo educacional.

También se vela por la sostenibilidad en la guía de buenas prácticas, y es de especial relevancia, porque el desarrollo y uso de sistemas de IA generan consumos de energía muy elevados, que de por sí son contrarios a los objetivos de la Agenda 2030. Sin embargo, con la guía se indica qué acciones se deben tomar para poder reducir el consumo de energía y hacer un uso más responsable, conectando directamente con el ***ODS 7. Energía asequible y no contaminante.*** Entre ellos, la optimización de código para evitar sobrecargas innecesarias durante el trabajo de la inteligencia artificial o el uso de hardware más modular, para que en el caso en el que se precise una sustitución, afecte al menor número de componentes y reduciendo la huella de carbono. Este último punto concuerda con el ***ODS 12. Producción y consumo responsable***, y en sí, al actuar sobre el resto de los objetivos, se traducirá en una menor emisión de gases de efecto invernadero que provocan el cambio climático, relacionándose con el ***ODS 13. Acción por el clima***.

Además, si combinamos las buenas prácticas sobre el medio ambiente con los avances de la propia inteligencia como la posibilidad de optimizar la gestión del tráfico, el propio consumo de energía y la planificación urbana para hacer las ciudades más sostenibles, se descubre la gran relación del ***ODS 11. Ciudades y comunidades sostenibles*** con el trabajo.

En un momento en el que la inteligencia artificial está comiendo terreno a los derechos fundamentales, causando malestar en la sociedad, es necesario aplicar también la guía de buenas prácticas en el ámbito social, afectando al ***ODS 3. Salud y bienestar***. Desarrollando la inteligencia artificial con visión antropológica prevalecerá el bienestar de las personas frente a los intereses individuales o empresariales. Además, en ese punto también se tendrá más precaución con el uso de sistemas de IA en el campo de la salud para no causar daños contraproducentes. Por esa misma razón, se tiene en cuenta también en la guía la preocupación del impacto social en cuanto a los puestos de trabajo, afectando al ***ODS 18. Trabajo decente y crecimiento económico***, pero se trata de tal manera que las personas sigan teniendo participación en mayor medida mientras se optimizan los gastos para maximizar los beneficios.

Por último, hay que destacar la innovación que produce la inteligencia artificial, relacionándolo con el ***ODS 19. Industria, innovación e infraestructuras***. Pese al pensamiento de que la regulación de la inteligencia artificial frenará su innovación, desde la Unión Europea con su reciente Reglamento de la IA han procurado que la velocidad de crecimiento de esta tecnología no disminuya y que siga evolucionando a la vez que se mantienen los derechos fundamentales de las personas.

ANEXO 2: Formulario de autoevaluación de sistemas de IA responsables

AFIRMACIONES	SÍ CUMPLE	NO CUMPLE
Visión antropológica		
1- El sistema contribuye al bienestar humano general		
2- El sistema busca tener un impacto positivo en la sociedad		
3- No se pretende ocasionar daños a personas físicas con el uso del sistema		
4- El sistema no toma el control sobre las decisiones de las personas		
Observaciones:		
Justicia e igualdad		
5- Los datos de entrenamiento incluyen todos los géneros		
6- Los datos de entrenamiento incluyen todas las etnias		
7- Los datos de entrenamiento incluyen al mayor número de grupos sociales posible y son diversos		
8- El sistema no ha desarrollado ningún sesgo ni discrimina		
9- Existe diversidad en el grupo de desarrollo		
Observaciones:		
Transparencia y explicabilidad		
10- Se mantiene actualizada la documentación del modelo durante el desarrollo		
11- En la documentación se exponen los datos y los algoritmos		
12- En la documentación se exponen la metodología de entrenamiento		
13- El sistema ofrece trazabilidad de decisiones		
14- Se comprenden las decisiones tomadas por el sistema		
15- El sistema advierte al usuario final que está interactuando con una inteligencia artificial		
Observaciones:		
Protección de datos, privacidad y seguridad		
16- Los datos recopilados son anonimizados		
17- Los datos no necesarios son eliminados		
18- Se pide consentimiento explícito al usuario final para el tratamiento de datos personales		
19- El sistema está preparado para que el usuario final pueda revocar su consentimiento en el tratamiento de datos		

