www.ismsforum.es

# INTRODUCCIÓN A LA IA PARA PROFESIONALES DE SEGURIDAD DE LA INFORMACIÓN

Una iniciativa de

isms forum

NOVIEMBRE 2023

# INTRODUCCIÓN A LA IA PARA PROFESIONALES DE SEGURIDAD DE LA INFORMACIÓN

# AUTORES


| | |
|---|---|
| **COORDINADORES** | Ángel Pérez |
| | Francisco Lázaro |
| **SUBCOORDINADORES** | Fernando Rubio |
| **PARTICIPANTES** | Alberto Bernaldez |
| | Gemma Deler |
| | Jorge Barrios |
| | Juan Carlos Aguero |
| | Juan Carlos Valle |
| | Mª Pilar Alanpont |
| **GESTOR DE PROYECTOS** | Beatriz García |
| **DISEÑO/MAQUETACIÓN** | Rim Souri |

# CONTENIDOS

# CONTENIDOS

# 1 INTRODUCCIÓN

> La inteligencia artificial no tiene como finalidad reemplazar a los humanos, sino mejorar significativamente las capacidades y contribuciones de estos.

El Informe sobre el Futuro de los Empleos 2023 del Foro Económico Mundial indica que se espera que el 44% de las competencias básicas de los trabajadores cambien en los próximos cinco años debido al aumento de las tareas realizadas por máquinas.

En cuanto a la Inteligencia Artificial Generativa o IAG (rama de la inteligencia artificial que se enfoca en la generación de contenido original a partir de datos existentes), un estudio reciente realizado por OpenAI concluye que aproximadamente el 80 % de la población activa de EE. UU. verá afectada una parte de su trabajo por los GPT (Generative Pre-trained Transformer, un modelo lingüístico capaz de producir texto similar al humano) y alrededor del 19% de los trabajadores verá afectada una parte significativa de su trabajo.

A lo largo de las páginas de este documento se tratan diversos conceptos y aspectos relacionados con la Inteligencia Artificial (IA en adelante), por eso debemos comenzar buscando una definición de IA que nos permita entender qué es.

EL MIT maneja como definición de IA: aquella entidad que realiza comportamientos que una persona podría calificar razonablemente de inteligentes si un humano hiciera algo similar.

La RAE por su parte define a la IA, como la disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.

Finalmente, el Reglamento de la UE define como Sistema de inteligencia artificial (sistema de IA): un programa informático desarrollado con una o varias de las técnicas y enfoques enumerados en el anexo I (del reglamento) y que puede, para un conjunto determinado de objetivos definidos por el ser humano, generar resultados tales como contenidos, predicciones, recomendaciones o decisiones que influyen en los entornos con los que interactúan.

Así pues, podríamos decir que la IA mediante sus algoritmos ejecutados en sistemas informáticos se centra en crear sistemas y programas capaces de realizar tareas, que normalmente requerirían de intervención humana, replicando o imitando la inteligencia humana; es decir, replicando la capacidad de razonamiento, aprendizaje, percepción y toma de decisiones que asociamos con los seres humanos. En principio La inteligencia artificial no tiene como finalidad reemplazar a los humanos, sino mejorar significativamente las capacidades y contribuciones de estos.

La IA está transformando las organizaciones al brindarles una ventaja competitiva, mejorar las experiencias de los clientes y mejorar la eficiencia en sus procesos internos.

# 1.1. UN POCO DE HISTORIA DE LA IA

Aunque podamos pensar que la IA es un fenómeno reciente, ya en 1950, Alan Turing con un enfoque totalmente humano de la IA diseñó la conocida "Prueba de Turing", la computadora pasa la prueba si una persona le plantea una serie de preguntas escritas y no puede conocer si las repuestas provienen de un humano o no.

El primer trabajo de IA suele atribuirse a Warren MCCulloch y Walter Pitts, en 1943, publicaron conjuntamente un artículo titulado "A Logical Calculus of Ideas Immanent in Nervous Activity", que sentó las bases teóricas para el desarrollo de las redes neuronales artificiales. Propusieron un modelo matemático simplificado de una neurona, conocido como la "neurona de McCulloch-Pitts". Esta neurona artificial era una abstracción de las neuronas biológicas del cerebro humano y se basaba en la idea de que las neuronas individuales pueden funcionar como elementos de procesamiento de información. Este modelo proporcionó un marco matemático para describir el funcionamiento de las redes neuronales y cómo pueden realizar operaciones lógicas y computacionales. Aunque la neurona de McCulloch-Pitts era un modelo simplificado y no tenía la capacidad de aprender, sentó las bases para futuros avances en la simulación de sistemas neuronales y el desarrollo de algoritmos de aprendizaje en redes neuronales artificiales.

Herbert A. Simon y Allen Newell (1952-1969) fueron dos destacados investigadores en el campo de la inteligencia artificial y la psicología cognitiva. Juntos, realizaron importantes contribuciones que ayudaron a sentar las bases de la inteligencia artificial moderna. Entre sus logros más destacados se encuentran los siguientes:

I **Lógica teórica y sistemas simbólicos:** Desarrollaron el primer programa de computadora capaz de demostrar teoremas matemáticos, su programa, llamado "Logic Theorist", fue desarrollado en 1956 y marcó un hito en la capacidad de las computadoras para razonar simbólicamente.

II **Teoría de la toma de decisiones:** Simon es conocido por sus contribuciones a la teoría de la toma de decisiones. Junto con Newell, desarrollaron el concepto de "racionalidad limitada", que postula que los seres humanos toman decisiones basadas en información incompleta y utilizando recursos cognitivos limitados. Su enfoque se basó en modelos computacionales y matemáticos para comprender cómo se toman las decisiones en condiciones de incertidumbre.

III **Arquitectura cognitiva:** Simon y Newell propusieron la noción de una arquitectura cognitiva basada en el procesamiento de información simbólica. Su modelo, llamado "GPS" (General Problem Solver), fue uno de los primeros en presentar una forma estructurada de resolver problemas a través de la manipulación de símbolos y reglas.

En conjunto, las contribuciones de Simon y Newell sentaron las bases teóricas y prácticas para el desarrollo de la inteligencia y la cognición artificiales. Sus investigaciones influyeron en el desarrollo de modelos computacionales y enfoques de razonamiento simbólico que aún son relevantes en el campo de la inteligencia artificial hoy en día.

En la década de 1960 fue desarrollado el programa DENDRAL (Edward A. Feigenbaum y Joshua Lederberg), que fue un proyecto de investigación pionero en el campo de la inteligencia artificial y la química computacional considerándose uno de los primeros sistemas expertos. El programa DENDRAL utilizaba técnicas de inteligencia artificial para analizar y deducir estructuras moleculares a partir de datos espectroscópicos. Se centraba en el problema de determinar la estructura química de compuestos orgánicos desconocidos utilizando información obtenida a partir de técnicas como la espectrometría de masas y la espectroscopia de resonancia magnética nuclear. Logró resultados destacados en la deducción de estructuras moleculares desconocidas, superando en muchos casos la capacidad de los químicos humanos para realizar la misma tarea.

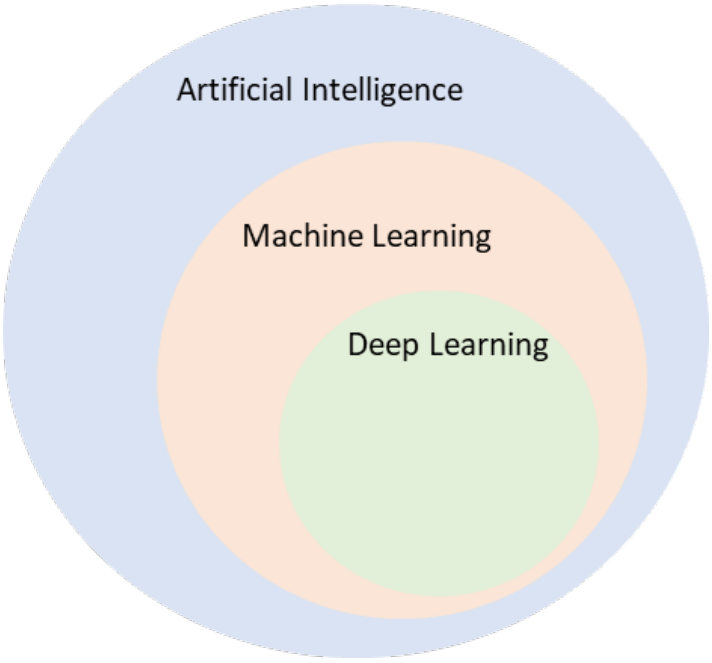
Ya entrado el siglo XXI, en el desafío de reconocimiento de imágenes ImageNet 2012, una red neuronal convolucional llamada AlexNet logró un error top-5 del 15.3%, más de 10.8 puntos porcentuales más bajo que el del segundo clasificado. Se suele identificar este hito como el inicio de la etapa moderna del aprendizaje profundo y fue en gran medida posible gracias al uso de nuevo hardware como GPUs para el entrenamiento de una red de muchas capas.

## 1.2. LA IA EN LA ACTUALIDAD

En la actualidad bajo el termino IA encontramos las siguientes variantes de modelos:

- Aprendizaje automático (Machine Learning, en adelante ML): las máquinas aprenden de los datos sin ser programadas. Los algoritmos de aprendizaje automático analizan patrones y construyen modelos predictivos o descriptivos basados en esos patrones. El ML diseña modelos predictivos que construyen por sí mismos la relación entre las variables a estudiar mediante el análisis de un conjunto inicial de datos, la identificación de patrones y el establecimiento de criterios de clasificación. El aprendizaje automático está, relacionado con las técnicas de minería de datos, optimización y big data.
- Aprendizaje profundo (Deep Learning) y Redes neuronales artificiales: Inspiradas en la estructura y funcionamiento del cerebro humano. Estas redes están compuestas por nodos interconectados llamados neuronas artificiales, que procesan la información y realizan cálculos complejos.
- Procesamiento del lenguaje natural: Se centra en permitir a las máquinas comprender y comunicarse con los humanos en lenguaje natural. Esto implica el procesamiento y análisis de texto, habla y otros tipos de datos lingüísticos.
- Visión por computadora: Es el campo que se ocupa de dotar a las máquinas de la capacidad de comprender y analizar imágenes y vídeos. Esto implica la detección de objetos, reconocimiento facial, seguimiento de movimiento, entre otras tareas relacionadas con la visión.

Y así la IA se ha convertido en una tecnología revolucionaria que está transformando rápidamente diversos sectores y aspectos de nuestras vidas. Desde los asistentes virtuales en nuestros teléfonos inteligentes hasta los algoritmos que impulsan los motores de búsqueda en Internet o herramientas como Chatgpt, Bard o Llama, la IA está presente en muchas aplicaciones cotidianas y empresariales.



## 1.3. EL FUTURO DE LA IA, INTELIGENCIA GENERAL Y SUPERINTELIGENCIA

Asistimos en la actualidad a la evolución exponencial de los usos de la inteligencia artificial, aprovechando la mejora en las capacidades de almacenamiento, proceso y conectividad de la última década para resolver problemas complejos y construyendo modelos que resuelvan tal y como lo hacemos los humanos.

A ello se ha unido, recientemente, una interfaz que nos permite la interacción de forma natural, de la misma manera que consultaríamos a nuestros colegas o preguntaríamos a un experto. Ello abre múltiples oportunidades y nuevos casos de uso que se ponen al servicio y facilitan los procesos de digitalización en los que estamos inmersos.

La combinación de la mejora en la capacidad y esta nueva interfaz nos coloca ante un fenómeno disruptivo que, muy probablemente, marcará un antes y un después tal y como pasó con el nacimiento de Internet.

En el futuro se debate hoy sobre si es de esperar que los sistemas de IA cada vez sean capaces de resolver un mayor número de tareas, hasta llegar a la Inteligencia Artificial General (AGI). AGI se refiere a una máquina que tiene la capacidad de comprender y aprender cualquier tarea intelectual que un humano puede realizar. Por ejemplo, una AGI podría ser un sistema capaz de realizar múltiples tareas y adaptarse a diferentes entornos. La AGI es un concepto teórico que aún no hemos logrado desarrollar, pero que podría tener grandes implicaciones para el futuro de la humanidad.

De igual modo se debate si en el futuro se encuentra la Inteligencia Artificial Superinteligente (ASI), es decir, una máquina que supera ampliamente a la inteligencia humana en todos los aspectos posibles. Por ejemplo, una ASI podría ser capaz de resolver problemas globales como el cambio climático o la desigualdad económica, desarrollar tecnologías y descubrimientos científicos inimaginables para nosotros, etc. La ASI es un concepto hipotético, que podría representar un riesgo o una oportunidad para la humanidad, dependiendo de cómo se diseñe y se controle.

Por ello, conscientes que estamos en los momentos iniciales de esta nueva tecnología y para que se pueda desarrollar en un entorno ético y seguro, es fundamental la involucración de todas las áreas de la sociedad, incluyendo por supuesto a los profesionales de la seguridad.

## 1.4. CONSIDERACIONES DE ÉTICA Y SEGURIDAD DE LA IA

La Inteligencia Artificial genera muchas dudas entre los usuarios, investigadores, especialistas, autoridades y la industria con relación a aspectos de cumplimiento normativo, respeto a los derechos de los interesados y seguridad jurídica de todos los intervinientes. También en cuanto a aspectos éticos y sociales, además como nueva tecnología introduce consideraciones de seguridad en su uso.

Así, entre estas consideraciones de seguridad está las diferentes amenazas que los sistemas pueden sufrir como envenenamiento de los datos con los que se entrenan los modelos, creación de puertas traseras, inferencia de los datos con los que ha sido entrenado el modelo a partir de sus respuestas, inversión del modelo por parte de un atacante mediante peticiones maliciosas, etc. Así mismo, la IA generativa introduce nuevos riesgos que deben ser convenientemente mitigados.

Igualmente existen consideraciones de privacidad, la protección de las personas físicas en relación con el tratamiento de datos personales es un derecho fundamental.

En definitiva, la IA es un fenómeno tecnológico que, si bien se utiliza desde hace mucho tiempo, en los últimos años se ha acelerado muchísimo su crecimiento, y vive hoy una nueva fase de adopción debido a las revoluciones del Deep Learning y la IA generativa. Los profesionales de la seguridad y el cumplimiento deben trabajar para que su adopción cumpla con los estándares de ética, privacidad y seguridad y se logre un uso beneficioso para la sociedad mitigando los principales riesgos.

# 2 SISTEMAS Y AGENTES INTELIGENTES

Los sistemas inteligentes se desarrollaron para automatizar y optimizar procesos. Tradicionalmente para automatizar tareas que requerían un esfuerzo físico. En la época digital comenzaron a incluir tareas que requerían un esfuerzo mental y sucesivamente han ido incluyendo elementos de inteligencia artificial, Big Data, robótica o visión artificial, que les permiten realizar cualquier función. En nuestros días un sistema inteligente es un sistema artificial que opera como un agente, es decir, el sistema percibe su entorno mediante sensores, toma decisiones inteligentes y finalmente actúa con el entorno a través de actuadores.

**PERCEPCIÓN MEDIANTE SENSORES**

Los sistemas inteligentes deben entender el entorno que les rodea ya que es un requisito fundamental para analizar la situación y tomar decisiones. Para ello, utilizan sensores que les permiten obtener los datos de su entorno. Estos sensores pueden tomar información muy diversa. El sistema inteligente mediante algoritmos crea y mantiene actualizada una representación del mundo.

**INTELIGENCIA Y DECISIÓN**

El aprendizaje es una de las características asociadas con la inteligencia. Existen dos formas de dotar inteligencia a los sistemas: un aprendizaje basado en refuerzo, recompensándole cuando realiza una tarea de forma óptima, y otro mediante la adquisición de conocimiento, en que el sistema aprende a base de alimentarle con información. En ambos casos los agentes se basan en modelos matemáticos que modifican sus características adaptándose a los cambios del entorno mediante un entrenamiento continuo.

Estos modelos y el entendimiento del entorno, dota al agente de la capacidad de análisis. Esta capacidad permite realizar inferencias sobre información recibida por los sensores en una situación que no se ha observado en el pasado y así poder tomar una decisión.

**ACTUADORES**

Los actuadores son los responsables de ejecutar acciones para modificar el entorno una vez que se ha tomado una decisión con la información adquirida. Estos actuadores pueden ser de dos tipos: reactivos, realiza una acción como respuesta a la información observada en tiempo o basados en objetivos, las acciones se basan en un razonamiento en función de objetivos.

