UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Creación de un informe de buenas prácticas para el uso profesional de la inteligencia artificial en el desarrollo de software

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Benavent García, Rafael

Tutor/a: Oltra Gutiérrez, Juan Vicente

CURSO ACADÉMICO: 2023/2024

Resumen

El empleo de distintas herramientas de inteligencia artificial en el día a día de un profesional informático supone la aparición de una serie de riesgos y problemas tanto legales como deontológicos que deben de ser cubiertos para lograr un uso ético de esta tecnología. El presente TFG tiene como objetivos la comprensión correcta de los fundamentos básicos de la IA, así como realizar una clasificación de las herramientas más empleadas, compilar las referencias legales más pertinentes y rastrear ejemplos de buen y mal uso para, con todo ello, construir un informe de buenas prácticas (BP) que sirva de apoyo en el ejercicio cotidiano de la profesión, en particular a lo relativo al desarrollo de software.

Para garantizar que la estructura y los contenidos sean lo más completos y actuales posibles, se ha investigado en profundidad tanto el nuevo reglamento de inteligencia artificial de la Unión Europea como diversos informes relevantes en la materia.

Palabras clave

IA, software, marco legal, deontología.

Abstract

The use of different artificial intelligence tools in the day-to-day work of a computer professional entails the appearance of a series of risks and problems, both legal and deontological, that must be covered in order to achieve an ethical use of this technology. The objectives of this thesis are to correctly understand the basic fundamentals of AI, as well as to classify the most commonly used tools, compile the most relevant legal references and trace examples of good and bad use in order to construct a best practices report that will serve as a support in the daily work of the profession, particularly in relation to software development.

To ensure that the structure and contents are as complete and up-to-date as possible, the new EU regulation on artificial intelligence as well as various relevant reports on the subject have been studied in depth.

Keywords

AI, software, legal framework, deontology.

Índice general

Índice de figuras

1 Introducción

La Inteligencia Artificial (IA), según John McCarthy, es "la ciencia y la ingeniería para crear máquinas inteligentes, especialmente programas informáticos inteligentes. Está relacionada con la tarea similar de utilizar ordenadores para comprender la inteligencia humana, pero la IA no tiene por qué limitarse a métodos que sean biológicamente observables"[1]. Explicado con otras palabras, la inteligencia artificial es una rama innovadora que combina la informática y las herramientas para examinar grandes cantidades de información, la cual tiene como función principal la resolución de problemas complejos, pero también puede ser aprovechada en tareas repetitivas grandes y simples para facilitar a los humanos ahorrarse trabajo y capital humano. Para lograr este objetivo, es necesario crear un sistema que pueda manejar grandes volúmenes de datos, estudiarlos y utilizar la información para tomar decisiones y resolverse de la misma manera que lo haría un ser humano, es decir, repita los mismos procesos de pensamiento que usamos nosotros.

Los sistemas inteligentes están empezando a tomar un papel muy importante en la vida diaria, algunos ejemplos que podemos utilizar para evidenciar esto son los asistentes virtuales o páginas como chatGPT o Copilot, las cuales nos ayudan a la hora de realizar actividades. Sin embargo, aplicar esto al desarrollo de software es un terreno relativamente reciente, inexplorado y lleno de posibilidades.

La creación de programas informáticos es un proceso complejo que requiere de muchas habilidades técnicas, así como una buena planificación y trabajo en equipo. La IA está en auge en este ámbito debido a sus mejoras significativas, como eficiencia, precisión, automatización e incluso la capacidad de anticipar y resolver problemas antes de que ocurran. Por esto es importante plantearse una serie de cuestiones. ¿Cómo podemos garantizar que las decisiones que toma sean justas y transparentes? ¿Cómo podemos asegurar que trate a todos por igual y no genere ningún tipo de sesgo? ¿Qué herramientas y conocimientos deben tener los desarrolladores para cumplir con todos los principios éticos que conlleva el uso de esta?

Este trabajo se propone abordar estas preguntas, así como otras que surgen al usar esta tecnología, desde una perspectiva deontológica, presentando un informe de buenas prácticas sobre el uso profesional de la inteligencia artificial en el desarrollo de software. Con este documento, se pretende otorgar a los desarrolladores las herramientas y el conocimiento necesario para utilizar los sistemas inteligentes de manera efectiva y ética.

Por último, todo lo que se trata en este trabajo no solo es aplicable en el área de la inge-

niería de software, sino que también se puede usar a otros entornos donde se este comenzando a utilizar la inteligencia artificial. Por lo tanto, se espera que este estudio sea de utilidad no sólo para los lectores principales hacía los que se ha desarrollado el informe, sino también para cualquier persona que tenga en mente el uso de esta herramienta en un entorno profesional.

1.1 Motivación

Motivación personal

Desde que empecé mis estudios en la carrera de ingeniería informática, siempre me ha fascinado el mundo de la inteligencia artificial. El simple hecho de que una máquina se pueda ir adaptando a los requisitos que se le pidan o incluso que se asemejen a nosotros en la capacidad de toma de decisiones me parece algo verdaderamente asombroso. Por esto, al ir profundizando en la carrera cada vez me he ido interesando más este campo, y al no haber podido acceder a la rama de computación, que es la que se especializa en este área, siempre se me ha quedado la espina clavada de haberle dedicado más tiempo en mis estudios. Por eso, decidí que mi Trabajo de Fin de Grado debía de centrarse en este tema.

Al tener claro a rasgos generales sobre que iba a hacer mi TFG, decidí echarle un vistazo a los trabajos que se ofertan de forma pública, para ver si así había alguno que me interesase y además me solucionaba el estar pensando la temática en concreto. Al mirar encontré la oferta que estoy desarrollando ahora mismo, la creación de un informe de buenas prácticas para el uso profesional de la inteligencia artificial en el desarrollo de software, el cual, al leerlo, lo puse casi de forma instantánea como mi primera opción. Esta decisión la tomé porque me parecía bastante interesante el mirar a la IA no desde un punto de vista informático o técnico, que es al que he estado acostumbrado en todos estos años de carrera, sino el verla desde un punto de vista deontológico, es decir, analizar de forma legal y ética el uso de esta en entornos profesionales.

Motivación profesional

En todos los aspectos de la vida, tanto diaria como profesional, el uso de la inteligencia artificial está creciendo a un ritmo sin precedentes, ya que su adopción en diversas industrias ha aumentado exponencialmente en los últimos años. Debido a que este mundo aún es re-

lativamente reciente, es normal que las empresas busquen actualmente formas de integrarla en sus proyectos y en sus formas de trabajo para mejorar la eficiencia y la productividad. En el libro "The Age of AI" (2021), Henry A. Kissinger, Daniel Huttenlocher y Eric Schmidt argumentan que la IA tiene el potencial de transformar profundamente la forma en que se trabaja y se vive, y que solo estamos al principio de esta revolución[2].

Sin embargo, aplicarla en entornos profesionales también plantea una serie de desafíos éticos y prácticos que requieren de ser resueltos de antemano para evitar problemas en el futuro. El objetivo principal de este Trabajo de Fin de Grado es contribuir a que el uso de los sistemas inteligentes se realice de forma fiable, ética y práctica. Además de esto, se espera que este trabajo ayude a establecer un conjunto de buenas prácticas para su uso profesional en la producción de programas informáticos, y que estas buenas prácticas puedan ser utilizadas por otros en el campo.

Por ello, la meta que se busca cumplir con este proyecto es de suma importancia en la actualidad, y, en adición a esto, aporta tanto a las personas que se encargan de su realización como a los lectores un gran conocimiento sobre este sector emergente, lo cual creemos que es motivación suficiente para invertir una gran cantidad de tiempo en su desarrollo.

1.2 Objetivos

El propósito de este trabajo es crear un informe de buenas prácticas para el uso profesional de la inteligencia artificial en el desarrollo de software, el cual tiene como objetivo principal proporcionar un conjunto de directrices y recomendaciones para ayudar a los profesionales a utilizar esta tecnología en la creación de programas de forma responsable, ética y eficaz. Este propósito se basa en la necesidad de que, a la hora de poner en uso esta herramienta, se sigan los principios establecidos por la Unión Europea (UE).

Para que cualquier profesional comprenda cómo iniciar el uso de la IA en su ámbito y conocer las directrices necesarias para un uso ético desde un enfoque deontológico, los siguientes objetivos específicos, que parten del general, son indispensables:

- **Comprender sus fundamentos:** Consiste en proporcionar una comprensión sólida de los conceptos básicos de la inteligencia artificial. Para lograr esto, se va a llevar a cabo una exploración de las definiciones y alcances, así como una discusión sobre sus

aplicaciones actuales y potenciales en diversos campos.

- **Analizar las técnicas, enfoques y utilidad:** Esta meta requiere de una investigación detallada de las diversas estrategias y métodos empleados en el desarrollo de software impulsado por los sistemas autónomos. Se evalúan distintas técnicas, como la lógica difusa, redes neuronales, algoritmos genéticos y algoritmos de aprendizaje automático, además de analizar cómo estas técnicas pueden usarse específicamente en este campo para mejorar la eficiencia, la precisión y la automatización de tareas.
- **Explorar consideraciones éticas y legales:** Se centra en las implicaciones éticas y legales asociadas con el uso de la inteligencia artificial en la creación de programas informáticos. Para ello, se estudian diversas cuestiones relacionadas con la privacidad de los datos, el sesgo algorítmico, la responsabilidad algorítmica, la equidad y la transparencia en los sistemas. Todo esto se realizará principalmente teniendo como base el nuevo reglamento del 13 de marzo de 2024 hecho por la Unión Europea[3].
- **Buscar ejemplos de buen y mal uso:** Identifica una serie de ejemplos de uso de esta tecnología para poder ver cuales han sido los errores que han llevado a un mal uso, y los aciertos que han conseguido lograr que esta se utilice de forma adecuada.

1.3 Estructura

En este apartado se enumeran y explican brevemente todos los capítulos que dividen este trabajo:

- **Introducción:** Tiene como meta contextualizar al lector, proporcionando una visión general de la investigación, su motivación, objetivos y, dentro de este marco, su estructura.
- **Estado del arte:** Este capítulo es el más importante de este trabajo, debido a que es en el que se procede a realizar toda la investigación que repercute en la redacción del informe. Esto significa que aquí se explican que proyectos previos nos han servido para inspirarnos en la redacción del informe y, además, que estudios se utilizan para poder justificar todas las afirmaciones que se han realizado.
- **Análisis del problema:** En primer lugar, se analizan una serie de soluciones previas para observar sus puntos positivos y negativos, para de esta forma tenerlos en cuenta

para nuestra redacción. Posteriormente se presenta la solución que se va a realizar junto a su estructura, planificación y estimación de esfuerzo.

- **Diseño y desarrollo del informe:** En esta sección se identifican que herramientas han sido de gran utilidad para la redacción del informe, se detalla más en profundidad la estructura final con la que ha quedado el informe y cómo ha evolucionado y si se ha cumplido con la planificación previa. Además, se incluye una herramienta para evaluar este documento.

- **Conclusión:** Una vez concluido el desarrollo del informe, se hace una vista hacía atrás para ver si todos los objetivos y problemas planteados se han solucionado y, en algunos casos, que desafíos se han encontrado durante su realización. En adición a esto, también se trata de los conocimientos que se han adquirido gracias a toda esta investigación, tanto personales como profesionales, finalizando con una relación la tarea con los estudios cursados, dejando claro que el contenido del TFG es conforme con estos.

- **Trabajos futuros:** Para terminar, es de suma importancia dejar claros que puntos de esta tarea se podrían haber tratado pero por problemas de complejidad o de tiempo no ha sido posible. Se explica también que nuevas líneas de desarrollo se abren para aplicar estos resultados a otras áreas y que mejoras aplicaría con toda la experiencia adquirida.

- **Referencias:** En esta sección se exponen detalladamente todas las fuentes utilizadas, haciendo posible de esta forma al lector poder ubicar toda la bibliografía si le es necesario o quiere comprobar la veracidad de ciertas secciones.

- **Anexos:** Aquí se encuentran dos documentos esenciales para este trabajo, siendo el primero de ellos el estudio que relaciona el informe con los objetivos de desarrollo sostenible de la agenda de 2030. El segundo de ellos es el propio informe al que hace referencia esta memoria: Informe de buenas prácticas para el uso profesional de la inteligencia artificial en el desarrollo de software.

2 Estado del arte

Para garantizar la relevancia y aplicabilidad de este documento, se requiere un estudio cuidadoso de la normativa vigente, opiniones éticas e informes similares en el campo. Para lograr esto, se analizan en este apartado varios estudios que, por su similitud e importancia en el contexto, forman la base de este informe.

Para comenzar, analizaremos las regulaciones y documentos propuestos por la Unión Europea (UE), que brindan unas valiosas directrices sobre el uso ético y responsable de la inteligencia artificial, además de ser los documentos más importantes en todo la evolución de este proyecto. Adicionalmente, revisaremos otros informes de otras organizaciones con objetivos similares a este para conocer sus conclusiones y recomendaciones, además de proporcionarnos una guía de como redactar el documento y que estructura debería tener.

Finalmente exploraremos el espacio de conocimiento y tecnología en el que reside este proyecto, además de incidir en que se va a diferenciar este estudio de los creados con anterioridad.

2.1 Regulaciones y normativas presentadas por la Unión Europea

Vamos a comenzar este apartado analizando el documento más importante en todo el desarrollo de este trabajo y que más relevancia tiene debido a lo reciente que es, el "Reglamento de Inteligencia Artificial", cuya resolución legislativa fue presentada el 13 de marzo de 2024[3] y posteriormente fue publicado el texto final el 13 de junio de 2024[4]. Esta normativa, establecida por el Parlamento Europeo, ha proporcionado una serie de reglas relacionadas con el sector de la IA y ha detallado en profundidad un marco que deben seguir los operadores de estos sistemas.

El objetivo principal de este documento (explicado en la página 4 del reglamento) es optimizar el mercado interno con un marco legal uniforme, esto para regular el desarrollo, la introducción en el mercado, la implementación y el uso de sistemas de inteligencia artificial dentro de la Unión Europea, promoviendo una tecnología centrada en las personas y fiable. Además de esto, también busca garantizar un alto nivel de protección de la salud, la seguridad y los derechos fundamentales establecidos en la Carta de Derechos Fundamentales[5], incluyendo aspectos como la igualdad, la privacidad y la conservación del medio ambiente.

Todo esto se hace para contrarrestar los impactos negativos de estos sistemas en la Unión y para apoyar la innovación.

Este reglamento surgió de la escritura de varios documentos anteriores, de los cuales cabe destacar dos bastante relevantes, la "Ley de Inteligencia Artificial" presentada en abril de 2021[6], que fue la primera de su tipo en el mundo, y las "Normas de Derecho Civil sobre Robótica" presentadas en febrero de 2017[7]. Estos informes sentaron las bases éticas y legales en la UE sobre el uso de sistemas inteligentes, además de que subrayaron la importancia de garantizar que estos productos respeten los derechos fundamentales de los seres humanos.

A continuación, abordaremos todos los puntos esenciales que deben conocerse para la redacción actual de un informe de buenas prácticas sobre este tema[1].

2.1.1. ¿Qué sistemas de IA son afectados por este reglamento?

Este conjunto de normas se dirige a una variedad de desarrolladores encargados de la implementación y despliegue de sistemas de inteligencia artificial en la Unión, entre los que se encuentran proveedores, responsables del despliegue, importadores, distribuidores, fabricantes de productos con esta tecnología, representantes autorizados de proveedores no establecidos en la Unión, y personas afectadas ubicadas en esta zona.

Los sistemas que presentan un alto riesgo, tal y como explicaremos más adelante, están sujetos a una serie de regulaciones específicas. Sin embargo, estas normas no se aplican a áreas que están más allá del alcance del Derecho de la Unión, ni afectan las competencias de los Estados miembros en materia de seguridad nacional, además de tampoco aplicarse a sistemas destinados exclusivamente a objetivos militares, de defensa o de seguridad nacional. En adición a esto, las autoridades públicas de países terceros y organizaciones internacionales están exentas si utilizan estos modelos en cooperación con la Unión para la aplicación de la ley y la cooperación judicial, siempre y cuando ofrezcan garantías suficientes de protección de los derechos y libertades fundamentales.

La normativa tampoco afecta la responsabilidad de los prestadores de servicios intermediarios, ni se aplica a modelos desarrollados y puestos en servicio exclusivamente para investigación y desarrollo científicos. Además, no se aplica a ninguna actividad de investigación,

[1]Cabe destacar que hay apartados del respectivo documento que no se han tratado aquí por su poca relación con la temática de este trabajo.

prueba o desarrollo antes de su introducción en el mercado o puesta en servicio, incluyendo también las obligaciones de los responsables del despliegue que sean personas físicas que utilicen sistemas de IA en el ejercicio de una actividad puramente personal y no profesional.

Sin embargo, si que se aplicarán a los divulgados con licencias libres y de código abierto, a no ser que se introduzcan en el mercado o se pongan en servicio como de alto riesgo o que entren en el ámbito de aplicación de las prácticas prohibidas o de las obligaciones de transparencia (en la figura 2.1 se pueden ver por encima aquellos modelos que pueden llegar a no estar afectados).

Figura 2.1: Sistemas no afectados por el reglamento

Además, el reglamento no impide que la Unión o los Estados miembros mantengan o introduzcan disposiciones más favorables a los trabajadores en lo que respecta a la protección de sus derechos respecto al uso de esta tecnología por parte de los empleadores. También se entiende sin perjuicio de las normas establecidas por otros actos jurídicos de la Unión relativos a la protección de los consumidores y a la seguridad de los productos.

Por último, el Derecho de la Unión que se ocupa de la protección de los datos personales, la privacidad y la confidencialidad de las comunicaciones, se aplicará a los datos personales que se procesen en relación con los derechos y obligaciones establecidos en este Reglamento.

2.1.2. Capacitación en inteligencia artificial

Todo aquel proveedor y responsable de un sistema de IA debe asegurarse de que toda persona relacionada con su funcionamiento y utilización tenga conocimiento suficiente en esta área. Para saber si una persona debe o no encargarse de este sistema, es necesario tener en cuenta sus conocimientos técnicos, su experiencia, su educación, su formación, el contexto en el que se va a usar y a quienes está destinado.

2.1.3. Prácticas prohibidas

El artículo 5 (capítulo II) establece una serie de prohibiciones que se deben tener en cuenta para cualquier desarrollo de IA, las cuales son las siguientes:

- Están prohibidas las técnicas subliminales o manipuladoras que alteran significativamente el comportamiento de una persona o grupo, reduciendo su capacidad para tomar decisiones informadas.
- No se pueden explotar las vulnerabilidades de una persona o grupo específico debido a su edad, discapacidad o situación socioeconómica para alterar significativamente su comportamiento.
- No se permite la evaluación o clasificación de individuos o grupos basándose en su comportamiento social o características personales, resultando en una puntuación ciudadana que provoca un trato perjudicial o desfavorable.
- Está prohibido realizar evaluaciones de riesgo de individuos con el fin de predecir la probabilidad de que cometan un delito penal, basándose únicamente en la elaboración de perfiles o la evaluación de rasgos y características de personalidad.
- No se puede crear o ampliar bases de datos de reconocimiento facial mediante la extracción indiscriminada de imágenes faciales de internet o de circuitos cerrados de televisión.
- No se pueden inferir en las emociones de una persona en lugares de trabajo y centros educativos, a menos que sea por motivos médicos o de seguridad.
- Está prohibido clasificar a las personas individualmente basándose en sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, orientación sexual o vida sexual.

- No se puede utilizar la identificación biométrica remota "en tiempo real" en espacios públicos con fines de aplicación de la ley, a menos que dicho uso sea estrictamente necesario para alcanzar uno o varios objetivos específicos, los cuales son la búsqueda de víctimas de delitos y de personas desaparecidas, prevención de amenazas a la vida y a la seguridad e identificación de sospechosos de delitos graves.

2.1.4. Sistemas de IA de alto riesgo

Según el artículo 6 (capítulo III, sección 1), un modelo se considerará de alto riesgo si cumple las siguientes dos condiciones:

- Si está diseñado para ser empleado como un elemento de seguridad en un producto que se encuentra bajo la jurisdicción de los actos legislativos de armonización de la Unión, o si el modelo en sí es uno de estos dispositivos.
- El producto, ya sea el sistema de inteligencia artificial en su totalidad o un elemento de seguridad dentro de él, debe someterse a una evaluación de conformidad llevada a cabo por una entidad independiente antes de su lanzamiento al mercado o su uso, en cumplimiento con las normativas legislativas de armonización de la Unión.

Esto aplica independientemente de si el dispositivo se ha introducido en el mercado o se ha puesto en servicio sin estar integrado en los elementos mencionados en las dos condiciones.

Además de estas dos condiciones, si consultamos el anexo III se enumeran una serie de condiciones extra por las cuales un modelo puede ser considerando de alto riesgo (figura 2.2):

Figura 2.2: Sistemas de alto riesgo del anexo III

- **Biometría:** Incluye sistemas de identificación biométrica remota, la categorización biométrica basada en atributos sensibles o protegidos, y el reconocimiento de emociones. No incluye sistemas de verificación biométrica cuyo único objetivo sea confirmar la identidad de una persona.
- **Infraestructuras críticas:** Incluye modelos utilizados como componentes de seguridad en la gestión y funcionamiento de infraestructuras digitales críticas, tráfico rodado, y suministro de agua, gas, calefacción o electricidad.
- **Educación y formación profesional:** Es aquella tecnología utilizada para determinar el acceso o admisión a centros educativos y de formación profesional, evaluar los resultados del aprendizaje, analizar el nivel de educación adecuado que recibirá una persona, y para el seguimiento y detección de comportamientos prohibidos durante los exámenes.
- **Empleo, gestión de trabajadores y acceso al autoempleo:** Se usan para la contratación o selección de personas, tomar decisiones que afecten a las condiciones laborales, asignación de tareas basadas en comportamientos individuales o rasgos personales, y

para supervisar y evaluar su rendimiento y comportamiento en el marco de relaciones laborales.

- **Acceso a servicios privados esenciales, y servicios y prestaciones públicos esenciales:** Sirven para evaluar la elegibilidad de individuos para recibir beneficios de asistencia pública y servicios de salud, así como para determinar su solvencia o calificación crediticia, con excepción de su uso en la detección de fraudes financieros. Además, se emplean en la evaluación de riesgos y fijación de precios en seguros de vida y salud, y en la clasificación de llamadas de emergencia para el envío prioritario de servicios de primera intervención en situaciones de crisis.
- **Aplicación de la ley:** Emplea diversas herramientas para evaluar el riesgo de que un ser humano sea víctima de un delito, así como la fiabilidad de las pruebas durante investigaciones o juicios. Estos métodos también se utilizan para determinar la probabilidad de que alguien cometa o reincida en un delito, así como para analizar rasgos de personalidad y comportamientos delictivos pasados, contribuyendo así a la elaboración de perfiles durante la detección, investigación o enjuiciamiento de crímenes.
- **Migración, asilo y gestión del control fronterizo:** Las autoridades públicas competentes usan instrumentos como polígrafos para evaluar diversos riesgos, como la seguridad, la salud o la migración irregular, asociados con individuos que intentan ingresar o ya han ingresado al territorio de un Estado miembro. Por ello, son empleadas para examinar solicitudes de asilo, visado o permiso de residencia, así como reclamaciones relacionadas, en el contexto de la migración, el asilo o la gestión del control fronterizo. Su aplicación tiene como objetivo detectar, reconocer o identificar a ciertas personas, excluyendo la verificación de documentos de viaje.
- **Administración de justicia y procesos democráticos:** Las herramientas de administración de justicia y procesos democráticos se utilizan para asistir a las autoridades judiciales en la investigación, interpretación y aplicación de la ley en casos específicos, así como en la resolución alternativa de disputas. Además, estas tienen un papel en influir en los resultados electorales o referendos, así como en el comportamiento de los votantes al ejercer su derecho de voto en dichos eventos. Sin embargo, se excluyen los modelos que no están directamente expuestos a individuos, como aquellas utilizadas para la organización, optimización o estructuración de campañas políticas desde una perspectiva administrativa o logística.

Hay que tener en cuenta que este anexo puede ser modificado por la Comisión para añadir

o modificar casos de uso de sistemas de IA de alto riesgo si estos están destinados a ser utilizados en ciertos ámbitos o si conllevan un riesgo significativo para la salud, la seguridad o los derechos fundamentales. Además de esto, también pueden eliminarlos de este estado si ya no representan un riesgo considerable para el bienestar integral.

2.1.5. Requisitos de los sistemas de IA de alto riesgo

Después de tener claro que sistemas son considerados de alto riesgo, es momento de analizar a que reglas y procedimientos deben estar sometidos[2]:

- **Sistema de gestión de riesgos:** Es un proceso cíclico al cual está sometido durante todo momento en que la IA este en funcionamiento. Este consta de 4 fases:
  - **Determinación y análisis:** Se identifican y estudian todos aquellos peligros que puede conllevar la utilización y puesta en servicio del respectivo sistema siendo usado de la forma contemplada inicialmente.
  - **Estimación y evaluación:** En esta fase se intenta prever no solo que riesgos pueden surgir usando el modelo de la forma intencionada, sino ir un paso más allá y ver que posibles riesgos pueden ocurrir si se utiliza con otras finalidades.
  - **Contemplación de otras vulnerabilidades:** Se analizan los impactos negativos a partir de un sistema de vigilancia poscomercialización, el cual está explicado en el artículo 72, pero de igual forma es tratado en este informe en el apartado "Vigilancia poscomercialización, intercambio de información y vigilancia del mercado".
  - **Adopción de medidas:** Se implementan las estrategias diseñadas para poder evitar todos los riesgos analizados en el primer apartado. Para determinarlas, se buscará eliminar o reducir todas las vulnerabilidades detectadas mientras sea técnicamente posible a través de un diseño y desarrollo adecuados. En el caso de que no sea posible eliminarlos en su totalidad, se implementarán medidas de control adecuadas, además de proporcionar la información necesaria y de impartir formación a todos los implicados en el despliegue.
- **Datos y gobernanza de datos:** Todos aquellos sistemas de alto riesgo que vayan a

[2]En la sección 2 del capítulo III se pueden encontrar todos aquellos requisitos que se han tratado en este apartado explicados de forma más profunda y extensa, recomendamos su lectura encarecidamente en caso de querer saber más sobre casos específicos.

ser entrenados con datos se deben de gestionar de forma adecuada teniendo en cuenta los siguientes puntos:

- Las elecciones adecuadas vinculadas al diseño del modelo.
- Se considerará el origen de esta información y, en el caso de que sea personal, la finalidad original de su recogida.
- Las operaciones adecuadas para su preparación, como anotación, etiquetado, depuración, actualización, enriquecimiento y agregación.
- Se formularán supuestos sobre la información que se supone que miden y representan.
- Se evaluará la disponibilidad, la cantidad y la adecuación de los conjuntos de elementos necesarios.
- Se examinarán posibles sesgos que puedan afectar a la salud y la seguridad de las personas, vulnerar derechos fundamentales u ocasionar algún tipo de discriminación prohibida por el Derecho de la Unión.
- Se tomarán medidas adecuadas para detectar, prevenir y reducir estos sesgos.
- Se detectarán lagunas o deficiencias que impidan el cumplimiento de la normativa, y, en caso de que las haya, se buscará la forma de subsanarlas.

En adición a esto, los conjuntos de datos deben ser pertinentes, representativos, libres de errores y completos en la medida de lo posible, teniendo en cuenta su uso final, además que deben de tener en cuenta las características o elementos particulares del entorno geográfico, contextual, conductual o funcional específico en el que se prevé utilizar. Si resulta imprescindible a la hora de corregir errores, los desarrolladores tienen la opción de manejar ciertos tipos específicos de datos personales en circunstancias excepcionales, siempre y cuando proporcionen las garantías necesarias para proteger los derechos y libertades fundamentales de las personas físicas.

- **Documentación técnica:** Todo sistema de IA de alto riesgo debe elaborar la documentación antes de que se ponga en marcha o se introduzca en el mercado y debe de asegurar que este cumple con toda normativa establecida, además de proporcionar a las autoridades y organismos notificados toda información vital para su evaluación [3]. Una vez este ya este en funcionamiento dentro de la UE, se elaborará una serie de textos especializados que contengan toda la información descrita anteriormente.

[3]Toda la documentación necesaria se explica en profundidad en el anexo IV del reglamento.

- **Conservación de registros:** Todo aquel modelo que implique un alto riesgo estará habilitado para registrar automáticamente eventos en todas las etapas de su funcionamiento. Para asegurar esto, las capacidades de registro posibilitarán la documentación de eventos relevantes para:

  - La detección de sucesos que hagan que este entorno represente una amenaza o una modificación sustancial.
  - El apoyo a la supervisión posterior a la comercialización (artículo 72, sección I, capítulo IX).
  - La supervisión del funcionamiento de estas tecnologías (apartado 6, artículo 26, sección 3, capítulo III).

  Estos eventos incluyen: un registro de la duración de cada utilización del sistema (registrando tanto el inicio como el fin de cada uso), la base de datos de comparación empleada para contrastar los datos de entrada, además de aquellos que han coincidido con la búsqueda, y la identificación de los individuos involucrados en la verificación de los resultados (tratado en este mismo apartado posteriormente).

- **Transparencia y comunicación de información a los responsables del despliegue:** Todas las arquitecturas de alto riesgo deben de diseñarse con la suficiente claridad para que toda aquella persona relacionada con el sistema pueda entender y usar correctamente la información de salida. Estos deberán tener unas instrucciones de uso para poder cumplir con lo mencionado anteriormente, las cuales deben contener información como la finalidad prevista del modelo y cualquier circunstancia que le puede llegar a afectar[4].

- **Vigilancia humana:** Todo modelo de IA de alto riesgo debe ser planteado de forma que pueda ser vigilado por personas físicas durante su funcionamiento, lo cual incluye proporcionarles las herramientas adecuadas para cumplir con este propósito. El objetivo de realizar esto es la reducción de todo posible riesgo que pueda ocurrir en el ciclo de vida del sistema y las estrategias de supervisión deben ajustarse en proporción a sus riesgos, grado de autonomía y entorno de uso.

  Es de vital importancia que todo aquel responsable de la vigilancia pueda comprender correctamente las capacidades y restricciones del sistema, entiendan plenamente toda la

[4]Si se desea conocer en mayor profundidad el contenido de estas instrucciones, consulta el apartado 3 del artículo 13, sección 2, capítulo III.

información de salida y tengan la capacidad de intervenir o detener su funcionamiento de manera segura.

- **Precisión, solidez y ciberseguridad:** Primero, gracias a la colaboración de la Comisión y de las organizaciones pertinentes, se establecerán una serie de directrices para determinar cómo se pueden medir y evaluar de manera efectiva la precisión y la solidez de estos modelos, lo que se incluirá en las instrucciones de uso mencionadas en apartados previos.

  Una vez hecha la medición, se debe asegurar que este salga al mercado con la mayor resistencia a errores y fallos posible. Además se debe asegurar que todo aquel sistema que siga aprendiendo y mejorando tras su puesta en marcha se desarrolle de forma que, en la medida de lo posible, se elimine o reduzca el riesgo de que la información de salida que pueda estar sesgada influya en la información de entrada de proyectos futuros.

  Por último, estos deberán ser inmunes a cualquier intento de individuos no autorizados para modificar su utilización, manipular la información que generan o afectar su funcionamiento mediante la explotación de posibles debilidades. Para lograrlo, es importante seguir una serie de medidas que eviten y prevengan todo aquel intento malicioso que conlleve a su corrupción.

2.1.6. Obligaciones de los proveedores y responsables del despliegue de sistemas de IA de alto riesgo y de otras partes

Una vez explicados todos los requisitos que debe cumplir todo modelo de alto riesgo, es imprescindible que todas aquellas personas nombradas en este subtítulo cumplan con ellas de forma correcta. En este apartado se tratan todos los artículos en relación a esto, para así garantizar el cumplimiento de las normas anteriormente mencionadas[5].

Primero que nada, es obligatorio que esta arquitectura sea administrada por un sistema de gestión de calidad (como se puede observar en la figura 2.3) para garantizar el cumplimiento de los requisitos, el cual deberá contener los siguientes aspectos:

- La metodología que se va a usar para el cumplimiento de la normativa.
- Métodos y protocolos sistemáticos para el diseño, supervisión, verificación y desarrollo.

[5]Todos estos artículos se encuentran en la sección 3 del capítulo III. En caso de querer saber más al respecto, recomendamos su lectura.