AFIRMACIONES	SÍ CUMPLE	NO CUMPLE
20- El sistema está preparado para eliminar datos personales tras una petición y para reentrenarse sin esos datos		
21- Existen medidas de seguridad especiales para la protección de datos		
22- El acceso a los datos del sistema está restringido y controlado		
23- Se han realizado pruebas de ataques contra el sistema y el sistema ha salido victorioso		
24- En caso de sufrir una brecha de seguridad, existe un protocolo de actuación a seguir incluyendo información a autoridades y usuarios		
Observaciones:		
Sostenibilidad		
25- Se ha estudiado la eficiencia del sistema para maximizarla		
26- Se utiliza energía renovable en los centros de datos utilizados		
27- Se realizan evaluaciones de impacto energético periódicamente		
28- Existe un protocolo de gestión responsable del ciclo de vida de los elementos hardware		
29- Se utilizan configuraciones de hardware modulares		
Observaciones:		
Supervisiones y responsabilidades		
30- Las decisiones críticas son supervisadas y aprobadas por un humano antes de ser ejecutadas		
31- Si el sistema es de alto riesgo, en caso de problema crítico existe un protocolo de actuación para informar a las autoridades reguladoras		
32- Existe una división clara de responsables en los diferentes aspectos		
33- Las intervenciones humanas quedan registradas		
34- Las políticas de la empresa junto con sus expectativas éticas y sociales se mantienen actualizadas		
35- Se tiene especial supervisión en aquellos aspectos controlados por las normativas vigentes		
36- Se está al corriente de las novedades legislativas		
Observaciones:		
Identificación de riesgos		
37- Previo al desarrollo, se identifican los posibles riesgos asociados a este		

AFIRMACIONES	SÍ CUMPLE	NO CUMPLE
38- Se utiliza una matriz de riesgos para evaluar los riesgos potenciales		
39- Se asignan puntuaciones a los riesgos y se clasifican		
40- Se priorizan los posibles riesgos de mayor nivel		
41- La matriz de riesgos es actualizada periódicamente		
Observaciones:		
Tratamiento de datos		
42- Se detalla al usuario el propósito del tratamiento de datos		
43- Se detalla qué tipos de datos son recopilados		
44- Se explica de qué manera son utilizados los datos recopilados		
45- Se detalla qué tecnologías se utilizan en el proceso		
46- Se toman medidas especiales para tratar los datos de forma segura		
Observaciones:		
Propiedad intelectual		
47- Los datos utilizados no tienen derechos de autor o si tiene la licencia pertinente para poder utilizarlos		
48- Se establecen pautas claras sobre el uso de contenido de terceros		
49- Se está al tanto de las normativas de propiedad intelectual del territorio y sus actualizaciones		
Observaciones:		
Riesgo alto y prácticas prohibidas		
50- Ante un cambio en la normativa que suponga un incumplimiento, existen protocolos de actuación para minimizar los riesgos de incumplimiento		
51- El sistema no manipula subliminalmente al usuario ni le hace comportarse de una manera que no harían normalmente		
52- El sistema no ejerce abuso de poder hacia personas o grupos en situación vulnerable		
53- El sistema no otorga puntuaciones para clasificación en base a características personales		
54- El sistema no hace uso de datos biométricos para controlar el estado emocional de las personas en el ámbito laboral o educativo		
55- El sistema no entrena reconocimiento facial a través de imágenes de circuito cerrado		

AFIRMACIONES	SÍ CUMPLE	NO CUMPLE
Observaciones:		
Diseño		
56- El sistema se desarrolla priorizando la seguridad de las personas		
57- El sistema se desarrolla priorizando la ética de las personas		
58- El sistema se desarrolla priorizando la privacidad de las personas		
59- El sistema se desarrolla priorizando los derechos de las personas		
60- El sistema está diseñado para mantener la seguridad y responsabilidad en todas las etapas de su ciclo de vida		
61- El sistema tiene un diseño resiliente que permite la continuidad del servicio pese a fallos técnicos, ataques o condiciones adversas		
62- Existen protocolos de respaldo, manuales o automáticos, para volver a un estado seguro		
63- Existen mecanismos de recuperación automática para volver a un estado seguro		

Figura 39: Formulario de autoevaluación de sistemas de IA responsables (Fuente: elaboración propia)