# 3 APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

## 3.1. ¿QUÉ ES EL APRENDIZAJE AUTOMÁTICO?

El aprendizaje automático es un campo de estudio que brinda a las computadoras la capacidad de aprender y mejorar automáticamente a partir de la experiencia. El aprendizaje automático vive en la intersección de la informática, la estadística y la ciencia de datos. Utiliza elementos de cada uno de estos campos para procesar los datos de una manera que pueda detectar y aprender de los patrones, predecir la actividad futura, o tomar decisiones. Se categoriza en función del tipo de aprendizaje: Supervisado, no supervisado y por refuerzo.

## 3.2. TIPOS DE APRENDIZAJE

### 3.2.1. APRENDIZAJE SUPERVISADO

El aprendizaje supervisado consiste en entrenar el modelo con datos etiquetados. Como el modelo se inicializa de forma aleatoria las primeras predicciones serán poco acertadas, sin embargo, al disponer de las etiquetas puede comprobar si sus predicciones son acertadas o no y adaptar el modelo con esta realimentación. Durante el entrenamiento el modelo se fija en las características más relevantes de cada entidad y las asocia a una clase. De esta manera podrá hacer una predicción en función de las características de cualquier dato de entrada.

Los tipos de aprendizaje supervisado son: regresión y clasificación. La regresión nos relaciona características independientes con una variable dependiente numérica. Por ejemplo, el precio del oro, variable dependiente, en función de variables independientes como la infracción, tipos de interés, etc. La clasificación es similar a la regresión, pero se utiliza cuando la variable dependiente es categórica. Por ejemplo, clasificar si un correo es spam o no, variable dependiente, en función de los parámetros de las cabeceras, variables independientes.

## 3.2.2. APRENDIZAJE NO SUPERVISADO

El aprendizaje no supervisado consiste en entrenar el modelo sin supervisión. El modelo se entrena con un conjunto de datos sin etiquetar con el objetivo de que el propio sistema reconozca y extraiga las relaciones ocultas entre las entidades y las agrupe en función de la similitud o discrepancia con los patrones observados.

Los tipos de aprendizaje no supervisado son: agrupación y asociación. Cada tipo de aprendizaje se utiliza en función de lo que queramos predecir. En el caso de los modelos basados en agrupación durante el entrenamiento se tratará de dividir los datos en grupos similares en función de sus características. En la definición del modelo se puede indicar el tipo de agrupación y se debe realizar en función del problema que queremos resolver; grupos homogéneos, agrupaciones naturales o detección de valores atípicos son algunos ejemplos. Por otro lado, en los modelos basados en asociación durante el entrenamiento se crearán reglas para encontrar las relaciones entre los elementos de un conjunto de datos, por ejemplo, las reglas para desarrollar sistemas de recomendación.

## 3.2.3. APRENDIZAJE POR REFUERZO

El aprendizaje por refuerzo consiste en entrenar el modelo de forma interactiva en formato prueba y error. La diferencia con el aprendizaje supervisado es que el aprendizaje por refuerzo utiliza un sistema de recompensas o castigos con el objetivo de encontrar el modelo que maximice la recompensa total acumulada. El aprendizaje por refuerzo se utiliza en problemas en los que se deben tomar decisiones de forma secuencial con dependencia del estado anterior.

Los tipos de aprendizaje por refuerzo son: refuerzo positivo y refuerzo negativo. En el refuerzo positivo se fomenta que se realicen acciones al ofrecer recompensas en base al resultado de la acción, maximiza el rendimiento. Por el contrario, el refuerzo negativo se elimina la recompensa al realizar acciones lo que fomenta el mínimo desempeño.

## 3.3. TÉCNICAS DE EVALUACIÓN Y VALIDACIÓN

Las técnicas de evaluación y validación nos van a permitir medir la validez de los modelos que hemos entrenado para una tarea. Para ello se define la función de pérdida (también conocida como función de coste) que va a medir el error producido y que determina cómo va a aprender el modelo. El objetivo durante el entrenamiento es minimizar dicha función.

Parte de los datos disponibles no se usarán para el entrenamiento y se dejarán para una segunda fase de pruebas en las que se valida que el modelo se comporta con la misma calidad con datos no usados durante el entrenamiento.

En clasificación (machine learning), las métricas más usadas son:

- **La Precisión *(precision)*** representa la calidad del modelo cuando hace una predicción. Indica qué porcentaje de las predicciones que hemos hecho han sido correctas.
- **La Exhaustividad *(recall)*** o sensibilidad nos informa sobre la cantidad de elementos que es capaz de clasificar correctamente el modelo.
- **La Exactitud *(accuracy)*** mide el porcentaje de casos que el modelo ha acertado. Es una métrica que puede llevar a confusión si las clases están desbalanceadas. Si simplemente predecimos la clase más probable, sin tener un modelo detrás, el resultado de la exactitud sería muy alto.

En regresión la métrica más usada es el MSE (error cuadrático medio), el valor medio de los cuadrados de la diferencia entre los valores predichos y los valores reales. Otras posibilidades habituales son el MAE (error medio absoluto), o el R2 (coeficiente de determinación.

A la hora de optimizar un modelo para un caso de uso concreto, es fundamental hacernos preguntas como: ¿Podemos asumir un alto ratio de falsos positivos? ¿Quizás sea mejor optimizar para una mayor exhaustividad?, etc. Es habitual que las diferentes métricas estén relacionadas, a menudo de manera inversa y que de manera sencilla podamos adaptar el modelo en función de nuestros requisitos hacia aquellas que más nos interesen.

## 3.4. DEGRADACIÓN DEL MODELO Y REENTRENAMIENTO

Las métricas permiten evaluar si el modelo es bueno y realizar modificaciones o volver a reentrenarlo si no cumple las expectativas. Se debe tener en cuenta que no se puede construir un modelo preciso al 100% y que es habitual que se degraden en el tiempo. Existen algunas recomendaciones que se pueden seguir para mejorar los modelos. Las fuentes de errores que debemos analizar son:

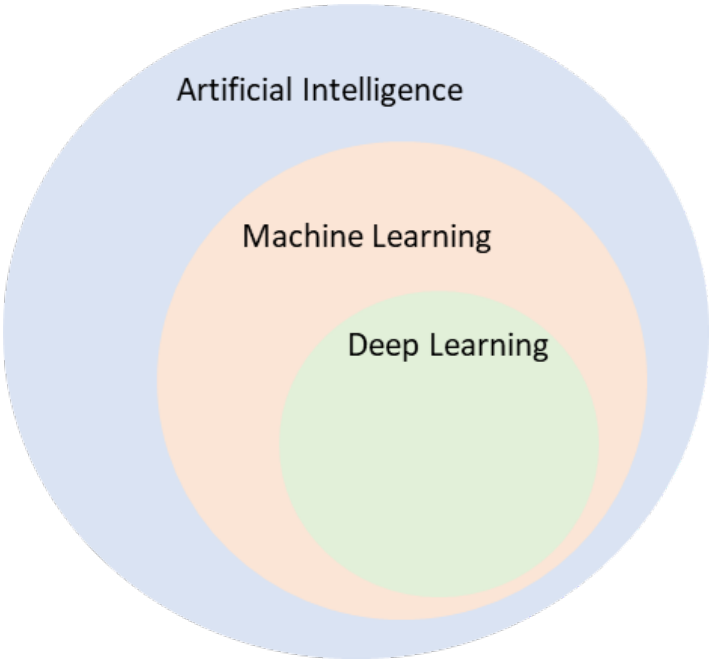
- **Error de varianza**: La función objetivo se estima durante el entrenamiento por lo que debemos asegurarnos de que los datos de entrenamiento no sean muy diferentes a los datos en test o producción.
- **Error de sesgo**: Este error se da cuando se utiliza un modelo demasiado simple que no es capaz de aprender las relaciones entre las variables de análisis y la variable a predecir. Usar regresión lineal con datos que tienen un patrón no lineal produciría este error.
- **Sobreajuste *(Overfitting)***: En el afán de querer obtener las mejores métricas durante el entrenamiento se pueden generar modelos tan adaptados a los datos de entrenamiento que recojan el ruido y las fluctuaciones aleatorias de estos datos. Este error se detecta cuando los valores de las métricas en entrenamiento y en test son muy diferentes. Este error se puede resolver con técnicas de validación cruzada o métodos de regularización.

# 4 APRENDIZAJE PROFUNDO (DEEP LEARNING)

El Deep Learning, o aprendizaje profundo en español, es un campo parte del aprendizaje automático dedicado al estudio y desarrollo de modelos computacionales, compuestos por múltiples capas de procesamiento, que le permite a una máquina aprender a partir de unos datos. Gracias a estas múltiples capas, la máquina aprende representaciones, patrones o características de los datos con varios niveles de abstracción. Esto ha permitido realizar grandes avances en diferentes campos como la visión artificial, la interpretación del lenguaje, la medicina o biología entre otras numerosas áreas de estudio.

## 4.1. REDES NEURONALES ARTIFICIALES: ASPECTOS CLAVES

El aprendizaje profundo se basa en redes neuronales artificiales. Una red neuronal es un algoritmo inspirado en el funcionamiento del cerebro humano. Se componen de múltiples neuronas artificiales interconectadas, organizadas en capas. Cada neurona toma entradas, realiza un cálculo utilizando pesos y funciones de activación, y produce una salida que se transmite a las neuronas de la capa siguiente. Así tendremos una capa de entrada, una o varias capas ocultas y una capa de salida.




*Ejemplo de red neuronal con una capa de entrada, una de salida y únicamente una capa oculta (fuente: Wikipedia)*

Una red neuronal aprende mediante un proceso de entrenamiento, en donde se presenta a la red un conjunto de datos de entrada junto con las salidas deseadas. El objetivo del entrenamiento es minimizar esta función de coste, ajustando los pesos de las conexiones, de manera que las predicciones de la red sean lo más cercanas posible a las salidas reales. Esto se logra utilizando algoritmos de optimización, el más famoso siendo el descenso del gradiente mediante retropropagación (backpropagation), que guían los ajustes de los parámetros en función del error cometido.

Hay que destacar que no es necesario entrenar todas las redes desde cero, se puede utilizar el enfoque de aprendizaje por transferencia (fine-tuning). En este se toma una red neuronal ya entrenada y se la entrena nuevamente en el nuevo conjunto de datos para adaptarla a tareas o dominios específicos. Esto permite aprovechar el conocimiento previo de la red y acelerar el proceso de entrenamiento en el nuevo dominio.

## 4.2. TIPOS DE REDES NEURONALES ARTIFICIALES MÁS COMUNES

Actualmente, existen diversos tipos de redes neuronales. Las más conocidas son las Redes Neuronales Convolucionales (Convolutional Neural Network – CNN) y las Redes Neuronales Recurrentes (Recurrent Neural Network – RNN) que aparecieron a principio de los 90. Estas redes están diseñadas para procesar datos que vienen en forma de matrices múltiples, por ejemplo, una imagen en color compuesta por tres matrices 2D que contienen intensidades de píxeles en los tres canales de color. Muchas modalidades de datos se pueden representar en forma de matrices múltiples: 1D para señales y secuencias, incluido el idioma; 2D para imágenes o espectrogramas de audio; y 3D para video o imágenes volumétricas.

Aparte de las comentadas anteriormente, también nos encontramos Redes Neuronales Recurrentes Bidireccionales (BRNN), que pueden observar tanto el pasado como el futuro. Estas procesan secuencias de datos en ambas direcciones, capturando así relaciones complejas en datos secuenciales. Esto es útil en tareas como reconocimiento de voz y análisis de secuencias de ADN. También existen las redes neuronales de tipo autoencoder, que toman una entrada de datos, la comprimen a una representación más pequeña (llamada codificación), y luego trata de reconstruir la entrada original a partir de esa codificación. Esto se utiliza por ejemplo para la compresión de datos, eliminación de ruido y detección de anomalías.

Actualmente existen dos tipos de redes que están transformando el panorama del aprendizaje profundo, estas son las redes neuronales transformer y las redes neuronales generativas adversariales (GAN). Las transformers son los superhéroes del procesamiento del lenguaje natural. Pueden analizar grandes cantidades de texto y capturar relaciones complejas entre las palabras. Son muy útiles en tareas de traducción automática, generación de texto y resumen de documentos.

Las redes GAN surgieron como respuesta a la pregunta de si las máquinas pueden crear contenido nuevo y original. A diferencia de las redes neuronales tradicionales que se centran en predecir y clasificar datos, las redes generativas se enfocan en generar datos nuevos que se asemejen a los datos de entrenamiento. Dichas redes, están revolucionaron la inteligencia artificial, dando lugar a una nueva rama, la IA generativa y están aportando grandes avances en la generación de imágenes, música, texto y video. Esta nueva IA generativa está abriendo nuevas posibilidades creativas y desafíos éticos. Por un lado, permite a los artistas y creadores explorar nuevas formas de expresión y generar contenido original. Por otro lado, plantea preguntas sobre la autenticidad y la responsabilidad ética al utilizar datos generados por máquinas.

## 4.3. CONCEPTO DE MULTIMODALIDAD

Por otro lado, actualmente está emergiendo el concepto de multimodalidad dentro del aprendizaje profundo. Esto se refiere a la capacidad de una red neuronal para procesar y fusionar diferentes tipos de información provenientes de múltiples fuentes, como texto, imágenes, audio, entre otros. Podría asemejarse a combinar los sentidos humanos para obtener una comprensión más completa. Por ejemplo, si queremos entender una historia, podemos leer el texto, ver imágenes relacionadas y escuchar el audio. De manera similar, las redes neuronales multimodales pueden tratar diferentes modalidades de datos para obtener una representación más rica y precisa. La multimodalidad, junto con la IA generativa puede que sean las áreas clave en donde se enfocarán los investigadores los próximos años.

## 4.4. PERSPECTIVAS DE FUTURO DEL APRENDIZAJE PROFUNDO

El campo del aprendizaje profundo ha experimentado avances revolucionarios en los últimos años, transformando la forma en que abordamos problemas complejos de inteligencia artificial. Las redes neuronales profundas han demostrado una capacidad única para aprender y resolver problemas complejos. A medida que la investigación y la tecnología continúan avanzando, se espera que las redes neuronales se vuelvan más eficientes, capaces de manejar conjuntos de datos aún más grandes y complejos (un ejemplo de ello son los grandes modelos del lenguaje, LLM en inglés). Además, se espera que haya mejoras en la interpretabilidad y la explicabilidad de los modelos, lo que permitirá una toma de decisiones más confiable y ética.

# 5 APLICACIONES EN IA

Prácticamente cualquier proceso puede beneficiarse del uso de la IA, dado que, con esta tecnología, se busca dotar de habilidad humana a los sistemas y procesos y hacer que respondan a situaciones reales sin requerir un nuevo desarrollo.

Como se ha mencionado anteriormente, aunque ya desde los años 50 se trabajaba en los modelos matemáticos asociados a la IA, su implementación se había visto limitada por el alto consumo de recursos. Actualmente, el despliegue de tecnologías para el tratamiento masivo de datos, las mejoras en el hardware que incrementan la velocidad de los algoritmos o el abaratamiento de los dispositivos de almacenamiento, están haciendo posible su uso en una multitud de campos, de los que destacamos:

## 5.1. VISIÓN POR COMPUTADOR Y PROCESAMIENTO DE IMÁGENES

Es una de las principales áreas de aplicación del aprendizaje profundo, tanto para reconocimiento de objetos como en el reconocimiento de caracteres y patrones.

La visión artificial es de especial interés en las aplicaciones de vigilancia y seguridad, al no ser sensible a la fatiga. También para el reconocimiento facial en imágenes, en la detección del movimiento, en la detección de objetos abandonados o que no debieran estar presentes en determinadas áreas. Otros ejemplos son en el guiado de carretillas robotizadas en plantas logísticas, la detección de anomalías en imágenes como la rotura de una tubería o la conducción autónoma.

Cabe mencionar que, en algunos casos, la visión artificial supera a aquella que se desea imitar, ofreciendo soluciones a retos complejos, no viables con visión tradicional por ejemplo cuando la captura va más allá del espectro visible.

## 5.2. RECONOCIMIENTO Y SÍNTESIS DE VOZ

El reconocimiento y tratamiento de la secuencia de palabras en una conversación es fundamental en los sistemas de traducción simultánea. La aplicación de la IA permite que, en un futuro próximo, sea posible mantener reuniones multi lenguaje, teniendo en cuenta las variantes locales, giros, refranes, modismos … que están presentes en el lenguaje natural.