- Procedimientos de revisión, prueba y validación que se llevarán a cabo antes, durante y después del desarrollo.
- Normas técnicas y estándares.
- Sistemas y pautas para la gestión de datos.
- El sistema de gestión de riesgos explicado en el anterior apartado.
- Una arquitectura de seguimiento post-venta.
- Recursos para informar de un incidente grave.
- Administración de la comunicación con las autoridades competentes, otros operadores, clientes y otras partes interesadas.
- Estructuras y rutinas para mantener un registro de toda la documentación e información relevante.
- Organización de recursos, incluyendo medidas de seguridad del suministro.
- Un marco de responsabilidad que defina las obligaciones del personal.

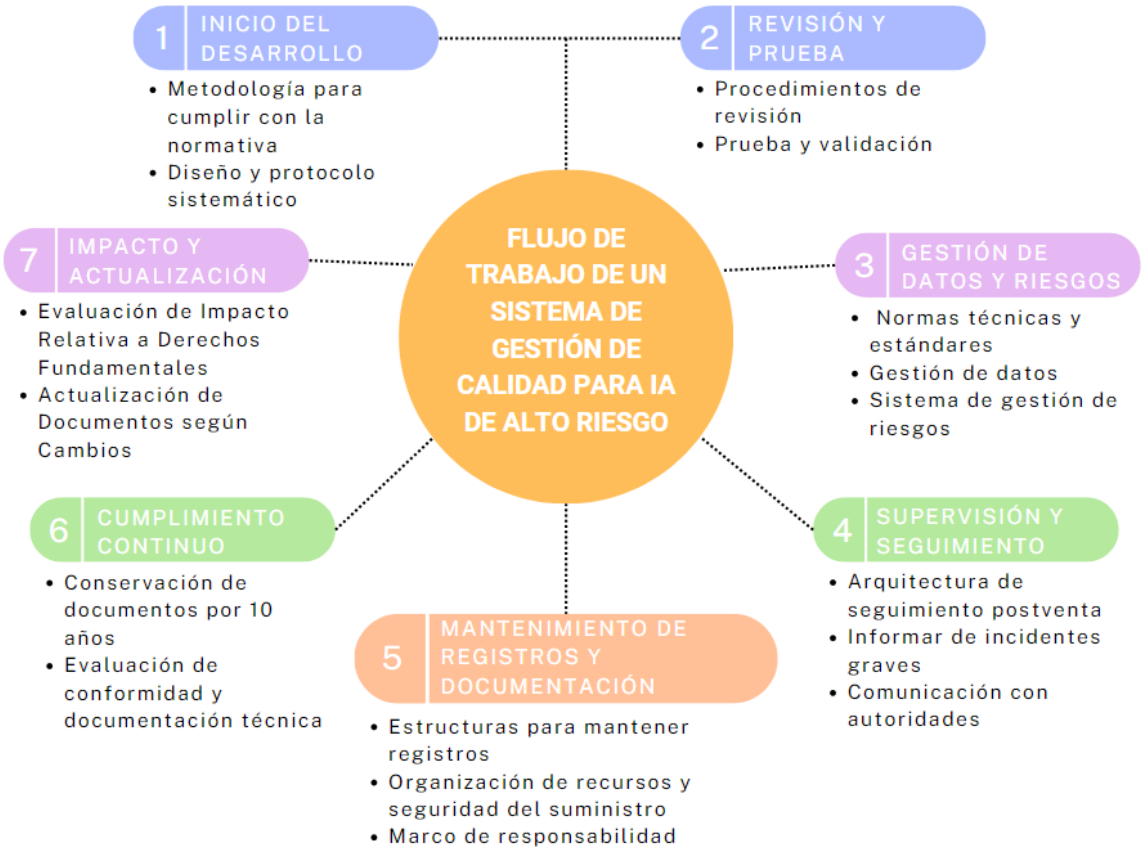

Figura 2.3: Flujo de trabajo de un sistema de gestión de calidad para IA de alto riesgo

Durante una década desde que se ha empezado a usar el modelo en el mercado, el proveedor está en la obligación de conservar una serie de documentos, entre los cuales están la documentación técnica del artículo 11, los que se refieren al sistema de gestión de calidad previamente mencionado,la declaración UE de conformidad del artículo 47 y los que tratan los cambios aceptados, decisiones y otros registros creados por los organismos notificados. Dependiendo de en que región de la UE se encuentre registrado, se establecerán una serie de condiciones para mantener esta documentación a buen recaudo, especialmente en casos donde la empresa responsable del despliegue del sistema haya cerrado o este en quiebra, además de que toda aquella que este relacionada con los requisitos descritos en la sección 2 del capítulo III deberá ser proporcionada a las autoridades competentes.

En caso de que los distribuidores tenga la creencia firme de que una arquitectura que se encuentra actualmente en el mercado no cumpla con las normativas previamente mencionadas,

deben tomar las medidas necesarias para que así sea, o en su defecto se puede retirar o desactivar, avisando a todos los responsables de este sobre la decisión tomada. En el caso de que presente un riesgo significativo a la salud, seguridad o derechos fundamentales, se debe analizar de inmediato las razones por las que está ocurriendo junto al responsable del despliegue que lo ha comunicado y notificará a los diferentes países en los que se encuentra en uso.

En caso de que se quiera lanzar esta tecnología desde un tercer país no perteneciente a la UE, se deberá nombrar a un representante autorizado en esta región antes de la puesta en marcha del servicio. Este será el responsable de cumplir con las mismas obligaciones a las que están sometidas las empresas de dentro de la Unión Europea, y en caso de que este no cumpla con ello, se debe terminar su mandato y notificar de ello a las autoridades y Estados miembros en los que se encuentra el producto.

A continuación, se va a explicar en profundidad todas las obligaciones a las que están sometidas las personas relacionadas con la puesta en marcha de un modelo de alto riesgo:

- **Obligaciones de los importadores y distribuidores:** En primer lugar, deben verificar que se cumpla el reglamento en su totalidad, se hayan realizado las evaluaciones de conformidad, preparado la documentación técnica y designado un representante autorizado. En caso de que no se cumpla alguno de los puntos mencionados previamente, ambos tienen que evitar introducir el sistema en el mercado e informar de inmediato a los responsables de este y a las autoridades competentes.

  En adición a esto, también han de indicar todos sus datos en el embalaje (solo es requerido para los importadores), así como garantizar que el transporte sea seguro y no comprometa la integridad, además de conservar toda la documentación necesaria y entregarla a las entidades pertinentes.

  Es de vital importancia recalcar que estos deben en todo momento cooperar con las autoridades nacionales para evitar y disminuir cualquier riesgo que pueda ocurrir en la UE por la importación de una tecnología potencialmente peligrosa.

- **Responsabilidades a lo largo de la cadena de valor:** Cualquier persona involucrada en la puesta en marcha del sistema de IA será considerada un proveedor cuando ponga su nombre o marca en uno ya introducido en el mercado, cuando lo modifique considerablemente o cuando provoque un cambio en la finalidad en uno que no era de alto riesgo, haciendo que ahora si lo presente. En el caso de que ocurra una de las tres

situaciones anteriores, el suministrador inicial dejará de ser considerado como tal y se delegará esa función en quienes la provocaron.

El proveedor y cualquier tercero que suministre dicho modelo deben detallar por escrito qué información y apoyo se requieren para que este pueda cumplir totalmente con las responsabilidades establecidas en el reglamento.

- **Obligaciones de los responsables del despliegue:** Deben adoptar una serie de medidas para garantizar que la tecnología se este usando acorde a las instrucciones de uso. Para ello, tienen que asegurarse de que haya una correcta supervisión humana y que los datos de entrada sean los adecuados considerando su finalidad, además de que están en la obligación de mantener todo archivo generado automáticamente en un plazo de seis meses. En adición a esto, antes de que se ponga en uso en un lugar de trabajo, tendrán que avisar a todo implicado de esta área de que estarán expuestos a su utilización.

  En caso de que el sistema presente un peligro considerable, estos han de asegurarse de informar a los proveedores del mismo, y dependiendo de la gravedad del riesgo, se podría llegar a suspender el funcionamiento de este. Si se trata de una entidad financiera, habrán cumplido con la vigilancia cuando sigan correctamente las regulaciones sobre cómo se gestionan, los sistemas y procesos internos, en línea con las leyes aplicables en el ámbito de los servicios financieros.

  En la situación de autoridades públicas o instituciones, deberán también cumplir con las normas de registro y solo podrán usar el sistema en caso de que este registrado en la base de datos.

  Si se está realizando una investigación con identificación biométrica en diferido, se deberá pedir una autorización, además de tener en cuenta que su uso solo se limitaría a la propia infracción. Por último, si estos sistemas involucran a personas físicas siempre se les deberá notificar de ello y absolutamente en cualquier situación deberán cooperar con las autoridades competentes para el cumplimiento de la normativa.

Para finalizar este apartado, es de vital importancia saber que antes de poner en funcionamiento cualquier modelo de inteligencia artificial, se deberá hacer una evaluación de impacto relativa a los derechos fundamentales. Ese documento constará de las siguientes partes: descripción de los casos donde se va a utilizar, periodo de tiempo y frecuencia, las categorías y riesgos de las personas que se verán involucrados en su uso, explicar como se supervisará y todas las pautas que se seguirán en caso de que los riesgos planteados se vuelvan realidad.

Este registro solo se debe de realizar en la primera puesta en servicio de la tecnología, por lo que en futuros casos donde el impacto sea muy similar se podrá utilizar el mismo. Al mismo tiempo, si estos cambian, se deberá de modificar la información para que sea acorde.

2.1.7. Normas, registros, certificados y evaluación de conformidad

Todo sistema de inteligencia artificial de alto riesgo que cumpla con las normas armonizadas (publicadas en el Diario Oficial de la Unión Europea de conformidad con el Reglamento (UE) n.º 1025/2012[8]) se considerará que está cumpliendo con todos los requisitos que se les requiere a estos modelos (descritos anteriormente). Además de esto, la Comisión solicitará unas peticiones de normalización para que se cubran todas las obligaciones de este, las cuales incluirán algunas solicitudes de mejora, como por ejemplo reducción del consumo de energía. Cuando dirija esta petición, deberá especificar cuales de los requisitos deben ser fáciles de entender y consistentes en su contexto, además de que solicitará una serie de pruebas para comprobar que todo lo nombrado anteriormente se haya abordado y cumplido.

La Comisión tiene la facultad de tomar medidas para establecer una serie de especificaciones comunes en caso de que se cumplan una serie de condiciones, como que las normas armonizadas existentes no cumplan con los posibles riesgos a los derechos fundamentales, o que las organizaciones no acepten la solicitud para elaborar una norma armonizada, entre otras. Previamente a elaborar este acto de ejecución, se debe contemplar si se cumplen con las condiciones necesarias, y en caso de que un modelo de IA cumpla con las especificaciones comunes se distará que es conforme con lo establecido. Si no se satisfacen estas características, se debe explicar debidamente que se han realizado una serie de acciones para el cumplimiento de todos los requisitos.

A la hora de realizar la evaluación de conformidad, si la tecnología forma parte del ámbito de la biometría, debe optar por dos procedimientos: el fundamentado en el control interno o el fundamentado en la evaluación del sistema de gestión de la calidad y la evaluación de la documentación técnica[6]. Esto dependerá de si se han aplicado las normas y especificaciones necesarias, además de otra serie de condiciones.

En el caso de que forme parte del resto de áreas mencionadas previamente como de alta vulnerabilidad, solo tendrán que acatar el procedimiento fundamentado en el control interno, en el cual no participa ningún organismo notificado. Si ya han sido sometidos por actos

[6]Estos dos procedimientos están mencionado en el anexo VI y VII del Reglamento.

legislativos de armonización de la Unión (incluyendo también biometría), deberá continuar con el proceso de evaluación descrito en dicho acto. Cabe destacar que si cualquier arquitectura que ya haya sido evaluada de conformidad ha sufrido un gran cambio se deberá de volver a someter a esta (no se tienen en cuenta los entornos de aprendizaje continuo ni si este ha sufrido cambios que ya estaban planeados desde un inicio y, por ello, notificados con anterioridad).

Una vez completado este protocolo se entregará un certificado por parte de los organismos notificados, el cual tendrá una duración máxima de cuatro años para todo aquel sistema mencionado en el anexo III y de cinco para el resto, pudiéndose duplicar en caso de que el proveedor lo solicite (deberá de someterse a una nueva evaluación). Al igual que con otras normas mencionadas anteriormente, si esta organización detecta que este ya no cumple con la normativa, se le retirará este certificado o, en su defecto, se le impondrán una serie de restricciones.

Es importante recalcar que hay un caso por el cual esta evaluación no sería necesaria de realizar, que es si un modelo tiene como objetivo la seguridad pública o la protección de las vidas y salud de las personas, el medio ambiente o recursos clave de la industria e infraestructuras. Esta autorización será temporal y con el paso del tiempo deberá someterse a la evaluación de igual forma.

En cualquier situación de extrema necesidad por casos de seguridad pública o en caso de una amenaza inminente y específica para el bienestar de los seres humanos, se podría poner en marcha la IA sin la autorización mencionada anteriormente, siempre que se solicite durante o después de su uso, y si se deniega, se deberá de suspender su uso inmediatamente.

En adición a lo anteriormente mencionado, el proveedor deberá redactar una declaración UE de conformidad para cada sistema de alto riesgo y tendrá que mantenerlo a disposición de las autoridades correspondientes como mínimo durante diez años desde su puesta en marcha. Todo el contenido que consta este escrito está explicado con detenimiento en el anexo V del reglamento y se tendrá que escribir en una lengua legible a la correspondiente a los Estados miembros donde se utilizará.

Esta tecnología específica también necesita obtener un certificado CE, el cual debe cumplir con los requisitos establecidos en el artículo 30 del Reglamento (CE) n.º 765/2008[9]. Si este es digital se usará un tipo de marcado digital siempre que sea de fácil acceso a través de su interfaz o código y en cualquier caso se debe colocar de forma visible, legible e imborrable, además de ir acompañado del número de identificación del organismo notificado.

Por último, es crucial subrayar que todo sistema de IA, sea o no de alto riesgo (exceptuando el referente a infraestructuras críticas), deberá ser registrado en la base de datos de la Unión Europea por parte del proveedor. Si este se usa en el ámbito de la biometría, aplicación de la ley y en la migración, asilo y gestión del control fronterizo, se registrará en una sección segura de este repositorio de información, de forma que permanecerá oculta para todo el mundo exceptuando la Comisión y las autoridades nacionales. En referencia a los usados en el área de las infraestructuras críticas, solo se deberán registrar a nivel nacional.

2.1.8. Requerimientos sobre transparencia a los proveedores y responsables del despliegue

Todos los individuos mencionados en este subapartado deben de informar a todas las personas físicas involucradas en el uso de un sistema de inteligencia artificial sobre que están interactuando con este. Esto no será necesario cuando resulte evidente la interacción o cuando la tecnología este diseñada para prevenir y evitar delitos.

Además de esto, también deben asegurarse de que todo dato de salida debe ser perfectamente legible, interpretable y marcado como generado o manipulado de forma artificial. Esta medida en concreto tampoco se aplicará a los modelos con fines penales, además de aquellos que realicen una función de apoyo a la edición estándar o que sus datos de entrada no sean prácticamente alterados.

Por último, es importante que estos profesionales también avisen de cuando su programa a generado o modificado contenido ultrafasificado, es decir, vídeos, grabaciones o imágenes falsas donde participan personas aparentemente reales. Esta medida se verá atenuada en el caso de que se use para fines creativos, cómicos o de ficción, informando únicamente de la existencia de este tipo de contenido, y no se aplicará en los utilizados en ámbitos delictivos y en todo aquel que haya sido revisado por una persona o editorial y que un profesional tenga la responsabilidad editorial del mismo.

2.1.9. Sistemas de inteligencia artificial de uso general

Antes de tratar los modelos definidos como de uso general, es importante describir cuando este es catalogado con riesgo sistémico. Esto se da cuando se considera que tiene una gran capacidad de provocar un impacto relevante medido, o bien a través de técnicas y herramientas

adecuadas, o a través de la decisión de la Comisión teniendo en cuenta los siguientes factores[7]:

- La cantidad de parámetros.
- La calidad y cantidad del conjunto de datos.
- El nivel de cálculo utilizado en el entrenamiento, en concreto, si este supera los 10^{25} FLOPs (operaciones de coma flotante por segundo) generalmente será ya calificado de este tipo.
- La cantidad de gente registrada.
- La forma en la que interactúa con los datos, es decir, propiedades como el manejo de diferentes tipos de información.
- Los parámetros de referencia que se hayan usado en su evaluación.
- El impacto que puede llegar a producir en el mercado interior, lo que se dará como cumplido si la cantidad de profesionales ubicados en la Unión Europea supera los 10.000.

Una vez detectado, en las posteriores dos semanas, los proveedores están en la obligación de comunicarlo a la Comisión aportando toda la documentación necesaria confirmando que se cumple este requisito. En el caso de que esto no se notificará, el organismo lo calificará como de riesgo sistémico por su propia cuenta (todo este proceso se puede observar en a figura 2.4).

Figura 2.4: Diagrama de flujo del proceso de clasificación y notificación de riesgo sistémico

A continuación, se describirá a que obligaciones están sometidos este tipo de sistemas:

[7]Todas estas características están descritas con mayor profundidad en el anexo XIII del Reglamento.

Obligaciones de la IA de uso general

Los distribuidores deben de realizar y mantener toda la documentación técnica[8] (incluyendo la información relevante sobre los procesos de prueba, entrenamiento y evaluación), además de proporcionar la información e informes pertinentes a otros profesionales que vayan a trabajar con su tecnología, para que así estos puedan entender sus capacidades y limitaciones para proceder con el cumplimiento de la normativa. En adición a esto, también tendrán que establecer una serie de directrices para acatar con la legislación de la UE y entregarán públicamente un resumen con todo lujo de detalles sobre el contenido usado en la fase de entrenamiento.

Estas obligaciones no se aplicarán a los modelos que se divulguen bajo código y licencia abierta, es decir, que se admita su acceso, alteración, uso y distribución además de que todos los parámetros y datos referentes a su arquitectura se compartan de forma colectiva.

Al igual que ocurre con los sistemas de alto riesgo, en este tipo si no están ubicados en la Unión Europea pero desean introducirse aquí, deberán nombrar a un representante que este establecido en esta zona. Este debe efectuar todas las obligaciones descritas en los sistemas de gran vulnerabilidad pero aplicándoles los requisitos pertinentes a estos modelos, teniendo en cuenta también si presentan un riesgo sistémico o no.

Obligaciones IA de uso general con riesgo sistémico

Este tipo de tecnología, aparte de tener que cumplir con las mismas obligaciones del apartado anterior, deberán de acatar las siguientes:

- Se debe realizar una evaluación siguiendo los protocolos y normas adecuados, lo que implica realizar pruebas de simulación de adversarios para identificar y disminuir cualquier peligro.
- Los proveedores están obligados a analizar y mitigar todos los riesgos que procedan de su desarrollo, puesta en marcha o utilización.
- Se tiene que vigilar, notificar y documentar a la Oficina de la IA cualquier riesgo o amenaza que haya surgido para proceder con su resolución.

[8] Si se desea saber en profundidad más acerca de esta documentación, se ubica en el anexo XI del Reglamento.

- Es preciso que se hayan cumplido con las medidas referentes a la ciberseguridad necesarias.

Códigos de buenas prácticas

Para poder cumplir con estas obligaciones, la Oficina de IA realizará una serie de documentos de buenas prácticas, las cuales incluyen medidas como medios para mantener actualizada toda la información respecto a la evolución del mercado y de la tecnología, el nivel de detalle que deben tener los datos utilizados en el entrenamiento y la identificación del tipo y naturaleza de los riesgos sistémicos, además de establecer una serie de procesos y medidas de evaluación y gestión de estos.

La Oficina de la IA, con ayuda del Comité, se asegurará de de que se cumplan con todos los requisitos descritos en este documento junto a revisar que este cumpla con las normas previamente explicadas. Además, deben garantizar que todos los participantes le informen periódicamente sobre su aplicación.

2.1.10. Supervisión post-venta, intercambio de información y monitorización del mercado

Los proveedores deberán crear un sistema de vigilancia que supervise el sistema de inteligencia artificial de alto riesgo una vez se haya puesto en marcha en el mercado, el cual se encargará de esto durante toda su vida útil. Esto se hace para evitar lo antes posible cualquier riesgo o amenaza que pueda surgir durante su utilización, además de que, en la medida de lo posible, también se controlará toda interacción con otros modelos de IA.

En caso de que ocurra un incidente grave, los distribuidores están en la obligación de comunicarlo a las autoridades pertinentes de cada Estado en el que se ubique. Esta notificación se tiene que hacer de forma inmediata una vez se identifique una relación entre el accidente y la tecnología, con un plazo máximo de quince días desde que se descubrió dicha conexión (exceptuando que ocurra en un espacio de acceso público, lo cual solo tendrá un plazo máximo de dos días, y en caso de fallecimiento, que tendrá un plazo de diez días). Una vez transmitido el aviso, se realizará una investigación adecuada junto a las autoridades y organismos correspondientes que establezca de forma más clara y precisa que impacto ha tenido el sistema con el incidente.

Todos los modelos de IA bajo los que actúa esta normativa se verán regulados por el Reglamento (UE) 2019/1020 [10], el cual se centra en la vigilancia del mercado y en asegurar que los productos cumplan con sus obligaciones. Además de esto, establece una serie de mecanismos para monitorear y garantizar que estos cumplan con todas las normativas de seguridad y calidad, asigna diferentes entidades responsables de estos procedimientos y define una estructura de vulnerabilidades y riesgos específicos.

Es importante recalcar que todo individuo u organización que participe en el cumplimiento de este reglamento deben de respetar la confidencialidad de la información y datos obtenidos. En concreto se deben respetar los derechos de propiedad intelectual e industrial, datos empresariales, secretos comerciales, la aplicación efectiva del Reglamento, los intereses de seguridad pública y nacional, el desarrollo de causas penales o procedimientos administrativos y la información clasificada.

2.1.11. Códigos de conducta y directrices

La Oficina de IA y los Estados miembros promoverán e impulsarán la elaboración de una serie de códigos de conducta para la aplicación voluntaria de los requisitos de los sistemas de alto riesgo para aquellos que no presentan el mismo peligro. Estos podrán llegar a ser desarrollados tanto por los proveedores como por las organizaciones que los representen, pudiendo colaborar con una gran cantidad de profesionales. Algunos elementos que se deberían incluir en este protocolo son:

- Las directrices éticas para una IA fiable[11].
- Un análisis del impacto y repercusión que pueda llegar a tener en el medio ambiente y, si es posible, minimizarlo para que sea lo más reducido posible.
- La promoción de la instrucción de todos los profesionales involucrados.
- La creación de un diseño inclusivo, es decir, que pueda llegar a ser utilizado por todo tipo de individuos independientemente de su condición.
- La evaluación y prevención de perjuicios hacia personas vulnerables o que por su condición física o mental pueden llegar a ser tratadas de forma distinta por el modelo.

En cuanto a las directrices sobre la aplicación práctica, las realizará la Comisión sobre estas secciones:

- La aplicación de los requisitos de los sistemas de alto riesgo y las responsabilidades a lo largo de la cadena de valor.
- Las prácticas prohibidas y obligaciones de transparencia descritas anteriormente.
- El manejo de los cambios importantes de este tipo de tecnología.
- La relación que tiene este documento con otras leyes de la UE.
- La definición de un sistema de inteligencia artificial proporcionada por la Unión Europea.

2.1.12. Sanciones

En caso de que no se cumplan con los requisitos y normas del presente reglamento, se le aplicará a las personas responsables una sanción adecuada a la infracción que han cometido. En caso de que no se respeten las prácticas de IA prohibidas referentes al artículo 5 (capítulo II) serán multados con hasta 35.000.000 de euros o, en caso de ser una empresa, hasta el 7 % del volumen de negocios mundial que posea, y en caso de que el incumplimiento sea distinto a los explicados en el artículo previo, llegará hasta 15.000.000 de euros o, si es empresa, el 3 % del volumen. Si se da la situación de que se entregase información errónea, se castigará con una multa de hasta 7.500.000 euros o, en caso de compañía o negocio, el 1 % de la facturación global.

Si el sistema es de uso general se le aplicarán sanciones económicas de hasta 15.000.000 de euros o el 3 % de los ingresos globales en caso de que incumpla alguno de los siguientes puntos:

- Han infringido alguna de las normativas que se les aplica a este tipo de tecnologías.
- No han entregado la documentación necesaria, o esta se ha realizado de forma inexacta o incompleta.
- Han incumplido alguna obligación solicitada por la Comisión.
- No han dado acceso al modelo a la Comisión para que pueda proceder a su evaluación.

2.1.13. Entrada en vigor

Para finalizar, es muy importante saber cuando se va a empezar a utilizar este reglamento, el cual entrará en vigor 20 días después de su publicación en el Diario Oficial de la UE y será aplicable dentro de dos años (concretamente, el 2 de agosto de 2026). A pesar de esto, hay una serie de excepciones que tienen un diferente rango de implementación respecto a la entra en vigor, las cuales son:

- El capítulo de disposiciones generales y las prácticas prohibidas serán aplicables seis meses después (2 de febrero de 2025).
- La sección referente a las autoridades notificantes y organismos notificados, modelos de IA de uso general, gobernanza y sanciones se ejercerán 12 meses después (2 de agosto de 2025).
- Las reglas por las cuales un sistema es clasificado como de alto riesgo y sus respectivas obligaciones se implementarán 36 meses después (2 de agosto de 2027).
- Los sistemas que sean componentes de otros modelos informáticos de mayor amplitud que se hayan puesto en marcha antes de 36 meses de la entrada en vigor deberán adecuarse al contenido del reglamento antes del 31 de diciembre de 2030.
- Toda aquella tecnología de uso general publicada 12 meses antes de la entrada en vigor deberá cumplir con las obligaciones 36 meses después (2 de agosto de 2027).

2.2 Principios éticos para una inteligencia artificial segura

Una vez explicado en detenimiento el nuevo reglamento sobre el que se deben regir los sistemas de IA en la Unión Europea, es crucial estudiar detenidamente otro documento: "Directrices éticas para una IA fiable", publicado por la Comisión Europea el 8 de abril de 2019 y elaborado por el Grupo de Expertos de Alto Nivel en Inteligencia Artificial[11]. Este busca incentivar a toda persona involucrada en el desarrollo de un modelo de inteligencia artificial para que esta sea fiable, es decir, que se sigan los principios legales, éticos y de robustez.

En concreto, aquí se tratan en profundidad solo los 2 últimos principios explicados anteriormente, ya que todo el tema referente a leyes se ha abordado en el apartado anterior. Para

poder organizarnos de forma sencilla, se dividirá esta sección en tres apartados, utilizados de forma similar a como se hacen en el propio informe (véase la figura 2.5 para poder ver detenidamente su estructura).

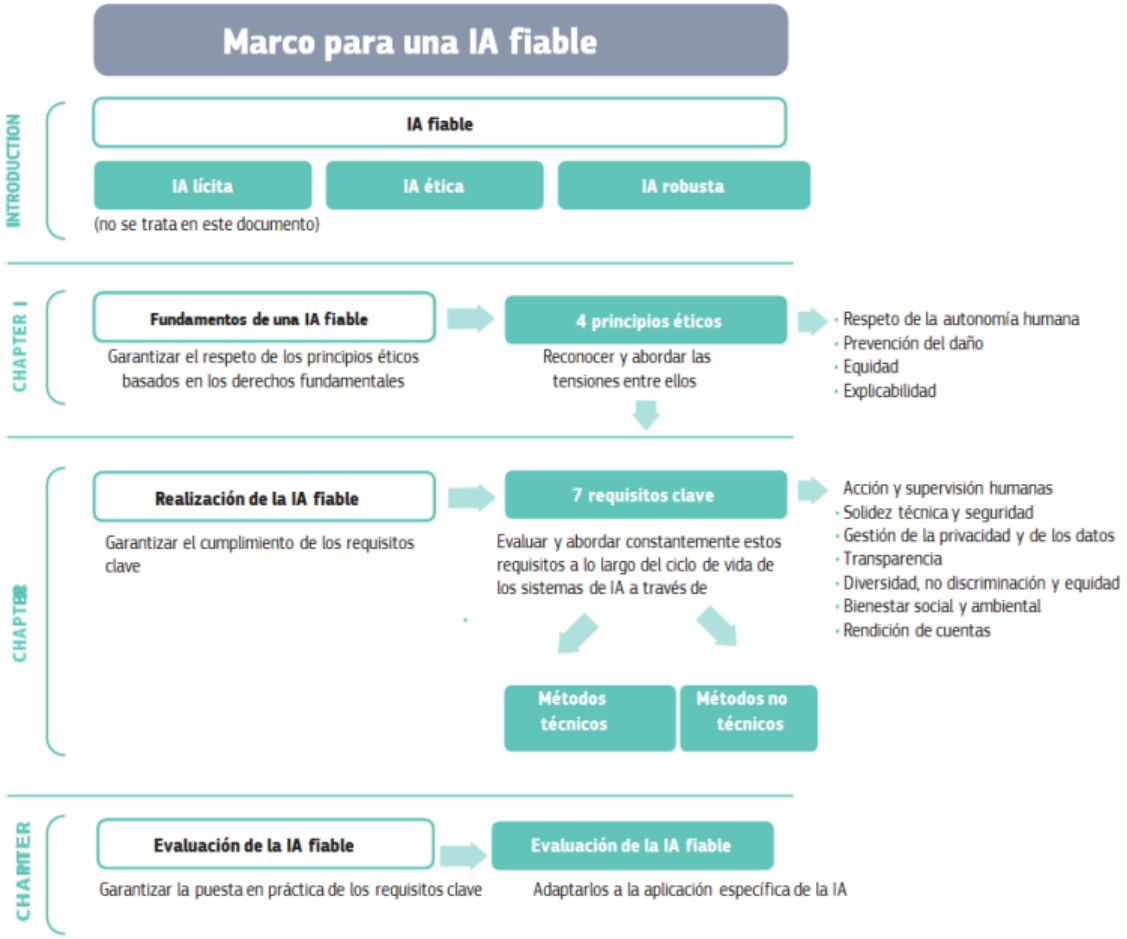

Figura 2.5: Estructura del documento "Directrices éticas para una IA fiable"

2.2.1. Principios éticos y orientaciones para el desarrollo

Primero que nada, es importante estudiar que derechos fundamentales nos proporciona la UE y, entre estos, cuales de ellos están relacionados con el uso de inteligencia artificial:

- **Respeto de la dignidad humana:** Todo ser humano tiene un "valor" que nunca se

debe de disminuir ni reprimir bajo ningún concepto. En este ámbito, todas las personas deben ser tratadas con el respeto que se debe, y no como simples objetos o herramientas que se pueden dirigir o manipular. Por ello estos sistemas se deben desarrollar para servirnos siempre, de forma que nunca estén por encima de nosotros.

- **Libertad individual:** Cualquier persona debe ser libre de tomar una decisión vital por sí mismo, es decir, una tecnología jamás debe interferir en su capacidad de elección y en su autonomía, sobretodo teniendo en cuenta cualquier individuo en riesgo que pueda llegar a ser manipulado con facilidad.

- **Respeto de la democracia, la justicia y el estado de Derecho:** Ninguna autoridad puede actuar arbitrariamente, es decir, todos sus actos deben de tener una base legal y restringirse según las normas y principios establecidos por la legislación. En caso de que la IA se use en el sector de la política, siempre se debe usar para favorecer los procesos involucrados en la democracia y no deben influir en el voto de los seres humanos.

- **Igualdad, no discriminación y solidaridad:** Toda persona tiene los mismos derechos y el mismo "valor" independientemente de su raza, género, orientación sexual, religión, nacionalidad, edad, discapacidad u otras características personales. En este contexto, esto implica que los modelos no deben generar resultados sesgados y siempre deben promover el respeto independientemente de las condiciones personales.

- **Derechos de los ciudadanos:** Debido al alto potencial que tiene la IA de mejorar la eficacia del gobierno en la entrega de servicios y bienes públicos, es importante evitar que vulneren los derechos de todo ser humano[9].

A partir de esto se han desarrollado cuatro principios éticos que deben cumplirse para el desarrollo de una IA fiable:

- **Respeto de la autonomía humana:** Toda persona que use el sistema está en la obligación de conservar plena autonomía y poder participar activamente en la democracia, además de no ser coaccionada, sometida, engañada y manipulada por este. En su lugar, se tiene que diseñar para que haya más facilidad y conocimiento a la hora de la toma de decisiones, lo cual requiere que esta tecnología sea supervisada para evitar poner en

[9]Es importante recalcar que a lo que se refiere el documento que estamos analizando por "derechos de las personas" no involucra únicamente a aquellas pertenecientes a la Unión Europea, sino también a los que se ubican en terceros países o los que se encuentran ilegalmente en esta región.

riesgo lo anteriormente descrito. En el ámbito laboral, debe de apoyar en las diferentes tareas y siempre favorecer a la creación de puestos de empleo, nunca disminuirlos.