Respecto a la síntesis de voz, cada vez encontramos más compañías que hacen uso del "Text-to speech", en la creación y edición de contenidos multimedia.

## 5.3. GENERACIÓN Y RESUMEN DEL LENGUAJE NATURAL

Los lenguajes naturales son ambiguos, lo cual dificulta su descripción formal. Para resolver con éxito este reto, tanto traductores como asistentes virtuales, entre otros, hacen uso de la IA.

Sistemas como ChatGPT o Google Bard han popularizado la IA generativa para un público general.

## 5.4. SISTEMAS DE RECOMENDACIÓN Y PERSONALIZACIÓN

Estos sistemas usan la IA para personalizar al máximo la oferta a cada individuo más allá del análisis de los datos de su consumo anterior.

Se recogen los datos proporcionados por los propios usuarios, sea de manera directa o indirecta, se analizan y contrastan junto con datos de colectivos con los que el usuario comparte características. Así permite crear patrones de los gustos y preferencias del individuo y generar recomendaciones a partir del patrón generado y de los contenidos consumidos por usuarios con un patrón similar.

## 5.5. APRENDIZAJE POR REFUERZO

Se trata de un área de aprendizaje automático en el que se aplica el Machine Learning para planear estrategias efectivas. Usa un sistema de recompensas positivas o negativas para buscar la conducta óptima y reforzar el aprendizaje.

Entre sus aplicaciones se encuentra el mantenimiento predictivo, la optimización de costes, la minimización de riesgos o la resolución de problemas complejos en sistemas dinámicos desde la experiencia, sin olvidar la aplicación en el caso de los juegos de ordenador.

## 5.6. IOT

La IA permite disponer de dispositivos inteligentes que se adapten al comportamiento de sus usuarios además de a los cambios en el entorno.

Algunos de los campos que se benefician son la domótica o las ciudades inteligentes (Smart cities) en aplicaciones que, por ejemplo, optimizan el consumo de energía o la recogida de residuos, o de transporte inteligente en la que se recoge información de sensores y se aplica la IA para predecir y optimizar el tráfico.

No son los únicos y cabe mencionar otros en los que el tratamiento con IA de los datos recogidos con los sensores representa una gran ventaja, como por ejemplo en aplicaciones de agricultura inteligente – los sensores IoT recogen datos sobre humedad del suelo, temperatura y otros factores ambientales para maximizar la producción y reducir el consumo de agua – o en el ámbito de la salud con sensores que monitorizan las constantes del paciente, detectando anomalías y alertando a los profesionales.

## 5.7. ROBÓTICA

Se trata de un campo estrechamente relacionado con la visión artificial y con el procesado de voz para poder recibir órdenes, atender a alarmas ...

Más allá del reconocimiento de defectos en el proceso de producción, la IA hace posibles aplicaciones de utilidad en entornos caóticos, como por ejemplo el bin picking en los que la visión artificial ayuda al robot a detectar el objeto, dónde está, la posición en la que está, decidiendo cual es la mejor forma para su recolección y agarre, lo que por ejemplo permite la separación y recuperación de residuos en plantas de reciclaje.

Además de los entornos industriales en su aplicación, otros se ven beneficiados como, por ejemplo, el agrícola, pudiéndose aplicar a la poda de vides, o al reconocimiento y tratamiento de plagas.

## 5.8. CIBERSEGURIDAD Y CIBERDELINCUENCIA

Esta área se beneficia de la incorporación de la IA en las soluciones de ciberseguridad. Los fabricantes incorporan estas técnicas en el reconocimiento de patrones, en la interrelación entre sucesos y en el análisis de situaciones para incrementar su efectividad tanto en prevención detección o repuesta.

El uso de la IA no queda restringido a aplicaciones lícitas y estas técnicas también son usadas por los ciberdelincuentes para propósitos no legítimos siendo un ejemplo el uso de la síntesis de voz o de imágenes para la creación de deepfakes, que buscan el engaño y que son, cada vez, más difíciles de detectar.

# 6 IMPACTO DE LA IA

## 6.1. IMPACTO EN LAS ORGANIZACIONES

### 6.1.1. EL VALOR DE LOS DATOS PARA LA IA, GOBIERNO DE DATOS Y SILOS DE DATOS

Los acontecimientos de los últimos años, sus consecuencias y los cambios sociales y económicos producidos (Pandemia, guerra Rusia-Ucrania, teletrabajo, Web 3.0, transformación digital de las empresas y de la sociedad, problemas en la cadena de suministro e inflación) han permitido que las organizaciones impulsadas por los datos (Data-Driven) estén en una posición de ventaja para sobrevivir en un entorno incierto y complejo.

Estos datos son fundamentales para que la IA ejerza un rol importante para predecir datos (pasados, presentes o futuros) que nos permitan tomar buenas decisiones:

3 factores pueden considerarse claves:

- La existencia de una cultura data-driven.
- La capacidad de conectar fuentes de datos a través de una variedad de activos, sistemas, nubes y dispositivos y en tiempo real.
- La democratización del dato: información accesible para todos los departamentos en todas las organizaciones (open data).

Esta gestión masiva de datos requiere de un Gobierno del dato maduro que permita identificar y gestionar: datos, roles, procesos, flujo de datos, métricas y herramientas. La ausencia de este gobierno puede dar lugar a silos de datos.

Los principales retos que abordar serán, por tanto:

- El control del acceso a los datos.
- La exactitud, consistencia y fiabilidad del dato.
- El almacenamiento del dato y su integración en diferentes infraestructuras (cloud, on premise, hibridas)
- Los incidentes de seguridad por movimientos no controlados dentro y fuera de la organización.
- La implementación del plan de Gobierno del dato.

## 6.1.2. MEJORA DE PROCESOS Y EFICIENCIA USANDO IA

La IA trae nuevas capacidades para los procesos de Negocio. Desde los bancos que tratan de mejorar la interacción con sus clientes, las compañías de seguros que automatizan la gestión de siniestros, las industrias que rediseñan sus procesos de mantenimiento e ingeniería hasta en las ciencias de la salud con el diagnóstico, tratamiento y la telemedicina.

Pero esto supone que las organizaciones necesiten repensar que tareas son necesarias, con qué frecuencia y quién las tiene que realizar. también se tienen que replantear quien debe de asumir la responsabilidad de la mejora de los procesos, que hasta ahora dependían de los managers de operaciones, o si es necesario incorporar a ingenieros de AI y managers de producto quienes pueden tener una visión end to end de los procesos. Incluso algunas organizaciones adoptan técnicas de "design thinking" para entender mejor como los flujos de trabajo y que actividades necesitan ser rediseñadas para alcanzar las necesidades de los clientes externos e internos.

En resumen, las compañías necesitan explorar donde se generan datos suficientes para extraer patrones que puedan ser usados en la toma de decisiones operacionales.

## 6.1.3. AUTOMATIZACIÓN PARCIAL VS TOTAL DE PROCESOS

La automatización del proceso de predicción depende de 2 factores: la calidad de los datos y la dificultad del juicio.

Si los datos son de alta calidad y el juicio es sencillo la automatización total es lo más adecuado.

Por el contrario, si los datos son de baja calidad o el juicio es complicado la automatización no será posible o supondrá un mayor riesgo.

La discusión entre manual o automático no debería reducirse a una decisión binaria de si o no. Se trata más de encontrar un balance entre los aspectos donde podemos encontrar el beneficio de la automatización frente a las actividades donde se debería mantener el sentido de la participación humana.

Una opción es la inclusión selectiva de la participación humana aprovechando la eficiencia de la automatización de la inteligencia manteniendo la retroalimentación humana y el sentido del propósito de la actividad.

De esta idea surge la metodología de "Human in the loop" usado en diferentes contextos y para referirse a diferentes personas pero que en su sentido amplio se refiere a los especialistas humanos que reducen el número de errores supervisando el proceso de las máquinas.

## 6.1.4 DISRUPCIÓN DE LA ESTRATEGIA EMPRESARIAL

Los últimos desarrollos en IA y robótica pueden generar una nueva era en la automatización de actividades que permitirán a las organizaciones mejorar o reinventar sus procesos, reducir sus errores, aumentar la velocidad de entrega de valor y en algunos casos superar las propias capacidades humanas.

Según McKinsey 5 son los factores que influirán en el ritmo de adopción:

- **La viabilidad técnica:** La tecnología tiene que ser inventada, integrada y adaptada a soluciones para cada caso de uso.
- **Costo del desarrollo y despliegue de las soluciones.**
- **Dinamismo del mercado de trabajo:** La oferta y la demanda y el coste de la fuerza de trabajo afecta a la forma en las actividades serán automatizadas.
- **Beneficios económicos:** Incluye mayor rendimiento y calidad y también el impacto en el desempleo.
- **Aceptación regulatoria y Social:** Aunque la automatización puede tener un sentido empresarial puede tener efectos negativos en la sociedad.

Las estrategias empresariales y económicas tendrán que ser capaces de equilibrar la automatización, la productividad y el empleo.

## 6.1.5. VISIÓN DE GARTNER SOBRE EL IMPACTO DE LA IA

En relación con las tendencias globales que guían el desarrollo de la innovación en el campo de la inteligencia artificial, la consultora Gartner ha identificado en 2023 que la IA generativa (GenIA) está dominando los debates sobre la IA y ha aumentado la productividad de los desarrolladores y profesionales de manera muy real. Esto ha provocado que las organizaciones y las industrias reconsideren sus procesos comerciales y el valor de los recursos humanos, llevando a la GenAI al pico de expectativas infladas en el ciclo de exageración.

En esta línea Gartner identifica:

- Innovaciones que serán impulsadas por la GenIA:
  - » Computer visión.
  - » Edge AI.
  - » Intelligent applications.
  - » MLOps.
  - » Prompt engineering.
  - » Synthetic data.
- Innovaciones que impulsarán a la GenAI:
  - » Causal AI.
  - » Data labeling an annotation.
  - » Responsible AI.
  - » Knowledge graphs.

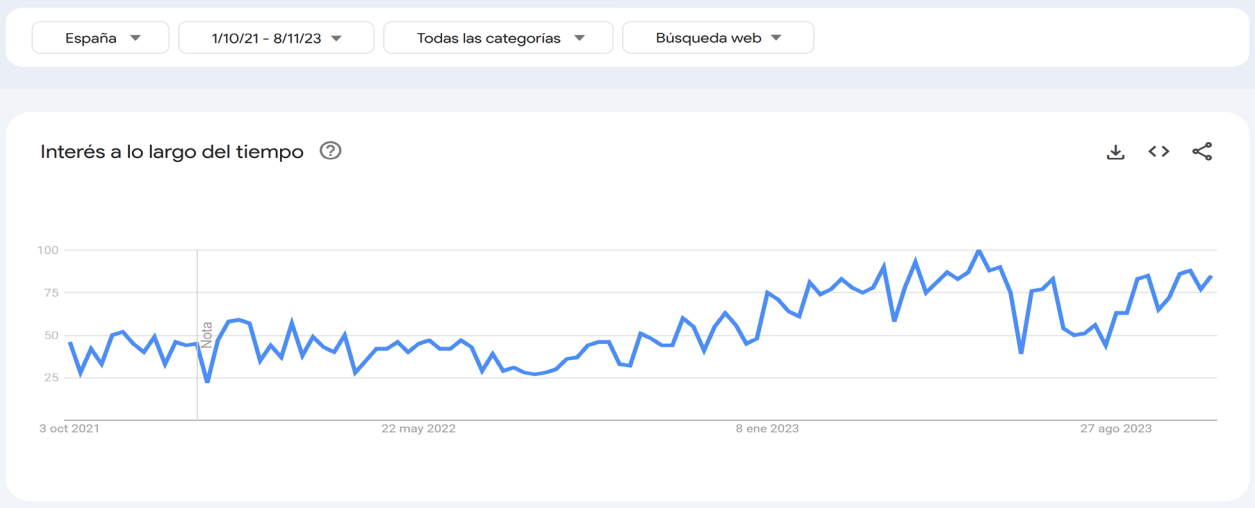
Más allá de la IA generativa, otras técnicas prometedoras como la simulación con IA, la IA causal (comprender las relaciones de causa y efecto) y el aprendizaje automático federado, están preparadas para perfeccionar las interacciones digitales con los clientes y conducir a decisiones comerciales mejor informadas.

Respecto a la seguridad, las tecnologías emergentes como AI TRISM tendrá una gran importancia dado que los humanos son el centro de la mayoría de las violaciones de seguridad, diseñar una seguridad y privacidad centradas en el ser humano tendrá una gran importancia. En esa línea se incluirá la gestión de confianza, riesgos y seguridad de la IA (TRiSM), IA responsable, ética digital y kits de enseñanza y creación de IA.

# 6.2. IMPACTO EN LA SOCIEDAD

Las bases de la inteligencia artificial se empezaron a forjar a mediados del siglo XX permeando en la sociedad de forma irregular hasta nuestros días. Un punto de inflexión, tecnológico y social, ha sido la introducción de las redes generativas, que representaron un hito en esta área ocasionando que la inteligencia artificial haya tenido más impacto en la sociedad hasta el punto de que la Real Academia Española la nombró "palabra del año 2022".

Si analizamos el interés mundial por la búsqueda del término "Artificial intelligence" podemos ver que el termino no suscitaba casi interés hasta el anuncio de GPT3 por parte de OpenAI, o incluso el lanzamiento de GPT4 que nos ofrece unas respuestas mucho más completas, con más detalles y riqueza, y con datos más actualizados. Estos nuevos modelos fundacionales marcaron una mejora cualitativa y cuantitativa en la respuesta de las aplicaciones




como traductor, asistentes, tratamiento de imágenes, etc. que permitió el auge en el uso de las aplicaciones basadas en IA y un mayor impacto en la sociedad.

En el sector servicios o administración la inteligencia artificial se están adoptando de forma activa; sanidad, educación, energía, gestión de residuos, transporte público, protección civil, etc. En el ámbito de la salud, por ejemplo, la IA tiene un impacto enorme en los diagnósticos médicos. La Sociedad Americana para la Mejora de la Diagnosis en Medicina cifró en 12 millones los pacientes americanos que sufrían errores diagnósticos cada año, con unos costes para el sistema sanitario estimados en 100 mil millones de dólares. Las decisiones diagnósticas basadas en inteligencia artificial (AI-DDS, por su nomenclatura inglesa) realizan diagnósticos más precisos y en etapas precoces de las enfermedades, mejorando las expectativas de curación. Sistemas basados en el análisis de imágenes (resonancias magnéticas, radiografías, etc.) permiten diagnosticar con mayor precisión y rapidez enfermedades.

Hoy en día la proporción de expertos en inteligencia artificial en las compañías es muy reducida, sin embargo, la adopción de la inteligencia artificial supone, de acuerdo con un estudio que publicó IndesIA a principios de 2022, que España va a generar una demanda de 90.000 profesionales en inteligencia artificial en los tres años siguientes.

# 7 GOBIERNO DE LA IA

El gobierno de la IA debe conseguir minimizar los impactos negativos de los sistemas basados en IA en la regulación, las leyes y en última instancia los derechos fundamentales de los individuos a la vez que es capaz de conseguir impactos positivos en la sociedad y en las organizaciones. Afrontar, documentar, informar y gestionar los riesgos de la IA permitirá una mayor confianza en los sistemas de IA.

El Gobierno de IT permite conseguir que los objetivos de IT estén alineados con los objetivos de la organización. Para ello se debe hacer foco en:

**LA ENTREGA DE VALOR:** Asegurando que IT genera los beneficios esperados por la estrategia de la organización optimizando los recursos IT.

**GESTIÓN DE RIESGOS: Teniendo en cuenta el apetito al riesgo de la organización, los requerimientos legales y regulatorios ofreciendo transparencia sobre los riesgos y las responsabilidades.**

**MEDIDA DEL RENDIMIENTO:** Monitorizando la estrategia, la implementación de proyectos, así como la utilización de recursos y la entrega de valor en forma de soluciones y servicios IT.

**ALINEAMIENTO ESTRATÉGICO:** Orientando la estrategia de IT para a impulsar la estrategia de la organización y al mantenimiento de la operación.

**GESTIÓN DE RECURSOS:** Optimizando la inversión y el uso de todos los recursos IT: sistemas, datos, infraestructuras y personas.

El gobierno de IA como parte del gobierno de IT tendría la función de gestionar y monitorizar las actividades y procesos de la IA en la organización para cumplir los 5 objetivos anteriores orientándose al ciclo de vida de la IA mediante el seguimiento y cobertura de las siguientes actividades:

- Integrar en el proceso de gestión de riesgo de las organizaciones durante el diseño, adquisición, desarrollo, despliegue, evaluación y explotación de los sistemas IA. Esto supone la identificación, análisis, evaluación y aplicación de planes de remediación ante los riesgos derivados del uso de la IA.
- Incorporar las actividades de IA dentro de los principios, políticas y estándares de la organización.
- Afrontar todo el ciclo de vida de la IA incluyendo los requisitos legales, regulatorios y de cumplimiento que la organización establezca, así como la relación con servicios de terceros.
- Incluir procesos de Gobierno de la IA integrados dentro del proceso general de Gobierno de la organización.
- Implementar marcos de trabajo estándar de la gestión de IA que facilite la supervisión interna y externa y que incorpore mejores prácticas del mercado.
- Formalizar las estructuras adecuadas para el gobierno y gestión de la IA dotándolas de la responsabilidad y rendición de cuentas apropiada, acordando los roles y perfiles de los equipos, consiguiendo un mapeo, medición y gestión adecuada de los riesgos.
- Categorizar los sistemas de IA documentando sus funciones, objetivos de uso y beneficios esperados frente al coste estimado, así como sus posibles riesgos. Quien, Que, Como y Cuando puede ser afectado.
- Medir: Implementando procedimientos de evaluación y de seguimiento de los procesos IA e incluyendo beneficios, costes, riesgos y recursos utilizados. Para ello habrá que establecer diferentes tipos de métricas desde las más operativas (KPIs) a las más estratégicas (OKRs) pasando por las relacionadas con los riesgos (KRIs).
- Reportar de forma clara a los diferentes interesados: Equipo de Dirección, Regulador y Áreas de Negocios.
- Actualizar los riesgos revisando con frecuencia su impacto y adaptando los controles.

Los principales riesgos vinculados con la IA pueden ser agrupados en 3 áreas:

I. Perjuicio en las personas. Afectando a los derechos fundamentales, la seguridad físico o lógica y en las oportunidades económicas. Discriminando a grupos de población. Perjudicando la participación política y el acceso a la educación.

II. Perjuicio en las organizaciones. Afectando a los negocios de las organizaciones, causando pérdidas económicas o fugas de datos o afectando a su reputación.

III. Perjuicio en los ecosistemas. Impacto ente sistemas y recursos interconectados. Impacto en los sistemas globales financieros y cadena de suministro, así como en los recursos naturales, el medio ambiente y el planeta.

Un gobierno de la IA efectivo evaluará las oportunidades y los riesgos buscando un equilibrio entre impacto y beneficio teniendo en cuenta las posibles consideraciones económicas, éticas y legales. Además, deberá ser operativo, eficiente, auditable y trazable.

# 8 ADOPCIÓN Y GOBIERNO DE LA IA

En 2023 ISMS ha lanzado la I Encuesta de Adopción y Gobierno de la IA en España, en este anexo se muestran los resultados estructurados en 4 bloques distintos.

| SEGMENTACIÓN | ADOPCIÓN DE LA IA GENERATIVA | GOBIERNO DE LA IA | RIESGO DE LA IA |
|---|---|---|---|

Desde que se inició el proyecto de realización de la encuesta hasta el momento de redacción de este informe el equipo de proyecto, y creemos que toda la sociedad, se ha visto sorprendido por el elevado volumen de novedades que, prácticamente cada semana, surgen al respecto de la IA. Esto nos da a entender que, una nueva edición de la encuesta muy probablemente cambiaría el peso de algunas de las respuestas obtenidas.

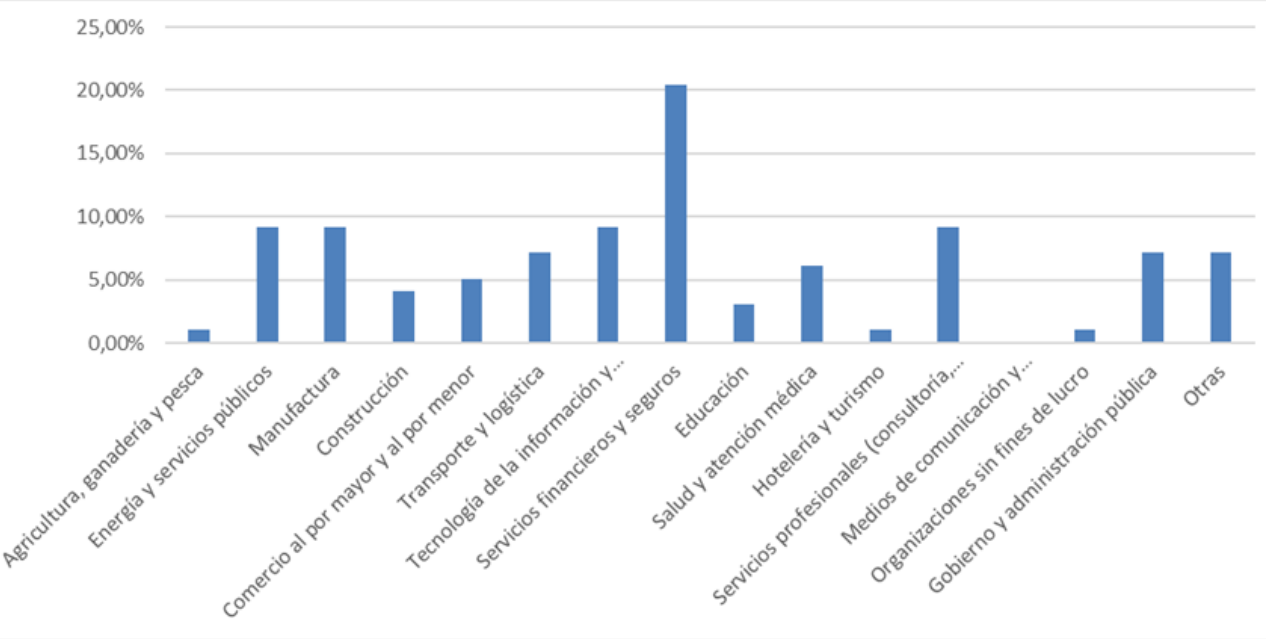
De forma resumida las principales conclusiones de la encuesta son:

- Existe una falta generalizada de gobierno de la IA, no sólo la generativa, las organizaciones apenas tienen roles dedicados.
- La mayoría de las organizaciones están implementando o considerando la implementación de sistemas de IA Generativa.
- Existe una importante mayoría que opta por la implementación o uso de algoritmos de terceros.
- La gran mayoría percibe que se va a incrementar el uso de la IA Generativa en sus organizaciones, más del 10% percibe que va a haber un incremento significativo o incluso un cambio radical por su adopción.
- La gran mayoría opina que hay riesgo en el uso de IA en las organizaciones y que, por tanto, deben desarrollarse marcos de control de esta. El riesgo viene tanto por posibles incidentes derivados de sesgo, regulatorios, o directamente por afectación a las actividades básicas de la organización.

# 8.1. SEGMENTACIÓN

**PREGUNTA 1: Indique el sector de actividad de su organización**

Los profesionales que han contestado se reparten en los siguientes sectores de actividad:

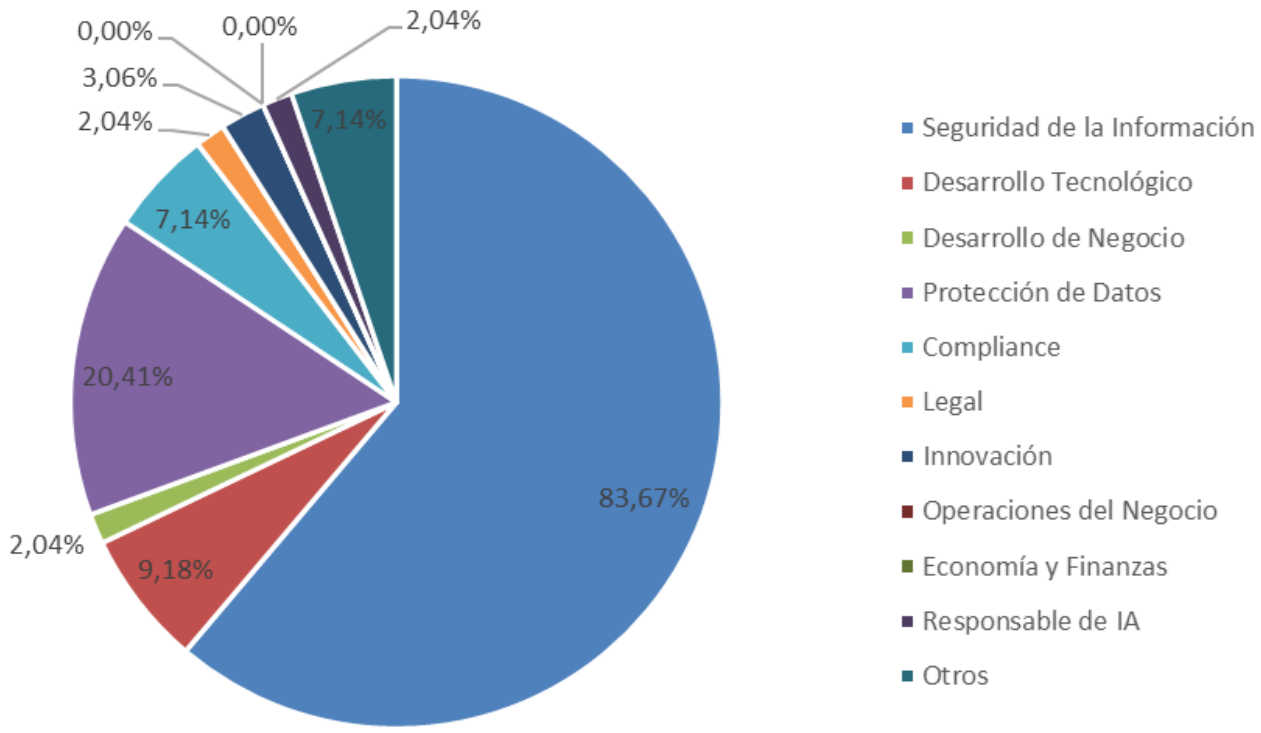



Es muy relevante observar la elevada respuesta del sector de Servicios Financieros y Seguros, puede ser explicable al tratarse de uno de los sectores donde la IA Generativa está generando mayores expectativas de adopción.

La siguiente, y última pregunta de segmentación, era relativa al rol en la organización, se trata de una pregunta de respuesta múltiple y se han observado varios datos previsibles:

- La mayoría de las respuestas tienen rol de Seguridad de la Información, El 83,6%.
- Un 20,41% tienen rol de Protección de Datos, bastantes de ellos de forma complementaria al rol anterior.
- Surge un porcentaje discreto del 2,04% que se declaran responsables de IA en sus organizaciones.

**PREGUNTA 2: Seleccione cuál es su rol en la organización (Selección Múltiple)**

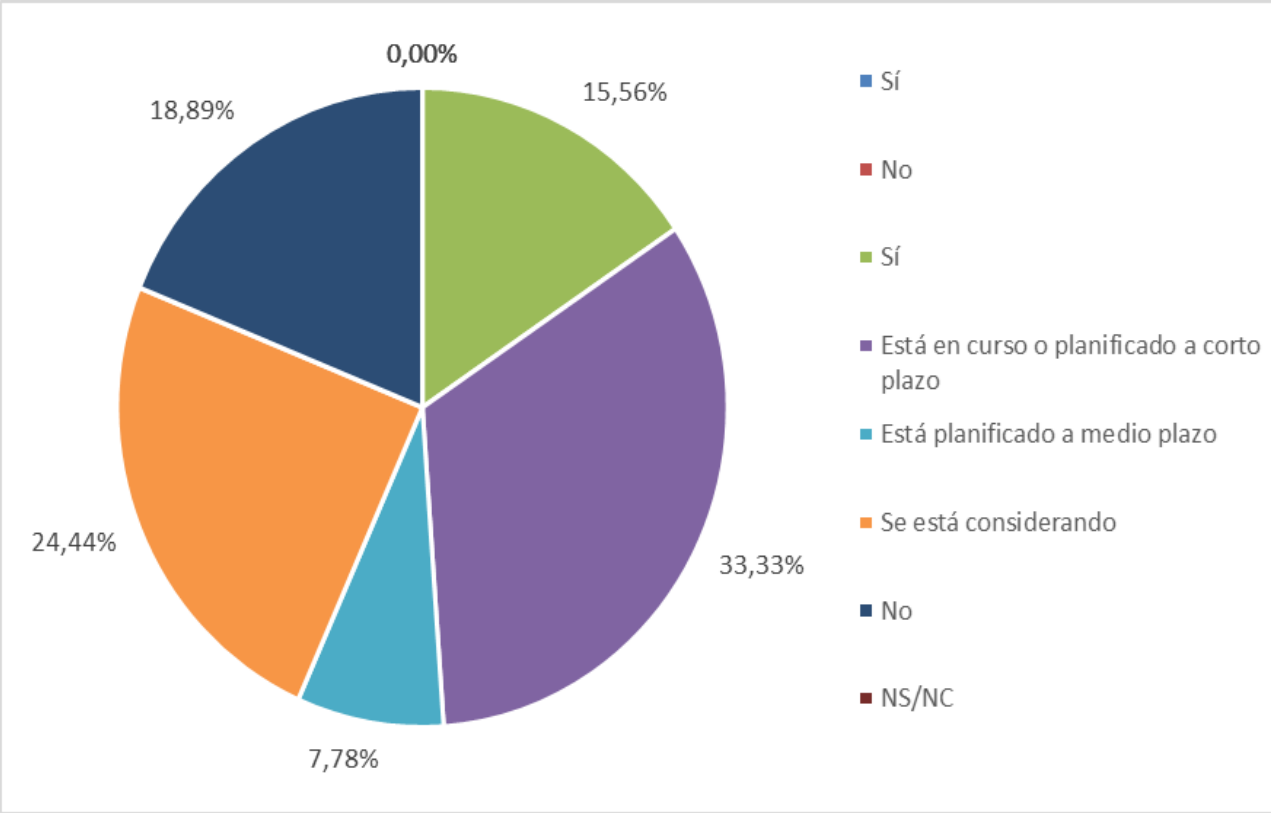



## 8.2. GRADO DE ADOPCIÓN DE LA IA

**PREGUNTA 3: ¿Su organización ha implementado sistemas basados en Inteligencia Artificial Generativa?"**

De las respuestas se deducen varias conclusiones:

- El 48,89% los ha implementado, lo está haciendo o lo tiene en el objetivo a corto plazo.
- Un 32,22% lo está considerando, a medio o largo plazo.
- Un 18,89% no lo considera en estos momentos.