- **Prevención del daño:** Nunca deben provocar daños, agravar los existentes y perjudicar de otras maneras a los individuos, lo que implica que se debe de proteger la dignidad humana y la integridad física y mental. En adición a esto, deben ser a prueba de fallos, no deben poder usarse nunca con fines malintencionados y deben prestar mayor atención a los seres humanos más vulnerables. Asimismo, no se ha de tener en cuenta solo a nosotros, sino también al medio ambiente y a todos los seres vivos.

- **Equidad:** Se debe de asegurar una distribución justa y equitativa de los costes y beneficios, y asegurar que nadie sufra un sesgo ni discriminación por alguna característica personal. Esto incluye fomentar uniformemente el acceso a la educación, la tecnología y los bienes de los servicios. La equidad no solo se refiere a esto, sino también a que los profesionales en su entorno deben entender que los métodos elegidos deben estar en proporción con el propio objetivo y deben equilibrar los diferentes intereses y metas que estén en juego. Por último, todos los individuos se pueden oponer a las decisiones tomadas por el sistema o por las personas que lo manejan, pudiendo identificar a los responsables y pedir una explicación sobre los procesos utilizados.

- **Explicabilidad:** Es crucial que los modelos sean transparentes, o dicho de otra forma, que todos los usuarios involucrados en su uso entiendan perfectamente sus capacidades, finalidad y decisiones. En los casos en los que no se pueda explicar detalladamente alguno de los puntos previos, será necesario adoptar otras medidas, como la trazabilidad, verificación y comunicación transparente.

Para finalizar este apartado, cabe destacar que hay apartados donde entran en conflicto estos puntos, por ejemplo, hasta que nivel un sistema de inteligencia artificial se puede usar para prevenir la delincuencia, porque por un lado puedes estar vulnerando la libertad de las personas y por otro no estás previniendo daños que podrían llegar a ocurrir, por lo que estos se deben tomar como una orientación y no como una solución concreta.

2.2.2. Requisitos

Todos los principios anteriormente nombrados deben de traducirse en una serie de requisitos (mostrados en la figura 2.6) para garantizar una IA fiable, los cuales están dirigidos a

los desarrolladores, responsables del despliegue y usuarios finales:

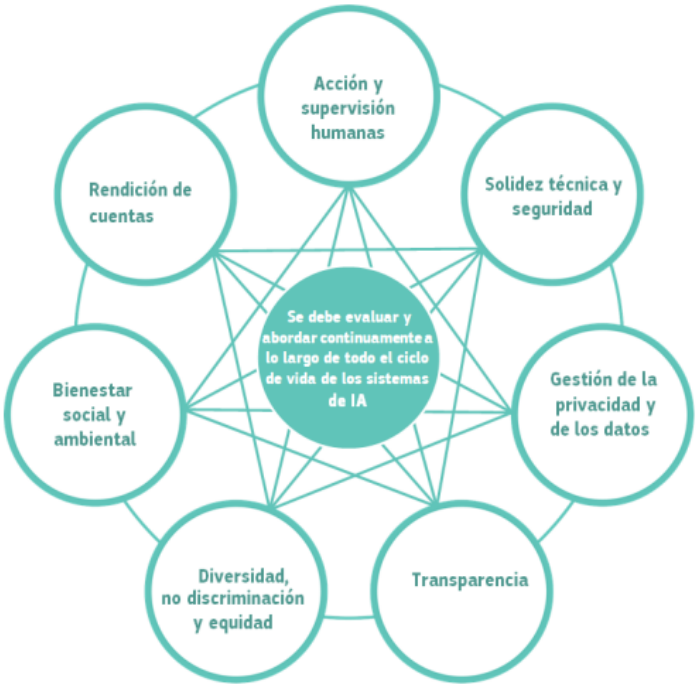

Figura 2.6: Requisitos de una IA fiable

- **Acción y supervisión humanas:** Esta tecnología puede afectar tanto positiva como negativamente a los derechos fundamentales, por lo que es de vital importancia realizar una evaluación sobre el impacto que puede tener hacía estos antes de su desarrollo, incluyendo una valoración de como reducir o justificar estos riesgos.

  A veces pueden influir en la toma de decisiones de las personas, por lo que es importante que el usuario reciba la información y las herramientas necesarias para interactuar de forma adecuada, además de facilitar las elecciones que hagan y que estás estén coordinadas con sus objetivos y pensamientos. Por ello, el principio de autonomía humana debe ser un pilar central del sistema, garantizando que los usuarios no sean controlados por decisiones automáticas que puedan tener un impacto significativo en sus vidas.

  Para poder evitar que esto ocurra se deben de supervisar estos dispositivos con humanos, lo que se llevará a cabo con mecanismos de gobernanza, es decir, haciendo que estos participen en todos los ciclos de decisión del modelo, intervengan durante el diseño y seguimiento, y vigilen y decidan cómo y cuándo utilizarlo.

- **Solidez técnica y seguridad:** Este requisito hace referencia al principio de prevención del daño, ya que estos sistemas se deben desarrollar de tal forma que se eviten la mayor cantidad de riesgos posibles. Por ello es crucial que este tenga un mecanismo de seguridad perfectamente diseñado, para así poder evitar que agentes exteriores manipulen cualquier característica haciendo que su finalidad e integridad estén en peligro (modificando su comportamiento, obteniendo datos confidenciales e incluso llegar a desconectarlo directamente) y que un mal uso o situaciones inesperadas lo corrompan. En el caso de que ocurra alguna de estas situaciones, se debe contar con un plan de resguardo para así prevenir que vaya a más.

  Se requiere además la garantía de que el dispositivo vaya a actuar siempre como se espera de él (sin causar daños al medio ambiente o a los seres vivos), realice juicios correctos y tome decisiones adecuadas basándose en datos o modelos, haciendo que en el situación de una predicción incorrecta su daño sea el mínimo posible (en caso de que no se puedan evitar, se tiene que indicar la probabilidad de que ocurran).

  Para finalizar con esta necesidad, es crucial que todos los resultados sean reproducibles y fiables, lo que es necesario para poder evaluar como se enfrentan a diversas situaciones y prevenir que ocurran daños involuntarios.

- **Gestión de la privacidad y datos:** Toda la información proporcionada por el usuario y generada sobre este se tiene que proteger para evitar que terceros puedan acceder a ella, además de garantizar que todo lo que se recabe durante su ciclo de vida no se vaya a usar contra estos (por ejemplo, características como la raza u orientación sexual).

  En adición a esto, se necesita comprobar de antemano que los datos de entrada utilizados no estén sesgados, sean imprecisos o contengan errores, ya que podría comprometer en el funcionamiento del dispositivo. Asimismo, hay que tener siempre claro que protocolos, que individuos y en que tipo de situaciones se tienen acceso a datos personales.

- **Transparencia:** Todos los datos y procesos utilizados para la generación del contenido de salida, incluyendo este, se deberán documentar correctamente para así poder aumentar su trazabilidad y transparencia, lo que permitirá identificar fácilmente la causa de los errores y poder evitarlos en un futuro.

  Las decisiones que tome este sistema también tendrán que ser entendibles y explicables a todas las partes involucradas en su uso, lo que incluye que si una persona ha sido afectada en cierta medida por este, pueda reclamar una explicación adecuada del proceso de toma de decisiones (la cual debe ser entendible también). Además de esto, se necesita informar

de que capacidades y limitaciones tiene y, en cuanto al usuario final, debe conocer si está interactuando con una IA o con un humano, pudiendo en algunos casos elegir por cual de los dos ser atendido (por ejemplo, en servicios de atención al cliente).

- **Diversidad, no discriminación y equidad:** Como se ha comentado anteriormente, hay que tener cuidado de que los datos de entrada no se encuentren sesgados, lo que puede dar lugar a discriminación y prejuicios. Este problema no se puede dar solo con esta situación, sino que si los desarrolladores o consumidores utilizan esta tecnología de modo inadecuado, o el propio modelo esta evolucionando de forma que se acentúan estos rasgos, probablemente lleve al mismo problema, por lo que hay que asegurarse que toda la información usada sea correcta y se utilicen métodos de supervisión apropiados.

  Es importante recalcar que el diseño de este tipo de modelos debe estar centrado principalmente en el usuario, de forma que permita a cualquier tipo de persona usarlo independientemente de sus capacidades o características (hay que tener en cuenta sobre todo a discapacitados). Por último, es recomendable pedir a las partes interesadas su opinión periódicamente para así poder ir modificando el diseño y funcionamiento para que este acorde con todo tipo de seres humanos.

- **Bienestar social y ambiental:** Al mismo que tiempo que estos sistemas abordan nuestras preocupaciones, se debe garantizar que no afecten negativamente al medio ambiente, por lo que todos los procesos involucrados en su desarrollo, puesta en servicio y utilización (como puede ser el uso de recursos o energía) deben de ser lo más respetuosos posibles.

  En cuanto al impacto social, se tiene que tener bastante precaución con que estos dispositivos no empeoren nuestras relaciones y vínculos sociales, lo que afecta negativamente al estado físico y mental. Además, no solo debemos pensar en cómo afecta a las personas individualmente, sino también en cómo impacta en toda la sociedad, por lo que se debe evaluar que efecto puede llegar a tener en las instituciones, la democracia y en la propia sociedad en su conjunto.

- **Rendición de cuentas:** Este requisito implica la necesidad de crear mecanismos que aseguren que se pueda responsabilizar y rendir cuentas por las acciones de los modelos de inteligencia artificial y sus resultados, cosa que debe aplicarse tanto antes de ponerlos en funcionamiento como después de hacerlo.

  El primero de ellos es la auditabilidad, que es la capacidad de evaluar los procesos, los datos y los algoritmos involucrados en el diseño. En este contexto implica que se tiene

que evaluar por parte de auditores externos e internos, creando una serie de documentos disponibles en todo momento para de esta manera garantizar la fiabilidad de la tecnología.

El segundo es la minimización y notificación de efectos negativos, método por el cual se asegura que se pueda informar sobre las acciones o decisiones que dan lugar a ciertos resultados de un sistema, así como responder a las consecuencias de estos. Identificar, evaluar, notificar y reducir estas repercusiones es especialmente crucial para todas las personas involucradas directa o indirectamente. Usar evaluaciones de impacto, como "equipos rojos"[10] o ciertos tipos de evaluaciones algorítmicas, antes y después de desarrollar, implementar y usar estos dispositivos, puede ayudar a reducir sus efectos negativos.

El tercero es la búsqueda de equilibrios, que se refiere a que en el momento de aplicar los requisitos mencionados, pueden surgir conflictos entre ellos, por lo que puede ser necesario encontrar un punto medio. Debido a esto es importante abordarlas de manera lógica y ordenada, teniendo en cuenta el nivel técnico actual. Si se presentan discrepancias, es crucial explicar cómo se intentó encontrar un equilibrio entre ellos y evaluar esta estabilidad en términos del riesgo para los principios éticos y los derechos fundamentales. Si no es posible encontrar una armonía éticamente aceptable, no se debe continuar con el desarrollo, implementación y uso del sistema de IA como estaba planeado.

Por último, cabe desatacar que cuando ocurran efectos negativos derivados del uso del modelo, deben de realizarse con anterioridad mecanismos que aseguren una compensación adecuada, lo cual garantiza y aumenta en gran medida la confianza de los usuarios.

En el propio documento se nos explican una serie de métodos técnicos y no técnicos para ayudarnos a cumplir con los requisitos anteriormente mencionados, entre los cuales cabe destacar los siguientes:

Métodos técnicos

- **Arquitecturas:** Los requisitos para una inteligencia artificial confiable deben incorporarse a la estructura de los sistemas mediante normas que definan comportamientos

[10] Los "equipos rojos" son grupos especializados que se emplean en diversas áreas, como la seguridad informática, la defensa nacional o la evaluación de riesgos, para llevar a cabo simulaciones de ataques o intrusiones con el fin de identificar vulnerabilidades en sistemas, infraestructuras o procesos.

adecuados ("lista blanca"), restricciones ("lista negra") y garantías sobre el comportamiento del sistema. Para modelos con capacidad de aprendizaje, es crucial integrar estos requisitos en cada etapa del ciclo "sentir-planificar-actuar". Durante la primera etapa ("sentir"), se debe de reconocer todos los elementos del entorno necesarios para cumplir las directrices, en la segunda ("planificar"), a partir de la información obtenida en la fase anterior, se considera únicamente aquellos planes que cumplan con los requisitos y, en la última ("actuar"), las acciones de la tecnología se limitan a las seleccionadas en la anterior etapa.

- **Ética y estado de derecho desde el diseño:** Cuando nos encontramos en la etapa de diseño es importante asegurarnos de que se cumplen con ciertos valores desde el inicio, lo que significa que los principios abstractos que queremos que el sistema siga deben estar claramente vinculados a las decisiones específicas sobre cómo se implementa y utiliza. En la actualidad es de vital relevancia que todo lo relacionado con el dispositivo este totalmente seguro, por lo que pensar desde el principio sobre como proceder a su protección, resistencia a fallos y apagado es crucial.

- **Métodos de explicación:** Como hemos mencionado en apartados previos, todas las personas involucradas en el uso de un sistema deben entender a la perfección cómo se comporta y porque ha llegado a cierta conclusión. Ahí entra un campo de investigación bastante actual, la inteligencia artificial explicable (XAI), la cual se centra en entender por qué se toman ciertas decisiones. Es un desafío importante, especialmente para modelos basados en redes neuronales, donde los resultados pueden ser difíciles de interpretar y pequeñas diferencias en los datos pueden llevar a interpretaciones completamente distintas. Los métodos de XAI son vitales no solo para explicar el comportamiento de esta tecnología a los usuarios, sino también para garantizar su fiabilidad.

- **Realización de ensayos y validación:** La elaboración de ensayos y la validación son cruciales debido a su naturaleza impredecible y la influencia del contexto en el que operan, ya que las pruebas convencionales no son suficientes porque algunos errores solo se revelan cuando el sistema interactúa con datos realistas. Por lo tanto, es necesario monitorear cuidadosamente la estabilidad, robustez y rendimiento del modelo durante su desarrollo y uso, asegurando que las decisiones tomadas sean validadas adecuadamente. Los ensayos y la validación deben llevarse a cabo lo antes posible y deben abarcar todos los aspectos de la arquitectura, incluyendo datos, modelos, entornos y comportamiento general. Es importante que estos procesos sean diseñados y ejecutados por un equipo diverso para obtener una variedad de perspectivas. Se deben considerar métodos como

prácticas contradictorias realizadas por equipos confiables, así como la recompensa para quienes encuentren y reporten errores y vulnerabilidades del sistema. Finalmente, es esencial garantizar que los productos o acciones derivadas de estos procesos estén en línea con las políticas establecidas previamente para evitar vulneraciones.

- **Indicadores de calidad del servicio:** Se pueden establecer indicadores específicos para evaluar la calidad del servicio de esta tecnologías, asegurando que se hayan tenido en cuenta las consideraciones de seguridad durante su desarrollo y ensayo. Estos indicadores podrían abarcar aspectos como la evaluación de las pruebas realizadas, la formación de algoritmos y los parámetros habituales de calidad del software, incluyendo su rendimiento, usabilidad, fiabilidad, seguridad y mantenimiento.

Métodos no técnicos

- **Normativas:** La legislación actual sobre seguridad de productos y marcos de responsabilidad proporcionan apoyo para la fiabilidad de la IA. Si es necesario adaptar o actualizar estas regulaciones, se debe considerar en futuras recomendaciones políticas relacionadas con este sector.
- **Códigos de Conducta:** Las organizaciones pueden adoptar las directrices anteriormente explicadas en sus políticas internas, como códigos de conducta y documentos de responsabilidad empresarial, para contribuir a la confiabilidad.
- **Normalización:** Establecer normas para el diseño y la fabricación de sistemas puede ayudar a promover una conducta ética en su uso y desarrollo.
- **Certificación:** La certificación de modelos transparentes, responsables y equitativos puede ser realizada por organizaciones especializadas, aunque esto no reemplaza la responsabilidad y debe ir acompañada de marcos de rendición de cuentas.
- **Gobernanza y rendición de cuentas:** Las organizaciones deben establecer marcos internos y externos de gobernanza para asegurar la responsabilidad ética en todas las etapas del desarrollo y uso de estas tecnologías.
- **Educación y concienciación:** La difusión del conocimiento sobre estos dispositivos y la promoción de la participación informada de todas las partes interesadas son esenciales para una IA ética y confiable.

- **Participación y diálogo social:** El debate abierto y la participación de las partes interesadas son fundamentales para evaluar y abordar los impactos y preocupaciones con este tipo de sistemas.
- **Diversidad e inclusión:** Los equipos que desarrollan estos modelos deben reflejar la diversidad de la sociedad para garantizar que se consideren diferentes perspectivas y se aborden diversas necesidades y preocupaciones.

2.2.3. Lista de evaluación para una IA fiable y relación con el reglamento

En caso de querer cumplir con mayor precisión con todos los requisitos descritos anteriormente, este documento presenta una lista de evaluación con el fin de poner en práctica el desarrollo de una IA fiable, aunque está en versión piloto (pag. 35 a 43)[11].

Para finalizar con esta sección, cabe destacar que hace este diferente al reglamento que se ha explicado anteriormente. En cuanto al "Reglamento de inteligencia artificial"[3], este se centra mucho más en todo el apartado legal que deben cumplir todos los sistemas de inteligencia artificial en la actualidad, mientras que este informe trata de explicar los problemas que puede llegar a traer este tipo de tecnología desde un punto de vista ético y deontológico, por lo que consideramos que para estar al tanto de todo los conocimientos que se deben tener a la hora de desarrollar una IA actualmente se deben de estudiar el contenido de ambos documentos.

2.3 Trabajos similares

Habiendo establecido los conocimientos necesarios para la redacción del informe, es fundamental explorar y analizar que otros trabajos existentes abordan las buenas prácticas en el uso de la inteligencia artificial. A continuación, se presenta una revisión de varios documentos clave que destacan las directrices y recomendaciones sobre el uso responsable y ético de la IA en distintos contextos, además de ofrecer una visión general de cada uno, destacando sus principales contribuciones y enfoques:

[11] No se ha descrito en este trabajo esa lista de evaluación porque consideramos que es muy redundante y que simplemente alargaría el contenido sobre temas que ya se han explicado anteriormente, pero en caso de querer profundizar más en este tema, recomendamos encarecidamente su lectura.

2.3.1. Aproximación a la Inteligencia Artificial y la ciberseguridad

El objetivo principal del informe "Aproximación a la Inteligencia Artificial y la ciberseguridad", creado por el ccn-cert en 2023[12], es el de proporcionar una visión integral de la interconexión entre la IA y la ciberseguridad, proporcionando consejos y métodos efectivos para garantizar que esta se utilice de manera segura y eficaz en el mundo digital. En concreto, se va a tratar detenidamente como se ha organizado todo el documento, ya que se tomará en parte como referencia para la posterior redacción de nuestro estudio.

Nada más comenzar se nos presentan los principios esenciales de la inteligencia artificial y la ciberseguridad, destacando sobre todo el papel crucial de la protección digital en este sector, y enfatiza la importancia de implementar protocolos de seguridad desde las primeras etapas de desarrollo de estas tecnologías para prevenir posibles brechas y ataques.

Se discute la evaluación de riesgos, detallando las amenazas potenciales y las vulnerabilidades específicas a las que pueden estar expuestos estos sistemas, junto a la descripción de una serie de métodos para evaluar y priorizar estos riesgos, incluyendo técnicas de modelado de amenazas y análisis de impacto, lo cual es fundamental para garantizar que esta tecnología sea robusta y segura.

Las mejores prácticas para el desarrollo seguro incluyen recomendaciones para integrar principios de seguridad desde la fase de diseño, asegurando que los sistemas sean resistentes a ataques, destacando también estrategias para realizar pruebas exhaustivas y continuas, tales como pruebas de penetración y simulaciones de ataques, así como procedimientos para velar por una buena protección mediante actualizaciones regulares y gestión de parches.

En cuanto a la protección de datos y privacidad, este documento ofrece directrices para la recolección, almacenamiento y procesamiento seguro de la información, asegurando la privacidad y confidencialidad. Además, se enfatiza el cumplimiento de normativas y estándares internacionales de su protección, como el "RGPD"[12], para garantizar que los sistemas de inteligencia artificial operen dentro del marco legal y ético. Además de esto, se subraya la importancia de desarrollar algoritmos que sean transparentes y cuyas decisiones puedan ser interpretadas y explicadas para mantener la confianza en este tipo de modelos. También

[12]El "RGPD", también conocido como el Reglamento General de Protección de Datos, es una legislación creada por la Unión Europea implementada a partir del 25 de mayo de 2018, cuyo propósito fundamental es salvaguardar la información personal de los habitantes de esta región y regular su tratamiento por parte de las organizaciones y empresas, así como garantizar la libre circulación de esos datos dentro del Espacio Económico Europeo (EEE).

se recomienda mantener una documentación detallada y precisa de estas infraestructuras, incluyendo decisiones de diseño, pruebas y actualizaciones.

Para finalizar, cabe destacar que se aborda la gestión de incidentes y la respuesta a estos, recomendando el desarrollo de planes y procedimientos para reaccionar eficazmente a incidentes de seguridad que involucren entornos de procesamiento inteligente, así como la implementación de estructuras de monitorización continua para detectar y responder a actividades sospechosas en tiempo real. Junto a esto se trata la importancia de la formación y la concienciación en ciberseguridad, sugiriendo la implementación de programas de formación para desarrolladores y operadores de este tipo de tecnología, así como iniciativas para promover una cultura de seguridad dentro de las organizaciones que utilizan inteligencia artificial.

2.3.2. GuIA de buenas prácticas en el uso de la inteligencia artificial ética

La "Guía de buenas prácticas en el uso de la inteligencia artificial ética"[13] es un documento colaborativo elaborado por OdiseIA, PwC, Google, Microsoft, IBM y Telefónica con el objetivo de proporcionar directrices y recomendaciones para el desarrollo y uso responsable de la IA con un enfoque ético.

En este informe se tratan varios aspectos relacionados con la ética en el sector de las tecnologías inteligentes, resaltando sobre todo la importancia de la transparencia, equidad, protección de la privacidad y responsabilidad. Se presentan también normas morales esenciales que deben orientar la creación, implementación y despliegue de sistemas tecnológicos avanzados, subrayando la necesidad de integrar protecciones morales desde las primeras fases del proceso.

Además, se proporcionan directrices para la evaluación y reducción de riesgos morales relacionados con el uso de modelos de IA, así como estrategias de gobernanza eficaces para asegurar la adhesión a normas morales y regulaciones pertinentes.

2.3.3. Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores

El informe "Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores", escrito por la dirección general de educación

en 2022[14], describe como deben usarse los sistemas de inteligencia artificial en el sector educativo. Se abordan, como en informes anteriores, que principios se deben aplicar a la hora de usar este tipo de modelos en la educación, destacando principalmente la protección de la privacidad de los estudiantes, la igualdad entre todos ellos independientemente de sus rasgos y pensamientos, y la fomentación de la transparencia a la hora de tomar decisiones.

Además de esto, se explican una serie de directrices orientadas a los educadores sobre como integrar este tipo de tecnología al ámbito de la enseñanza y del aprendizaje, incluyendo lo importante que es que estas personas estén ampliamente formadas y continuamente aprendiendo de este área. Al final, se trata la necesidad de que exista un pensamiento crítico y un cierto grado de alfabetización digital entre los estudiantes, para que así puedan entender y analizar perfectamente todo el impacto que provocan estos sistemas en su educación y desarrollo.

2.4 Propuesta

Este trabajo llena el espacio de conocimiento referente al buen uso de la inteligencia artificial en el sector del desarrollo de software en términos legales y éticos. Dentro de este, se proporcionarán una clasificación de las herramientas más utilizadas, se explicarán las referencias legislativas y morales involucradas en este área y se rastrearán ejemplos de buen y mal uso para poder construir una guía adecuada.

Diferencia y valor añadido

A continuación se van a enumerar y explicar una serie de puntos que hacen que este informe sea diferente de la mayoría creados con anterioridad:

- **Actualización de la normativa:** Todos los estudios realizados previamente se basan o en leyes antiguas o en un marco ético no desarrollado en profundidad, por lo que este estudio usará las últimas normativas vigentes y los análisis deontológicos para poder crear una guía lo más actualizada posible.
- **Enfoque global e integrador:** Aunque hay ya algunos documentos con abordan los mismos aspectos explicados anteriormente, este se distingue por su enfoque global e integrador, ya que combina elementos éticos, legales, de ciberseguridad y metodológicos

en una sola guía coherente y práctica, además de centrarlo en el ámbito general del desarrollo del software (evitando así centrarnos en sectores más pequeños y que por ello sirvan a menos gente).

- **Propuesta innovadora:** Más que una simple recopilación de conceptos ya conocidos, este trabajo propone la combinación de soluciones ya existentes de manera original, presentando ejemplos prácticos y estudios de casos de uso que ilustran cómo implementar estas buenas prácticas en proyectos reales con IA.
- **Contexto académico:** En la Universitat Politècnica de València (UPV) hasta ahora no se ha desarrollado un informe específico que trate de forma integral la temática de este documento, por lo que llenaría ese vacío y proporcionaría una referencia académica para futuros estudios y proyectos de la institución.

3 Análisis del problema

Una vez analizada en profundidad toda la información actual más relevante para la escritura del informe, toca realizar una investigación del problema que trata de resolver este trabajo, el cual es la falta de un marco claro y unificado de buenas prácticas sobre el uso de IA en el desarrollo del software. Dentro de este apartado se evaluarán algunas de las soluciones existentes, se identificarán las posibles soluciones junto a sus ventajas y desventajas, y se seleccionará la más adecuada para cumplir con nuestro objetivo. Una vez seleccionada, se detallará en mayor profundidad y se establecerá un plan de trabajo para la elaboración de este documento.

3.1 Identificación y análisis de las soluciones posibles

En esta sección, se identifican y analizan distintas soluciones existentes para abordar la necesidad de establecer buenas prácticas en el uso de IA en el software. El análisis se basa en una revisión exhaustiva de otros informes que han sido redactados con propósitos similares al nuestro, con el objetivo de seleccionar la solución más adecuada y aplicable. Estos documentos son los analizados previamente en el apartado del "Estado del arte", pero es necesario volver a echarles un vistazo para ver que puntos positivos y negativos tienen y así mejorarlos en nuestro informe:

Revisión de soluciones existentes:

1. **Aproximación a la Inteligencia Artificial y la ciberseguridad[12]:**
    - **Descripción:** Aborda la relación entre la IA y la ciberseguridad, destacando la importancia de proteger este tipo de sistemas contra ataques y asegurando la integridad y confidencialidad de los datos.
    - **Relevancia:** La ciberseguridad es un componente crítico para el uso seguro y confiable de esta tecnología, por lo que las recomendaciones proporcionadas son esenciales para cualquier organización que las desarrolle o utilice.
    - **Aplicabilidad:** Facilita estrategias y recomendaciones específicas para proteger los sistemas, lo que es crucial para garantizar la seguridad de los datos y la infraestructura.

2. **Guía de buenas prácticas en el uso de la inteligencia artificial ética[13]:**
    - **Descripción:** Brinda un marco para el desarrollo y uso ético de la IA, incluyendo recomendaciones prácticas sobre transparencia, responsabilidad, privacidad y equidad.
    - **Relevancia:** Desarrollada por una asociación de líderes de la industria, ofrece un conjunto de mejores prácticas respaldadas por la experiencia práctica y el consenso de expertos.
    - **Aplicabilidad:** Sus recomendaciones prácticas son fácilmente implementables por organizaciones de distintos tamaños y sectores, proporcionando un modelo adaptable para diversas realidades empresariales.
3. **Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores[14]:**
    - **Descripción:** Ofrece directrices específicas para el uso de IA en el ámbito educativo, enfatizando la importancia de la ética, la protección de datos y la mejora de la calidad educativa mediante el uso responsable de esta.
    - **Relevancia:** Aunque centradas en el ámbito educativo, estas pautas ofrecen principios éticos y prácticas que pueden ser adaptados a otros contextos.
    - **Aplicabilidad:** Proporciona ejemplos claros de cómo aplicar principios éticos en la práctica, lo que puede ser adaptado para otros sectores y contextos empresariales.

Análisis de pros y contras:

1. **Aproximación a la Inteligencia Artificial y la ciberseguridad:**
    - **Pros:** Ofrece estrategias claras para proteger los sistemas de IA contra amenazas cibernéticas, lo que es crucial para cualquier implementación segura.
    - **Contras:** Está muy centrado en aspectos técnicos de ciberseguridad, sin cubrir completamente otras áreas éticas y legales necesarias para un enfoque integral de buenas prácticas.
2. **Guía de buenas prácticas en el uso de la inteligencia artificial ética:**
    - **Pros:** Proporciona un marco comprensivo y práctico desarrollado por líderes de la industria, lo que asegura que las recomendaciones sean aplicables y efectivas.

- **Contras:** A pesar de lo descrito anteriormente, puede llegar a existir variabilidad en la implementación debido a diferencias contextuales entre organizaciones, lo que puede requerir adaptaciones específicas.

3. **Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores:**
    - **Pros:** Facilita principios éticos claros y ejemplos prácticos que son adaptables a diversos contextos más allá del educativo.
    - **Contras:** Aunque estas pautas pueden ser adaptables, su enfoque central sigue siendo el entorno educativo, por lo que si queremos adaptar su información a otros contextos habría que realizar una serie de ajustes.

3.2 Solución propuesta

Para solucionar la problemática planteada, se ha optado por realizar un informe que integre principalmente el nuevo reglamento de la Unión Europea junto a las directrices de la IA fiable de 2021, además de que estará estructurado de forma que aborde la mayor cantidad de aspectos éticos, legales y técnicos.

3.2.1. Estructura

A continuación se describen brevemente todos los apartados principales del informe junto a una breve descripción:

- **Introducción:** Definiciones básicas, papel de la inteligencia artificial en el software y objetivo e importancia del documento.
- **Principales técnicas y enfoques de inteligencia artificial utilizados en el desarrollo de software:** Descripción de diferentes técnicas como el aprendizaje automático, el procesamiento del lenguaje natural y los sistemas expertos.
- **Herramientas más utilizadas:** Presentación de las herramientas y plataformas de desarrollo más populares en el sector.

- **Ética y consideraciones legales en el desarrollo de software con IA:** Análisis y explicación de todas las normas legales y principios éticos que se deben seguir para un uso correcto de esta tecnología.
- **Ejemplos de uso:** Se muestran aplicaciones de IA en diferentes sectores, destacando tanto las buenas decisiones que han llevado a un uso correcto, como los problemas generados por una utilización indebida.

3.3 Plan de trabajo

Para finalizar, se definirá un plan de trabajo detallado, usando técnicas de planificación, estimación de esfuerzo y presupuesto aproximado[1]:

3.3.1. Planificación:

- **Fase de definición y objetivos:** Establecer el propósito y estructura del informe, y reunir la información básica junto a escribir todo el apartado de introducción.
- **Fase de investigación técnica:** Investigar las principales técnicas y enfoques de la IA y describir las herramientas más utilizadas en este campo.
- **Fase de evaluación ética y legal:** Analizar y detallar las consideraciones éticas y legales más importantes, y evaluar el impacto que tiene esta tecnología en diferentes ámbitos de la sociedad.
- **Fase de aplicaciones y conclusiones:** Recopilar todo tipo de ejemplos de uso de la IA en el desarrollo del software y redactar las conclusiones.
- **Fase de revisión y edición:** Revisar todo el documento completo para que no tenga faltas de ortografía ni problemas en la coherencia, precisión y claridad. Además hay que asegurarse que todas las secciones están integradas de forma correcta y se siga un flujo lógico.

[1]Cabe destacar que todo este esfuerzo es solamente el relacionado con la redacción del informe, por lo que también habría que añadir todo lo relacionado con la escritura de esta memoria para reflejarlo de manera real.

3.3.2. Estimación de esfuerzo:

- **Fase de definición y objetivos:** 30 horas (1-2 semanas)
- **Fase de investigación técnica:** 50 horas (2-3 semanas)
- **Fase de evaluación ética y legal:** 80 horas (3-4 semanas)
- **Fase de aplicaciones y conclusiones:** 20 horas (1 semana)
- **Fase de revisión y edición:** 20 horas (1 semana)

3.3.3. Presupuesto:

Según las estimaciones previas, el informe costaría de redactar unas 200 horas aproximadamente y el rango de precio por hora para realizar esto oscilaría los 13,85 euros la hora. Con estos datos, se podría estimar que el precio necesario para la escritura de este informe (únicamente teniendo en cuenta el sueldo que habría que pagar a los trabajadores) sería de unos 2.770€.