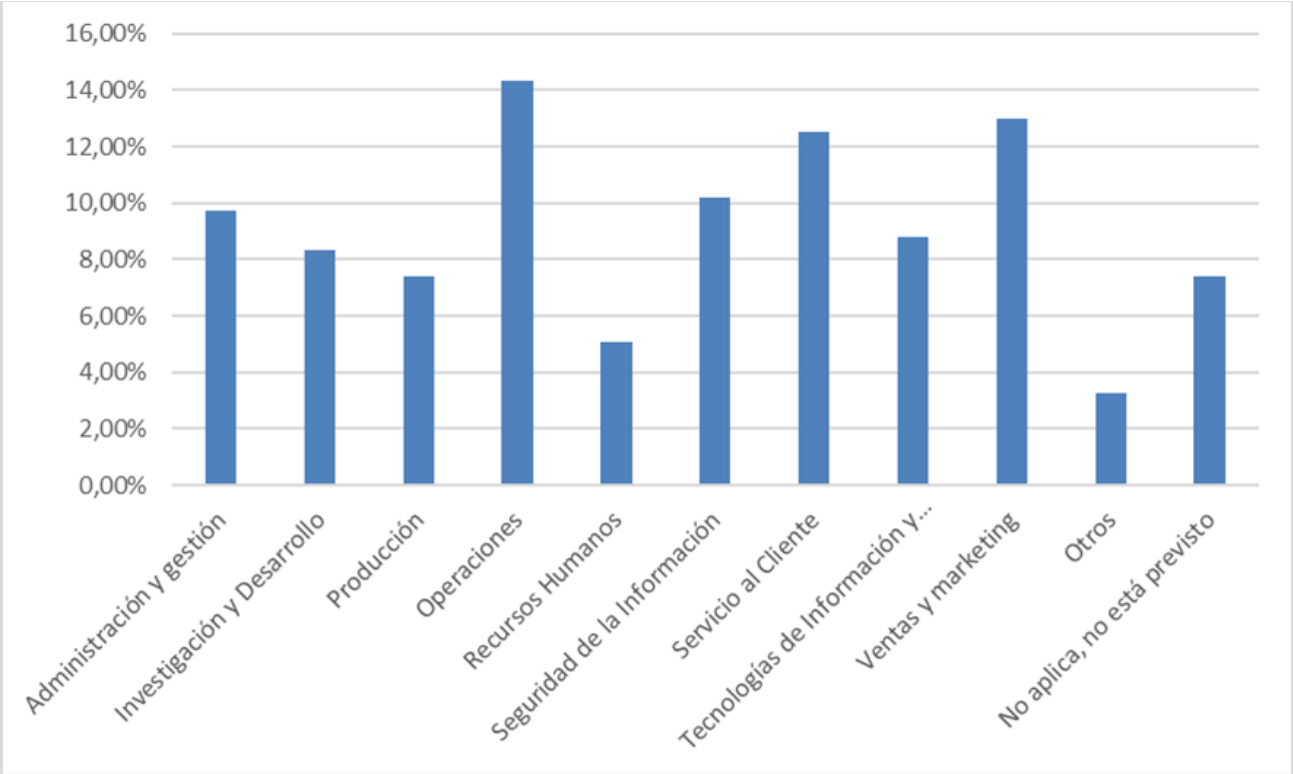



Sobre el 18,89% de organizaciones que no consideran la adopción de la IA Generativa se ha creado cierto debate entre el equipo que ha elaborado este informe, pues la mayoría entiende pasará algo similar a la adopción de IA basada en el comportamiento, tecnología sobre la que prácticamente todas las organizaciones tienen alguna implementación a través de la cadena de suministro, en el ámbito de la Seguridad de la Información un ejemplo muy claro son los sistemas EDR (End Point Detection & Response), que han abandonado el sistema tradicional de detección de malware basado en IoCs (Indicadores de Compromiso) y enfocan su tarea a la detección de patrones de comportamientos anómalos, tarea que requiere cierto nivel de machine learning.

Con IA Generativa se observa que determinados servicios de atención al cliente prestados por proveedores implementan tecnologías de IA Generativa, también para servicios legales de redacción de contratos, para servicios de seguros de generación de pólizas…

**PREGUNTA 4: ¿Para qué áreas de actividad? (selección múltiple)**

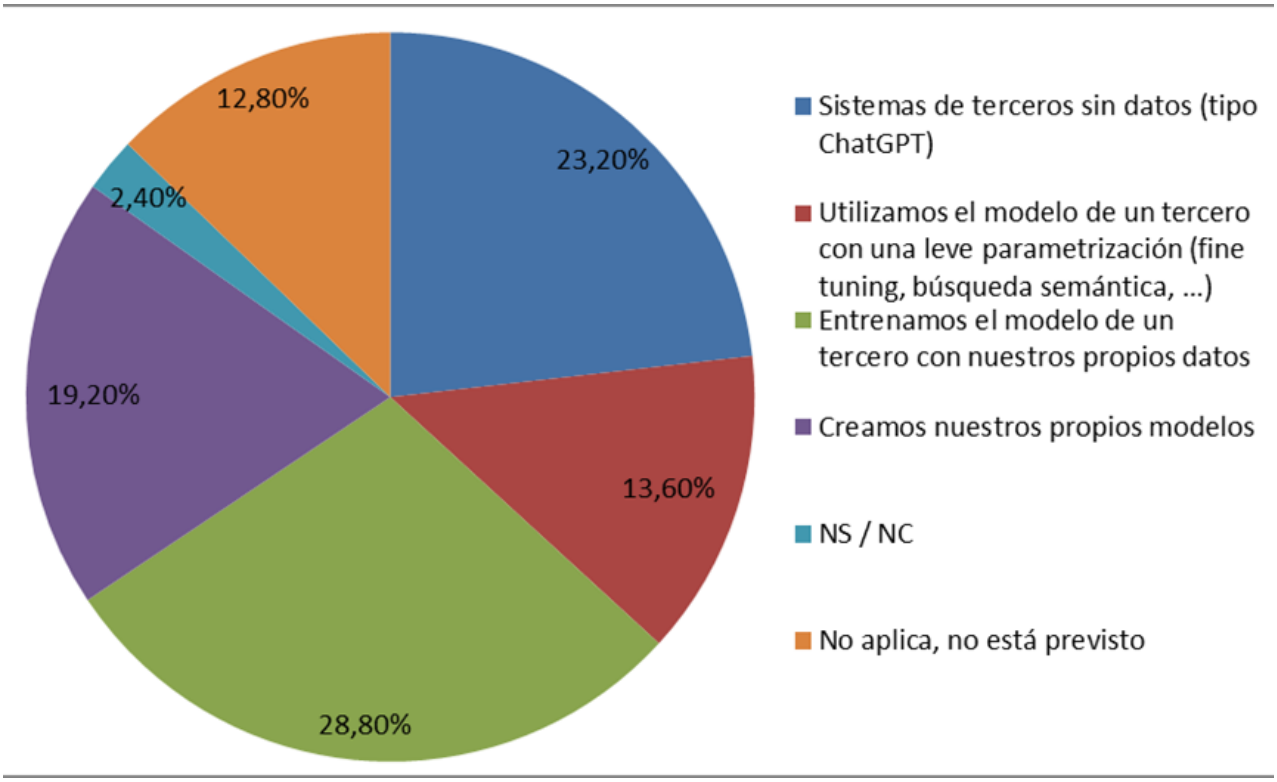
De las respuestas obtenidas se observa un abanico amplio de implementaciones realizadas o en curso.




El ámbito de operaciones es el que destaca con un mayor grado de interés seguido del Servicio al Cliente y Ventas y Marketing. Si bien los dos grupos anteriores eran previsibles sorprende el elevado foco en el sector de operaciones de negocio. Una interpretación de esta respuesta puede estar en la apuesta de relevantes proveedores de servicios de ofimática colaborativa e IA de proporcionar asistentes capaces de indexar los contenidos corporativos y ofrecer respuestas en base a preguntas de lenguaje natural.

**PREGUNTA 5: ¿Qué tipo de Sistemas?**

Como era previsible la mayoría de las respuestas (un 65,5%), apuntan a la utilización de sistemas de terceros, bien sin alimentación de datos propios, con una leve parametrización de estos y, de este grupo, la mayoría apunta al entrenamiento de estos sistemas con datos propios de la organización (el 28,8%).




A su vez el 12,8% de respuestas indican que "no aplica", razonablemente coherente con las respuestas de que "no está prevista esta implementación" a la pregunta 3.

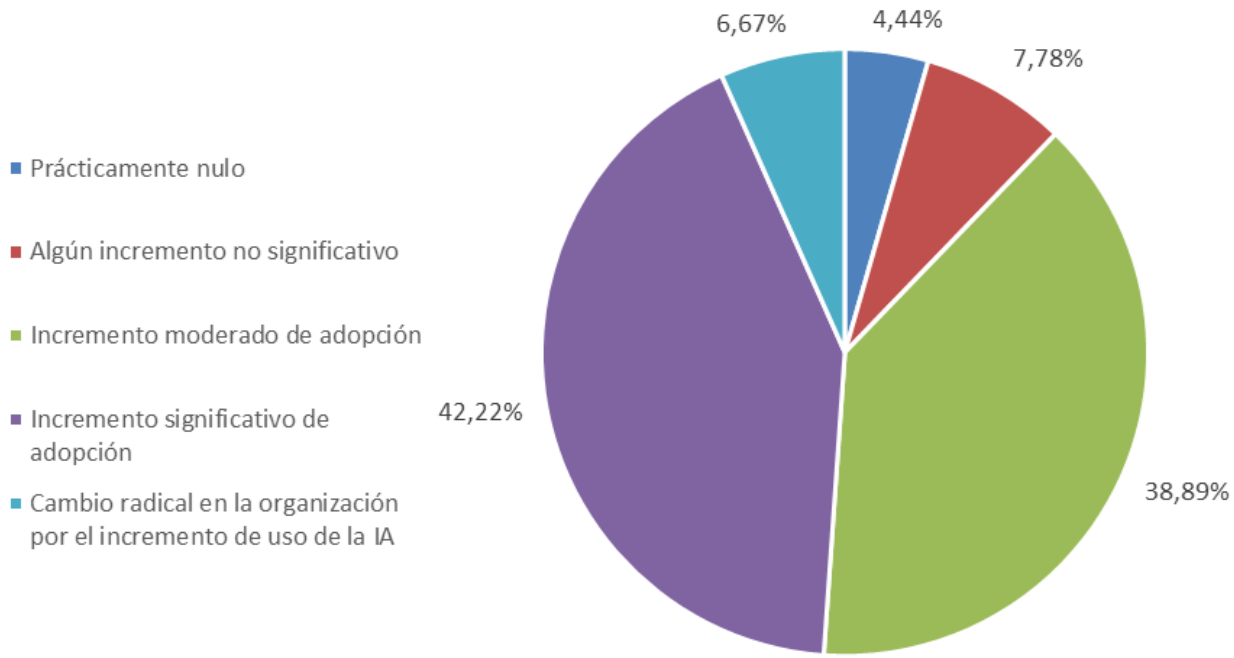
Queremos hacer foco en el elevado volumen de respuestas "Creamos nuestros propios modelos", con un 19,2% de respuestas. Es un porcentaje más alto del esperado inicialmente por la elevada dificultad que plantea desarrollar estos sistemas, siendo una tendencia observada la contratación a terceros de estos. Es un punto que observar a futuro.

**PREGUNTA 6: ¿Qué previsión de incremento de uso de la IA Generativa cree que va a pasar en su organización en los próximos 3 años?**

De las respuestas obtenidas se pueden obtener ciertas conclusiones:

» Los extremos son muy parecidos: un 4,44% considera que el incremento será prácticamente nulo frente al 6,67% que plantea un cambio radical en la organización.

» El gran grueso detecta incremento (88,89%), con un menor peso en "Algún incremento no significativo" (7,78%) frente a "Incremento significativo de adopción" (42,22%) en el otro lado.



## 8.3. GOBIERNO DE LA IA

**PREGUNTA 7: ¿En su organización hay algún rol o equipo responsable de definir la estrategia a seguir con la adopción y despliegue de la IA?**

Observando las respuestas, y complementándolo con las respuestas de la pregunta 2 sobre el rol de los profesionales que contestaron, de los que cabe destacar los siguientes:

» La mayoría de las ubican el rol en "Seguridad de la Información" (83,67%).

» "Protección de Datos" (20,41%).

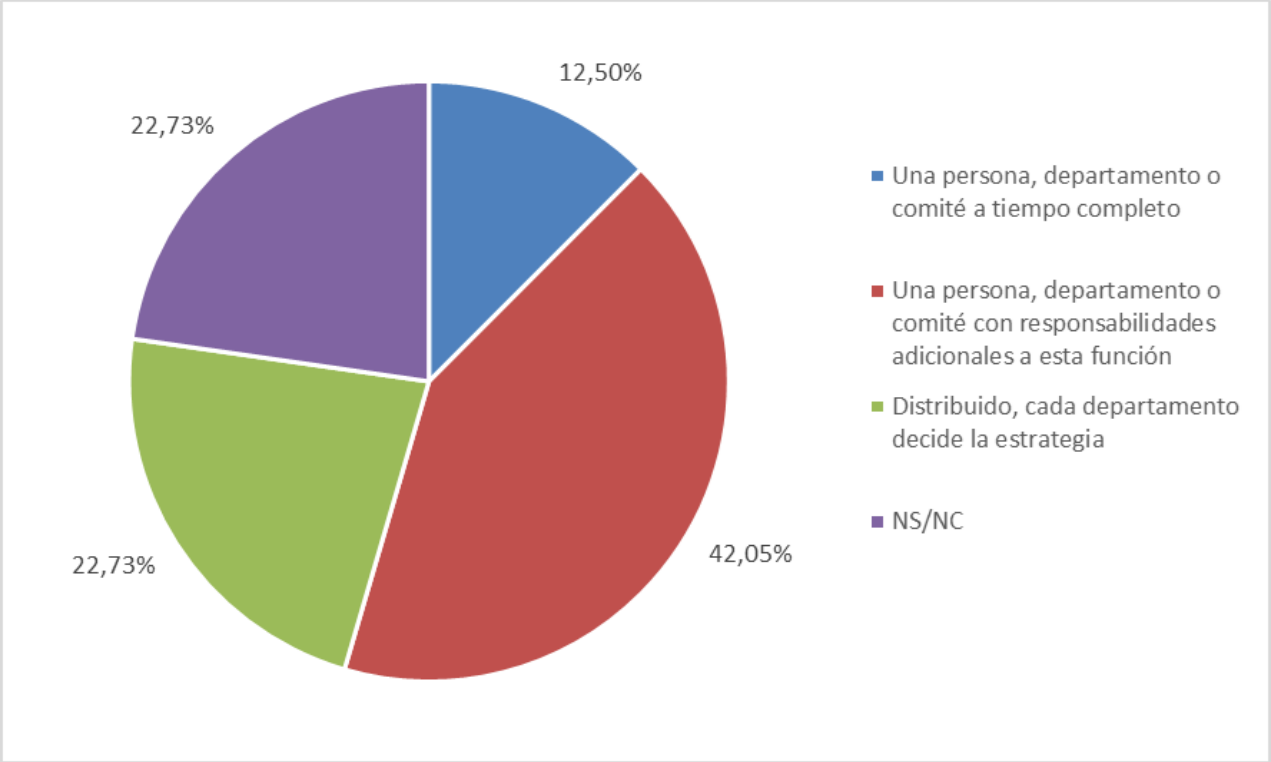
» "Responsable de IA" (2,04%).

Podemos ver dos respuestas que refuerzan las conclusiones de los estudios complementarios de este grupo de trabajo que apuntan a la necesidad urgente de definir roles que definan la estrategia de adopción y despliegue de la IA:

- Sorprende ver que un 22,73% responde "NS/NC", interpretamos que se desconoce.
- Por otro lado, y también con un porcentaje similar de respuestas, se apuntan a que cada departamento decide la estrategia con la IA.

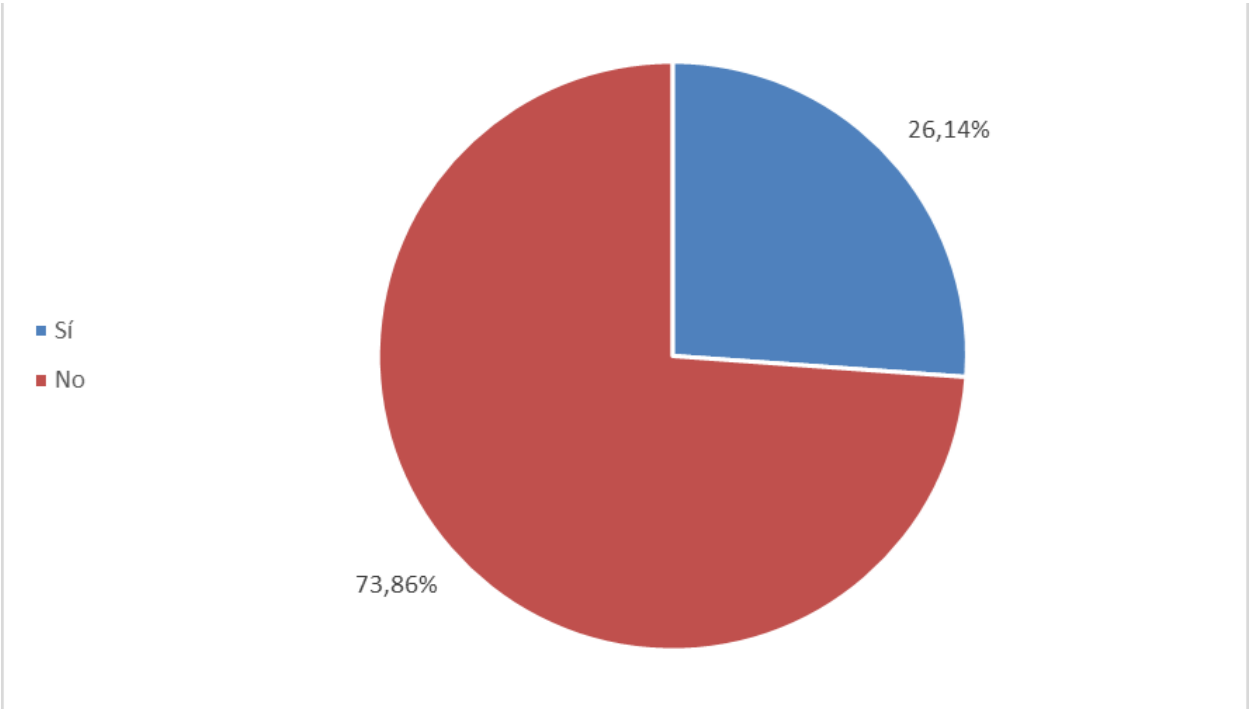
La primera conclusión relevante por tanto es que, tal y como hace ya años surgió la necesidad en las organizaciones de definir una estrategia de Data Management, es preciso que las organizaciones asuman una estrategia de AI Management.

Por otro lado, la mayoría de las respuestas, el 42,05% apunta que sí existe "Una persona, departamento o comité con responsabilidades adicionales a esta función" con la misión descrita, mientras que un esperanzador 12,5% apunta a que ya hay roles dedicados a tiempo completo a esta misión.




**PREGUNTA 8: ¿En su organización se ha definido un modelo de uso y despliegue de la IA (cómo se introduce en la organización, qué ámbitos se priorizan, ...)?**

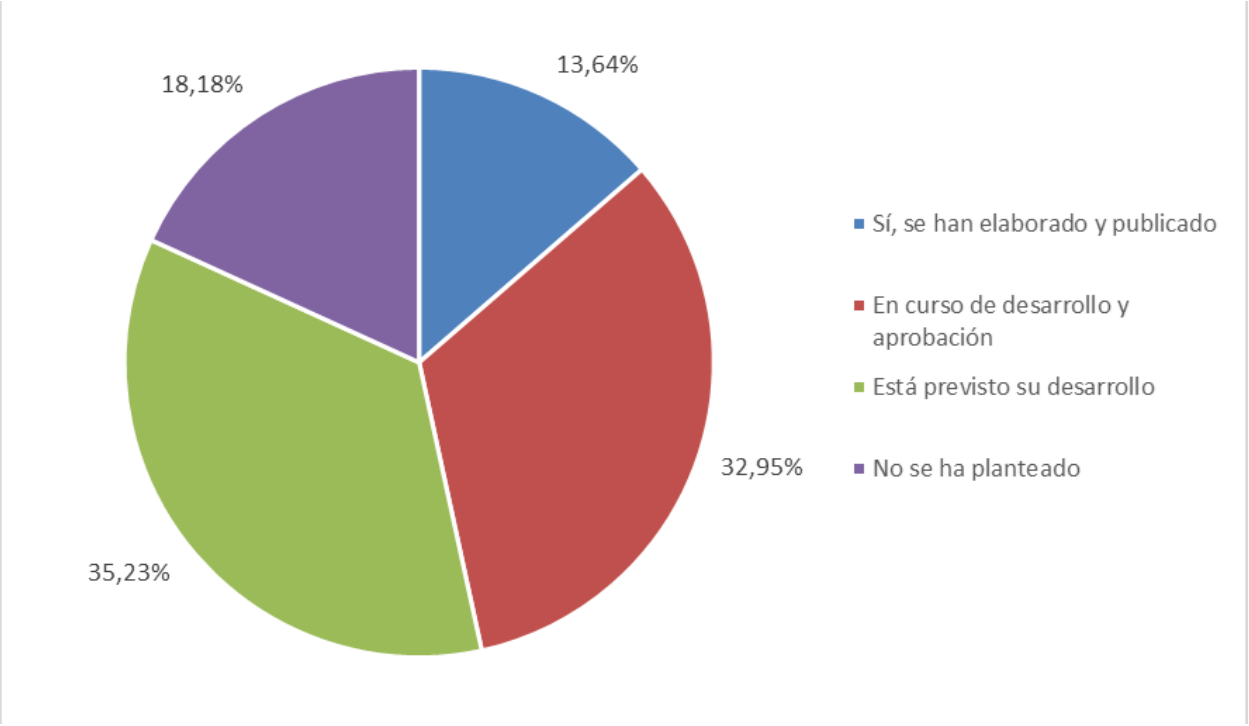
Los resultados son contundentes y apuntan a las mismas conclusiones de la Pregunta 7: es preciso que las organizaciones adopten una estrategia de "IA Management".



**PREGUNTA 9: ¿Se han establecido políticas y procedimientos específicos para garantizar la seguridad de los sistemas IA en su organización?**

Las respuestas obtenidas indican que un 81,82% las tiene, está desarrollándolas o bien lo tiene previsto, frente a un importante 32,95% que no se lo ha planteado.

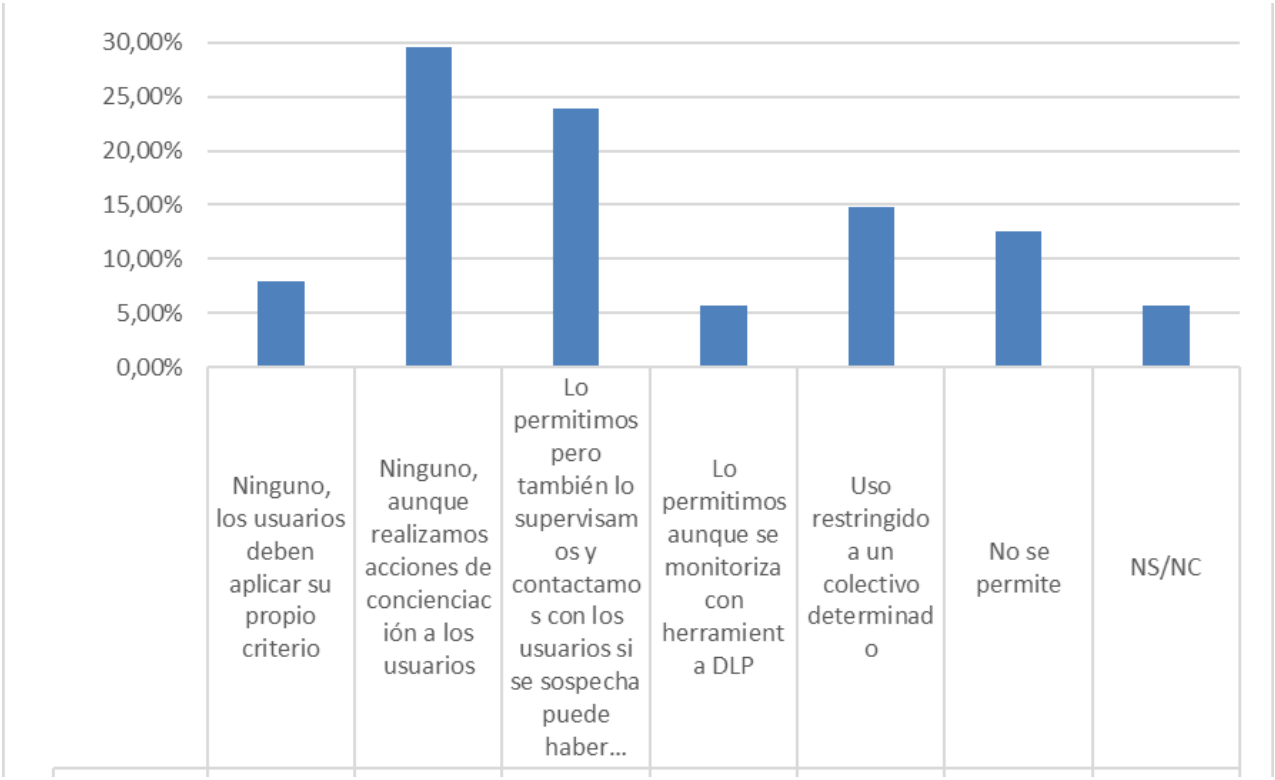
Es probable que un cierto número de respuestas del bloque que contesta en sentido afirmativo apunten a que o bien se refieren a políticas y procedimiento de amplio espectro en las que también tiene cabida la IA, o bien cubran partes muy específicas.




**PREGUNTA 10: ¿Qué grado de supervisión y control aplica sobre el uso de herramientas públicas de IA tipo ChatGPT con equipos corporativos?**

Nuevamente se aprecia cercanía entre los extremos, frente al 7,95% que no responde no tener ningún control sobre estos sistemas, un 12,5% declara bloquearlos de forma explícita, mientras que una amplia mayoría (el 53,41%) apela a la responsabilidad de los usuarios, mediante campañas de concienciación y/o auditoría de accesos realizados.

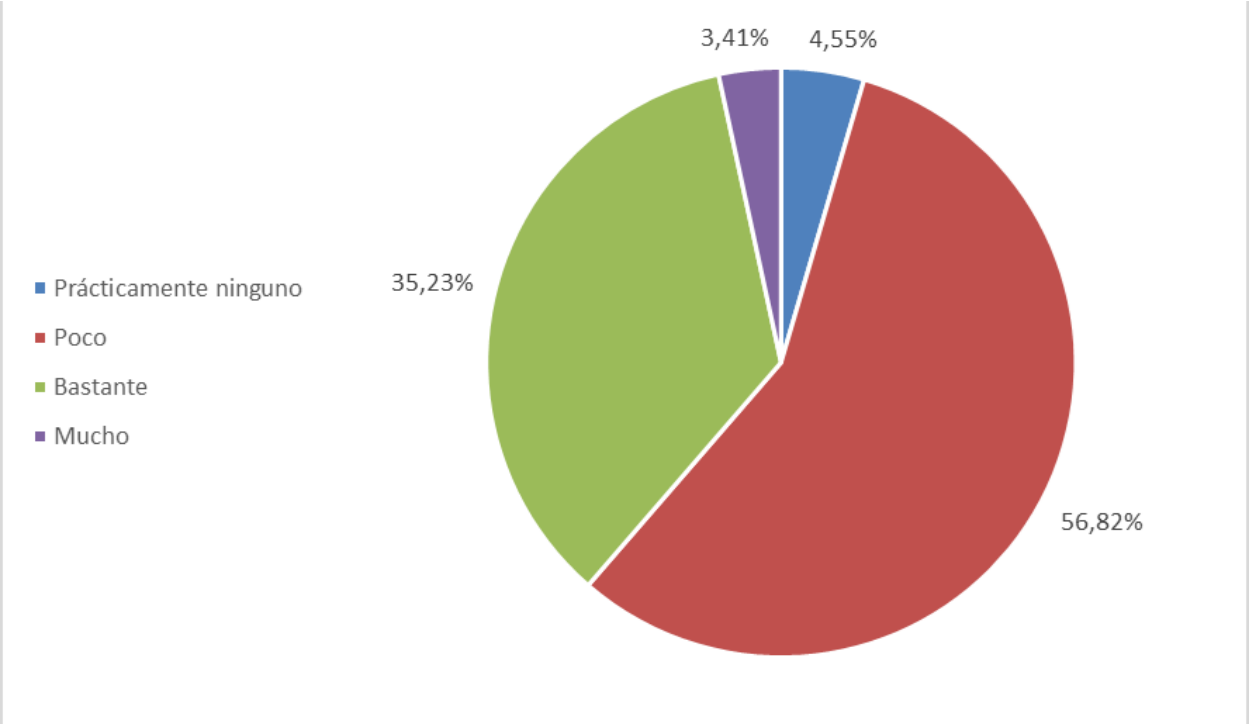
Con relación a las acciones de concienciación en el uso de herramientas de IA, le sugerimos consulte la guía de ISMS *"Disclaimer uso de IA en las organizaciones "* que propone medidas de concienciación al respecto.




## 8.4. RIESGOS ASOCIADOS A LA IA

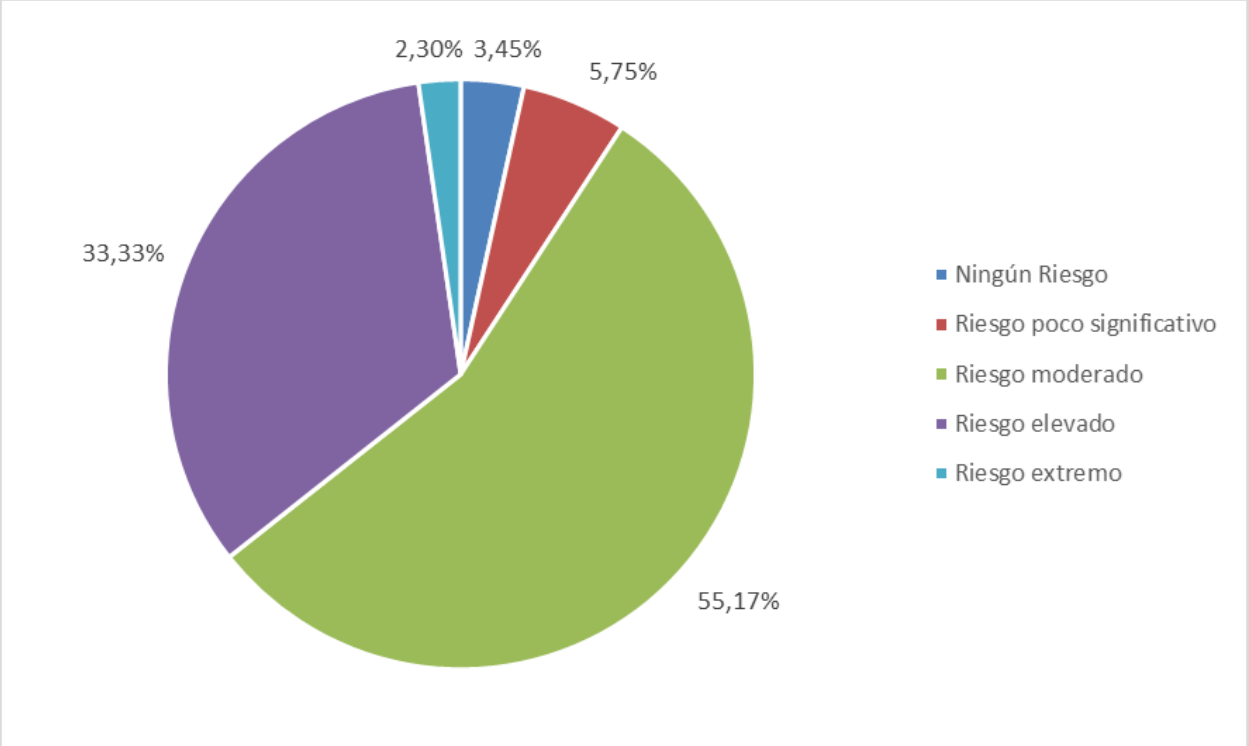
**PREGUNTA 11: ¿Cómo calificaría su nivel de conocimiento de IA Generativa?**

Pese al elevado ritmo de adopción de estas tecnologías más de la mitad de los profesionales que respondieron (el 61%), indicaron que tienen ninguno o poco conocimiento sobre esta tecnología.




**PREGUNTA 12: ¿En qué medida cree que hay potencial riesgo en el uso de IA en su organización?**

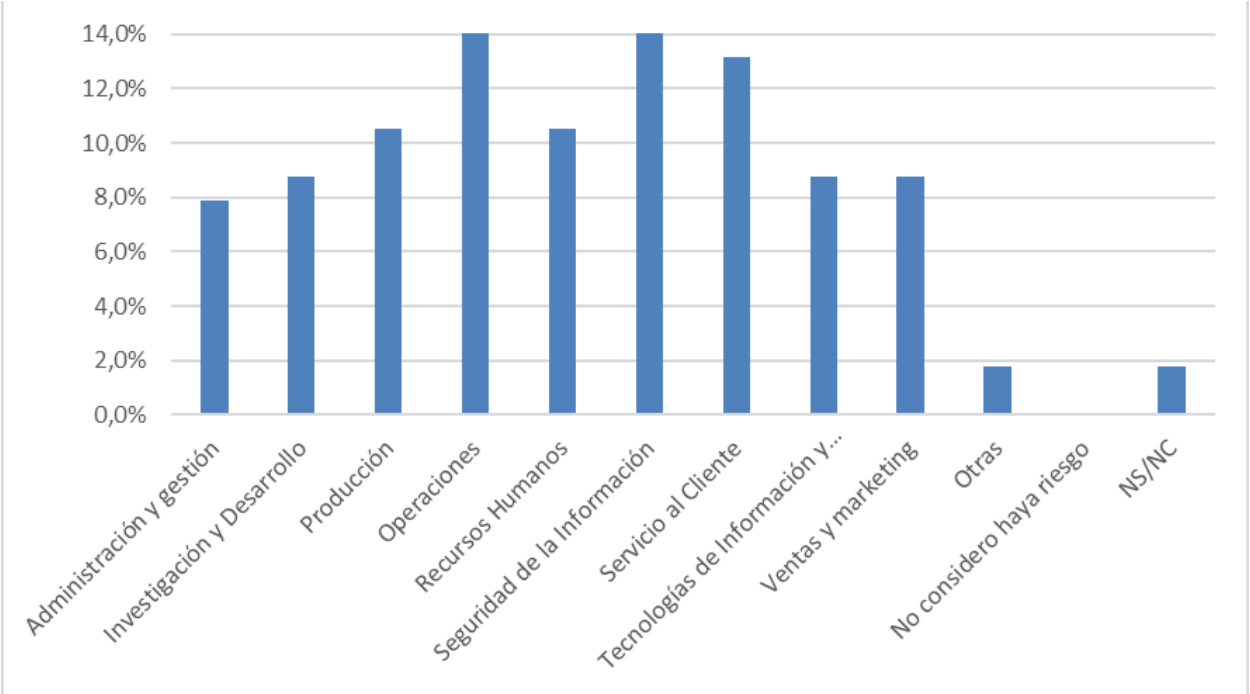
Las respuestas son contundentes, el 90,8% indican que existe riesgo, desde el 55% que indica moderado hasta el 2,3% que apunta un riesgo extremo en el uso de la IA en sus organizaciones.