4 Diseño y desarrollo del informe

En este apartado se tratarán todas aquellas secciones relevantes relacionadas con el propio proceso de redacción del informe de buenas prácticas. En concreto, se analizarán aquellas tecnologías y aplicaciones utilizadas que han ayudado en la propia escritura, la estructura final y detallada con la que nos hemos quedado, y una breve narración de cómo ha sido este proceso y que diferencias han habido con la planificación original. También se incluye en el último apartado una pequeña herramienta de evaluación para saber si el documento ha cumplido con los requisitos de las buenas prácticas.

4.1 Aplicaciones y tecnologías usadas

A la hora de redactar y desarrollar este informe, se han empleado una gran variedad de herramientas que han facilitado el proceso de creación de contenido, estructuración, diseño y revisión. A continuación, se muestran cada una de estas tecnologías con su respectiva descripción:

- **Overleaf:** Es una plataforma de redacción de documentos en línea que utiliza LaTex, un sistema de composición tipográfica de alta calidad. Ha sido esencial para poder crear tanto el informe como este propio documento, permitiendo una estructura más dinámica y profesional, gestionar versiones de forma más eficiente e integrar de manera sencilla tanto imágenes como referencias en el texto[15].
- **Google Scholar y RiuNet:** El primero es un motor de búsqueda especializado en literatura académica que permite encontrar trabajos de diversas disciplinas científicas, y el segundo es el repositorio institucional de la UPV, el cual alberga una gran variedad de documentos académicos. Ambas han sido de gran ayuda en la búsqueda de otros estudios similares al que se está realizando en este trabajo[16][17].
- **Canva:** Ha sido un entorno en línea especialmente útil para poder crear las imágenes que hay en ambos documentos, ya que su interfaz intuitiva y sus amplias bibliotecas de plantillas y recursos han permitido la creación de todo tipo de gráficas y elementos visuales[18].
- **Scribbr:** Está plataforma ha sido usada principalmente para la generación de citas y

referencias de forma automática, pudiendo así seguir los estilos adecuados de citación y asegurando una correcta atribución de fuentes en trabajos académicos[19].

4.2 Estructura final del informe

En la sección anterior se ha explicado un poco por encima la estructura general que iba a tener el informe, pero ahora que está ya más desarrollado se puede explicar más en detenimiento cada apartado (estos se pueden observar en la figura 4.1):

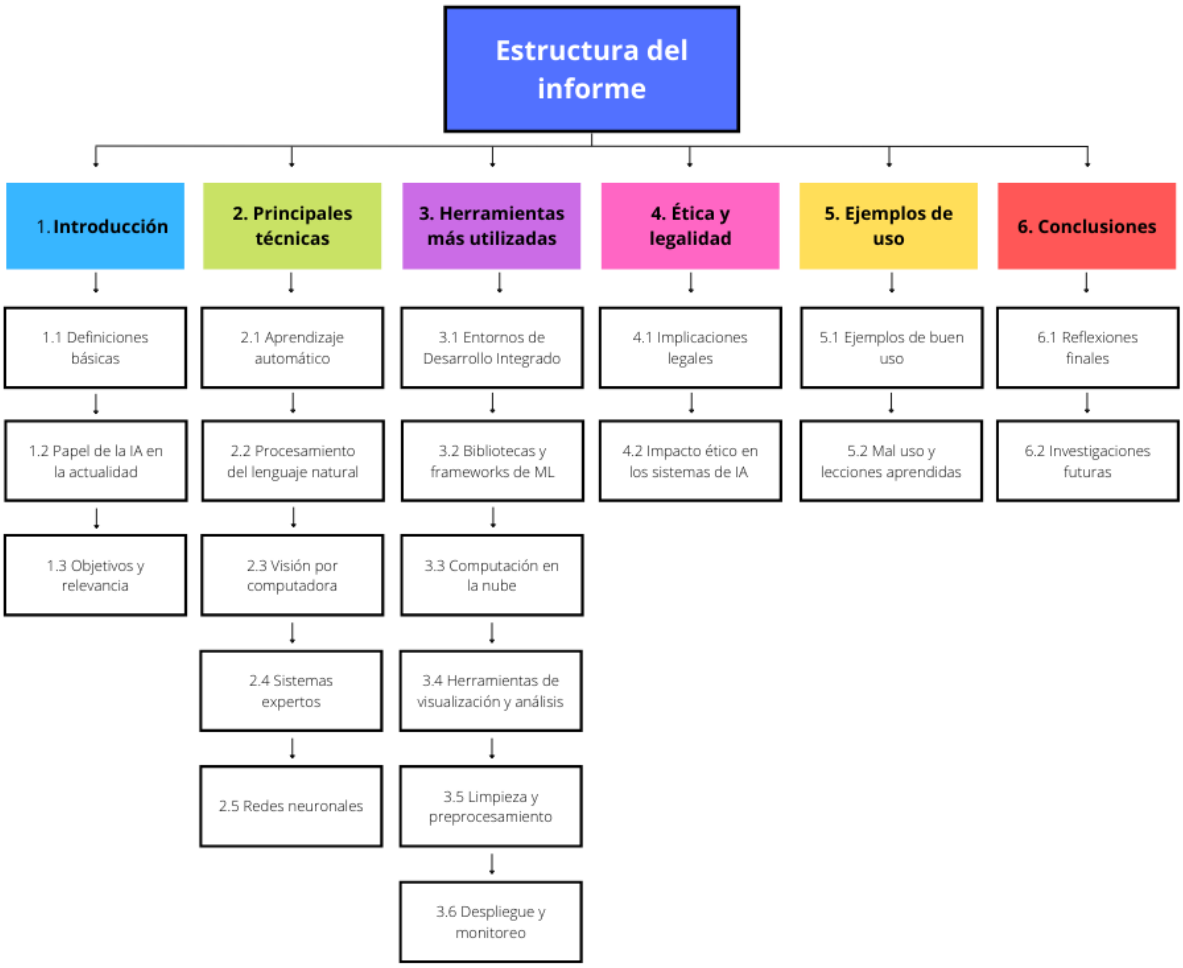

Figura 4.1: Estructura final del informe

- **Introducción:** Se exploran, a parte de una pequeña introducción inicial, una serie de

definiciones básicas, el papel de la IA en el desarrollo de software y el objetivo e importancia de este documento. Esto se hace para que el lector tenga un poco de contexto sobre la problemática que se intenta resolver y conocimiento sobre las tecnologías implicadas.

- **Principales técnicas y enfoques de inteligencia artificial utilizados en el desarrollo de software:** Se explican las principales técnicas usadas en IA, las cuales son el aprendizaje automático, el procesamiento del lenguaje natural, la visión por computadora, los sistemas expertos y las redes neuronales.
- **Herramientas más utilizadas:** Al igual que el anterior punto, se detallan las herramientas más usadas en el sector. Concretamente son los Entornos de Desarrollo Integrado(IDEs), las bibliotecas y frameworks de Machine Learning, las plataformas de computación en la nube, las herramientas de visualización y análisis de datos, las de limpieza y preprocesamiento, y las de despliegue y monitoreo de sistemas.
- **Ética y consideraciones legales en el desarrollo de software con IA:** En esta sección solo hay dos apartados, el de implicaciones legales y el impacto ético. Son los más extensos de todo el informe, ya que tratan sobre todos los aspectos jurídicos y morales que se deben seguir en la Unión Europea en la actualidad para cualquier desarrollo relacionado con la inteligencia artificial y el software.
- **Ejemplos de uso:** Se clasifican y analizan una serie de ejemplos de buen y mal uso en sectores como medicina, finanzas y TI (Tecnologías de la Información). Esto se hace con el propósito de señalar cuales han sido las buenas practicas identificadas, los errores cometidos y las lecciones aprendidas de estos.
- **Conclusiones:** Se presenta un breve resumen de los temas tratados en todo el documento, junto con reflexiones finales sobre los posibles trabajos futuros derivados de este y el desarrollo actual y futuro del software con IA.

4.3 Evolución y cumplimiento de la planificación

El proyecto ha avanzado de acuerdo al plan establecido previamente, cumpliendo en gran parte con la planificación original. A continuación, se presenta un breve resumen de cada fase y su evolución:

- **Fase de definición y objetivos:** Esta fase, que estaba estimada en 30 horas, se completó en el tiempo previsto. Se estableció el propósito y la estructura del informe y se reunió la información básica necesaria, redactando el apartado de introducción.
- **Fase de investigación técnica:** Con una estimación inicial de 50 horas, se investigaron las principales técnicas y enfoques de la inteligencia artificial, además de describir las herramientas más usadas del sector, todo esto dentro del plazo propuesto.
- **Fase de evaluación ética y legal:** Esta fase, que requería de unas 80 horas, se llevó a cabo en el tiempo establecido. Se analizaron y detallaron las consideraciones éticas y legales más importantes, para así esclarecer que pautas se deben seguir para lograr un buen uso de esta tecnología.
- **Fase de aplicaciones y conclusiones:** Con una duración estimada de 20 horas, se recopilaron una serie de ejemplos de uso de la IA en el desarrollo de software y se redactaron las conclusiones según como se había previsto.
- **Fase de revisión y edición:** Estimada en 20 horas, se completó en el tiempo programado. Se revisó el documento para corregir faltas de ortografía y asegurar la coherencia, precisión y claridad, integrando todas las secciones correctamente y manteniendo un flujo lógico.

4.4 Herramienta de evaluación

Una vez finalizado el informe, es de vital importancia saber si este ha cumplido con las buenas prácticas, es decir, si ha tratado la mayoría de problemáticas que afectan a la IA en el mundo del desarrollo del software. A continuación, se muestra una tabla donde se señala si se han abordado esos puntos:

Número	**Descripción**	**¿Cumple?**
1	Transparencia en el Algoritmo	Sí
2	Protección de Datos y Privacidad	Sí
3	Seguridad del Sistema	Sí
4	Prevención de Sesgo y Discriminación	Sí
5	Validación y Verificación de Modelos	Sí
6	Explicabilidad y Interpretabilidad	Sí
7	Responsabilidad y Seguimiento	Sí
8	Gestión de Riesgos	Sí
9	Uso Ético de IA	Sí
10	Eficiencia y Rendimiento	Sí
11	Documentación y Registro	Sí
12	Capacitación y Educación Continua	Sí
13	Pruebas y Evaluación Continua	Sí

Cuadro 4.1: Cumplimiento de las Buenas Prácticas en el Uso de IA

Como se puede observar, el informe ha tratado todos los riesgos y puntos clave, por lo que se puede afirmar que ha cumplido con las buenas prácticas.

5 Conclusión

En la elaboración del informe de buenas prácticas para el uso profesional de inteligencia artificial en el desarrollo del software se han alcanzado todos los objetivos proporcionados inicialmente. El primer objetivo, comprender los fundamentos de la IA, se ha logrado proporcionando una serie de conceptos básicos y aplicaciones. En cuanto al análisis de técnicas y enfoques, se investigaron diversas estrategias y métodos, como la visión por computadora y el aprendizaje automático, evaluando su utilidad en este campo. La exploración de consideraciones éticas y legales también se cumplió, analizando todas las implicaciones del uso de esta tecnología, basándose principalmente en el reglamento de la UE del 13 de marzo de 2024. Finalmente, se identificaron una serie de ejemplos concretos de buen y mal uso, destacando prácticas recomendadas, errores a evitar y lecciones aprendidas.

Durante la escritura del informe, se encontraron una serie de desafíos, entre los que cabe destacar la falta de consenso en la industria, la complejidad técnica de algunos conceptos y la falta de flexibilidad en determinadas normativas. Esto se solucionó consultando una gran cantidad de fuentes, señalando un conjunto de directrices modulares y empleando un enfoque detallado para asegurar la accesibilidad y aplicabilidad en diferentes niveles de experiencia. Cabe destacar que también se encontraron diversos problemas que se resolvieron mediante la consulta de información, cómo errores en la interpretación de algunas normativas, los cuales se podrían haber evitado con una investigación inicial más profunda.

5.1 Relación del trabajo desarrollado con los estudios cursados

Para poder realizar todo el informe han sido de gran ayuda todos los conocimientos adquiridos durante los estudios en ingeniería informática. Entre estos cabe destacar asignaturas relacionadas con inteligencia artificial y deontología, ya que han proporcionado conceptos y técnicas clave para abordar todos los puntos del proyecto.

La gestión de proyectos y el desarrollo de software también fueron de gran importancia, ya que proporcionaron habilidades en organización, planificación y diseño que permitieron estructurar el informe de manera eficiente y desarrollar recomendaciones prácticas y aplicables. En resumen, se ha podido integrar gran cantidad de los conocimientos adquiridos durante la carrera, demostrando de esta forma la relevancia de los estudios cursados y la capacidad para enfrentar desafíos reales en el mundo laboral.

6 Trabajos futuros

En el desarrollo del informe hubo varios aspectos que no se pudieron abordar por limitaciones de tiempo, como la inclusión de casos de estudio más detallados, un mayor énfasis en la ciberseguridad y la evaluación de herramientas emergentes del sector. Para futuros trabajos se podría explorar hacía varías direcciones, entre las que cabe destacar la aplicación de las directrices a otros sectores tecnológicos y la integración de las recomendaciones con metodologías ágiles para mejorar la flexibilidad en entornos en constante cambio.

En cuanto a mejoras, se podría aumentar la eficiencia de las pautas proporcionadas mediante herramientas de automatización y enriquecer el informe con algunas secciones más interactivas para personalizar el contenido según las necesidades específicas. Sin embargo, hay que evitar enfoques demasiado técnicos que puedan alejarse de la práctica y no enfocarse únicamente en tecnologías emergentes sin saber en profundidad cómo se pueden adoptar.

Bibliografía

[1] IBM. *¿Qué es la Inteligencia Artificial (IA)? | IBM*. URL: https://www.ibm.com/es-es/topics/artificial-intelligence (visitado 21-04-2024).

[2] Henry A Kissinger, Eric Schmidt y Daniel Huttenlocher. *The Age of AI. "THE BOOK WE ALL NEED"*. Hachette UK, 16 de nov. de 2021.

[3] Parlamento Europeo. *Reglamento de Inteligencia Artificial*. 13 de mar. de 2024. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf (visitado 25-04-2024).

[4] *Reglamento - UE Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo*. L 689. 27 de jul. de 2024. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AL_202401689.

[5] *Carta de los Derechos Fundamentales de la Unión Europea*. Actualizado por última vez el 14 de enero de 2022. URL: https://eur-lex.europa.eu/ES/legal-content/summary/charter-of-fundamental-rights-of-the-european-union.html (visitado 25-04-2024).

[6] Comisión Europea. *Reglamento del parlamento europeo y del consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión*. 21 de abr. de 2021. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206 (visitado 25-04-2024).

[7] Parlamento Europeo, Comisión de Asuntos Jurídicos, Comisión de Empleo y Asuntos Sociales, Comisión de Industria, Investigación y Energía, Comisión de Libertades Civiles, Justicia y Asuntos de Interior, Comisión de Medio Ambiente, Salud Pública y Seguridad Alimentaria, Comisión de Mercado Interior y Protección del Consumidor, Comisión de Transportes y Turismo. *Normas de derecho civil sobre robótica. Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre Normas de Derecho Civil sobre Robótica (2015/2103(INL))*. 2015/2103(INL). 16 de feb. de 2017. URL: https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52017IP0051 (visitado 18-04-2024).

[8] *Reglamento (UE) n.º 1025/2012 del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, sobre la normalización europea, por el que se modifica la Directiva 89/686/CEE del Consejo y las Directivas 93/15/CEE, 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE y 2009/105/CE del Parlamento*

Europeo y del Consejo y se derogan la Decisión 87/95/CEE del Consejo y la Decisión n.º 1673/2006/CE del Parlamento Europeo y del Consejo. 1025/2012. 14 de nov. de 2012. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32012R1025 (visitado 17-05-2024).

[9] *Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos y se deroga el Reglamento (CEE) n.º 339/93*. 765/2008. 13 de ago. de 2008. URL: https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=celex%3A32008R0765 (visitado 17-05-2024).

[10] Parlamento Europeo y Consejo de la Unión Europea. *Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo relativo a la vigilancia del mercado y la conformidad de los productos Y por el que se modifican la Directiva 2004/42/CE y los reglamentos (CE) n.° 765/2008 y (UE) n.° 305/2011*. 2019/1020. 20 de jun. de 2019. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32019R1020 (visitado 17-05-2024).

[11] Dirección General de Redes de Comunicación, Contenido y Tecnologías (Comisión Europea) y Grupo de expertos de alto nivel en inteligencia artificial. *Directrices éticas para una IA fiable*. 2019. DOI: 10.2759/14078. URL: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (visitado 18-04-2024).

[12] Carlos Galán Cordero y Carlos M. Galán Pascual. *Aproximación a la inteligencia artificial y la ciberseguridad. Informe de Buenas Prácticas BP/30*. 4 de dic. de 2023. URL: https://www.ccn-cert.cni.es/es/seguridad-al-dia/novedades-ccn-cert/12852-nuevo-informe-de-buenas-practicas-bp-30-sobre-aproximacion-a-la-inteligencia-artificial-y-la-ciberseguridad.html (visitado 12-02-2024).

[13] OdiseIA, PwC, Google, Microsoft, IBM y Telefónica. *GuIA de buenas prácticas en el uso de la inteligencia artificial ética*. 17 de feb. de 2022. URL: https://www.pwc.es/es/publicaciones/tecnologia/odiseia-pwc-guia-responsable-ia.html (visitado 20-04-2024).

[14] Dirección General de Educación, Juventud, Deporte y Cultura (Comisión Europea). *DIRECTRICES ÉTICAS SOBRE EL USO DE LA INTELIGENCIA ARTIFICIAL (IA) y LOS DATOS EN LA EDUCACIÓN y FORMACIÓN PARA LOS EDUCADORES*. 2022. DOI: 10.2766/898. URL: https://op.europa.eu/es/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1 (visitado 28-04-2024).

[15] *Overleaf, Editor de LaTeX online*. URL: https://es.overleaf.com/.

[16] *Google Scholar*. URL: https://scholar.google.es/.

[17] *RiuNet repositorio UPV*. URL: https://riunet.upv.es/.

[18] *Canva*. URL: https://www.canva.com/.

[19] *Scribbr - Generador de citas*. URL: https://www.scribbr.es/citar/generador/.

A Anexos

A.1 Objetivos de Desarrollo Sostenible

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

Objetivos de Desarrollo Sostenible	Alto	Medio	Bajo	No Procede
Fin de la pobreza.			X	
Hambre cero.			X	
Salud y bienestar.		X		
Educación de calidad.		X		
Igualdad de género.				X
Agua limpia y saneamiento.				X
Energía asequible y no contaminante.			X	
Trabajo decente y crecimiento económico.	X			
Industria, innovación e infraestructuras.	X			
Reducción de las desigualdades.		X		
Ciudades y comunidades sostenibles.		X		
Producción y consumo responsables.			X	
Acción por el clima.			X	
Vida submarina.				X
Vida de ecosistemas terrestres.				X
Paz, justicia e instituciones sólidas.		X		
Alianzas para lograr objetivos.	X			

Cuadro A.1: Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS)

Reflexión sobre la relación del TFG con los ODS

Este trabajo, como se ha podido observar en el apartado anterior, se relaciona con varios Objetivos de Desarrollo Sostenible (ODS). A continuación, se analiza cómo se alinea con algunos de estos objetivos:

1. **Educación de calidad (ODS 4):** Con la creación de un informe de buenas prácticas se está contribuyendo a la educación continua de los profesionales, además de que sirve como material educativo para cursos de formación y programas de capacitación.

2. **Trabajo decente y crecimiento económico (ODS 8):** Promueve el uso ético y responsable de la inteligencia artificial, lo que contribuye en la creación de empleos de calidad en el sector tecnológico. Al proporcionar diferentes directrices para el desarrollo e implementación se fomenta un entorno laboral innovador y eficiente, impulsando de esta forma el crecimiento económico.

3. **Industria, innovación e infraestructuras (ODS 9):** Este trabajo está relacionado con el fortalecimiento de la industria mediante el uso de prácticas innovadoras en el desarrollo de software. La implementación de estas tecnologías mejora significativamente las infraestructuras digitales y promueve un desarrollo industrial sostenible.

4. **Reducción de las desigualdades (ODS 10):** La adopción de prácticas éticas y responsables ayuda en gran parte a reducir desigualdades, especialmente en términos de acceso a tecnología y oportunidades. En este documento se abordan cuestiones de equidad y justicia, proponiendo soluciones que eviten sesgos y promuevan la inclusión.

5. **Ciudades y comunidades sostenibles (ODS 11):** Se contribuye al desarrollo de ciudades más inteligentes y sostenibles, además de que las directrices propuestas en el informe fomentan el uso de IA para mejorar la eficiencia energética, la gestión de recursos y la calidad de vida en las comunidades urbanas.

6. **Paz, justicia e instituciones sólidas (ODS 16):** Como se ha mencionado anteriormente, este trabajo trata en profundidad diversos temas legales y éticos, lo que puede fortalecer la confianza en las instituciones y promover la justicia. Esto se debe a que estas organizaciones son en parte responsables de que las tecnologías se desarrollen y utilicen de manera responsable y transparente.

7. **Alianzas para lograr objetivos (ODS 17):** El seguimiento de las pautas proporcionadas implica la colaboración entre expertos en diversas áreas, ya que sin esto sería mucho más complicado alcanzar los objetivos propuestos y la promoción de un uso más amplio y ético de la IA.

A.2 Informe de buenas prácticas

INFORME DE BUENAS PRÁCTICAS PARA EL USO PROFESIONAL DE LA INTELIGENCIA ARTIFICIAL EN EL DESARROLLO DE SOFTWARE

ESCRITO POR
RAFA BENAVENT GARCÍA

SEPTIEMBRE 2024

Índice general

Índice de figuras

CAPÍTULO 1

Introducción

En la era actual de la digitalización, la inteligencia artificial (IA) emerge como una fuerza transformadora en múltiples áreas, y el desarrollo de software no es la excepción. Este informe trata básicamente sobre las mejores prácticas en la creación de programas informáticos con IA, sector el cual ha evolucionado constantemente estos últimos años.

Los sistemas autónomos, en su esencia, simulan la inteligencia humana en máquinas configuradas para emular el pensamiento y acciones de las personas. Esta tecnología aplicada al desarrollo informático abre nuevas puertas para la automatización, eficiencia e innovación, aunque también trae consigo una serie de desafíos significativos y responsabilidades éticas que hay que abordar.

El principal objetivo de este documento es otorgar al lector una visión completa de las prácticas recomendadas en el desarrollo de software con sistemas inteligentes, donde se analizarán desde definiciones básicas hasta su impacto actual, pasando por las principales técnicas y enfoques, su utilidad en el campo, herramientas comunes, consideraciones éticas y legales, y casos de estudio relevantes. En concreto, este prólogo sitúa el escenario para el informe y establece una base sólida para una comenzar a entender sobre esta tecnología y su importancia, además de servir como marco útil para los lectores, permitiéndoles familiarizarse con el tema.

1.1. Definición de software e inteligencia artificial

Para comprender plenamente el contenido de este informe, es esencial tener al menos un conocimiento superficial sobre el software y la inteligencia artificial, ya que más adelante se explicará más en detenimiento la relación que hay entre ambos (además de que es necesario para entender de mejor forma los siguientes apartados).

Un sistema informático se divide principalmente en dos elementos: el hardware y el software. A diferencia del hardware, que abarca todo los elementos físicos, el software comprende todo el conjunto de instrucciones y programas que guían al ordenador en sus funciones. Dentro de este conjunto, se encuentran herramientas comunes de uso cotidiano, como puede ser el sistema operativo, los navegadores web, las aplicaciones de mensajería y las redes sociales[1].

La Inteligencia Artificial es una herramienta indispensable a la hora de crear sistemas que sean capaces de llevar a cabo tareas que normalmente requieren de una intervención humana. Según la Unión Europea (UE), estos sistemas se caracterizan por su capacidad para actuar con cierto grado de autonomía e independencia del control humano, además de ser capaces de aprender, razonar, modelar y operar según objetivos definidos por los seres humanos, ya sean implícitos o explícitos[2].

En adición a esto, también pueden generar contenidos, predicciones, recomendaciones o decisiones que impactan en el entorno real o virtual en el que operan. La UE ha establecido normativas claras para regular el uso de esta tecnología, con el fin de garantizar la seguridad y el respeto de los derechos de los ciudadanos en la Unión Europea[3].

1.2. Papel de la inteligencia artificial en el desarrollo de software

Una vez ya explicadas las terminologías anteriores, es de vital importancia resaltar como estas están estrechamente relacionadas. A continuación, se muestran una serie de puntos que justifican el porque la IA fue, es y será esencial en el campo del desarrollo del software[1]:

- **Revolución tecnológica:** La irrupción de esta tecnología ha supuesto un gran cambio en el mundo del software, ya que ha abierto nuevas puertas y ha provocado una gran revolución. Gracias a esto, los desarrolladores ahora tienen la capacidad de diseñar sistemas que aprenden y se adaptan con el tiempo, lo que permite una mayor personalización y eficiencia.
- **Impacto en el desarrollo del software:** La IA ha dejado una huella permanente en esta

[1]Para realizar estas afirmaciones nos hemos basado en dos artículos: "Inteligencia artificial en desarrollo de software: Tendencias emergentes y futuro"[4], escrito por Pablo Huet (experto en FrontEnd); y "La Inteligencia Artificial en el Desarrollo de Software: Impulsando la Innovación y la Eficiencia"[5], escrito por Sandra Cabrera García (Software Developer).

área, permitiendo la creación de aplicaciones más eficientes, precisas y personalizadas. Para poner un ejemplo, los algoritmos de aprendizaje automático pueden analizar grandes cantidades de datos para identificar patrones y tendencias, lo que puede ser utilizado para mejorar la funcionalidad y la experiencia del usuario.

- **Cambio de paradigma:** Los sistemas inteligentes han cambiado la forma en que se crean los programas, dado que antes el desarrollo de software se basaba en un enfoque manual y basado en reglas, pero ahora, gracias a su evolución, se ha adoptado un enfoque más automatizado y basado en datos. Esto significa que los modelos pueden aprender de los datos y mejorar con el tiempo, en lugar de seguir un conjunto de reglas predefinidas.
- **Evolución:** Esta herramienta ha recorrido un largo camino en la creación de aplicaciones, el cual comprende desde cuando se utilizaba principalmente para tareas simples y basadas en reglas hasta la actualidad, que es usada para una amplia gama de tareas complejas, como el procesamiento del lenguaje natural, la visión por computadora y el aprendizaje profundo. Esto ha logrado que los desarrolladores puedan crear aplicaciones más sofisticadas y eficientes.
- **Importancia:** Por último, cabe destacar que esta es de vital importancia en el desarrollo de software, y se espera que siga creciendo con el paso del tiempo. Con la constante disponibilidad de datos y la necesidad de sistemas más inteligentes y personalizados, se está convirtiendo en una herramienta esencial para cualquier profesional en este campo, además de que a medida que la tecnología sigue avanzando, es probable que veamos aún más aplicaciones innovadoras de la inteligencia computacional.

1.3. Objetivo e importancia del desarrollo del informe

Como se ha mencionado anteriormente, el objetivo principal de este informe es ofrecer una visión detallada y exhaustiva de las buenas prácticas en el desarrollo de software con inteligencia artificial (IA), sirviendo como guía para aquellos interesados en el ámbito, ya sean profesionales experimentados o principiantes que deseen adentrarse en este campo. La relevancia de este estudio radica en su enfoque en el campo de la deontología y la ética, ya que en un mundo cada vez más digitalizado, donde la automatización evoluciona en diversos ámbitos, resulta fundamental que los diseñadores de software comprendan y apliquen estrategias efectivas para asegurar resultados fiables y éticos.

Dentro de este se tratan en profundidad diversos aspectos que son cruciales para lograr nuestro objetivo, los cuales son las principales técnicas y enfoques usados en este sector, las herramientas mas utilizadas, un análisis legal y ético de los documentos más relevantes dentro de la Unión Europea, y una serie de ejemplos de uso con casos reales (tanto beneficiosos como perjudiciales).

CAPÍTULO 2

Principales técnicas y enfoques de inteligencia artificial utilizados en el desarrollo de software

Como se ha explicado anteriormente, la Inteligencia Artificial (IA) está provocando un gran impacto en el sector del desarrollo del software, ya que ofrece una serie de nuevas herramientas y procedimientos que mejoran, optimizan y automatizan significativamente todas las tareas. Esto permite abordar desafíos más complejos de una forma más eficiente y efectiva, facilitando a los desarrolladores crear aplicaciones más adaptativas y robustas.

En este capítulo en concreto se analizarán las principales técnicas y enfoques de este área, cada uno de los cuales proporciona una serie de capacidades que pueden ser aprovechadas para mejorar una amplia gama de aspectos del desarrollo, que pueden ir desde la propia programación hasta la implementación y el mantenimiento. Dentro de cada subapartado, se explorará una visión detallada de cada método, destacando en particular que aplicaciones, beneficios y utilidades pueden ofrecer. Al terminar de comprender y aplicar estas estrategias, los desarrolladores deberían tener una visión mucho más clara y completa sobre como aprovechar al máximo esta tecnología y como crear soluciones innovadoras y de alta calidad. En concreto, se explorarán las siguientes cinco técnicas (véase la figura 2.1):

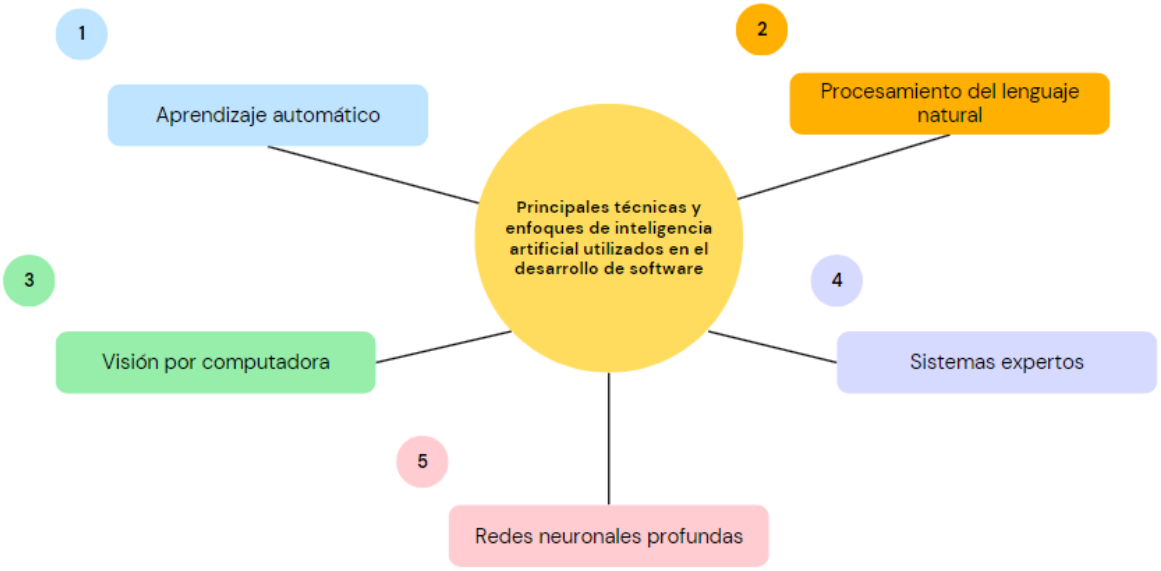

Figura 2.1: Técnicas principales de la inteligencia artificial

2.1. Aprendizaje automático

El aprendizaje automático (Machine Learning o, por sus siglas, ML) es un procedimiento por el cual las computadoras aprenden a hacer sus tareas sin que nadie les diga exactamente cómo hacerlo, haciendo que en vez de seguir instrucciones precisas, usen datos y descubran patrones por su cuenta.

Esto funciona mediante un extenso análisis de información (como pueden ser imágenes o textos) para encontrar una serie de patrones, usándolos luego para crear una especie de "modelo" que les sirve para hacer predicciones o tomar decisiones. Esto implica que cuantos más datos reciben, mejor se harán estas predicciones (de manera similar a como las personas nos adaptamos a realizar una tarea mediante la práctica).

Esto es de gran utilidad cuando las muestras y las situaciones varían mucho o cuando es de gran dificultad escribir instrucciones detalladas para cada posible escenario. Algunos ejemplos de esto serían el reconocimiento facial, la predicción del clima y recomendar series o películas.