**PREGUNTA 13: En caso de que así lo crea, ¿en qué áreas cree que habrá mayor riesgo en el uso de la IA?**

Profundizando en la cuestión anterior sobre percepción de riesgo, la pregunta busca identificar áreas del negocio donde se percibe mayor riesgo.

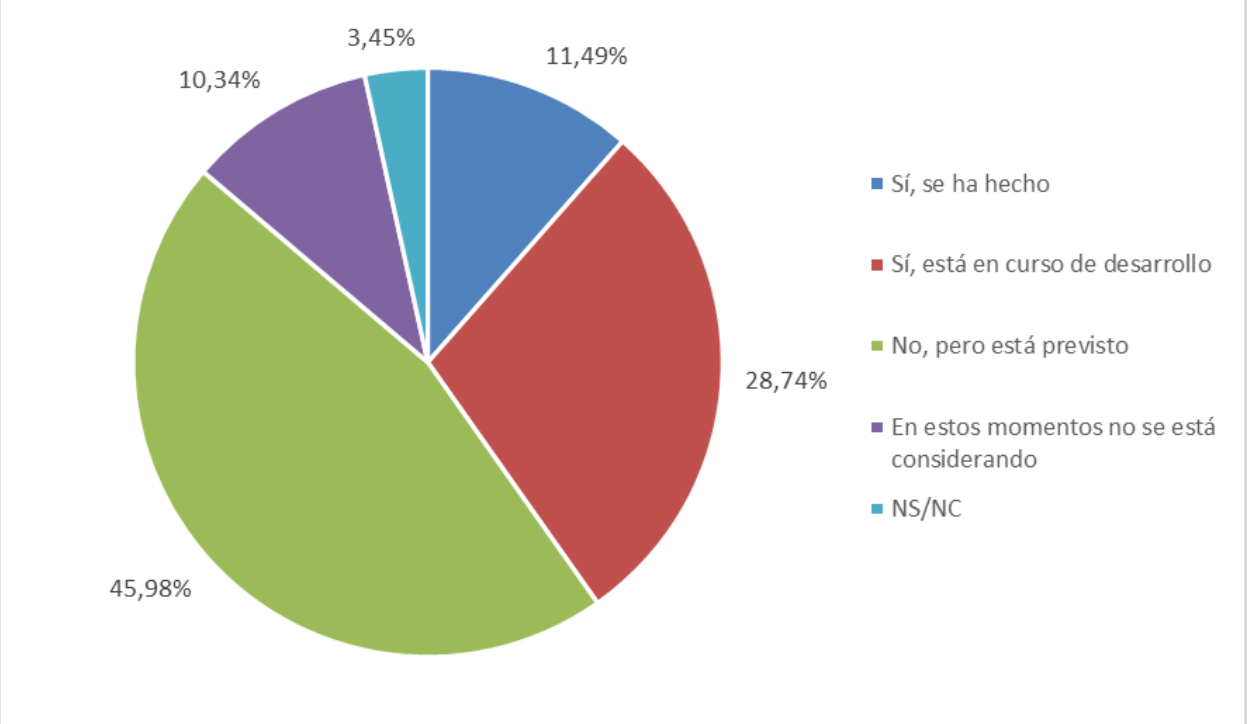
Las principales percepciones de riesgos se encuentran en las áreas de Operaciones y Seguridad de la Información, seguidas muy de cerca por Servicio al Cliente.




**PREGUNTA 14: ¿Se han identificado y evaluado los riesgos de seguridad asociados con la implementación de sistemas de IA?**

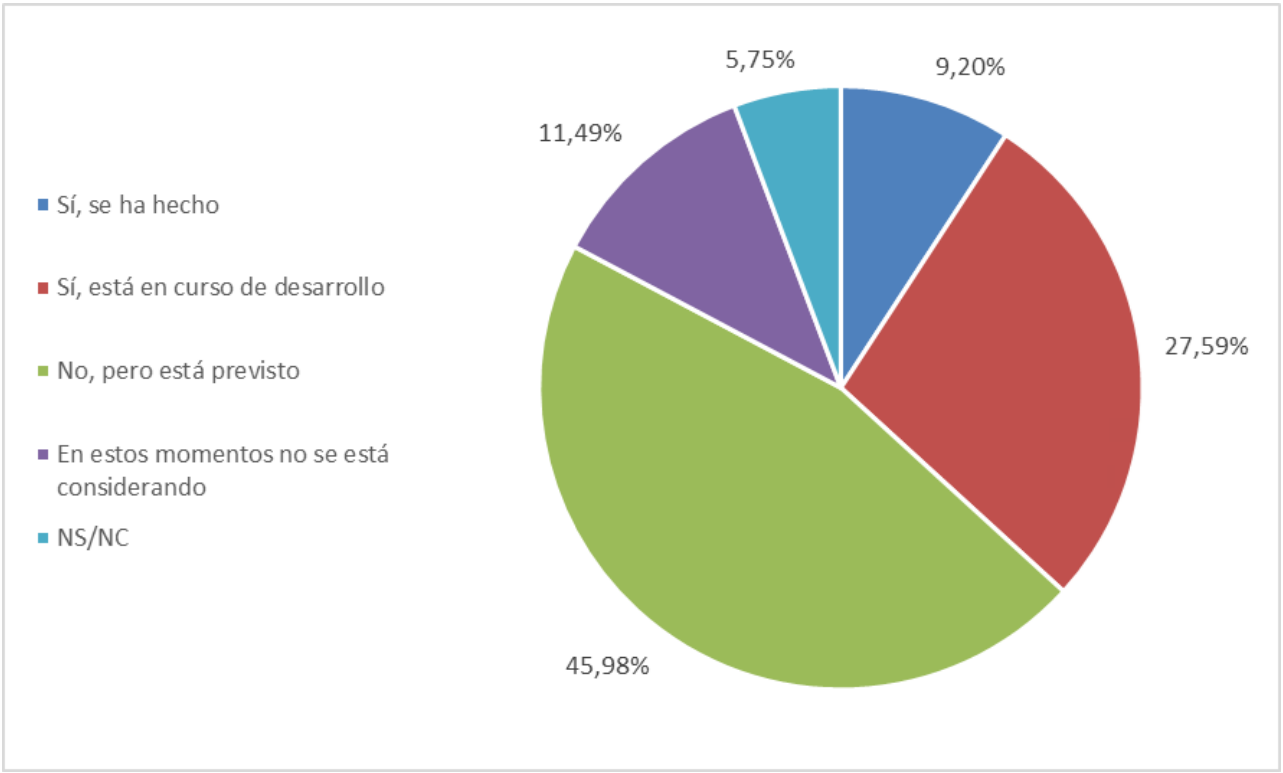
Las respuestas obtenidas son una llamada de atención, el 59,77% no lo ha hecho, únicamente un 11,49% manifiesta haber realizado la evaluación de riesgos de seguridad.

En la parte positiva también cabe destacar que un 45,98% manifiesta no haberlo hecho, aunque lo tiene previsto.



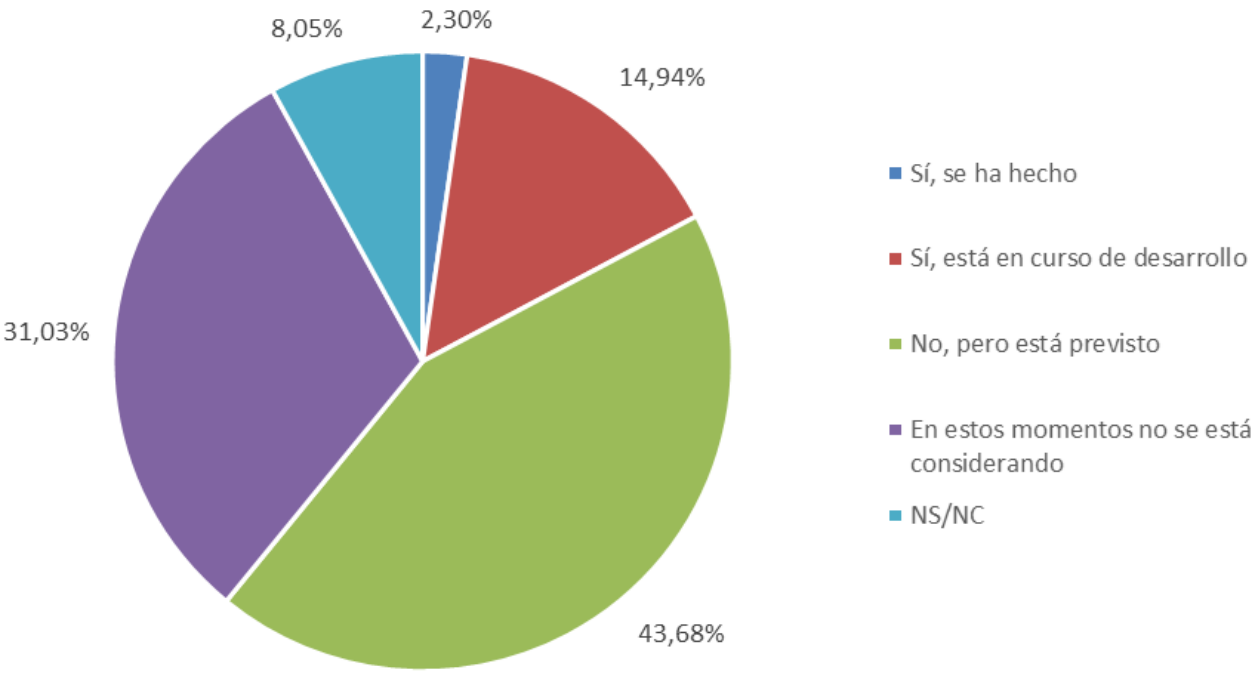

**PREGUNTA 15: ¿Se han realizado evaluaciones de impacto en la privacidad y se han implementado salvaguardias adecuadas para proteger los datos personales en sistemas de IA?**

Como era previsible las respuestas obtenidas son muy parecidas a las obtenidas en la pregunta 14, un 9,2% lo ha hecho y un 27,59% está en ello.



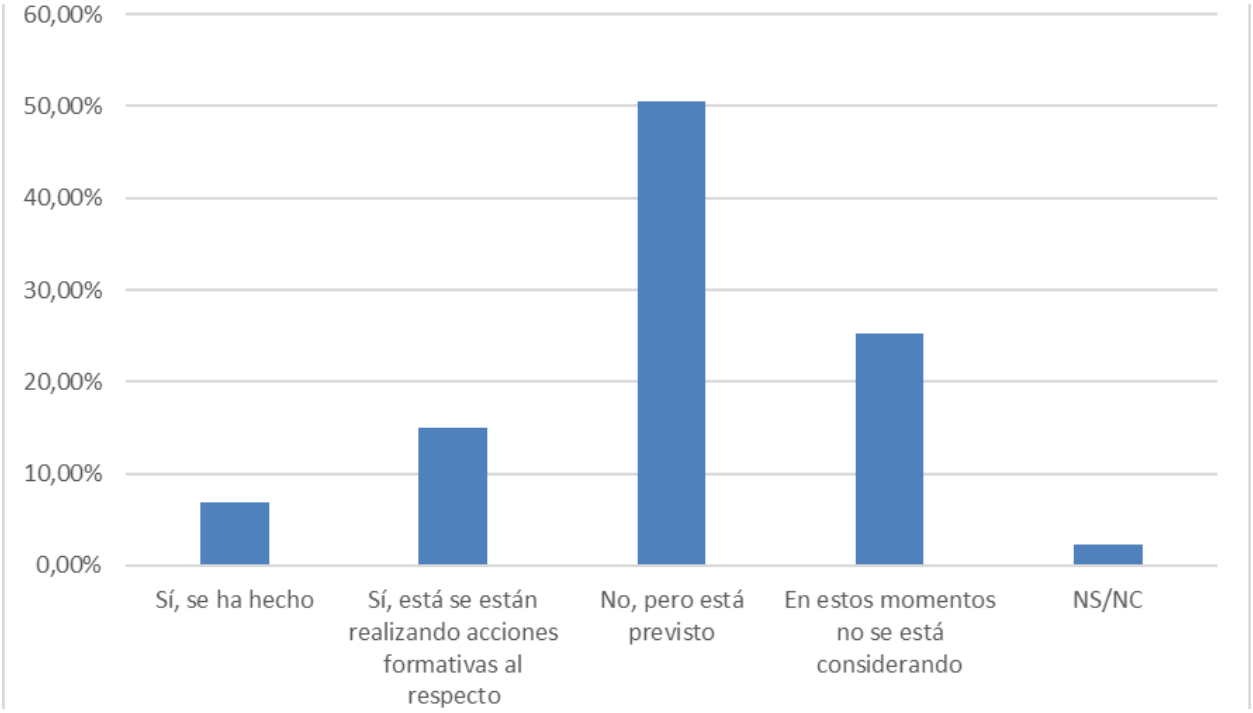
**PREGUNTA 16: ¿Existe un proceso para auditar y revisar regularmente los sistemas de IA implementados para identificar posibles vulnerabilidades o debilidades de seguridad?**

Nuevamente se percibe coherencias con las respuestas anteriores, un 2,3% declara haberlo hecho y un 14,94% declara tenerlo en curso de desarrollo, en el lado opuesto algo más del 31% no lo está considerando.




**PREGUNTA 17: ¿Se ha capacitado al personal en temas de seguridad relacionados con la IA y la detección de posibles amenazas?**

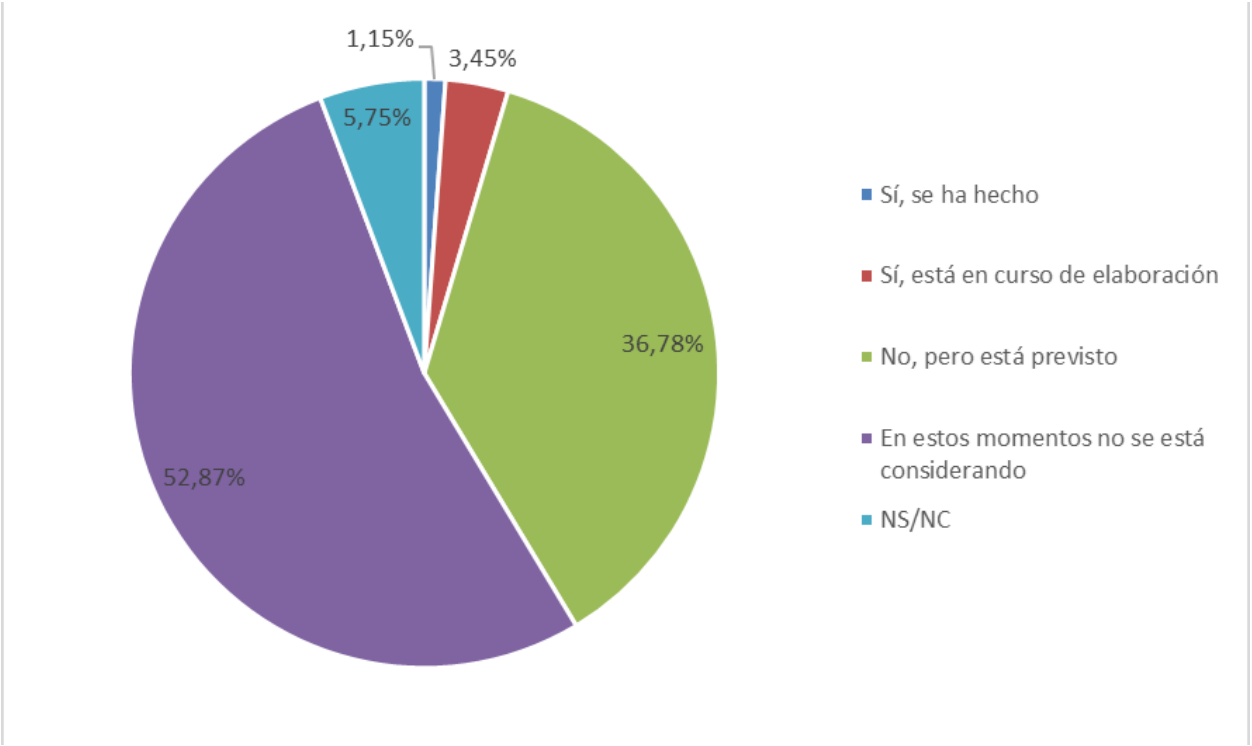
Los resultados apuntan a que existe la preocupación por este tema, pero aún no hay la suficiente capacitación profesional en este ámbito, cerca del 22% han hecho o están haciendo acciones formativas al respecto, y el 50% declara tenerlo previsto, pero aún no se ha iniciado. En el lado opuesto un 25% no lo está considerando.