Dentro de la inteligencia artificial, el aprendizaje automático se considera un subconjunto de este y le permite al sistema realizar tareas que normalmente requieren de inteligencia humana, como pueden llegar a ser el reconocimiento de objetos y la toma de decisiones. Al igual que como se ha explicado anteriormente, gracias a esta técnica la IA será capaz de aprender cosas por su

propia cuenta mediante el uso de información y ejemplos (en lugar de instrucciones detalladas). Una de las maneras para poder lograr esto es mediante el uso de redes neuronales, las cuales imitan el funcionamiento del cerebro de los seres humanos[6].

2.1.1. Tipos de aprendizaje automático

Hay cuatro tipos principales de aprendizaje automático, cómo se puede observar en la figura 2.2[7]:

Figura 2.2: Tipos de aprendizaje automático

- **Supervisado:** El modelo aprende gracias a la intervención de un "maestro", el cual es un científico de datos que le proporciona una serie de ejemplos ya clasificados y etiquetados. Un ejemplo de esto seria (en el caso del reconocimiento facial) entrenar una arquitectura a base de otorgarle una gran cantidad de fotos de caras de personas, clasificando cada una por los rasgos de esa persona (sexo, color de los ojos, forma de la nariz, etc.), haciendo luego que la computadora estudie estos ejemplos y aprenda a reconocer a los seres humanos por su cuenta.
- **No supervisado:** A diferencia del anterior tipo, en este el sistema aprende por sí mismo sin la necesidad de etiquetas o respuestas dadas previamente, proporcionándole una gran cantidad

de información sin clasificar y pidiéndole que encuentre patrones por sí mismo. Básicamente lo que hará con esto es agrupar los datos similares y le asignará sus propias etiquetas.

- **Semisupervisado:** Como se puede intuir por su propio nombre, combina el uso de información etiquetada y no etiquetada, en concreto primero usa una serie de datos marcados para así aprender a identificarlos y luego con ese conocimiento adquirido clasifica los que no lo están. Es de gran utilidad cuando hay una gran cantidad de registros pero pocos de ellos están identificados.
- **Por refuerzo:** A través de este método el sistema aprende a tomar decisiones a base de prueba y error, haciendo que cuando acierte reciba una "recompensa" y cuando falle sea "penalizado". Esto hará que con el tiempo los juicios tomados sean en su mayoría correctos para maximizar estas "recompensas", lo que se suele utilizar en situaciones donde se requiere aprender a través de la experiencia directa y la retroalimentación continua.

Dependiendo de las necesidades, la finalidad y la información que disponga cada tipo de negocio se debe analizar cual de estos casos es de más utilidad y en cual se requiere el menor uso de recursos posibles[8].

2.1.2. Ventajas del aprendizaje automático

En base a lo explicado previamente, podemos afirmar que el aprendizaje automático nos garantiza los siguientes beneficios[6]:

- **Reducción de costos:** Automatiza tareas permitiendo que los desarrolladores se enfoquen en otros procesos más importantes.
- **Identificación de la información:** Facilita el reconocimiento de patrones y estructuras de diferentes tipos de datos, ayudando así a interpretar su significado.
- **Mitigación de riesgos:** Detecta y se amolda a nuevas tácticas de fraude, previniendo este tipo de actividades de manera efectiva.
- **Optimización de la experiencia del usuario:** Mejora la interacción con los clientes mediante el uso de interfaces adaptativas, contenido personalizado, chatbots y asistentes virtuales.
- **Predicción del comportamiento del cliente:** Analiza los datos para poder detectar patrones y tendencias, mejorando las recomendaciones de productos y la satisfacción del cliente.
- **Mejora de la calidad de los datos:** Perfecciona de manera continua la minería de datos, incrementando su calidad y precisión con el tiempo.

2.2. Procesamiento del lenguaje natural

El procesamiento del lenguaje natural (NLP) es una rama de la inteligencia artificial que permite a los sistemas la capacidad de interpretar y trabajar con el lenguaje humano, combinando el mundo de la ciencia de la computación, los idiomas y la IA (figura 2.3). Hoy en día las empresas cuentan con una cantidad de datos bastante amplia en formato de voz y texto (como pueden ser los emails, redes sociales, etc.), por lo que utilizan esta tecnología para analizar esta información, determinando la intención o el sentimiento detrás del mensaje y respondiendo de forma acorde[9].

Figura 2.3: Fusión de computación, IA y lenguaje en el procesamiento del lenguaje natural

2.2.1. Tareas del procesamiento del lenguaje natural

Nuestro lenguaje posee una serie de características que implican una gran dificultad de análisis por parte de los programas informáticos. Algunas de estas son el uso del sarcasmo, las metáforas y las variaciones en las estructuras de las oraciones, las cuales deben ser consideradas por los desarrolladores para enseñar a estos modelos desde el inicio para que sean efectivas. Para poder realizar esto, el NLP (por sus siglas en inglés) divide los datos de texto y voz en tareas específicas que permiten una mayor compresión. Algunas de estas son[10]:

- **Desambiguación de palabras:** Este proceso es el encargado de diferenciar con que sentido o significado se usa una palabra en caso de que esta tenga varios de estos. Como muestra estaría la palabra “banco”, la cual tiene tres significados principales (institución financiera, mueble para sentarse y conjunto de peces) y se debe identificar cual de ellos es el que esta presente en cada contexto.

- **Resolución de correferencias:** Es la actividad de reconocer cuándo diferentes palabras o frases en un texto se refieren a la misma cosa o persona. Un ejemplo de esto sería en la frase: "Juan fue al parque. Allí, él encontró a su amigo y jugaron al fútbol", esta tarea ayuda a entender que "él" se refiere a "Juan". Hay que tener especial cuidado aquí con el uso de metáforas o modismos, ya que por ejemplo en la frase "El profesor es un faro de sabiduría para los estudiantes" hay que relacionar "faro de sabiduría" con "profesor".
- **Conversión de voz a texto:** En el caso de la existencia de programas que solo siguen comandos de voz o responden preguntas habladas es de vital importancia transcribir los datos de voz a texto y viceversa. Aquí nos encontramos con problemas que puede llegar a acarrear la propia voz, como puede ser la velocidad, la gramática incorrecta y el acento.
- **Generación de lenguaje natural:** Es la tarea opuesta del reconocimiento de voz, ya que convierte información estructurada (datos de fácil entendimiento por la inteligencia artificial) a lenguaje natural.
- **Etiquetado gramatical:** Determina la función gramatical que tienen las palabras en determinados contextos, es decir, identifica si una palabra es un verbo, sustantivo o adjetivo cuando se puede usar en varios de estos términos.
- **Reconocimiento de entidades nombradas:** Es importante saber cuando nos referimos a alguna entidad o organización importante, por lo que esta actividad se encarga de reconocer este tipo de expresiones, como podrían ser "Google" como nombre de una empresa existente o "Italia" como nombre de un país.
- **Análisis de sentimientos:** Por último, para poder lograr ciertos objetivos es muy importante detectar que sienten las personas a la hora de escribir un texto, por lo que esta labor se encarga de extraer emociones, actitudes y otras cualidades subjetivas en estas situaciones.

2.2.2. Ejemplos de uso

Algunos ejemplos de uso son los que van a continuación, que son de gran utilidad por:

- **Atención al cliente:** Responder de forma automática preguntas frecuentes y ayudar a los clientes en tiempo real gracias a los chatbots y asistentes virtuales, reduciendo los tiempos de espera y aumentando la eficiencia, y evaluar comentarios y reseñas en distintos sitios web para medir el grado de satisfacción y detectar problemas.
- **Gestión de correos electrónicos:** Organizar y priorizar los correos entrantes para redirigirlos al departamento o a la persona adecuada, además de generar respuestas automáticas a todos los mensajes genéricos o rutinarios.

- **Marketing y análisis de mercado:** Monitorear redes sociales para entender como los clientes perciben la marca, los productos o servicios y con ello adaptar los anuncios y campañas de publicidad según sus preferencias.
- **Reclutamiento y gestión de recursos humanos:** Filtrar y analizar de manera autónoma currículums para identificar los mejores candidatos para cada vacante e interpretar los resultados de encuestas para medir el compromiso y satisfacción de los trabajadores.
- **Seguridad y cumplimiento:** Detectar transacciones y comunicaciones para la detección de fraudes y vigilar que las comunicaciones cumplan con las regulaciones y políticas internas.
- **Investigación y desarrollo:** Extraer información relevante de grandes documentos científicos para acelerar investigaciones y ayudar a redactar y mejorar este tipo de informes.

2.3. Visión por computadora

La visión por computadora es un área de la inteligencia artificial que permite el entendimiento y procesamiento de los datos a partir de imágenes y vídeos, ya que permite a los sistemas ver, analizar y comprender el mundo visual.

Esta tecnología imita en gran parte el funcionamiento del ojo de los seres humanos, pero con una diferencia clave, y es que las personas tienen muchos años de experiencia en el aprendizaje del reconocimiento de objetos, movimientos, distancias y anomalías en imágenes, mientras que la visión artificial debe ser entrenada mucho más rápido usando cámaras, datos y algoritmos. A pesar de ello, esta ya más que comprobado que estos modelos pueden ser entrenados para inspeccionar productos o supervisar procesos de forma mucho más eficiente que nosotros, analizando miles de elementos por minuto para detectar cualquier error o fallo que a nosotros se nos podría pasar desapercibido[11].

Al igual que el aprendizaje automático, hay dos métodos de aprendizaje por los cuales un sistema puede llegar a desarrollar esta función, los cuales son la enseñanza guiada y la enseñanza autónoma. En el primer caso el modelo recibe una serie de instrucciones previamente etiquetadas y analizadas para que se acostumbre a determinados tipos de imagen y vídeo y así poder responder antes ellos más efectivamente, y en el segundo solamente se le asignan unas pautas básicas para que por su propia cuenta reconozca patrones o interprete este contenido de manera efectiva (normalmente es usado en entornos más complejos)[12].

2.3.1. Ventajas y casos de uso

Las principales ventajas que aporta este tipo de tecnología con las siguientes:

- **Automatización:** Permite la autonomización de diferentes tareas que requieren la interpretación visual, lo que aumenta la eficiencia y reduce la necesidad de la intervención humana. Por ejemplo, en una fábrica se pueden llegar a encargar de inspeccionar las piezas a medida que salen de producción detectando si tienen algún defecto y clasificándolos según sus características.
- **Precisión:** Normalmente este tipo de tecnología posee un alto grado de precisión, superando con creces la humana en labores de reconocimiento y clasificación. Un ejemplo claro de esto es en el campo de la medicina, donde tiene la capacidad de analizar imágenes médicas y detectar anomalías que una persona no podría sin hacer las pruebas adecuadas.
- **Velocidad:** Como se ha dicho anteriormente, también poseen la capacidad de realizar estas tareas a una velocidad que nos supera con creces, lo que es de gran utilidad en sistemas de vigilancia en tiempo real, permitiendo un procesamiento rápido de las imágenes y vídeos para poder tomar una decisión inmediata.
- **Funcionamiento continuo:** A diferencia de nosotros, este tipo de modelos pueden funcionar correctamente sin necesidad de descanso alguno, cosa que es especialmente útil en entornos de vigilancia y monitorización. Por ejemplo, en la seguridad de unas instalaciones este tipo de sistemas pueden controlar continuamente toda la zona alertando al personal de seguridad cuando se detecte una actividad sospechosa.
- **Sustitución de humanos en entornos peligrosos:** Implica una amplia mejora en la seguridad de entornos industriales, en especial en aquellos que implican el uso de maquinaria de alto riesgo. Esto ayuda a que haya un menor riesgo de accidentes y a evitar la contaminación cruzada, facilitando una limpieza y esterilización más eficaz al haber menos personas involucradas.

2.4. Sistemas expertos

Los sistemas expertos son un tipo de programa diseñado para imitar el conocimiento y el proceso de toma de decisiones de un profesional en su determinado campo, es decir, proporcionan a los usuarios recomendaciones, soluciones y diagnósticos del mismo nivel que lo haría un experto en su propia área.

Esta tecnología funciona a través de dos componentes, la base de conocimientos y el motor de inferencia. El primero se encarga del almacenamiento de toda la información y reglas (proporcionadas por expertos) que se necesitan para la toma de decisiones, mientras que el segundo usa estos datos y los aplica en determinadas situaciones para la resolución de problemas o la recomendación de medidas.

2.4.1. Tipos de sistemas expertos

Esta tecnología se puede clasificar según la manera en que funcionan y cómo su motor toma las decisiones (figura 2.4[13]). A continuación se explican tres principales[14]:

Figura 2.4: Tipos de sistemas expertos

- **Basados en reglas (RBR):** Usan reglas (suelen estar basadas en lógica difusa) establecidas de antemano para llegar a las conclusiones, empezando con una posible solución y buscando pruebas que lo reafirmen.
- **Basados en casos (CBR):** Parten de información existente y la utilizan para la toma de decisiones, analizando problemas similares resueltos previamente y adaptando esas soluciones a la situación actual.
- **Basados en redes bayesianas:** A partir de principios estadísticos y el teorema de Bayes[1] hacen predicciones, clasificaciones y diagnósticos, los cuales son de gran utilidad en el campo de la medicina para predecir y diagnosticar enfermedades.

2.4.2. Ventajas y casos de uso

Las principales ventajas con las que cuentan este tipo de sistemas son[15]:

[1]El teorema de Bayes es una herramienta clave en la probabilidad y la estadística que ayuda a ajustar nuestras creencias sobre la probabilidad de algo cuando recibimos nueva información. Básicamente, nos permite recalcular la probabilidad de que algo sea verdad, utilizando tanto lo que ya sabíamos como la nueva información que obtenemos.

- **Claridad en el razonamiento:** Una característica a destacar es su capacidad para explicar el proceso por el cual ha pasado la solución hasta su etapa final, detallando el motivo de sus conclusiones y facilitando la comprensión y confianza en el sistema. Un ejemplo de esto sería en las tareas de diagnóstico, donde pueden detallar el proceso que siguieron para determinar una enfermedad, explicando cómo los datos de los síntomas y antecedentes llevaron al diagnóstico final.
- **Adaptabilidad a nuevas situaciones:** A diferencia de los seres humanos, estos modelos se pueden adaptar con mucha rapidez a nuevas condiciones y requisitos, teniendo la capacidad de incorporar nueva información y convertirla en reglas de inferencia, lo que lo hace de gran utilidad en trabajos de planificación y programación.
- **Uniformidad en las decisiones:** Tienen la capacidad de proporcionar soluciones consistentes en tareas repetitivas y en decisiones similares, lo que garantiza que en el caso de que aparezcan problemas parecidos se resuelvan de forma homogénea, asegurando uniformidad y fiabilidad en los resultados. Una situación en la que nos pueda servir esto es en la supervisión, donde se compararían datos en tiempo real para detectar condiciones de alarma, garantizando que estas se resuelvan siempre de la misma manera.
- **Operación sin restricciones humanas:** Al igual que ocurre con la visión por computadora y otros tipos de tecnología relacionados con esta, puede funcionar de forma continua sin las limitaciones físicas o cognitivas que nosotros tenemos, permitiendo un uso constante en la búsqueda de soluciones. Además de esto, al almacenar la información proporcionada por profesionales, actúan como un depósito permanente de datos valiosos, accesibles y actualizados. Esto es especialmente útil en la tarea de interpretación de datos complejos, como los procedentes de sensores en el ámbito del IoT (Internet of Things[2]), donde el sistema puede operar 24/7, analizando y procesando información sin descanso.

2.5. Redes neuronales

Las redes neuronales son métodos de aprendizaje automático (Machine Learning) que toman decisiones de manera similar a como lo haría el cerebro humano, funcionando como una red de neuronas artificiales que procesan los datos para identificar patrones, evaluar opciones y llegar a conclusiones. Estas redes están formadas por varias capas de nodos: la de entrada, una o más ocultas y la de salida (figura 2.5[16]). Cada nodo de una capa está conectado a los nodos de la capa siguiente y tiene una ponderación y un umbral. En el caso de que la salida de un nodo supere su umbral, este se activa y envía la transmisión a la capa siguiente (si no ocurre no pasa nada)[17].

[2]El Iot es una red de dispositivos físicos interconectados que pueden recopilar, compartir y actuar sobre datos a través de internet. Estos incluyen desde electrodomésticos y vehículos hasta sensores industriales, y están equipados con tecnología que les permite comunicarse entre sí y con sistemas centralizados, mejorando la automatización y el control en diversos entornos.

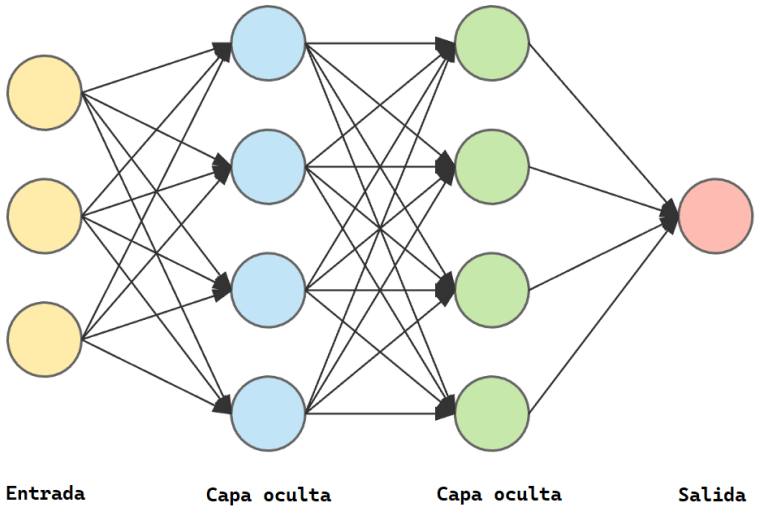

Figura 2.5: Capas de las redes neuronales

Esta tecnología aprende y mejora su precisión mediante el entrenamiento de los datos, aumentando la eficacia en tareas como el reconocimiento de voz e imágenes y procesando la información mucho más rápido que nosotros. Esto es muy similar a diferentes técnicas que se han nombrado anteriormente, y es que para poder lograr la visión por computadora, el procesamiento del lenguaje natural, el reconocimiento de voz y las recomendaciones personalizadas hay que hacer uso de las redes neuronales[18].

2.5.1. Tipos de redes neuronales

A continuación se explican por encima los tipos principales:

- **Redes Neuronales Artificiales (ANN):** Son su tipo más simple y funcionan exactamente como se ha explicado en el apartado anterior, es decir, los nodos de cada capa están completamente conectados a los de la siguiente y existe una capa de entrada, otra de salida y una o más ocultas. Se caracterizan por ser muy versátiles y por usarse en una amplia variedad de aplicaciones que tienen como finalidad la clasificación y la regresión.
- **Redes Neuronales Convolucionales (CNN):** Se utilizan principalmente para procesar datos estructurados en forma de rejilla (pueden organizarse bidimensionalmente), usado capas de convolución para aplicar filtros que detectan cualidades importantes, seguidas de capas de pooling que reducen la dimensión de la información mientras conservan las características importantes. Esto las hace increíblemente útiles en tareas de reconocimiento de imágenes, detección de objetos y análisis de vídeos.

- **Redes Neuronales Recurrentes (RNN):** Son adecuadas para el manejo de información secuencial donde su orden es de vital importancia. Su rasgo principal es que poseen conexiones que permiten la persistencia de los datos y que estos sean usados en el procesamiento de entradas posteriores, aunque cabe destacar que tienen dificultades con las dependiencias a largo plazo debido a problemas como el desvanecimiento del gradiente.

- **Redes de Memoria a Largo Plazo (LSTM):** Son una mejora de las redes neuronales recurrentes, ya que estas si que están diseñadas para mantener los registros por largos periodos de tiempo. Usan celdas de memoria y mecanismos de puertas para regular el flujo de transmisión, lo que las hace ideales para tareas como la traducción automática, el reconocimiento de voz y el análisis de series temporales complejas.

- **Redes Generativas Antagónicas (GAN):** Consisten en dos redes que compiten continuamente entre sí, donde una de ellas genera datos falsos y la otra intenta distinguir entre cuales de ellos son falsos y ciertos. Gracias a esto ambas mejoran progresivamente en su campo y son conocidas por su capacidad de generación de imágenes, vídeos y otros tipos de contenidos sintéticos de alta calidad.

- **Autoencoders:** Se basan en el aprendizaje de la compresión de la información de entrada a una representación de menor dimensión y luego en su reconstrucción a partir de esta última. Se logra mediante una capa de codificación que reduce la dimensionalidad y otra de decodificación que intenta recuperar su forma inicial, y se usan principalmente para la eliminación del ruido y la generación de contenido sintético.

- **Tranformers:** Utilizan un mecanismo de atención que permite a la propia red enfocarse en diferentes partes de una secuencia al mismo tiempo, permitiendo un manejo a largo plazo de las dependencias más efectiva. Este tipo de tecnología es la base de sistemas más avanzados como GPT-3 y BERT, que se usan principalmente en la traducción automática, el resumen de textos y la generación del lenguaje natural.

CAPÍTULO 3

Herramientas más utilizadas

El grado de éxito de las técnicas explicadas anteriormente depende en gran medida de las herramientas que usemos durante su ciclo de desarrollo, facilitando etapas que comprenden desde la conceptualización y prototipado hasta la implementación y mantenimiento.

En este apartado se exploran los principales tipos utilizados en la creación y puesta en marcha de la inteligencia artificial, destacando en concreto aquellas que son más efectivas y populares en el sector. A través de una revisión detallada de entornos de desarrollo integrado (IDE), bibliotecas y frameworks de machine learning, plataformas de computación en la nube, herramientas de análisis y visualización de datos, plataformas de AutoML, frameworks de implementación de modelos y herramientas de preprocesamiento y limpieza de datos, se proporcionará una visión comprensiva de las tecnologías que están impulsando la innovación en la IA.

Cada una de estas juega un papel crucial en el desarrollo, y su uso correcto puedo marcar la diferencia entre el éxito y el fracaso de un proyecto, por lo que además de lo anteriormente mencionado, también se proveerán una serie de ejemplos específicos de cada una, destacando su utilidad e importancia en este área.

3.1. Entornos de Desarrollo Integrado(IDEs)

Un entorno de desarrollo integrado (IDE) es una aplicación que mejora significativamente el proceso de programación, permitiendo a los propios programadores escribir código de forma más efectiva, sencilla y organizada. Esto es posible ya que esta herramienta combina en una sola aplicación diversas funciones esenciales, como puede ser la edición de código, la compilación, las pruebas y el empaquetado del software, haciendo que en vez de tener que configurar y manejar múltiples programas por separado, se pueda empezar a desarrollar nuevas aplicaciones rápidamente, centrándose en una única plataforma.

Uno de los principales beneficios que aporta esta tecnología es la automatización de la escritura y la edición de código, ya que esta sugiere y completa esta tarea automáticamente, agilizando el trabajo del programador. Además de esto, realzan la sintaxis mediante el uso de colores y formatos distintos para cada tipo de palabra, lo que no solo mejora la legibilidad, sino que también proporciona retroalimentación instantánea sobre posibles errores sintácticos.

Otra ventaja importante a destacar es la manera en la que ofrece sugerencias inteligentes mientras se está trabajando, es decir, que a la hora de empezar a escribir una línea del programa, este ofrece sugerencias para terminarla más rápidamente y evitando que te equivoques al escribir (como se puede observar en la figura 3.1). En adición a esto, también tienen la capacidad de refactorizar, permitiendo la reestructuración del código para mejorar su entendimiento y eficiencia sin alterar su funcionalidad original, lo que beneficia sobre todo a que otros miembros del equipo que no han participado en su desarrollo comprendan y colaboren en el propio proyecto.

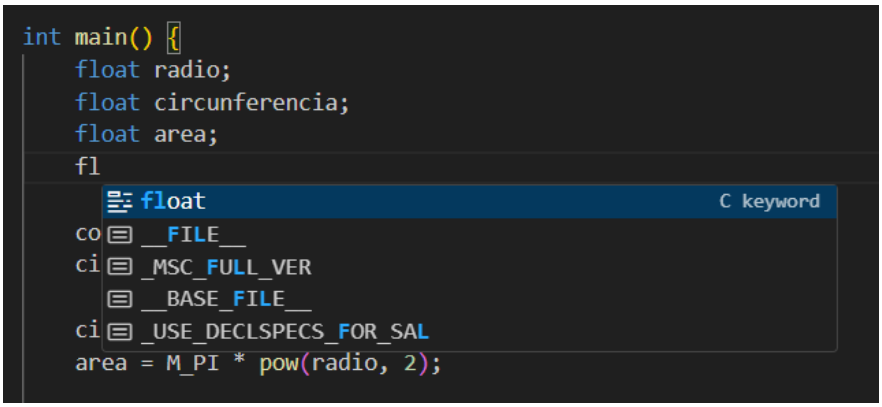

Figura 3.1: Sugerencias de Visual Studio Code

Los IDEs también pueden automatizar una serie de tareas recurrentes, como puede ser la compilación, la cual convierte el programa a un formato que el sistema operativo pueda ejecutar (en concreto se suele utilizar la compilación "justo a tiempo", la cual funciona constantemente a tiempo real). Además pueden automatizar las pruebas unitarias, asegurando que todo funcione correcta-

mente antes de su integración con el trabajo de otros desarrolladores y posibilitando la realización de pruebas más complejas.

Como última característica de esta herramienta cabe destacar la depuración, la cual permite seguir el código línea por línea mientras se está ejecutando, ayuda a identificar y corregir los errores que surjan mientras se trabaja, y hace que sean más fáciles de detectar los fallos en tiempo real (incluso mientras se está escribiendo), lo cual es fundamental para mantener la calidad y funcionalidad del propio programa[19].

3.1.1. Tipos y ejemplos de Entornos de Desarrollo Integrados

A continuación se presentarán los principales tipos de IDEs junto con algún ejemplo real:

- **IDEs Generales:** Son los más versátiles y pueden ser usados para una amplia gama de lenguajes de programación. En concreto cabe destacar Visual Studio Code (VS Code), el cual es el más popular y es altamente personalizable, ya que a través del uso de extensiones permite una gran variedad de lenguajes, como Python, C++, JavaScript y muchos más.
- **IDEs Específicos para un Lenguaje:** Están optimizados para un lenguaje de programación en específico, proporcionando una serie de herramientas y funciones especializadas en este. Un ejemplo de esto es PyCharm, que está diseñado específicamente para Python, incluyendo características avanzadas para el desarrollo de aplicaciones en este entorno e incluyendo soporte para frameworks como Django y Flask.
- **IDEs para Desarrollo Web:** Se especializan unicamente en el diseño de aplicaciones web, ofreciendo cualidades como la previsualización en tiempo real, la depuración integrada y herramientas para trabajar con HTML, CSS y JavaScript. Un IDE muy popular en este sector es Atom, conocido ampliamente por su flexibilidad y amplia gama de extensiones.
- **IDEs para Desarrollo de Aplicaciones Móviles:** Están dirigidos a la creación de aplicaciones móviles para plataformas como Android e IOs, y poseen una gran cantidad de recursos para trabajar en este ámbito. Android Studio es el mejor ejemplo para describir esta tecnología, ya que es el IDE oficial para desarrollar en Android y ofrece una serie de características especializadas en este sistema.
- **IDEs para Ciencia de Datos y Análisis:** Son los diseñados concretamente para esta área, debido a que ofrecen una serie de herramientas para trabajar con datos, realizar análisis estadísticos y crear modelos de Machine Learning (ML). Concretamente Jupyter Notebook ofrece un entorno interactivo en estos campos, permitiendo la creación y el intercambio de documentos que contienen código, visualizaciones y narrativas explicativas.

- **IDEs para Desarrollo de Juegos:** Están optimizados para el desarrollo de videojuegos, pudiendo trabajar concretamente con gráficos, fisicas, sonido y otros aspectos del sector. El más importante y relevante es Unity, usado en la creación de videojuegos en 2D y 3D.

3.2. Bibliotecas y frameworks de Machine Learning

Las bibliotecas y frameworks de inteligencia artificial desempeñan un papel muy importante a la hora de impulsar soluciones inteligentes, ofreciendo un enfoque estructurado y eficiente para construir sistemas sofisticados y confiables.

A la hora de tener que elegir que framework usar, hay que considerar cuales son las necesidades de tu proyecto, la experiencia de tu equipo, los recursos de los que se disponen y la escalabilidad requerida. Son un conjunto de herramientas que aceleran el trabajo, simplifican procesos complejos y permiten reusar componentes de forma efectiva. Además de esto, se pueden usar con una gran variedad de lenguajes de programación[20].

Esta tecnología se divide en tres componentes, los algoritmos de aprendizaje automático, los modelos pre-entrenados y las herramientas de visualización. Los primeros son la base fundamental, permitiendo al sistema aprender patrones y realizar predicciones a partir de los datos proporcionados. Los segundos son redes neuronales u otros modelos de Machine Learning que han sido entrenado previamente en grandes conjuntos de información y son de gran utilidad porque pueden ser ajustados para tareas específicas con conjuntos más pequeños. Los últimos permiten a los desarrolladores y científicos visualizar los datos, los resultados y otros aspectos del proceso, lo que incluye características como los gráficos, diagramas de dispersión y mapas de calor (como se puede observar en la figura 3.2[21]).

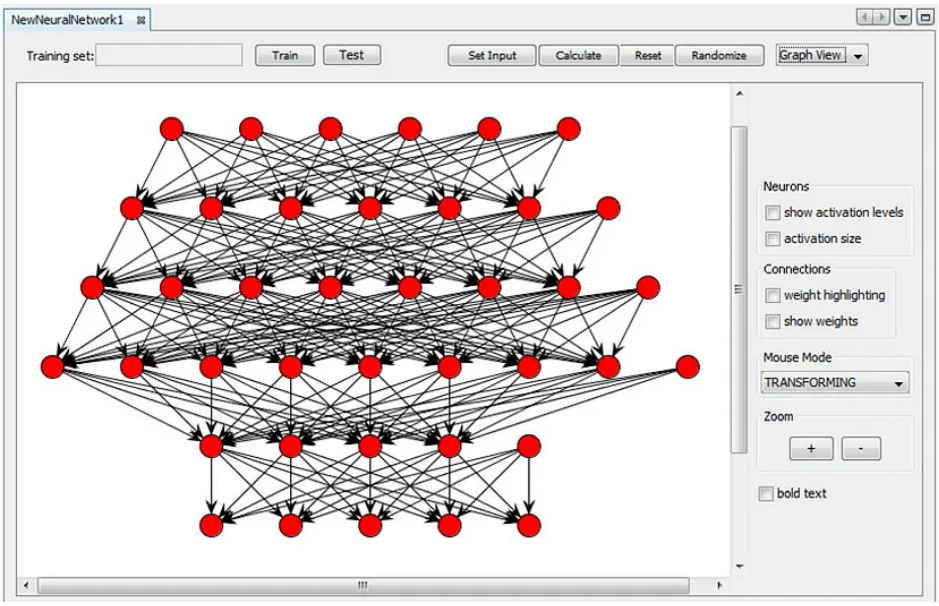

Figura 3.2: Herramienta de visualización en frameworks de Machine Learning

3.2.1. Tipos y ejemplos de bibliotecas y frameworks de Machine Learning

En este apartado se trataran los diferentes tipos de bibliotecas y frameworks junto con algunos ejemplos de aplicaciones reales:

- **Por el lenguaje de programación:** Al igual que los IDEs mencionados anteriormente, están diseñados para trabajar con un lenguaje de programación en particular. Algunos ejemplos son TensorFlow y Keras en Python y Weka en Java.
- **Por el enfoque de aprendizaje:** Las herramientas de aprendizaje automático pueden diferir en el tipo de algoritmos y técnicas que admiten, ya que algunas se centran en el aprendizaje supervisado, otras en el no supervisado o el aprendizaje profundo, y algunas son más versátiles y admiten múltiples enfoques.
- **Por su propósito y dominio de aplicación:** Están diseñados específicamente para aplicarse en determinadas áreas, como puede ser en la visión por computadora, el procesamiento del lenguaje natural y análisis financieros. Un claro caso de uso es OpenCV para visión por computadora y NLTK para procesamiento del lenguaje natural.
- **Por la escalabilidad y el rendimiento:** Algunas están optimizadas para manejar grandes volúmenes de datos y ejecutar cálculos de manera eficiente en paralelo, lo que las hace perfectas en entornos empresariales. Algunas herramientas usadas en este ámbito serían Apache Spark MLlib y H2O.ai.

- **Por la facilidad de uso y la curva de aprendizaje:** Por último, estos son usados o para ser accesibles por novatos y proporcionar una curva de aprendizaje más suave, o ser más complejos y ofrecer más flexibilidad a los usuarios experimentados. En el caso de las herramientas fáciles de usar por principiantes podemos encontrar ejemplos como Scikit-learn y RapidMiner.

3.3. Plataformas de computación en la nube

Las plataformas de gestión en la nube ofrecen una variedad de soluciones, como las de Plataforma como Servicio (PaaS) e Infraestructura como Servicio, las cuales proporcionan un panel de control unificado para supervisar y gestionar todas las cargas de trabajo y recursos en la nube. Son fundamentales para que las diferentes organizaciones de TI (tecnologías de la información) puedan administrar los servicios en la nube de forma eficiente, optimizar recursos, virtualizar servidores y cumplir con las normativas. Además de esto, muchas soluciones están basadas en software de código abierto, lo que permite evitar las dependencias con un solo proveedor[22].

Esta tecnología opera mediante la creación de un conjunto virtual de recursos compartidos, que ofrecen servicios de computación, almacenamiento y redes a través de Internet. Los usuarios tienen acceso a estos medios según sus requerimientos y generalmente solo pagan por lo que utilizan. Como se ha mencionado anteriormente, usan herramientas de virtualización que permiten la creación de múltiples máquinas virtuales en un único servidor físico, lo que posibilita la ejecución de diferentes sistemas operativos y aplicaciones independientes para cada cliente en el mismo servidor.

Estos servicios facilitan a las organizaciones el almacenamiento y respaldo de datos, el desarrollo y pruebas de aplicaciones, el acceso a bases de datos en la nube, el análisis de grandes volúmenes de datos, la distribución global de software bajo demanda, el acceso a herramientas de inteligencia empresarial y la creación de aplicaciones nativas de la nube. Por último, es importante resaltar que se dividen principalmente en dos tipos: las plataformas públicas y las privadas (como se puede observar en la figura 3.3)[23].

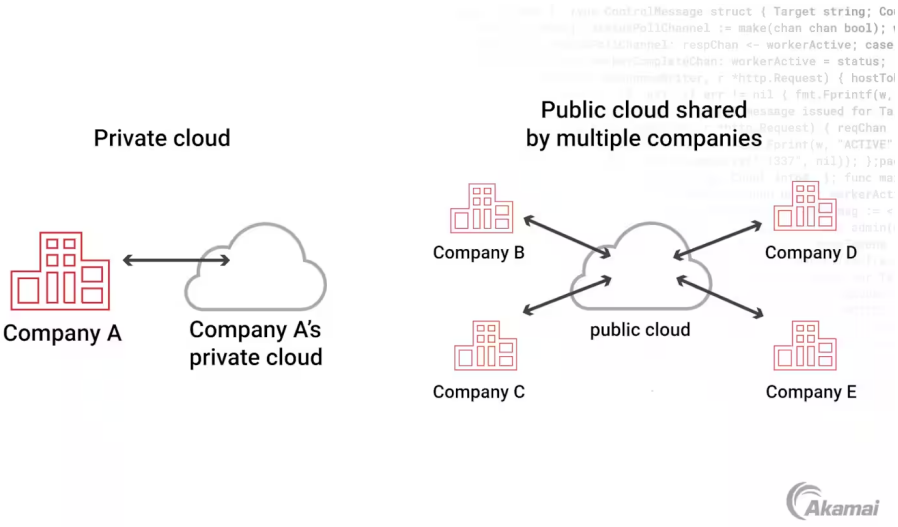

Figura 3.3: Plataformas de computación en la nube privadas y públicas

3.3.1. Ejemplos de plataformas de computación en la nube

Se presentarán una serie de herramientas populares del sector de la inteligencia artificial explicando sus particularidades y beneficios, para que dependiendo de las necesidades técnicas y empresariales se elija la que mejor se adapte:

- **Amazon Web Services (AWS):** AWS es una de las plataformas de nube más completas y adoptadas mundialmente, ya que su servicio de IA (Amazon SageMaker) facilita a los profesionales la creación, entrenamiento y despliegue de modelos de aprendizaje automático a gran escala. Además de esto también simplifica cada paso del flujo de trabajo de esta tecnología, proporcionando herramientas para la gestión de datos, entrenamiento y despliegue. Sus principales ventajas de uso son la escalabilidad, la amplia variedad de servicios integrados y la flexibilidad de pagar solo por lo que se usa.

- **Microsoft Azure:** Ofrece servicios con un fuerte enfoque en la inteligencia artificial a través de su palaforma Azure AI. Dentro de este cabe destacar Azure Machine Learning, ya que permite a sus usuarios construir, entrenar y desplegar este tipo de modelos, ofreciendo al mismo tiempo herramientas para la automatización de procesos y una integración completa con otros servicios de Microsoft. Se caracteriza principalmente por una integración profunda con otros productos de la misma organización y opciones híbridas que combinan servicios en la nube con infraestructura local.

- **Google Cloud Platform (GCP):** Es conocido ampliamente por sus avanzadas capacidades en análisis de datos y aprendizaje automático. En concreto (para el sector donde estamos

ubicados actualmente) su servicio Google AI Platform ofrece medios para construir y desplegar arquitecturas de ML y, junto al uso de TensorFlow (mencionada en el anterior apartado), permite a los usuarios diseñar modelos complejos y escalarlos para grandes volúmenes de datos. Además de esto proporciona servicios gestionados como AutoML, el cual concede un entrenamiento de alta calidad con un mínimo de experiencia en IA. Entre sus principales utilidades cabe destacar su uso en big data y un fuerte enfoque en la eficiencia y escalabilidad.

- **IBM Cloud:** Integra sus capacidades con IBM Watson, la cual es una plataforma que ofrece soluciones de IA para empresas. Esta suministra una serie de utilidades para el procesamiento del lenguaje natural, la visión por computadora, y el análisis predictivo, permitiendo a los profesionales construir modelos personalizados para sus necesidades específicas y desplegarlos en la nube. Sus principales ventajas son su integración con soluciones empresariales de IBM y su planteamiento en la seguridad y la confiabilidad.

- **Oracle Cloud:** Proporciona una variedad de recursos de computación en la nube a través de Oracle AI, admitiendo a las empresas desarrollar aplicaciones inteligentes utilizando herramientas como Oracle Machine Learning y Oracle Digital Assistant. Estas ayudan en la automatización de tareas, análisis de datos y sugieren una serie de recomendaciones basadas en este proceso. Se destaca por su robusta infraestructura y sus avanzadas capacidades en análisis de información y Machine Learning.

3.4. Herramientas de visualización y análisis de datos

La visualización de datos se refiere a la técnica de convertir esta información en representaciones más visuales como gráficos, mapas y otros elementos visuales, ayudando a desglosar los conjuntos complejos y facilitando la identificación de patrones, relaciones y tendencias. El propósito fundamental es transformar los datos sin procesar en información clara y usable (en la figura 3.4[24] se puede observar un ejemplo de esto).

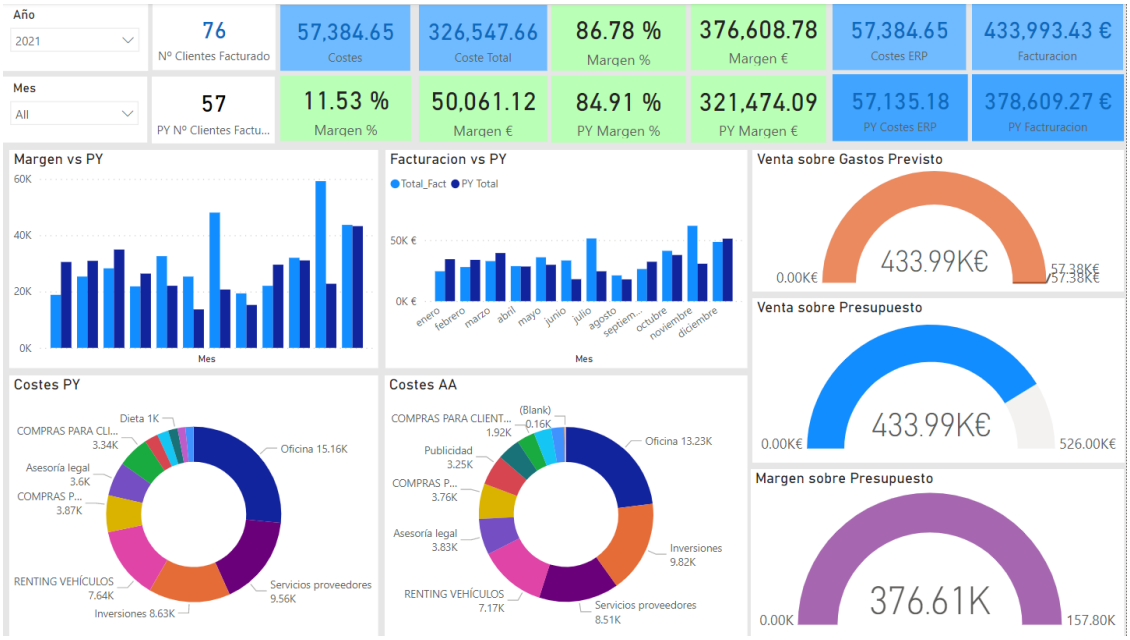

Figura 3.4: Interfaz de Power BI

En concreto, las herramientas de visualización y análisis de datos son software diseñado para permitir la creación y presentación visual de los registros, procesándolos de forma bruta, convirtiéndolos en representaciones más entendibles y favoreciendo el tratamiento de estos. Además de esto ofrecen una serie de funcionalidades, entre las cuales están la integración y organización de información hasta la creación de gráficos interactivos, paneles de control, reportes y la colaboración entre distintos usuarios, lo que permite su procesamiento a tiempo real, facilita su exploración y genera conocimientos compartidos. Por último, cabe destacar que garantiza la seguridad y cumplimiento de las normas en este tipo de actividades[25].

3.4.1. Ejemplos de herramientas de visualización y análisis de datos

Se mostrarán diversas herramientas destacadas en el campo de la inteligencia artificial, detallando sus características y ventajas individuales lo que permitirá seleccionar la más adecuada según las preferencias que se deseen:

- **Tableau:** Es un programa de visualización de datos muy reconocida por su habilidad para convertir información compleja en gráficos y paneles interactivos. Su integración con IA se realiza a través de funcionalidades como Tableau Prep, que ayuda a preparar y limpiar estos conjuntos automáticamente, y Explain Data, que proporciona insights[1] sobre anomalías

[1]Los "insights" son percepciones, entendimientos o revelaciones obtenidas a partir del análisis de datos o de la observación de situaciones.

y tendencias. Los beneficios de usar esta tecnología incluyen su facilidad de uso, una variedad extensa de conexiones disponibles y la capacidad de generar visualizaciones altamente interactivas sin necesidad de conocimientos avanzados en programación.

- **Power BI:** Es una plataforma que permite a los usuarios crear informes e interfaces interactivas, ofrecer análisis automatizados y modelos de aprendizaje automático pre-entrenados gracias al uso de AI Insights. Esta herramienta se integra perfectamente con otros productos de Microsoft, como Excel y Azure, lo que facilita su uso en entornos empresariales que ya han adoptado estas utilidades previamente. Las organizaciones suelen usar este programa por su interfaz intuitiva, las potentes capacidades de análisis que posee y su integración con el ecosistema de Microsoft.

- **Google Data Studio:** Es una tecnología gratuita de Google para la creación de dashboards y visualizaciones de datos, el cual (junto con Google Analytics, BigQuery y otros servicios de Google Cloud) permite a los profesionales aprovechar modelos de aprendizaje automático para el análisis de información. Destaca por su facilidad de uso, la capacidad de integrar registros de diferentes fuentes, la opción de compartir informes dinámicos con otras personas y su potencial de gestionar grandes conjuntos de manera eficiente.

- **Qlik Sense:** Es una herramienta que utiliza inteligencia aumentada para ayudar a descubrir insights de manera más eficiente. Su motor asociativo único facilita la exploración de cualquier tipo de relación en los datos e incluye funcionalidades como las sugerencias automáticas de gráficos y análisis, y la capacidad de realizar preguntas en lenguaje natural para obtener respuestas directas. Las principales ventajas que derivan de su utilización son su potente motor de búsqueda, la facilidad para descubrir relaciones ocultas y una experiencia de usuario intuitiva y visualmente atractiva.

- **IBM Cognos Analytics:** Es un programa de inteligencia de negocios que ofrece una inspección automatizada y genera recomendaciones, al igual que permite la exploración de datos mediante lenguaje natural. Sobresale por su habilidad para manejar grandes volúmenes de datos, su integración con otros productos de IBM y su destreza para proporcionar insights accionables a través de su avanzada inteligencia artificial. Los principales beneficios de uso que aporta incluyen una robusta infraestructura de evaluación, capacidades avanzadas de personalización y seguridad, y herramientas colaborativas.

- **Looker:** Es reconocido por permitir a las empresas explorar, analizar y compartir información en tiempo real, lo cual se logra mediante la utilización de modelos creados por los usuarios para ofrecer una perspectiva unificada de los datos en toda la organización. Esta tecnología se integra con BigQuery y otras soluciones de Google Cloud para aprovechar todos los beneficios posibles de la IA y el Machine Learning, entre los que destacan su flexibilidad para personalizar sistemas, potentes capacidades de integración y una experiencia de usuario centrada en la colaboración y el descubrimiento de insights.

3.5. Herramientas de preprocesamiento y limpieza de datos

El preprocesamiento y la limpieza de datos son dos etapas bastante importantes en cualquier proyecto relacionado con la inteligencia artificial. Como se ha explicado anteriormente, la información utilizada por estos sistemas deben estar en el formato adecuado para poder ser tratada correctamente por los algoritmos de aprendizaje automático. Algunos problemas comunes que pueden tener son valores faltantes, duplicados, errores y ruidos, así como formatos incorrectos, lo que afecta negativamente al rendimiento y a la precisión[26].

Este tipo de herramientas se utilizan para transformar los datos crudos en un formato limpio y estructurado para que pueda ser usado para el análisis y la construcción de modelos predictivos. También permiten a los profesionales realizar una gran variedad de tareas, tales como la eliminación de registros faltantes, la corrección de errores y ruidos, la normalización y escalado, la conversión y codificación, la detección y eliminación de duplicados, y la transformación y agregación de información (como se puede observar en la figura 3.5[27]).

Figura 3.5: Tareas de preprocesamiento y limpieza de datos

3.5.1. Ejemplos de herramientas de preprocesamiento y limpieza de datos

En esta sección se tratarán las principales herramientas destinadas a las tareas mencionadas previamente, destacando sus principales ventajas de uso en cada caso:

- **Pandas:** Es una biblioteca de Python que ofrece estructuras y herramientas de análisis de datos de alto rendimiento y de fácil uso. Es especialmente utilizada en áreas donde se emplean grandes volúmenes de información y donde se realizan operaciones complejas, ya que permiten

la manipulación y el estudio de datos tabulares, similar a como se hace en las hojas de cálculo de Excel pero con mayor flexibilidad y potencia.

Sus principales ventajas de uso son la capacidad de gestionar registros, filas y columnas faltantes, la normalización y escalado, la transformación mediante codificación, la agregación y agrupación para el análisis estadístico, y la compatibilidad con otras bibliotecas de IA.

- **NumPy:** Es una tecnología utilizada para el procesamiento numérico y la manipulación de matrices. A pesar de que no está diseñada plenamente para la limpieza, proporciona las bases para muchas operaciones de preprocesamiento que se realizan en las etapas iniciales del desarrollo de inteligencia artificial.

  Las claves que hacen esta aplicación tan usada son la creación y manipulación eficiente de arrays, su amplia variedad de funciones matemáticas y estadísticas, y su rendimiento.

- **Dask:** Es otra biblioteca de Python diseñada concretamente para la computación paralela y distribuida. Se utiliza para escalar el procesamiento de información en múltiples núcleos de CPU o clústeres completos, lo que lo hace perfecto para el manejo de grandes volúmenes de datos que no pueden ser alojados en la memoria de un solo equipo. Se destaca principalmente por su capacidad de trabajar en paralelo y su compatibilidad e integración con otras aplicaciones de Python.

- **OpenRefine:** Es una herramienta de código abierto especialmente útil para datos desordenados y heterogéneos. Permite a los usuarios explorar grandes conjuntos de registros, identificar y corregir errores, transformar estos en diferentes formatos, y realizar operaciones complejas gracias a una interfaz gráfica interactiva.

  Cabe destacar que sus capacidades de clustering ayudan a agrupar y corregir los valores similares de forma automática, facilitando la normalización de información inconsistente. Su principal ventaja respecto a otras tecnologías es su facilidad de uso para aquellos que no están familiarizados en el sector, aunque no por ello perjudica a los profesionales más experimentados, ya que también ofrece capacidades avanzadas a través de su lenguaje de expresión (GREL), permitiendo realizar transformaciones complejas y personalizadas.

3.6. Herramientas de despliegue y monitoreo de sistemas

En el ámbito de la inteligencia artificial, el desarrollo de sistemas es solo una parte del proceso del ciclo de vida de este tipo de tecnologías, ya que una vez creado y validado, es de vital importancia desplegarlo en un entorno de producción donde pueda ofrecer un valor real. Sin embargo, esto tampoco finaliza aquí, ya que una vez acabado el despliegue, se debe de monitorear adecuadamente para asegurar su precisión, rendimiento y reducir los riesgos de posibles fallos que pueda cometer.

Estas herramientas están diseñadas para gestionar estas etapas del ciclo de vida de los modelos de IA, permitiendo a los profesionales automatizar y simplificar el proceso de puesta en producción, así como garantizar que estos funcionen de forma óptima y sin interrupciones. Además de esto, también ayudan a proporcionar un entorno estandarizado y escalable para habilitar su ejecución en varios entornos, ya sea en la nube o en local.

Una vez desplegado, el monitoreo de estos sistemas es esencial para detectar y corregir cualquier degradación en el rendimiento, que puede ser causada por cambios en la información de entrada, problemas de drift (desviación) de los datos o fallos en la infraestructura. Estas aplicaciones solucionan todos estos riesgos rastreando métricas clave, realizando auditorías y gestionando las versiones, asegurando su fiabilidad y precisión.

3.6.1. Ejemplos de herramientas de despliegue y monitoreo

A continuación se muestran las cinco aplicaciones mas usadas en el despliegue y la monitorización de sistemas:

- **Docker:** Es una plataforma que facilita a los desarrolladores empaquetar aplicaciones y sus dependencias en contenedores ligeros y portátiles. Estos se pueden ejecutar en cualquier entorno, ya que esta aplicación asegura que haya consistencia entre todos los entornos de desarrollo, prueba y producción. Dentro del sector de la tecnología cognitiva permite a los equipos el empaquetamiento de modelos de aprendizaje automático junto a todas sus dependencias, lo que simplifica enormemente el despliegue y reduce los problemas relacionados con la configuración del entorno. Además de esto, también mejora la escalabilidad horizontal al autorizar la ejecución de múltiples instancias de un contenedor para manejar mayores volúmenes de solicitudes.
- **Kubernetes:** Automatiza la implementación, el escalado y la gestión de aplicaciones de contenedores. Es especialmente útil para la implementación de modelos de IA gracias a su capacidad para gestionar grandes clústeres de contenedores, proporcionando alta disponibilidad y capacidad de recuperación automática ante fallos, junto al beneficio de su capacidad de escalado automático, permitiendo ajustar los recursos de forma dinámica según la carga de trabajo. Cabe destacar que también facilita las actualizaciones continuas y la gestión de versiones, pudiendo así mantener los sistemas actualizados y mejorar su rendimiento con nuevas versiones.
- **MLflow:** Es una aplicación que facilita la gestión del ciclo de vida de los modelos de Machine Learning, gracias a que incluye componentes para el seguimiento de experimentos, la gestión de modelos y su puesta a punto. MLflow Tracking permite el empaquetado de código en un formato reproducible, y MLflow Models permite la exportación de sistemas en un formato

estándar que se puede usar con diversas herramientas de despliegue. Una de las principales ventajas que aporta esta tecnología es su capacidad de integración con múltiples frameworks de aprendizaje automático junto a su habilidad de poner en marcha estructuras en diferentes entornos, incluyendo plataformas de nube como AWS, Azure y Google Cloud, así como en servidores locales y Kubernetes.

- **Prometheus:** Es una herramienta que se utiliza para recopilar y almacenar métricas en tiempo real. En el área en la que se sitúa este informe, se puede utilizar para vigilar el rendimiento de los sistemas, incluyendo métricas como el tiempo de respuesta, el número de solicitudes procesadas y el uso de recursos. En adición a esto, usa un lenguaje de consulta flexible llamado PromQL para extraer y analizar datos, y puede integrarse con plataformas de visualización como Grafana para crear paneles interactivos que muestran el estado y rendimiento de los modelos en producción. También cabe destacar que la capacidad de esta tecnología para establecer alertas basadas en condiciones definidas por el usuario permite a los equipos de operaciones reaccionar rápidamente ante problemas de rendimiento o fallos.

- **TensorFlow Serving:** Es una aplicación diseñada específicamente para desplegar y servir estructuras de Machine Learning en producción. Es parte del ecosistema de TensorFlow y está optimizado para servir infraestructuras entrenadas con esta plataforma, aunque también puede extenderse para servir otras de diversos frameworks. Es importante resaltar que facilita el despliegue sin interrumpir el servicio, permitiendo actualizaciones suaves y minimizando el tiempo de inactividad, además de que sus capacidades de monitoreo y logging integradas permiten rastrear el rendimiento y diagnosticar problemas de manera eficiente. El principal beneficio de su uso es su fuerte integración con TensorFlow, lo que simplifica el proceso de despliegue para los modelos desarrollados en este framework.

CAPÍTULO 4

Ética y consideraciones legales en el desarrollo de software con IA

Una vez explicado en detenimiento todas las ventajas y oportunidades que nos puede brindar el uso de la inteligencia artificial en el sector del desarrollo del software, es de vital importancia resaltar que retos significativos nos causa en términos éticos y legales. La complejidad de estas tecnologías obliga a adoptar unos fuertes enfoques responsables y alineados tanto con los principios deontológicos como con las normativas vigentes. Esta sección en concreto explora las buenas practicas sobre el uso de IA enfocándose en dos áreas de suma importancia: la legalidad y la ética.

Dentro de este nos encontramos con dos apartados, uno por cada área mencionada anteriormente. En el primero de ellos se examina el "Reglamento de Inteligencia Artificial" del Parlamento Europeo, aprobado en marzo de 2024[28], el cual proporciona un marco jurídico que regula el uso de IA en la Unión Europea, asegurando que se respeten los derechos fundamentales y se minimicen todos los riesgos que puedan presentar estas tecnologías. En el segundo se analiza el documento "Directrices éticas para una IA fiable", elaborado por la Dirección General de Redes de Comunicación, Contenido y Tecnologías de la Comisión Europea en colaboración con el Grupo de expertos de alto nivel en inteligencia artificial[29], el cual establece todos los principios deontológicos que deben guiar el desarrollo y la implementación de estos sistemas, promoviendo, entre otras cosas, la transparencia, la justicia y la responsabilidad.

El estudio y análisis de ambos documentos son fundamentales para garantizar un uso que no solo cumpla con las regulaciones, sino que también respete los valores y principios de la sociedad. En este contexto, la deontología juega un papel crucial al orientar las prácticas profesionales hacia

un comportamiento moralmente responsable y socialmente beneficioso en el desarrollo de este tipo de modelos.

4.1. Implicaciones legales

En esta sección se analizará el documento que más relevancia tiene en cuanto a la legalidad de los sistemas de inteligencia artificial, ya que es el más reciente, completo y elaborado hasta el momento. Este es el "Reglamento de Inteligencia Artificial" presentado el 13 de marzo de 2024[28], el cual tiene como objetivo principal armonizar el mercado interno mediante un marco legal coherente, con la finalidad de regular el desarrollo, la comercialización, la implementación y el uso de modelos en la Unión Europea.

Se busca con esto fomentar una tecnología confiable y centrada en los seres humanos, garantizando al mismo tiempo una protección robusta de la salud, la seguridad y los derechos esenciales, abarcando temas como la igualdad, la privacidad y la protección del medio ambiente. En adición a esto, se busca mitigar todos los efectos negativos que puedan surgir con el uso de estos sistemas y promover su innovación.

Anteriormente se ha mencionado que este reglamento era el más completo actualmente, y eso es porque este ha surgido de la escritura de varios documentos anteriores, entre los cuales cabe destacar la "Ley de Inteligencia Artificial" presentada en abril de 2021[30], que fue la primera de su tipo en el mundo, y las "Normas de Derecho Civil sobre Robótica" presentadas en febrero de 2017[31]. Son los más destacables ya que sentaron las bases éticas y legales en la UE sobre el uso de esta tecnología, además de que subrayaron la importancia de garantizar que estos productos respeten los derechos fundamentales de las personas.

En los siguientes subapartados se mencionarán todos los puntos relevantes que hay que tener en mente a la hora de usar modelos de IA en el desarrollo del software dentro de la UE:

4.1.1. ¿Que personas y sistemas están afectados por el reglamento?

Este marco normativo está destinado a todos los desarrolladores que están involucrados en la implementación y despliegue de sistemas de inteligencia artificial en la UE, lo que incluye a proveedores, responsables de implementación, importadores, distribuidores, fabricantes de productos con esta tecnología, representantes autorizados de proveedores fuera de la Unión y personas afectadas dentro de esta zona.

Más adelante se mencionarán tanto los modelos de alto riesgo como los de uso general y que obligaciones deben de acatar cada uno de ellos, pero primero es de vital importancia mencionar

cuales no son afectados por este documento y que, por ello, no deben de seguir las normas descritas en las siguientes secciones. Los entornos no afectados son los siguientes (en la figura 4.1 se pueden ver los que quedan exentos):

- Áreas que están fuera del ámbito del Derecho de la Unión.
- Competencias de los Estados miembros relacionadas con la seguridad nacional.
- Sistemas destinados exclusivamente a fines militares, de defensa o seguridad nacional.
- Autoridades públicas de terceros países y organizaciones internacionales que usan estos modelos en cooperación con la Unión para la aplicación de la ley y la cooperación judicial, siempre que estos ofrezcan garantías adecuadas para la protección de derechos y libertades fundamentales.
- Proveedores de servicios intermediarios.
- Plataformas desarrolladas exclusivamente para la investigación y el desarrollo científico.
- Actividades de investigación, pruebas o desarrollo antes de la introducción al mercado o puesta en servicio.
- Tecnologías usadas únicamente para actividades personales no profesionales por parte de personas físicas.
- Sistemas que se distribuyen con licencias libres y de código abierto, en caso de que se comercialicen o se utilicen de manera que los convierte en sistemas de alto riesgo, o si se usan de formas que están específicamente prohibidas por los artículos 5 o 50 del reglamento[1], entonces se les aplicarán reglas especiales más estrictas (las cuales no son las que se van a mencionar posteriormente).

[1]Estos artículos tratan sobre las prohibiciones específicas de ciertos usos de la IA y cuestiones relacionadas con la transparencia, la responsabilidad y la rendición de cuentas.

Figura 4.1: Sistemas no afectados por el reglamento

Por último, es importante recalcar que este reglamento permite que la Unión Europea y sus países miembros puedan crear o mantener leyes que protejan mejor los derechos de los trabajadores, manteniendo el respeto por otras normativas que salvaguardan a los consumidores y aseguran la integridad de los productos. Además, se aplicarán las leyes sobre la protección de datos personales, la privacidad y la confidencialidad de las comunicaciones a los datos personales tratados según lo establecido en este documento.

4.1.2. Prácticas prohibidas

A continuación se establecen una serie de prohibiciones que se deben tener en cuenta para cualquier desarrollo de IA:

- Están prohibidas las técnicas subliminales o manipuladoras que alteran significativamente el comportamiento de una persona o grupo, reduciendo su capacidad para tomar decisiones informadas.
- No se pueden explotar las vulnerabilidades de una persona o grupo específico debido a su edad, discapacidad o situación socioeconómica para alterar significativamente su comportamiento.
- No se permite la evaluación o clasificación de individuos o grupos basándose en su comportamiento social o características personales, resultando en una puntuación ciudadana que

provoca un trato perjudicial o desfavorable.

- Está prohibido realizar evaluaciones de riesgo de individuos con el fin de predecir la probabilidad de que cometan un delito penal, basándose únicamente en la elaboración de perfiles o la evaluación de rasgos y características de personalidad.
- No se puede crear o ampliar bases de datos de reconocimiento facial mediante la extracción indiscriminada de imágenes faciales de internet o de circuitos cerrados de televisión.
- No se pueden inferir en las emociones de una persona en lugares de trabajo y centros educativos, a menos que sea por motivos médicos o de seguridad.
- Está prohibido clasificar a las personas individualmente basándose en sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, orientación sexual o vida sexual.
- No se puede utilizar la identificación biométrica remota "en tiempo real" en espacios públicos con fines de aplicación de la ley, a menos que dicho uso sea estrictamente necesario para alcanzar uno o varios objetivos específicos, los cuales son la búsqueda de víctimas de delitos y de personas desaparecidas, prevención de amenazas a la vida y a la seguridad e identificación de sospechosos de delitos graves.

4.1.3. Que sistemas son catalogados como de alto riesgo y que requisitos deben cumplir

Un sistema se considera de alto riesgo si cumple dos condiciones. Primero, debe estar diseñado para ser utilizado como un elemento de seguridad en un producto que esté bajo la normativa de armonización de la Unión Europea, o el propio modelo debe ser uno de esos dispositivos. Segundo, la plataforma, ya sea en su totalidad o como parte de seguridad dentro de un producto, debe ser evaluado por una entidad independiente para asegurar su conformidad antes de su lanzamiento al mercado o uso, de acuerdo con las normativas de este sector.

Esto aplica independientemente de si el dispositivo se ha introducido en el mercado o se ha puesto en servicio sin estar integrado en los elementos mencionados en las condiciones anteriores. Además de esto, existe una lista de condiciones adicionales bajo las cuales un modelo puede ser considerado de alto riesgo (como se puede ver en la figura 4.2):

Figura 4.2: Sistemas de alto riesgo del anexo III

- **Biometría:** Incluye todas las tecnologías de identificación biométrica remota, la categorización basada en atributos sensibles o protegidos, y el reconocimiento de emociones. Dentro de estos queda excluido únicamente aquellos que son aplicados para verificar si una persona es realmente quien afirma ser.
- **Infraestructuras críticas:** Son todos los modelos utilizados como componentes de seguridad en la gestión y funcionamiento de infraestructuras digitales críticas, tráfico rodado, y suministro de agua, gas, calefacción o electricidad.
- **Educación y formación profesional:** Es todo aquel sistema usado para determinar el acceso o admisión a centros educativos y de formación profesional, evaluar los resultados del aprendizaje, analizar el nivel de educación adecuado que recibirá una persona, y para el seguimiento y detección de comportamientos prohibidos durante los exámenes.
- **Empleo, gestión de trabajadores y acceso al autoempleo:** Se emplean principalmente para la contratación o selección de personas, tomar decisiones que afecten a las condiciones laborales, asignación de tareas basadas en comportamientos individuales o rasgos personales, y para supervisar y evaluar el rendimiento y comportamiento de los individuos en el marco de relaciones laborales.
- **Acceso a servicios privados esenciales, y servicios y prestaciones públicos esen-**

ciales: Sirven para evaluar la elegibilidad de las personas para recibir beneficios de asistencia pública y servicios de salud, así como para determinar su solvencia o calificación crediticia, con excepción de su uso en la detección de fraudes financieros. Además, se emplean en la evaluación de riesgos y fijación de precios en seguros de vida y salud, y en la clasificación de llamadas de emergencia para el envío prioritario de servicios de primera intervención en situaciones de crisis.

- **Aplicación de la ley:** Son empleados para evaluar el riesgo de que un ser humano sea víctima de un delito, así como la fiabilidad de las pruebas durante investigaciones o juicios. Estos métodos también se utilizan para determinar la probabilidad de que alguien cometa o reincida en un delito, así como para analizar rasgos de personalidad y comportamientos delictivos pasados, contribuyendo así a la elaboración de perfiles durante la detección, investigación o enjuiciamiento de crímenes.

- **Migración, asilo y gestión del control fronterizo:** Las autoridades públicas competentes usan instrumentos como polígrafos para evaluar diversos riesgos, como la seguridad, la salud o la migración irregular, asociados con individuos que intentan ingresar o ya han ingresado al territorio de un Estado miembro. Por ello, son empleadas para examinar solicitudes de asilo, visado o permiso de residencia, así como reclamaciones relacionadas, en el contexto de la migración, el asilo o la gestión del control fronterizo. Su aplicación tiene como objetivo detectar, reconocer o identificar a ciertas personas, excluyendo la verificación de documentos de viaje.