**PREGUNTA 18: ¿Su organización tiene un plan de respuesta a incidentes de seguridad específico para sistemas de IA?**

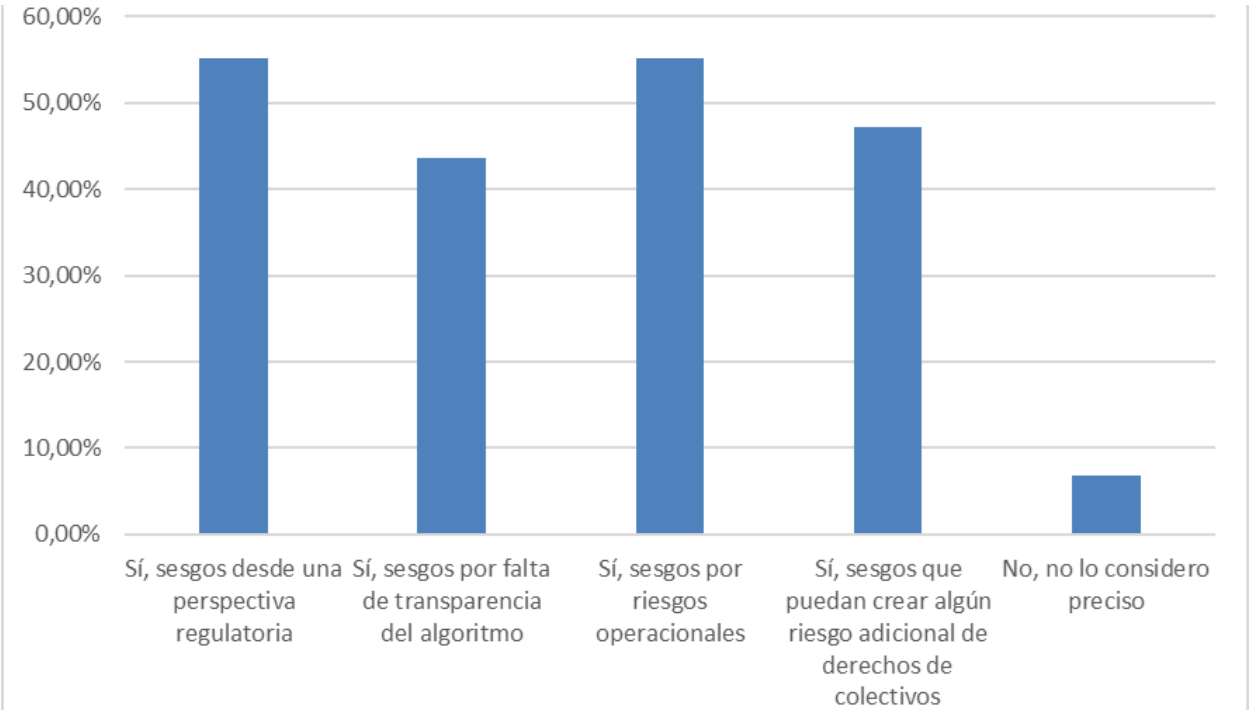
Únicamente el 1,15% declara tenerlo, y un 3,45% estar en curso de elaboración. En el lado contrario casi el 53% ni se lo plantea.

Esta pregunta abre un debate que también se ha vivido dentro del equipo que ha elaborado la encuesta, ¿tiene sentido un plan de respuesta a incidentes de seguridad específico para sistemas de IA? Como siempre la respuesta será "depende", si está utilizando un sistema de IA desarrollado por un tercero probablemente le baste con establecer unas cláusulas contractuales "habituales" que obliguen al tercero a adoptar controles proactivos y reactivos específicos para el sistema de IA; en cambio sí está desarrollando un algoritmo interno de IA y forma parte de un proceso crítico de su organización con toda probabilidad deberá hacer foco en incidentes de seguridad en este ámbito.



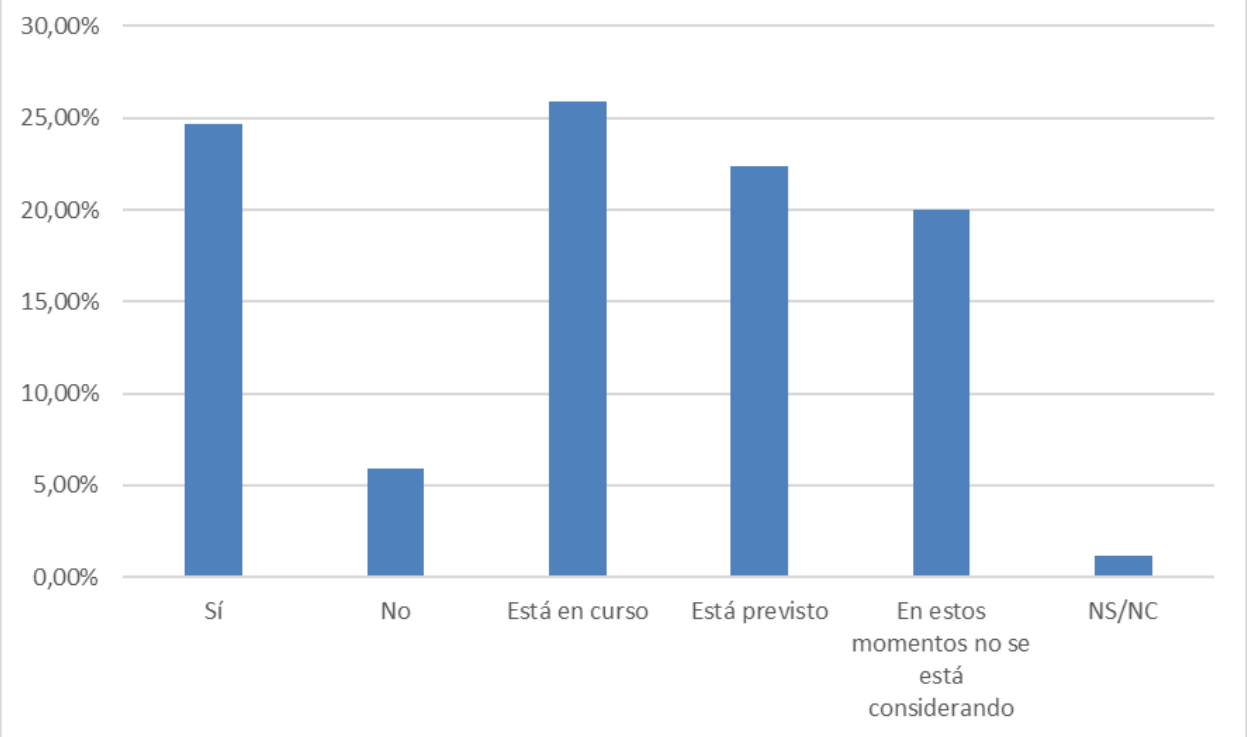

**PREGUNTA 19: ¿Considera preciso implementar en su organización algún tipo de control sobre los posibles sesgos de estos sistemas?**

Las respuestas no plantean dudas, el motivo principal de establecer controles sobre los sesgos se plantea tanto desde perspectiva regulatoria como perspectiva operacional, seguido muy de cerca por la posible vulneración de derechos de colectivos y, por último, por problemas de falta de transparencia del algoritmo.



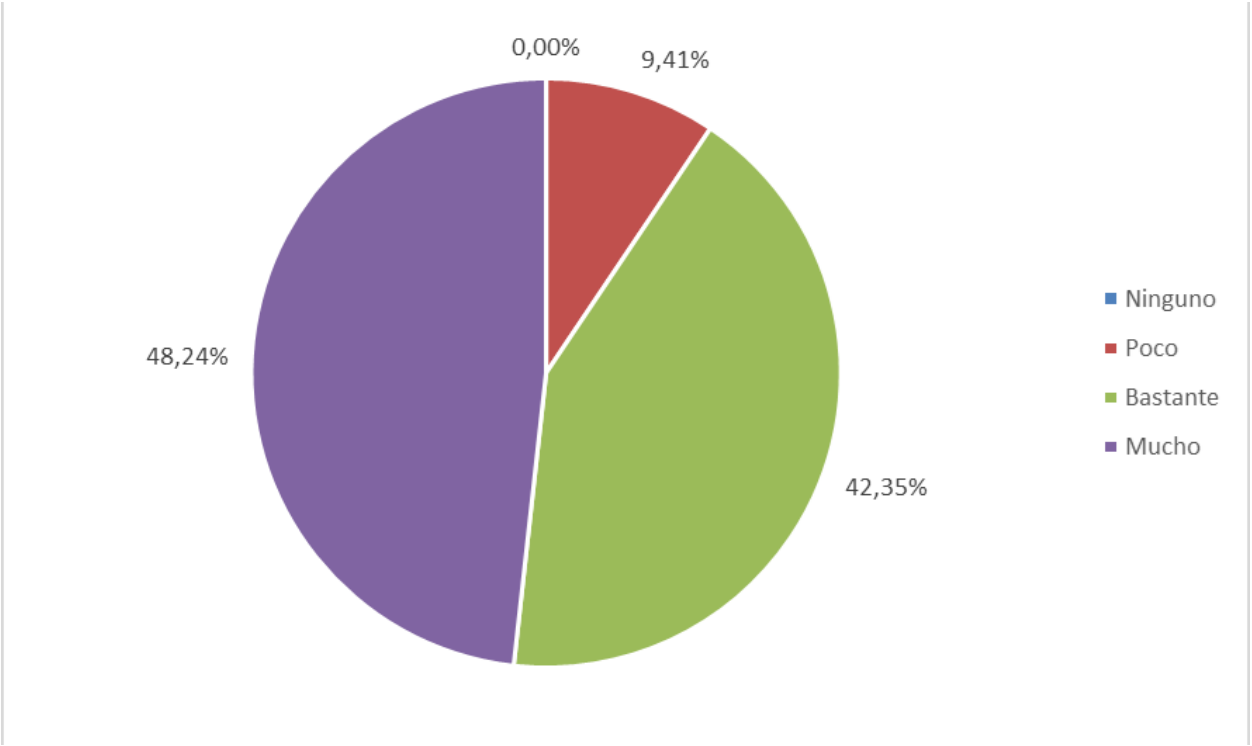
**PREGUNTA 20: ¿Ha considerado los aspectos éticos y legales relacionados con el uso de sistemas de IA en su organización?**

Nuevamente las respuestas son explícitas, casi el 73% declara haberlo hecho o haciéndolo o previsto, en la parte opuesta casi un 26% no lo ha hecho o no lo está considerando.




**PREGUNTA 21: ¿En qué medida considera que es necesario que se establezcan controles en el uso de herramientas de IA en las organizaciones?**

No hay discusión, todas las respuestas apuntan a que es necesario, cambian los matices, frente al 9,41% que declara es poco necesario, algo más del 90% lo considera bastante o muy necesario.




**PREGUNTA 22: ¿Qué otros aspectos considera que debería tratarse con relación a riesgos en la adopción y gobierno de la IA?** ***(pregunta abierta)***

En esta pregunta abierta la práctica totalidad de las consideraciones se estructuran en cuatro ámbitos principales, algunos ya tratados en los puntos anteriores:

- Necesidad de disponer de marcos regulatorios para el uso de la IA.
- Riesgo de terceros en el uso de estos sistemas.
- Preocupación sobre la privacidad y derechos digitales.
- Preocupación sobre sesgos no válidos y necesidad de disponer mecanismos de monitorización, detección y corrección ágil de anomalías.

INFORME "INTRODUCCIÓN A LA IA"


# 9 ANEXO – REFERENCIAS


**CAPÍTULO INTRODUCCIÓN**

https://learn.microsoft.com/en-us/training/modules/introduction-to-ai-technology/

https://mark-riedl.medium.com/a-very-gentle-introduction-to-large-language-models-without-the-hype-5f67941fa59e

http://aima.cs.berkeley.edu/

Dictamen 8/2014 sobre la evolución reciente de la Internet de los objetos ***(GRUPO DE TRABAJO SOBRE PROTECCIÓN DE DATOS DEL ARTÍCULO 29)***

Guía AEPD Adecuación al RGPD de Tratamientos que incorporan Inteligencia Artificial

Blog de la AEPD: "IoT (I): Qué es IoT y cuáles son sus riesgos"

https://www.europarl.europa.eu/portal/es

**CAPÍTULO MACHINE LEARNING**

The Royal Society (2017). Machine learning: the power and promise of computers that learn by example ,978-1-78252-259-1, 419.

**CAPÍTULO APRENDIZAJE PROFUNDO**

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

E. Vicente Cestero, & A. Mateos Caballero (2023). Inteligencia Artificial: Fundamentos matemáticos, algorítmicos y metodológicos. 978-84-09-46911-6.

Szandała, T. (2021). Review and comparison of commonly used activation functions for deep neural networks. Bio-inspired neurocomputing, 203-224.

Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., ... & Morimoto, J. (2022). Deep learning, reinforcement learning, and world models. Neural Networks.

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights, 1(1), 100004.

## CAPÍTULO IMPACTO DE LA IA

https://www.dataversity.net/using-ai-and-machine-learning-with-data-governance/

https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-i-d171b867b047

https://towardsdatascience.com/introducing-the-ai-project-canvas-e88e29eb7024

https://www.mckinsey.com/featured-insights/artificial-intelligence/how-artificial-intelligence-and-data-add-value-to-businesses

https://www.simplilearn.com/ai-is-transforming-real-time-data-governance-article

https://www.dataversity.net/developing-a-functional-data-governance-framework/

https://medium.com/louis-dorard/from-data-to-ai-with-the-machine-learning-canvas-part-i-d171b867b047

https://learn.microsoft.com/en-us/ai-builder/prediction-overview

https://insightsoftware.com/blog/top-5-predictive-analytics-models-and-algorithms

https://www.techtarget.com/searchbusinessanalytics/tip/Four-challenges-to-successful-predictive-analytics-models

https://hbr.org/2018/04/a-simple-tool-to-start-making-decisions-with-the-help-of-ai

https://direct.mit.edu/isec/article/46/3/7/109668/Prediction-and-Judgment-Why-Artificial

https://hbr.org/2023/03/how-ai-is-helping-companies-redesign-processes

https://enterprise-information-management.cioreview.com/cioviewpoint/using-artificial-intelligence-for-process-improvement-nid-27748-cid-184.html

https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems

https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Driving%20impact%20at%20scale%20from%20automation%20and%20AI/Driving-impact-at-scale-from-automation-and-AI.ashx

## PARTE GOBIERNO DE LA IA

https://www.isaca.org/resources/isaca-journal/issues/2021/volume-2/it-governance-framework-for-artificial-intelligence-in-marketing

https://www.ibm.com/docs/es/cloud-paks/cp-data/4.6.x?topic=governance-ai

https://planderecuperacion.gob.es/noticias/conoce-Estrategia-Nacional-Inteligencia-Artificial-ENIA-IA-prtr

INFORME "INTRODUCCIÓN A LA IA"


https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Roadmap for the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) | NIST

OWASP AI Security and Privacy Guide | OWASP Foundation

# INTRODUCCIÓN A LA IA PARA PROFESIONALES DE SEGURIDAD DE LA INFORMACIÓN

www.ismsforum.es
info@ismsforum.es
(+34) 915 63 50 62



Una iniciativa de

isms forum