- **Administración de justicia y procesos democráticos:** Las herramientas de administración de justicia y procesos democráticos se utilizan para asistir a las autoridades judiciales en la investigación, interpretación y aplicación de la ley en casos específicos, así como en la resolución alternativa de disputas. Además, estas tienen un papel en influir en los resultados electorales o referendos, así como en el comportamiento de los votantes al ejercer su derecho de voto en dichos eventos. Sin embargo, se excluyen los modelos que no están directamente expuestos a individuos, como aquellas utilizadas para la organización, optimización o estructuración de campañas políticas desde una perspectiva administrativa o logística.

Una vez explicado en detenimiento cuando un sistema es de alto riesgo, es necesario saber cuales son los requisitos que estos deben de cumplir:

- **Sistema de gestión de riesgos:** Debe estar sometido a un proceso cíclico de gestión de riesgos durante todo su ciclo de vida, el cual incluye cuatro fases principales. Primero, en la fase de determinación y análisis, se identifican y estudian todos los peligros asociados con su uso en su forma contemplada inicialmente. Luego, en la fase de estimación y evaluación, se prevén no solo los riesgos de uso intencionado, sino también aquellos que puedan surgir a partir de una

utilización no prevista. La tercera fase implica la contemplación de otras vulnerabilidades a través de un sistema de vigilancia poscomercialización (esto se explica en un apartado posterior). Finalmente, en la fase de adopción de medidas, se implementan estrategias para eliminar o reducir las vulnerabilidades detectadas mediante un diseño y desarrollo adecuados. Si no es posible eliminar todos los riesgos, se aplicarán medidas de control y se proporcionará la información y formación necesarias a los implicados en el despliegue.

- **Datos y gobernanza:** Todos aquellos sistemas de alto riesgo que vayan a ser entrenados con datos se deben de gestionar de forma adecuada teniendo en cuenta los siguientes puntos:
  - Las elecciones adecuadas vinculadas al diseño del modelo.
  - Se considerará el origen de esta información y, en el caso de que sea personal, la finalidad original de su recogida.
  - Las operaciones adecuadas para su preparación, como anotación, etiquetado, depuración, actualización, enriquecimiento y agregación.
  - Se formularán supuestos sobre la información que se supone que miden y representan.
  - Se evaluará la disponibilidad, la cantidad y la adecuación de los conjuntos de elementos necesarios.
  - Se examinarán posibles sesgos que puedan afectar a la salud y la seguridad de las personas, vulnerar derechos fundamentales u ocasionar algún tipo de discriminación prohibida por el Derecho de la Unión.
  - Se tomarán medidas adecuadas para detectar, prevenir y reducir estos sesgos.
  - Se detectarán lagunas o deficiencias que impidan el cumplimiento de la normativa, y, en caso de que las haya, se buscará la forma de subsanarlas.

  En adición a esto, los conjuntos de datos deben ser pertinentes, representativos, libres de errores y completos en la medida de lo posible, teniendo en cuenta su uso final, además que deben de tener en cuenta las características o elementos particulares del entorno geográfico, contextual, conductual o funcional específico en el que se prevé utilizar. Si resulta imprescindible a la hora de corregir errores, los desarrolladores tienen la opción de manejar ciertos tipos específicos de datos personales en circunstancias excepcionales, siempre y cuando proporcionen las garantías necesarias para proteger los derechos y libertades fundamentales de las personas físicas.

- **Documentación técnica:** Antes de la puesta en marcha o introducción en el mercado, se debe elaborar una documentación técnica que asegure el cumplimiento de la normativa establecida, la cual tiene que proporcionar a las autoridades y organismos notificados toda la información vital para su evaluación. Una vez en funcionamiento dentro de la UE, se elaborarán textos especializados que contengan todo lo necesario para la evaluación continua.

- **Conservación de registros:** Estos modelos están obligados a contar con capacidades de registro automático de eventos en todas las etapas de su funcionamiento, permitiendo de esta forma la detección de sucesos que puedan representar una amenaza o modificación sustancial, apoyar la supervisión posterior a la comercialización y supervisar el funcionamiento de estas tecnologías. Los eventos registrados incluyen la duración de cada uso del sistema, la base de datos utilizada para la comparación de información de entrada, y la identificación de las personas involucradas en la verificación de los resultados.

- **Transparencia y comunicación de información:** Estas arquitecturas tienen que diseñarse con suficiente claridad para que cualquier persona relacionada con este pueda entender y usar correctamente los datos de salida, además de que las instrucciones de uso deben contener información sobre la finalidad prevista del modelo y cualquier circunstancia que pueda afectarlo.

- **Vigilancia humana:** Es necesario que estas tecnologías sean vigiladas por personas físicas durante todo su funcionamiento, proporcionando las herramientas adecuadas para ello. El objetivo es reducir los posibles riesgos durante el ciclo de vida de la arquitectura, ajustando las estrategias de supervisión según sus amenazas, grado de autonomía y entorno de uso. Es vital que los responsables de la vigilancia comprendan sus capacidades y restricciones, entiendan toda la información de salida y puedan intervenir o detener su funcionamiento de manera segura.

- **Precisión, solidez y ciberseguridad:** En colaboración con la Comisión y organizaciones pertinentes, se establecerán una serie de directrices para medir y evaluar la precisión y solidez de estos sistemas, lo cual se incluirá en las instrucciones de uso. Es crucial que además de lo anterior, estos salgan al mercado con la mayor resistencia posible a errores y fallos, y que aquellos que continúan aprendiendo tras su puesta en marcha se desarrollen de manera que minimicen el riesgo de sesgo en la información de salida. Por último, deben ser, en la medida de lo posible, inmunes a intentos de manipulación por individuos no autorizados, mediante acciones que eviten y prevengan cualquier intento malicioso de corromper el modelo.

4.1.4. Obligaciones de los proveedores y responsables del despliegue de sistemas de IA de alto riesgo y de otras partes

Una vez establecidos los requisitos que deben cumplir los modelos de alto riesgo, es esencial que las personas mencionadas cumplan con ellos adecuadamente. A continuación se explicarán todas las implicaciones legales para asegurar el cumplimiento del reglamento.

Primero, es obligatorio que estos sistemas sean administrados mediante una arquitectura de gestión de calidad (como se puede observar en la figura 4.3) que garantice el cumplimiento de los

requisitos, el cual tiene que incluir aspectos como la metodología para cumplir con la normativa, métodos y protocolos sistemáticos para el diseño, supervisión, verificación y desarrollo, y procedimientos de revisión, prueba y validación antes, durante y después del desarrollo. Además, debe contener normas técnicas y estándares, sistemas y pautas para la gestión de datos, un modelo de gestión de riesgos, una arquitectura de seguimiento postventa, recursos para informar de incidentes graves, administración de la comunicación con autoridades, operadores, clientes y otras partes interesadas, estructuras y rutinas para mantener registros de documentación e información relevante, organización de recursos (incluyendo medidas de seguridad del suministro) y un marco de responsabilidad que defina las obligaciones del personal.

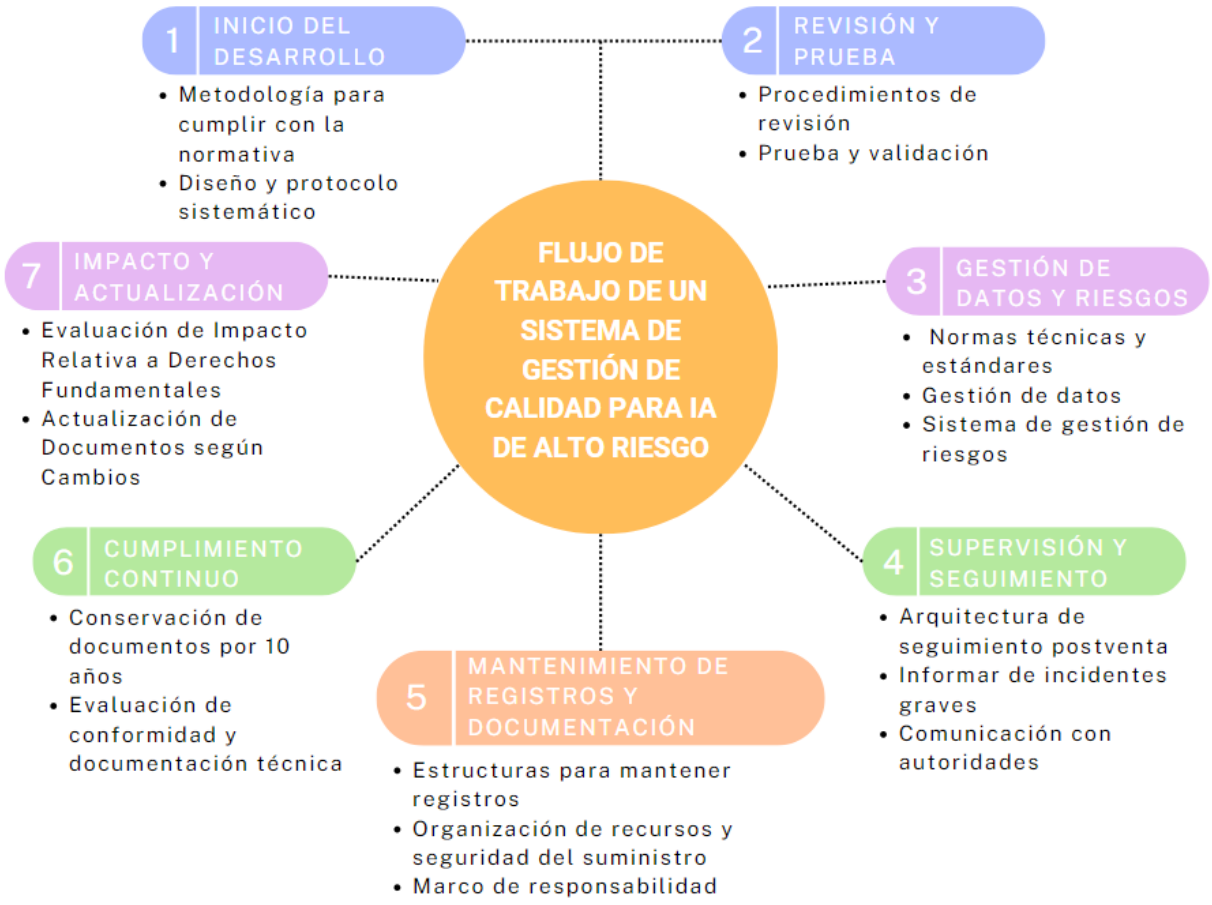

Figura 4.3: Flujo de trabajo de un sistema de gestión de calidad para IA de alto riesgo

Durante una década desde que se ha introducido el sistema en el mercado, el proveedor debe conservar documentos como la documentación técnica, los registros de gestión de calidad, la declaración UE de conformidad, y registros de cambios, decisiones y otros documentos creados por organismos notificados. Dependiendo de la región de la Unión Europea en la que se encuentre registrado, se establecerán condiciones para mantener esta información, especialmente si la empresa responsable

cierra o entra en quiebra. Toda la documentación relacionada con los requisitos descritos ha de ser proporcionada a las autoridades competentes.

Si los distribuidores creen que un modelo en el mercado no cumple con las normativas, deben tomar medidas para asegurar su cumplimiento o retirarlo del mercado, informando a los responsables y autoridades. En caso de que presente un riesgo significativo para la salud, seguridad o derechos fundamentales, se tiene que analizar de inmediato las causas y notificar a los diferentes países donde se utiliza.

Para lanzar esta tecnología desde un país fuera de la UE, se ha de nombrar un representante autorizado en la región antes de la puesta en marcha del servicio, quien será responsable de cumplir con las mismas obligaciones que las empresas de dentro de esta región. Si no cumple, se debe terminar su mandato y notificar a las autoridades y Estados miembros.

En adición a todo lo anteriormente mencionado, los importadores y distribuidores tienen que verificar que se cumpla el reglamento en su totalidad, realizar las evaluaciones de conformidad, preparar la documentación técnica y designar un representante autorizado. Si no se cumplen estos puntos, se ha de evitar introducir el modelo en el mercado e informar a los responsables y autoridades competentes. También deben indicar sus datos en el embalaje (solo los importadores), garantizar un transporte seguro, conservar toda la información necesaria y cooperar con las autoridades nacionales para evitar riesgos.

Cualquier persona involucrada en la puesta en marcha de un sistema de IA será considerada un proveedor si pone su nombre o marca en uno ya introducido, lo modifica considerablemente, o cambia su finalidad. En estos casos, el proveedor original deja de ser considerado como tal. El proveedor y cualquier tercero que suministre el modelo deben detallar por escrito qué información y apoyo se requieren para cumplir con las responsabilidades del reglamento.

Los responsables del despliegue han de adoptar medidas para garantizar el uso adecuado de la tecnología según las instrucciones de uso, lo que incluye asegurar una supervisión humana adecuada, usar datos de entrada apropiados y mantener archivos generados automáticamente por seis meses. Se tiene que informar a los proveedores si la arquitectura presenta un peligro considerable, y dependiendo de la gravedad, suspender su funcionamiento. Si se trata de una entidad financiera, habrán cumplido con la vigilancia cuando sigan correctamente las regulaciones sobre cómo se gestionan en línea con las leyes aplicables en el ámbito de los servicios financieros. En la situación de autoridades públicas o instituciones, deberán también cumplir con las normas de registro y solo podrán usar el modelo en caso de que este registrado en la base de datos.

Si se realiza una investigación con identificación biométrica en diferido, se debe pedir autorización y limitar su uso a la infracción específica. Si el sistema involucra a personas físicas, se debe notificarles y cooperar con las autoridades competentes.

Por último, antes de poner en funcionamiento cualquier infraestructura de este tipo, se ha de realizar una evaluación de impacto relativa a los derechos fundamentales, que incluirá la descripción de los casos donde se va a utilizar, periodo de tiempo y frecuencia, las categorías y peligros de las personas que se verán involucrados en su uso, explicar como se supervisará y todas las pautas que se seguirán en caso de que los riesgos planteados se vuelvan realidad. Este registro se realiza solo en la primera puesta en servicio, pero debe actualizarse si cambian las circunstancias.

4.1.5. Normas, registros, certificados y evaluación de conformidad

En este apartado se detallan las obligaciones y procesos que deben seguirse para asegurar la conformidad de estos sistemas, incluyendo el uso de normas armonizadas, la obtención de certificados de conformidad y el registro adecuado de las tecnologías en la base de datos de la UE:

- **Normas armonizadas**
  - Asegurarse de que los sistemas de IA de alto riesgo cumplan con las normas armonizadas publicadas en el Diario Oficial de la Unión Europea, en conformidad con el Reglamento (UE) n.º 1025/2012[32].
  - La Comisión puede solicitar mejoras como la reducción del consumo de energía y pruebas para verificar el cumplimiento de todos los requisitos.
- **Especificaciones Comunes**
  - La Comisión puede establecer especificaciones comunes si las normas armonizadas no cumplen con los riesgos a los derechos fundamentales o si las organizaciones no aceptan la solicitud para elaborar una norma armonizada.
  - Evaluar si se cumplen las condiciones necesarias antes de elaborar este acto de ejecución.
- **Evaluación de Conformidad**
  - Si la tecnología incluye biometría, optar entre el control interno o la evaluación del sistema de gestión de la calidad y la documentación técnica.
  - Para otras áreas de alta vulnerabilidad, solo se requiere el procedimiento de control interno.
  - Si el sistema ha sido evaluado previamente y sufre un cambio significativo, debe someterse a una nueva evaluación.
- **Certificado de Conformidad**
  - Una vez completado el protocolo de evaluación, se emitirá un certificado con una duración de cuatro años para los modelos de alto riesgo y cinco años para el resto, con posibilidad de renovación.

- Si el sistema ya no cumple con la normativa, el certificado puede ser retirado o se pueden imponer restricciones.

- **Excepciones a la Evaluación**
  - No es necesario realizar la evaluación si el modelo tiene como objetivo la seguridad pública, la protección de vidas y salud, el medio ambiente o recursos clave de la industria e infraestructuras, aunque esta autorización es temporal.

- **Declaración UE de Conformidad**
  - El proveedor debe redactar y mantener una declaración de conformidad disponible para las autoridades durante al menos diez años desde su puesta en marcha.

- **Certificado CE**
  - Obtener un certificado CE que cumpla con los requisitos del artículo 30 del Reglamento (CE) n.º 765/2008[33].
  - Si es digital, debe ser accesible y visible a través de su interfaz o código, acompañado del número de identificación del organismo notificado.

- **Registro en la Base de Datos de la Unión Europea**
  - Registrar todos los sistemas de IA, sean o no de alto riesgo (excepto los relacionados con infraestructuras críticas).
  - Sistemas de biometría, aplicación de la ley, migración, asilo y gestión del control fronterizo deben registrarse en una sección segura, visible solo para la Comisión y autoridades nacionales.
  - Sistemas usados en infraestructuras críticas deben registrarse a nivel nacional.

4.1.6. Transparencia y capacitación

Es fundamental que los proveedores y responsables del despliegue de sistemas de inteligencia artificial mantengan altos niveles de transparencia hacia todas las personas involucradas en su uso. En primer lugar, deben informar a todas las personas físicas de que están tratando con una tecnología de este tipo, aunque hay que tener en cuenta que esta obligación no aplica cuando la interacción con la IA es evidente, como en el caso de un chatbot claramente identificado, o cuando la tecnología está diseñada específicamente para prevenir y evitar delitos. Además, todo dato de salida generado o manipulado debe ser claramente legible e interpretable, y debe estar marcado como generado o manipulado artificialmente (esta medida no se aplica a modelos con fines penales, ni a aquellos que realizan funciones de apoyo a la edición estándar o donde los datos de entrada no son sustancialmente alterados). Por último, es crucial que los profesionales notifiquen cuando

esta tecnología ha generado o modificado contenido ultrafalsificado, es decir, vídeos, grabaciones o imágenes con personas aparentemente reales. Esta obligación se atenúa si este contenido se usa con fines creativos, cómicos o de ficción, en cuyo caso basta con informar de su existencia (esta medida tampoco se aplica en contextos delictivos o cuando el contenido ha sido revisado por una persona o editorial y un profesional tiene la responsabilidad editorial del mismo).

En cuanto a la capacitación, se debe garantizar un uso efectivo y seguro de los sistemas de IA, siendo crucial que los proveedores y responsables aseguren que todas las personas relacionadas con el funcionamiento y uso de estos posean conocimientos adecuados en la materia. Es esencial evaluar si una persona está capacitada para encargarse de este tipo de modelos considerando su conocimiento técnico, experiencia, educación y formación en el área, así como el contexto en el que se va a usar y a quiénes está destinado. Además, se debe proporcionar formación continua a todas las personas involucradas en su utilización y manejo para mantener sus habilidades y conocimientos actualizados. Esta capacitación debe incluir tanto aspectos técnicos como éticos, asegurando que los profesionales comprendan el funcionamiento del sistema y las implicaciones de su función. Asimismo, es importante asegurarse de que toda la documentación necesaria esté disponible y sea accesible para los usuarios y operadores, proporcionando recursos adicionales como manuales, tutoriales y soporte técnico para resolver dudas y problemas que puedan surgir durante su operatividad.

4.1.7. IAs de uso general y obligaciones que deben cumplir

Antes de abordar los modelos de IA de uso general, es crucial definir cuándo estos se catalogan como de riesgo sistémico. Un sistema se considera de este tipo cuando tiene una capacidad significativa para provocar un impacto relevante, medido ya sea a través de técnicas y herramientas adecuadas o por decisión de la Comisión Europea, teniendo en cuenta varios factores: la cantidad de parámetros, la calidad y cantidad del conjunto de datos, el nivel de cálculo utilizado en el entrenamiento (considerando como referencia aquellos que superan los 1025 FLOPs, operaciones de coma flotante por segundo), la cantidad de personas registradas, la forma en la que interactúa con los datos (incluyendo el manejo de diferentes tipos de datos), los parámetros de referencia usados en su evaluación, y el impacto potencial en el mercado interior, particularmente si el número de profesionales involucrados en la Unión Europea supera los 10,000.

Una vez que se ha identificado como tal, los proveedores tienen la obligación de notificar a la Comisión Europea en un plazo de dos semanas, proporcionando toda la documentación necesaria que confirme esta clasificación. Si no se realiza esta notificación, la Comisión puede calificar el sistema por su cuenta (todo este proceso se puede observar en a figura 4.4).

Figura 4.4: Diagrama de flujo del proceso de clasificación y notificación de riesgo sistémico

Obligaciones para la IA de uso general

Los distribuidores deben crear y mantener toda la documentación técnica, incluyendo información detallada sobre los procesos de prueba, entrenamiento y evaluación. Estos datos han de estar disponibles para otros profesionales que trabajen con la tecnología, para que puedan comprender sus capacidades y limitaciones y cumplir con la normativa vigente. Además, los distribuidores tienen que establecer directrices que aseguren el cumplimiento de la legislación de la UE y publicar un resumen detallado sobre los datos y contenidos usados durante la fase de entrenamiento.

Estas obligaciones no se aplican a los modelos distribuidos bajo código y licencia abierta, es decir, aquellos que permiten el acceso, alteración, uso y distribución libre, y comparten todos los parámetros y datos referentes a su arquitectura de manera colectiva. Al igual que los sistemas de alto riesgo, si estos modelos desean operar en la Unión Europea pero no están ubicados en la región, deben nombrar un representante establecido en esta región, el cual debe cumplir con todas las obligaciones descritas para los sistemas de alta vulnerabilidad, adaptadas a los requisitos específicos de estos modelos, considerando también si presentan un riesgo sistémico.

Obligaciones adicionales para IA de uso general con riesgo sistémico

Además de las obligaciones mencionadas anteriormente, los sistemas de IA de uso general clasificados como de riesgo sistémico tienen que cumplir con las siguientes medidas adicionales:

- Realizar una evaluación exhaustiva siguiendo protocolos y normas adecuados, incluyendo pruebas de simulación de adversarios para identificar y mitigar cualquier peligro potencial.
- Analizar y reducir todos los riesgos asociados con el desarrollo, implementación y uso del modelo.
- Monitorizar, notificar y documentar cualquier riesgo o amenaza a la Oficina de IA para proceder con su resolución adecuada.

- Cumplir con las medidas necesarias de ciberseguridad para asegurar la protección contra amenazas y vulnerabilidades.

4.1.8. Supervisión post-venta, intercambio de información y monitorización del mercado

Implementación de un sistema de vigilancia

Los proveedores de sistemas de inteligencia artificial de alto riesgo deben establecer un sistema de vigilancia que supervise el desempeño y uso del modelo una vez esté en el mercado. Este debe funcionar durante toda la vida útil del producto, con el objetivo de identificar y mitigar cualquier riesgo o amenaza que surja durante su utilización, además de que se controlará, en la medida de lo posible, su interacción con otras arquitecturas similares.

Gestión de incidentes graves

En caso de que ocurra un incidente grave, los distribuidores están obligados a notificarlo a las autoridades pertinentes en cada país donde se esté operando. Esta notificación debe realizarse de inmediato una vez se identifique una conexión entre el incidente y la tecnología, con un plazo máximo de quince días desde el descubrimiento de dicha conexión. Si el incidente ocurre en un espacio de acceso público, el plazo de notificación se reduce a dos días, y en caso de fallecimiento, a diez días. Posteriormente, se llevará a cabo una investigación adecuada en colaboración con las autoridades y organismos correspondientes para determinar con mayor claridad y precisión el impacto del modelo en el incidente.

Cumplimiento del Reglamento (UE) 2019/1020

Todos los modelos de IA bajo esta normativa estarán regulados por el Reglamento (UE) 2019/1020[34], que se centra en la vigilancia del mercado y en asegurar que los productos cumplan con sus obligaciones. Este reglamento establece mecanismos para monitorear y garantizar el cumplimiento de todas las normativas de seguridad y calidad, asignando a diversas entidades la responsabilidad de estos procedimientos y definiendo una estructura para gestionar vulnerabilidades y riesgos específicos.

Confidencialidad y protección de información

Es crucial que todas las personas y organizaciones que participen en el cumplimiento de esta normativa respeten la confidencialidad de la información y los datos obtenidos. En particular, se deben proteger los derechos de propiedad intelectual e industrial, los datos empresariales, los secretos comerciales, la efectiva aplicación del Reglamento, los intereses de seguridad pública y nacional, el desarrollo de causas penales o procedimientos administrativos y la información clasificada.

4.1.9. Sanciones

En caso de que no se cumplan con los requisitos y normas del presente reglamento, se aplicará a las personas responsables una sanción adecuada a la infracción cometida. En particular, las infracciones relacionadas con las prácticas de IA prohibidas serán multadas con hasta 35.000.000 de euros o, en caso de ser una empresa, hasta el 7 % del volumen de negocios mundial. Si el incumplimiento es distinto a los mencionados en el artículo previo, la multa podrá alcanzar hasta 15.000.000 de euros o, si es empresa, el 3 % del volumen de negocios mundial.

En los casos donde se entregue información errónea, la sanción será una multa de hasta 7.500.000 euros o, en caso de compañía o negocio, el 1 % de la facturación global. Para los sistemas de uso general, se aplicarán sanciones económicas de hasta 15.000.000 de euros o el 3 % de los ingresos globales si se incumple alguno de los siguientes puntos:

- Han infringido alguna de las normativas aplicables a este tipo de tecnologías.
- No han entregado la documentación necesaria, o esta se ha realizado de forma inexacta o incompleta.
- Han incumplido alguna obligación solicitada por la Comisión.
- No han dado acceso al modelo a la Comisión para que pueda proceder a su evaluación.

4.1.10. Entrada en vigor

Es crucial conocer el momento en que comenzará a aplicarse este reglamento. Entrará en vigor 20 días después de su publicación en el Diario Oficial de la UE y será aplicable dentro de dos años. No obstante, existen excepciones con diferentes plazos de implementación respecto a la entrada en vigor, las cuales son:

- El capítulo de disposiciones generales y las prácticas prohibidas serán aplicables seis meses después.

- La sección referente a las autoridades notificantes y organismos notificados, modelos de IA de uso general, gobernanza y sanciones se ejercerán 12 meses después.
- Las reglas por las cuales un sistema es clasificado como de alto riesgo y sus respectivas obligaciones se implementarán 36 meses después.
- Los sistemas que sean componentes de otros modelos informáticos de mayor amplitud, que se hayan puesto en marcha antes de 36 meses de la entrada en vigor, deberán adecuarse al contenido del reglamento antes del 31 de diciembre de 2030.
- Toda tecnología de uso general publicada 12 meses antes de la entrada en vigor deberá cumplir con las obligaciones 36 meses después.

4.2. Impacto ético en los sistemas de IA

La integración de sistemas de inteligencia artificial en el desarrollo de software implica una serie de consideraciones éticas que deben ser examinadas minuciosamente para asegurar que su uso sea tanto legalmente adecuado como moralmente responsable. En concreto, para realizar todas las afirmaciones siguientes, se ha estudiado en detenimiento el siguiente documento: "Directrices éticas para una IA fiable", publicado por la Comisión Europea el 8 de abril de 2019 y elaborado por el Grupo de Expertos de Alto Nivel en Inteligencia Artificial[29].

Este apartado aborda esta tecnología desde un punto de vista deontológico, desglosando los principios fundamentales que deben guiar su desarrollo y uso, los requisitos éticos que tienen que ser cumplidos, y los métodos técnicos y no técnicos que pueden ser empleados para asegurar que estos sean implementados de manera correcta. A través de estos subapartados, se busca proporcionar un marco comprensivo y práctico para integrar consideraciones éticas en todas las etapas del ciclo de vida de los modelos de IA.

4.2.1. Pincipios éticos

A continuación se examinan los principios éticos esenciales que han de ser considerados al diseñar e implementar sistemas de inteligencia artificial, asegurando que estos sean utilizados de manera justa, segura y transparente:

- **Respeto de la autonomía humana:** Toda persona que use el sistema está en la obligación de conservar plena autonomía y poder participar activamente en la democracia, además de no ser coaccionada, sometida, engañada y manipulada por este. En su lugar, se tiene que diseñar para que haya más facilidad y conocimiento a la hora de la toma de decisiones, lo

cual requiere que esta tecnología sea supervisada para evitar poner en riesgo lo anteriormente descrito. En el ámbito laboral, debe de apoyar en las diferentes tareas y siempre favorecer a la creación de puestos de empleo, nunca disminuirlos.

- **Prevención del daño:** Nunca deben provocar daños, agravar los existentes y perjudicar de otras maneras a los individuos, lo que implica que se debe de proteger la dignidad humana y la integridad física y mental. En adición a esto, deben ser a prueba de fallos, no deben poder usarse nunca con fines malintencionados y deben prestar mayor atención a los seres humanos más vulnerables. Asimismo, no se ha de tener en cuenta solo a nosotros, sino también al medio ambiente y a todos los seres vivos.
- **Equidad:** Se debe de asegurar una distribución justa y equitativa de los costes y beneficios, y asegurar que nadie sufra un sesgo ni discriminación por alguna característica personal. Esto incluye fomentar uniformemente el acceso a la educación, la tecnología y los bienes de los servicios. La equidad no solo se refiere a esto, sino también a que los profesionales en su entorno deben entender que los métodos elegidos deben estar en proporción con el propio objetivo y deben equilibrar los diferentes intereses y metas que estén en juego. Por último, todos los individuos se pueden oponer a las decisiones tomadas por el sistema o por las personas que lo manejan, pudiendo identificar a los responsables y pedir una explicación sobre los procesos utilizados.
- **Explicabilidad:** Es crucial que los modelos sean transparentes, o dicho de otra forma, que todos los usuarios involucrados en su uso entiendan perfectamente sus capacidades, finalidad y decisiones. En los casos en los que no se pueda explicar detalladamente alguno de los puntos previos, será necesario adoptar otras medidas, como la trazabilidad, verificación y comunicación transparente.

Por último, cabe destacar que hay apartados donde entran en conflicto estos puntos, por ejemplo, hasta que nivel un sistema de inteligencia artificial se puede usar para prevenir la delincuencia, porque por un lado puedes estar vulnerando la libertad de las personas y por otro no estás previniendo daños que podrían llegar a ocurrir, por lo que estos se deben tomar como una orientación y no como una solución concreta.

4.2.2. Requisitos

Todos los principios anteriormente nombrados deben de traducirse en una serie de requisitos (mostrados en la figura 4.5) para garantizar una IA fiable, los cuales están dirigidos a los desarrolladores, responsables del despliegue y usuarios finales:

Figura 4.5: Requisitos de una IA fiable

- **Acción y supervisión humanas:** Esta tecnología puede afectar tanto positiva como negativamente a los derechos fundamentales, por lo que es de vital importancia realizar una evaluación sobre el impacto que puede tener hacía estos antes de su desarrollo, incluyendo una valoración de como reducir o justificar estos riesgos.

  A veces pueden influir en la toma de decisiones de las personas, por lo que es importante que el usuario reciba la información y las herramientas necesarias para interactuar de forma adecuada, además de facilitar las elecciones que hagan y que estás estén coordinadas con sus objetivos y pensamientos. Por ello, el principio de autonomía humana debe ser un pilar central del sistema, garantizando que los usuarios no sean controlados por decisiones automáticas que puedan tener un impacto significativo en sus vidas.

  Para poder evitar que esto ocurra se deben de supervisar estos dispositivos con humanos, lo que se llevará a cabo con mecanismos de gobernanza, es decir, haciendo que estos participen en todos los ciclos de decisión del modelo, intervengan durante el diseño y seguimiento, y vigilen y decidan cómo y cuándo utilizarlo.

- **Solidez técnica y seguridad:** Este requisito hace referencia al principio de prevención del daño, ya que estos sistemas se deben desarrollar de tal forma que se eviten la mayor cantidad de riesgos posibles. Por ello es crucial que este tenga un mecanismo de seguridad perfectamente

diseñado, para así poder evitar que agentes exteriores manipulen cualquier característica haciendo que su finalidad e integridad estén en peligro (modificando su comportamiento, obteniendo datos confidenciales e incluso llegar a desconectarlo directamente) y que un mal uso o situaciones inesperadas lo corrompan. En el caso de que ocurra alguna de estas situaciones, se debe contar con un plan de resguardo para así prevenir que vaya a más.

Se requiere además la garantía de que el dispositivo vaya a actuar siempre como se espera de él (sin causar daños al medio ambiente o a los seres vivos), realice juicios correctos y tome decisiones adecuadas basándose en datos o modelos, haciendo que en el situación de una predicción incorrecta su daño sea el mínimo posible (en caso de que no se puedan evitar, se tiene que indicar la probabilidad de que ocurran).

Para finalizar con esta necesidad, es crucial que todos los resultados sean reproducibles y fiables, lo que es necesario para poder evaluar como se enfrentan a diversas situaciones y prevenir que ocurran daños involuntarios.

- **Gestión de la privacidad y datos:** Toda la información proporcionada por el usuario y generada sobre este se tiene que proteger para evitar que terceros puedan acceder a ella, además de garantizar que todo lo que se recabe durante su ciclo de vida no se vaya a usar contra estos (por ejemplo, características como la raza u orientación sexual).

  En adición a esto, se necesita comprobar de antemano que los datos de entrada utilizados no estén sesgados, sean imprecisos o contengan errores, ya que podría comprometer en el funcionamiento del dispositivo. Asimismo, hay que tener siempre claro que protocolos, que individuos y en que tipo de situaciones se tienen acceso a datos personales.

- **Transparencia:** Todos los datos y procesos utilizados para la generación del contenido de salida, incluyendo este, se deberán documentar correctamente para así poder aumentar su trazabilidad y transparencia, lo que permitirá identificar fácilmente la causa de los errores y poder evitarlos en un futuro.

  Las decisiones que tome este sistema también tendrán que ser entendibles y explicables a todas las partes involucradas en su uso, lo que incluye que si una persona ha sido afectada en cierta medida por este, pueda reclamar una explicación adecuada del proceso de toma de decisiones (la cual debe ser entendible también). Además de esto, se necesita informar de que capacidades y limitaciones tiene y, en cuanto al usuario final, debe conocer si está interactuando con una IA o con un humano, pudiendo en algunos casos elegir por cual de los dos ser atendido (por ejemplo, en servicios de atención al cliente).

- **Diversidad, no discriminación y equidad:** Como se ha comentado anteriormente, hay que tener cuidado de que los datos de entrada no se encuentren sesgados, lo que puede dar lugar a discriminación y prejuicios. Este problema no se puede dar solo con esta situación, sino que si los desarrolladores o consumidores utilizan esta tecnología de modo inadecuado, o el propio modelo esta evolucionando de forma que se acentúan estos rasgos, probablemente

lleve al mismo problema, por lo que hay que asegurarse que toda la información usada sea correcta y se utilicen métodos de supervisión apropiados.

Es importante recalcar que el diseño de este tipo de modelos debe estar centrado principalmente en el usuario, de forma que permita a cualquier tipo de persona usarlo independientemente de sus capacidades o características (hay que tener en cuenta sobre todo a discapacitados). Por último, es recomendable pedir a las partes interesadas su opinión periódicamente para así poder ir modificando el diseño y funcionamiento para que este acorde con todo tipo de seres humanos.

- **Bienestar social y ambiental:** Al mismo que tiempo que estos sistemas abordan nuestras preocupaciones, se debe garantizar que no afecten negativamente al medio ambiente, por lo que todos los procesos involucrados en su desarrollo, puesta en servicio y utilización (como puede ser el uso de recursos o energía) deben de ser lo más respetuosos posibles.

  En cuanto al impacto social, se tiene que tener bastante precaución con que estos dispositivos no empeoren nuestras relaciones y vínculos sociales, lo que afecta negativamente al estado físico y mental. Además, no solo debemos pensar en cómo afecta a las personas individualmente, sino también en cómo impacta en toda la sociedad, por lo que se debe evaluar que efecto puede llegar a tener en las instituciones, la democracia y en la propia sociedad en su conjunto.

- **Rendición de cuentas:** Este requisito implica la necesidad de crear mecanismos que aseguren que se pueda responsabilizar y rendir cuentas por las acciones de los modelos de inteligencia artificial y sus resultados, cosa que debe aplicarse tanto antes de ponerlos en funcionamiento como después de hacerlo.

  El primero de ellos es la auditabilidad, que es la capacidad de evaluar los procesos, los datos y los algoritmos involucrados en el diseño. En este contexto implica que se tiene que evaluar por parte de auditores externos e internos, creando una serie de documentos disponibles en todo momento para de esta manera garantizar la fiabilidad de la tecnología.

  El segundo es la minimización y notificación de efectos negativos, método por el cual se asegura que se pueda informar sobre las acciones o decisiones que dan lugar a ciertos resultados de un sistema, así como responder a las consecuencias de estos. Identificar, evaluar, notificar y reducir estas repercusiones es especialmente crucial para todas las personas involucradas directa o indirectamente. Usar evaluaciones de impacto, como "equipos rojos"[2] o ciertos tipos de evaluaciones algorítmicas, antes y después de desarrollar, implementar y usar estos dispositivos, puede ayudar a reducir sus efectos negativos.

  El tercero es la búsqueda de equilibrios, que se refiere a que en el momento de aplicar los requisitos mencionados, pueden surgir conflictos entre ellos, por lo que puede ser necesario en-

[2]Los "equipos rojos" son grupos especializados que se emplean en diversas áreas, como la seguridad informática, la defensa nacional o la evaluación de riesgos, para llevar a cabo simulaciones de ataques o intrusiones con el fin de identificar vulnerabilidades en sistemas, infraestructuras o procesos.

contrar un punto medio. Debido a esto es importante abordarlas de manera lógica y ordenada, teniendo en cuenta el nivel técnico actual. Si se presentan discrepancias, es crucial explicar cómo se intentó encontrar un equilibrio entre ellos y evaluar esta estabilidad en términos del riesgo para los principios éticos y los derechos fundamentales. Si no es posible encontrar una armonía éticamente aceptable, no se debe continuar con el desarrollo, implementación y uso del sistema de IA como estaba planeado.

Por último, cabe desatacar que cuando ocurran efectos negativos derivados del uso del modelo, deben de realizarse con anterioridad mecanismos que aseguren una compensación adecuada, lo cual garantiza y aumenta en gran medida la confianza de los usuarios.

4.2.3. Métodos técnicos y no técnicos

Para poder cumplir con todos los requisitos descritos anteriormente se van a establecer a continuación una serie de requisitos técnicos y no técnicos:

Métodos técnicos

- **Arquitecturas:** Los requisitos para una inteligencia artificial confiable deben incorporarse a la estructura de los sistemas mediante normas que definan comportamientos adecuados ("lista blanca"), comportamientos no adecuados ("lista negra") y garantías sobre el comportamiento del sistema. Para modelos con capacidad de aprendizaje, es crucial integrar estos requisitos en cada etapa del ciclo "sentir-planificar-actuar". Durante la primera etapa ("sentir"), se debe de reconocer todos los elementos del entorno necesarios para cumplir las directrices, en la segunda ("planificar") se tiene que considerar únicamente aquellas estrategias que las realicen de forma adecuada y, en la última ("actuar"), las acciones de la tecnología se limitan a las que acatan las características anteriores correctamente.
- **Ética y estado de Derecho desde el diseño:** Cuando nos encontramos en la etapa de diseño es importante asegurarnos de que se cumplen con ciertos valores desde el inicio, lo que significa que los principios abstractos que queremos que el sistema siga deben estar claramente vinculados a las decisiones específicas sobre cómo se implementa y utiliza. En la actualidad es de vital relevancia que todo lo relacionado con el dispositivo este totalmente seguro, por lo que pensar desde el principio sobre como proceder a su protección, resistencia a fallos y apagado es crucial.
- **Métodos de explicación:** Como hemos mencionado en apartados previos, todas las personas involucradas en el uso de un sistema deben entender a la perfección cómo se comporta y porque ha llegado a cierta conclusión. Ahí entra un campo de investigación bastante actual, la inteligencia artificial explicable (XAI), la cual centra en entender por qué se toman ciertas

decisiones. Es un desafío importante, especialmente para modelos basados en redes neuronales, donde los resultados pueden ser difíciles de interpretar y pequeñas diferencias en los datos pueden llevar a interpretaciones completamente distintas. Los métodos de XAI son vitales no solo para explicar el comportamiento de esta tecnología a los usuarios, sino también para garantizar su fiabilidad.

- **Realización de ensayos y validación:** La elaboración de ensayos y la validación son cruciales debido a su naturaleza impredecible y la influencia del contexto en el que operan, ya que las pruebas convencionales no son suficientes porque algunos errores solo se revelan cuando el sistema interactúa con datos realistas. Por lo tanto, es necesario monitorear cuidadosamente la estabilidad, robustez y rendimiento del modelo durante su desarrollo y uso, asegurando que las decisiones tomadas sean validadas adecuadamente. Los ensayos y la validación deben llevarse a cabo lo antes posible y deben abarcar todos los aspectos de la arquitectura, incluyendo datos, modelos, entornos y comportamiento general. Es importante que estos procesos sean diseñados y ejecutados por un equipo diverso para obtener una variedad de perspectivas. Se deben considerar métodos como prácticas contradictorias realizadas por equipos confiables, así como la recompensa para quienes encuentren y reporten errores y vulnerabilidades del sistema. Finalmente, es esencial garantizar que los productos o acciones derivadas de estos procesos estén en línea con las políticas establecidas previamente para evitar vulneraciones.
- **Indicadores de calidad del servicio:** Se pueden establecer indicadores específicos para evaluar la calidad del servicio de esta tecnologías, asegurando que se hayan tenido en cuenta las consideraciones de seguridad durante su desarrollo y ensayo. Estos indicadores podrían abarcar aspectos como la evaluación de las pruebas realizadas, la formación de algoritmos y los parámetros habituales de calidad del software, incluyendo su rendimiento, usabilidad, fiabilidad, seguridad y mantenimiento.

Métodos no técnicos

- **Normativas:** La legislación actual sobre seguridad de productos y marcos de responsabilidad proporcionan apoyo para la fiabilidad de la IA. Si es necesario adaptar o actualizar estas regulaciones, se debe considerar en futuras recomendaciones políticas relacionadas con este sector.
- **Códigos de Conducta:** Las organizaciones pueden adoptar las directrices anteriormente explicadas en sus políticas internas, como códigos de conducta y documentos de responsabilidad empresarial, para contribuir a la confiabilidad.
- **Normalización:** Establecer normas para el diseño y la fabricación de sistemas puede ayudar a promover una conducta ética en su uso y desarrollo.

- **Certificación:** La certificación de modelos transparentes, responsables y equitativos puede ser realizada por organizaciones especializadas, aunque esto no reemplaza la responsabilidad y debe ir acompañada de marcos de rendición de cuentas.
- **Gobernanza y rendición de cuentas:** Las organizaciones deben establecer marcos internos y externos de gobernanza para asegurar la responsabilidad ética en todas las etapas del desarrollo y uso de estas tecnologías.
- **Educación y concienciación:** La difusión del conocimiento sobre estos dispositivos y la promoción de la participación informada de todas las partes interesadas son esenciales para una IA ética y confiable.
- **Participación y diálogo social:** El debate abierto y la participación de las partes interesadas son fundamentales para evaluar y abordar los impactos y preocupaciones con este tipo de sistemas.
- **Diversidad e inclusión:** Los equipos que desarrollan estos modelos deben reflejar la diversidad de la sociedad para garantizar que se consideren diferentes perspectivas y se aborden diversas necesidades y preocupaciones.

CAPÍTULO 5

Ejemplos de uso

Como se ha podido ver anteriormente, la inteligencia artificial presenta una gran cantidad de riesgos que deben ser cubiertos para evitar un impacto negativo en los proyectos de las empresas y en la sociedad. Por ello, en este apartado se van a identificar y analizar ejemplos concretos de buen y mal uso de herramientas de IA para comprender tanto los aspectos positivos como los negativos. A través de estos casos se busca resaltar concretamente las buenas prácticas que se han realizado y las lecciones aprendidas del uso indebido, las cuales han generado problemas éticos, legales y técnicos.

Concretamente, los ejemplos de buen uso demuestran como una correcta aplicación puede llegar a optimizar procesos, mejorar la calidad de los programas y facilitar el trabajo a los profesionales, y, por otro lado, los de mal uso nos enseñan todos los problemas generados debido a no seguir tanto las leyes descritas en la normativa como las pautas éticas. Cada uno de estos casos se ha seleccionado siguiendo criterios de relevancia, actualidad, diversidad y disponibilidad de información detallada, además de que están ubicados en una amplia diversidad de áreas, entre las que se incluyen aplicaciones automatizadas de pruebas de software, análisis predictivo en mantenimiento, asistentes de codificación, sistemas de contratación, sesgos algorítmicos y algoritmos de trading.

5.1. Ejemplos de buen uso

- **Automatización de pruebas de software**

- **Descripción:** Uso de herramientas de IA para automatizar pruebas de regresión en aplicaciones.
- **Impacto positivo:** Aumento de la eficiencia, reducción de errores humanos y disminución del tiempo empleado por parte de los desarrolladores.
- **Ejemplo real:** La empresa Capgemini utilizó Tricentis Tosca, una herramienta de testeo de software que utiliza IA, para automatizar pruebas de regresión en proyectos de sus clientes, lo que resultó en una reducción significativa del tiempo de prueba y un aumento de la cobertura de estas. Al aplicar estas soluciones automatizadas, se logró identificar y solucionar problemas de software más rápidamente, lo que mejoró la calidad y la estabilidad del producto final[35].
- **Buenas prácticas identificadas:** Una de las primeras buenas prácticas identificadas en el uso de esta tecnología es la automatización de tareas repetitivas, permitiendo a los profesionales centrarse en aspectos más creativos y complejos del desarrollo. Además, implementar herramientas que puedan ejecutar múltiples escenarios de prueba de forma más rápida y con menos fallos garantiza una mayor cobertura, lo que es de vital importancia para la detección temprana de fallos y la mejora continua.

  También cabe resaltar que su integración facilita una retroalimentación rápida sobre el estado del software, permitiendo correcciones y mejoras en etapas tempranas del desarrollo, resultando en un ciclo más eficiente y en productos finales de mayor calidad. Además de lo anteriormente mencionado, también asegura que cada cambio en el código sea aprobado exhaustivamente antes de ser implementado, minimizando los errores y aumentando la confiabilidad.

- **Análisis predictivo en mantenimiento preventivo**
  - **Descripción:** Utilización de IA para predecir fallos en servidores y sistemas de red.
  - **Impacto positivo:** Reducción de tiempos de inactividad, ahorro en costos de reparación y mejora en la fiabilidad del sistema.
  - **Ejemplo real:** IBM Watson ha aplicado esta tecnología con éxito en el mantenimiento predictivo de equipos industriales, ya que con la capacidad de analizar grandes volúmenes de datos en tiempo real, puede identificar patrones y señales que indican un posible fallo antes de que ocurra. Esto permite a las empresas realizar un mantenimiento proactivo, programando reparaciones y reemplazos antes de que los problemas se conviertan en fallos graves, lo que resulta en una significativa reducción de costos y en la mejora de la fiabilidad y disponibilidad de los sistemas[36].
  - **Buenas prácticas identificadas:** En el contexto del análisis predictivo para el mantenimiento preventivo, una buena práctica clave es la integración de IA para monitorear constantemente el estado de los sistemas. La habilidad de esta tecnología para evaluar datos en tiempo real y anticipar posibles fallos permite a las organizaciones disminuir

significativamente los tiempos de inactividad, ya que los problemas se resuelven antes de que se conviertan en errores graves. Esto no solo aumenta la fiabilidad, sino que también contribuye de manera considerable a la reducción de costos de reparación, pudiendo organizar el mantenimiento de forma más eficiente.

Además, este tipo de implementación facilita una gestión de recursos más eficaz, ya que se pueden priorizar y programar las intervenciones de mantenimiento según la criticidad y la urgencia de los fallos previstos. Esta estrategia no solo optimiza lo anteriormente mencionado, sino que también mejora la sostenibilidad operativa al evitar interrupciones no planificadas y minimizar el desgaste innecesario de los equipos.

- **Asistentes de codificación**
  - **Descripción:** Uso de asistentes de codificación basados en IA para ayudar a los profesionales a escribir código.
  - **Impacto positivo:** Mejora en la productividad de los desarrolladores, reducción de errores y aceleración del desarrollo de software.
  - **Ejemplo real:** GitHub Copilot, desarrollado conjuntamente por GitHub y OpenAI, es un ejemplo destacado de asistente de codificación basado en IA. Este asistente sugiere líneas de código y funciones completas a partir de comentarios e información previamente escrita por el usuario. Al analizar el contexto de este en tiempo real, GitHub Copilot proporciona sugerencias precisas y relevantes que ayudan a los desarrolladores a escribir de manera más rápida y con menos fallos. Este enfoque no solo acelera el proceso de desarrollo de software, sino que también mejora la calidad del código al reducir la probabilidad de errores humanos[37].
  - **Buenas prácticas identificadas:** En el uso de asistentes de codificación basados en IA, una de las buenas prácticas más importantes es la integración de estos en el flujo de trabajo diario de los profesionales. La capacidad de estas herramientas para proporcionar sugerencias contextuales en tiempo real contribuye significativamente a la mejora de la productividad, facilitando el centrarse en aspectos más complejos del código mientras la IA se encarga de las tareas más repetitivas o de sugerir mejoras. Además de esto, la reducción de errores es otro beneficio crucial, ya que las sugerencias suelen basarse en patrones de buenas prácticas y estándares de la industria, lo que contribuye a la creación de código más limpio y robusto.

    La aceleración del desarrollo de software es otro impacto positivo, permitiendo a los profesionales completar sus tareas de manera más rápida y eficiente, lo que a su vez ayuda en gran medida a los equipos a cumplir con plazos de entrega estrictos y a adaptarse rápidamente a las demandas del mercado. La incorporación de estas herramientas también promueve un ambiente de aprendizaje continuo, ya que los desarrolladores pueden aprender de las sugerencias generadas por la IA, mejorando así sus habilidades de programación.

5.2. Ejemplos de mal uso y lecciones aprendidas

- **Discriminación algorítmica en contratación**
  - **Descripción:** Uso de un sistema de IA en procesos de contratación que resultó en discriminación por género.
  - **Impacto negativo:** Exclusión injustificada de candidatos, daño a la imagen de la empresa y posibles acciones legales.
  - **Ejemplo real:** El sistema de contratación automatizado de Amazon fue diseñado para revisar currículos y seleccionar a los mejores candidatos, aunque se descubrió más adelante que discriminaba sistemáticamente contra las mujeres. El algoritmo, entrenado con datos históricos, reflejaba y perpetuaba los sesgos presentes en esos datos, favoreciendo a los candidatos masculinos. Como resultado, Amazon abandonó el uso de esta tecnología, pero no sin antes enfrentar críticas públicas y un daño significativo a su reputación[38].
  - **Lecciones aprendidas:** Este ejemplo destaca la necesidad de crear y entrenar sistemas de IA utilizando datos variados y representativos para prevenir la reproducción de prejuicios. Las empresas deben realizar auditorías regulares de sus algoritmos para identificar y corregir cualquier sesgo, y deben incluir revisiones humanas en el proceso de toma de decisiones para garantizar la equidad. Además, es crucial que las organizaciones sean transparentes en el uso de esta y tomen medidas proactivas para mitigar cualquier impacto negativo potencial en los usuarios y en la sociedad.
- **Errores en diagnóstico por sesgo algorítmico**
  - **Descripción:** Uso de un sistema de IA para predecir el riesgo de sepsis en pacientes hospitalizados.
  - **Impacto negativo:** El sistema mostró sesgo contra pacientes de raza negra, resultando en una menor probabilidad de recibir atención oportuna y adecuada.
  - **Ejemplo real:** En 2020, un hospital en EE.UU. implementó un modelo de IA para predecir qué pacientes estaban en riesgo de desarrollar sepsis. Sin embargo, este exhibió un sesgo significativo contra pacientes de raza negra, lo que llevó a una menor probabilidad de que estos pacientes recibieran la atención oportuna y adecuada que necesitaban, pudiendo haber contribuido a resultados adversos en su salud[39].
  - **Lecciones aprendidas:** Este ejemplo destaca la necesidad de garantizar que estas tecnologías en el ámbito médico sean desarrollados y entrenados con datos representativos de todas las poblaciones que van a servir. Es de vital importancia realizar pruebas exhaustivas y auditorías regulares para identificar y corregir cualquier sesgo presente en los algoritmos. Además, los profesionales de la salud deben estar involucrados en el proceso de desarrollo y revisión de estos sistemas para asegurar que las decisiones algorítmicas sean validadas y que los pacientes reciban la mejor atención posible. La transparencia en

el uso de IA y la implementación de mecanismos de supervisión humana son esenciales para evitar la perpetuación de desigualdades en la atención médica.

- **Errores críticos en sistemas automatizados**
  - **Descripción:** Implementación de IA en sistemas críticos que resultó en errores graves.
  - **Impacto negativo:** Daños financieros, riesgo para la seguridad y fallos operacionales.
  - **Ejemplo real:** En 2012, la firma de trading Knight Capital experimentó una grave crisis por un fallo en su sistema de IA de trading. Esto ocurrió debido a un problema en el algoritmo que causó transacciones masivas de compra y venta de acciones, llevando a la empresa a perder $440 millones en apenas 45 minutos, lo que no solo causó un impacto financiero devastador para la empresa, sino que también puso en riesgo la estabilidad del mercado y la seguridad de las transacciones financieras[40].
  - **Lecciones aprendidas:** Este incidente subraya la importancia de una rigurosa verificación y validación de los modelos de IA antes de su implementación en entornos críticos. Es fundamental llevar a cabo pruebas detalladas para detectar y solucionar fallos potenciales en los algoritmos, así como establecer sistemas de monitoreo y control para identificar y corregir rápidamente cualquier error que pueda presentarse. Las organizaciones deben adoptar un enfoque proactivo en la gestión de riesgos, incluyendo planes de contingencia para responder a fallos operacionales, y mantener la transparencia y la responsabilidad en el uso de esta tecnología para asegurar la confianza del público y la estabilidad de los sistemas críticos.

CAPÍTULO 6

Conclusión

En este capítulo final, se resumen las reflexiones finales sobre el desarrollo actual y futuro del software con inteligencia artificial desde una perspectiva deontológica, además de que se subraya la importancia de estas prácticas para garantizar la calidad, la eficiencia y la ausencia de riesgos en los proyectos. En adición a esto, se sugieren áreas de investigación futura que pueden contribuir al fortalecimiento y a la mejora continua de este campo.

6.1. Reflexiones finales sobre el desarrollo actual y futuro de software con IA

La implementación de buenas prácticas en el desarrollo de software con IA es de vital importancia para garantizar que los proyectos sean de la mayor calidad posible y que no vulneren contra ningún derecho fundamental de las personas. A lo largo de este informe, se han mostrado las pautas que se deben seguir para lograr un buen uso de esta tecnología, manteniendo un enfoque centrado en el usuario y en los valores éticos y legales. Es fundamental adoptar un enfoque proactivo hacia la mejora continua, aprendiendo de los errores cometidos en el pasado y estando abiertos a la evolución de las respectivas normativas y principios que se vayan aplicando, ya que estos no solo aseguran un mayor éxito técnico, sino también la aceptación social y la sostenibilidad a largo plazo.

6.2. Investigaciones futuras

Hay varias áreas que se podrían llegar a investigar a futuro después de realizar este documento, como podría ser la integración de los principios éticos y de responsabilidad social en la inteligencia artificial, asegurando que se usen de forme justa y equitativa. También es crucial desarrollar herramientas y metodologías que faciliten la implementación efectiva de estas prácticas, haciendo que sean aún más fáciles de aplicar por los profesionales. Un último estudio que se podría llegar a realizar es cómo las buenas prácticas en este área pueden llegar a adaptarse y aplicarse a diferentes contextos y sectores industriales, maximizando su impacto y beneficios potenciales.

Bibliografía

[1] Colaboradores de Wikipedia. *Software*. 8 de abr. de 2024. URL: https://es.wikipedia.org/wiki/Software (visitado 25-04-2024).

[2] Garrigues Digital. «Pongamos orden en las definiciones de inteligencia artificial: así la define el reglamento de la UE que la regula». En: (1 de sep. de 2023). URL: https://www.garrigues.com/es_ES/garrigues-digital/pongamos-orden-definiciones-inteligencia-artificial-asi-define-reglamento-ue (visitado 25-04-2024).

[3] *Las claves de la nueva ley de Inteligencia Artificial*. 25 de ene. de 2024. URL: https://spain.representation.ec.europa.eu/noticias-eventos/noticias-0/las-claves-de-la-nueva-ley-de-inteligencia-artificial-2024-01-25_es (visitado 25-04-2024).

[4] Pablo Huet. *Inteligencia artificial en desarrollo de software: Tendencias emergentes y futuro*. 1 de mar. de 2024. URL: https://openwebinars.net/blog/inteligencia-artificial-en-desarrollo-de-software/ (visitado 26-04-2024).

[5] Sandra Cabrera. *La Inteligencia Artificial en el Desarrollo de Software: Impulsando la Innovación y la Eficiencia*. 17 de abr. de 2024. URL: https://itequia.com/es/inteligencia-artificial-desarrollo-software/ (visitado 26-04-2024).

[6] Microsoft Azure. *¿Qué es el aprendizaje automático?* URL: https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-machine-learning-platform (visitado 30-05-2024).

[7] *La importancia del aprendizaje automático en el desarrollo de aplicaciones - Auto WP*. 26 de jun. de 2023. URL: https://autowp.es/la-importancia-del-aprendizaje-automatico-en-el-desarrollo-de-aplicaciones/ (visitado 30-05-2024).

[8] Oracle. *¿Qué es el aprendizaje automático?* URL: https://www.oracle.com/es/artificial-intelligence/machine-learning/what-is-machine-learning/ (visitado 30-05-2024).

[9] Amazon. *¿Qué es el procesamiento de lenguaje natural? - Explicación del procesamiento de Lenguaje Natural - AWS*. URL: https://aws.amazon.com/es/what-is/nlp/ (visitado 30-05-2024).

[10] IBM. *¿Qué es el procesamiento del lenguaje natural (PLN)?* URL: https://www.ibm.com/es-es/topics/natural-language-processing (visitado 30-05-2024).

[11] IBM. *What is Computer Vision?* URL: https://www.ibm.com/es-es/topics/computer-vision (visitado 30-05-2024).

[12] Marketing. *Visión por Computador. Qué es, Aplicaciones y Objetivos*. 30 de mayo de 2022. URL: https://www.edsrobotics.com/blog/vision-computador-que-es/#:~:text=visi%C3%B3n%20por%20computador-,%C2%BFQu%C3%A9%20es%20la%20visi%C3%B3n%20por%20computador%3F,trav%C3%A9s%20de%20ellas%20(an%C3%A1lisis). (visitado 30-05-2024).

[13] Jean Carlos Santos y Pedro Anchundia. *Tipos de sistemas expertos | Sistemas expertos*. URL: https://pedro-94eg.wixsite.com/sistemas-expertos/tipos-de-sistemas-expertos (visitado 02-06-2024).

[14] *Sistemas Expertos: el impulso de la Inteligencia Artificial*. 24 de ago. de 2023. URL: https://www.santanderopenacademy.com/es/blog/sistemas-expertos.html (visitado 31-05-2024).

[15] Redacción InnovaciónDigital. *Sistemas expertos, guía completa: qué es, para qué sirven y clasificación*. 26 de feb. de 2024. URL: https://www.innovaciondigital360.com/i-a/sistemas-expertos-que-son-su-clasificacion-como-funcionan-y-para-que-se-utilizan/ (visitado 31-05-2024).

[16] Pablo Huet. *Qué son las redes neuronales y sus aplicaciones*. 24 de oct. de 2023. URL: https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/ (visitado 02-06-2024).

[17] *¿Qué es una red neuronal? | IBM*. URL: https://www.ibm.com/es-es/topics/neural-networks (visitado 31-05-2024).

[18] *¿Qué es una red neuronal? - Explicación de las redes neuronales artificiales - AWS*. URL: https://aws.amazon.com/es/what-is/neural-network/ (visitado 31-05-2024).

[19] *¿Qué es un IDE? - Explicación de los entornos de Desarrollo Integrado - AWS*. URL: https://aws.amazon.com/es/what-is/ide/#:~:text=Un%20entorno%20de%20desarrollo%20integrado%20(IDE)%20es%20una%20aplicaci%C3%B3n%20de,una%20aplicaci%C3%B3n%20f%C3%A1cil%20de%20usar. (visitado 06-06-2024).

[20] Chernandez. *Frameworks de IA: El arte de saber cual elegir y utilizar - OpenSistemas*. 14 de mar. de 2024. URL: https://opensistemas.com/elegir-y-utilizar-los-frameworks-de-ia/ (visitado 06-06-2024).

[21] José Manuel Ortega. «Frameworks de machine learning Open Source - José Manuel Ortega - medium». En: (15 de mayo de 2018). URL: https://jmortegac.medium.com/frameworks-de-machine-learning-open-source-5cbac67d38e5.

[22] Intel. *Elección de una plataforma de gestión en la nube*. URL: https://www.intel.la/content/www/xl/es/cloud-computing/cloud-management-platforms.html (visitado 07-06-2024).

[23] Akamai. *¿Qué es una plataforma en la nube?* URL: https://www.akamai.com/es/glossary/what-is-a-cloud-platform (visitado 07-06-2024).

[24] *Dashboard con Power BI - EVOTIC | Power BI*. 26 de sep. de 2023. URL: https://evotic.es/business-intelligence-bi/dashboard-con-power-bi/ (visitado 09-06-2024).

[25] Cristina Ortega. *Herramientas de visualización de datos: ¿Cuál elegir?* 6 de oct. de 2023. URL: https://www.questionpro.com/blog/es/herramientas-de-visualizacion-de-datos/ (visitado 07-06-2024).

[26] Admin. *¿Cómo limpiar y preprocesar datos antes del análisis?* 22 de ago. de 2023. URL: https://digitaltech180.com/big-data/analisis-de-datos/how-do-i-clean-and-preprocess-data-before-analysis/ (visitado 08-06-2024).

[27] FasterCapital. *Técnicas De Limpieza De Datos Para El Preprocesamiento*. URL: https://fastercapital.com/es/tema/t%C3%A9cnicas-de-limpieza-de-datos-para-el-preprocesamiento.html (visitado 08-06-2024).

[28] Parlamento Europeo. *Reglamento de Inteligencia Artificial*. 13 de mar. de 2024. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf (visitado 25-04-2024).

[29] Dirección General de Redes de Comunicación, Contenido y Tecnologías (Comisión Europea) y Grupo de expertos de alto nivel en inteligencia artificial. «Directrices éticas para una IA fiable». En: *Oficina de Publicaciones de la Unión Europea* (8 de nov. de 2019). DOI: 10.2759/14078. URL: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.

[30] Comisión Europea. *Reglamento del parlamento europeo y del consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión*. 21 de abr. de 2021. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206 (visitado 25-04-2024).

[31] *Textos aprobados - Normas de Derecho civil sobre robótica*. 16 de feb. de 2017. URL: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_ES.html (visitado 18-04-2024).

[32] *Reglamento (UE) n.º 1025/2012 del Parlamento Europeo y del Consejo, de 25 de octubre de 2012, sobre la normalización europea, por el que se modifica la Directiva 89/686/CEE del Consejo y las Directivas 93/15/CEE, 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE y 2009/105/CE del Parlamento Europeo y del Consejo y se derogan la Decisión 87/95/CEE del Consejo y la Decisión n.º 1673/2006/CE del Parlamento Europeo y del Consejo.* 1025/2012. 14 de nov. de 2012. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32012R1025 (visitado 17-05-2024).

[33] *Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo, de 9 de julio de 2008, por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos y se deroga el Reglamento (CEE) n.º 339/93.* 765/2008. 13 de ago. de 2008. URL: https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=celex%3A32008R0765 (visitado 17-05-2024).

[34] Parlamento Europeo y Consejo de la Unión Europea. *Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo relativo a la vigilancia del mercado y la conformidad de los productos Y por el que se modifican la Directiva 2004/42/CE y los reglamentos (CE) n.° 765/2008 y (UE) n.° 305/2011.* 2019/1020. 20 de jun. de 2019. URL: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32019R1020 (visitado 17-05-2024).

[35] Tricentis. *Capgemini - Tricentis.* 24 de abr. de 2024. URL: https://www.tricentis.com/partners/capgemini (visitado 30-06-2024).

[36] Erika Ulring. *How predictive maintenance improves efficiencies across five industries - IBM Blog.* 2 de mayo de 2023. URL: https://www.ibm.com/blog/predictive-maintenance-efficiencies-client-case-studies/ (visitado 30-06-2024).

[37] Eirini Kalliamvakou. *Research: quantifying GitHub Copilot's impact on developer productivity and happiness - The GitHub Blog.* 21 de mayo de 2024. URL: https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/ (visitado 30-06-2024).

[38] Rachel Goodman. «Why Amazon's automated hiring tool discriminated against women | ACLU». En: (27 de feb. de 2023). URL: https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against.

[39] Luna Wolfe. «Medical AI Misuse Could Cause Harm to Patients, Researchers Warn». En: (31 de oct. de 2023). URL: https://medium.com/@lunawolfe01/medical-ai-misuse-could-cause-harm-to-patients-researchers-warn-3b255bd43404.

[40] Henrico Dolfing. *Case Study 4: The $440 Million Software Error at Knight Capital.* 5 de jun. de 2019. URL: https://www.henricodolfing.com/2019/06/project-failure-case-study-knight-capital.html (visitado 30-06-2024).