



# Modelos de Inteligencia Artificial Generativa: riesgos y oportunidades para las empresas

Implicaciones de la utilización de ChatGPT, DALL-E 2, Bard y otras aplicaciones en las organizaciones.



KPMG Internacional

---

Mayo de 2023



**José Ángel Alonso**

Socio Lighthouse  
(CoE Analítica Avanzada)  
KPMG Asesores



Es necesario que los sistemas de IA suficientemente complejos se diseñen de forma que puedan explicar en lenguaje entendible cómo han llegado a los resultados que nos ofrecen, ofreciendo información comprensible sobre la lógica empleada, además de la relevancia y consecuencias previstas del procesamiento realizado sobre los datos. Desde KPMG estamos invirtiendo en el desarrollo de soluciones de IA que faciliten la explicabilidad y transparencia de estos modelos.



**Noemí Brito**

Directora Legal Operations  
and Transformation Services  
KPMG Abogados



En paralelo al despliegue de estas tecnologías emergentes, resulta clave contar con Estrategias, Políticas y Soluciones de IA Responsable, capaces de mitigar al máximo posible los riesgos legales y éticos asociados, permitir demostrar cumplimiento proactivo de la normas aplicables, así como de generar la necesaria confiabilidad en los usuarios, potenciando de esta forma las oportunidades inherentes a esta nueva revolución tecnológica.



**María Barcenilla**

Directora Smart Forensic  
(Ethics & Compliance)  
KPMG Asesores



Las soluciones de IA generativa son ya una realidad y poseen un gran poder para transformar la forma en que trabajamos y nos relacionamos, pero este gran poder requiere también abordar con seriedad los riesgos potenciales derivados de su uso.

La imagen de portada y las imágenes presentes en este informe se han diseñado utilizando DALL-E 2, un generador de arte basado en la IA que crea imágenes a partir de descripciones textuales.

Las indicaciones para la imagen de portada fueron: abstracto fluido, columna ondulada en azul y morado, salpicadura, gotas, fondo morado.

Si bien DALL-E 3 genera un contenido visual atractivo, no ha sido entrenado en las directrices de marca de KPMG ni dispone de la experiencia o el ingenio de los humanos para comprender el posicionamiento de marca de KPMG. Así pues, estas imágenes no se consideran pertenecientes a la marca y se utilizan únicamente con fines ilustrativos con permiso especial de la Marca Global de KPMG.

# Contents

Resumen ejecutivo	04
Descripción general del mercado	05
¿Qué son los modelos de IA generativa?	06
Cómo funcionan los modelos de IA generativa	06
Posibles oportunidades y casos prácticos	07
Consideraciones actuales	10
¿Qué nos depara el futuro?	13
Cómo puede ayudar KPMG	14
Personas de contacto	15

# 01

## Resumen ejecutivo

Los modelos de Inteligencia Artificial (IA) generativa tienen la capacidad de transformar las empresas mediante la automatización y ejecución de determinadas tareas con una velocidad y eficiencia sin precedentes. Este es especialmente el caso en situaciones en las que la experiencia y el ingenio de los humanos se combinan con una comprensión profunda en la manera de utilizar estas tecnologías y aprovechar sus capacidades con efectividad.

No obstante, requerirá tiempo y será necesario que la experiencia humana libere todo su potencial de forma

responsable, fiable y segura. Si está considerando el uso de IA generativa, es importante establecer una serie de procesos y controles internos a cumplir dentro de la organización.

En este informe, analizamos casos prácticos y oportunidades, así como aquellos aspectos a tener en cuenta si se está planteando usar aplicaciones de IA generativa, como ChatGPT, en su organización.

### A continuación, exponemos diez aspectos que debe saber sobre la IA generativa:

- 1 Las soluciones de IA generativa más comunes pueden dividirse, en términos generales, en cinco categorías: generadores de contenido en texto, imagen y vídeo, extractores de contenido, *chatbots* cognitivos, traductores de idiomas y generadores de código software.
- 2 Los modelos de IA generativa pueden resumir artículos, redactar correos electrónicos y producir imágenes y vídeos. Entrenados por humanos, algunos modelos de IA generativa pueden tener habilidades de conversación para, por ejemplo, responder preguntas aclaratorias, admitir errores, cuestionar hipótesis erróneas y filtrar o rechazar peticiones inapropiadas.
- 3 ChatGPT es un *chatbot* entrenado con instrucciones humanas. Su primer modelo de lenguaje extenso subyacente, GPT-3, tenía 175.000 millones de parámetros y fue entrenado con más de 1 millón de conjuntos de datos o 500.000 millones de tokens (palabras o fragmentos de palabras). GPT-3 no estaba conectado a Internet y se alimentó con datos de hasta septiembre de 2021. GPT-4, el nuevo modelo multimodal extenso de OpenAI, ha evolucionado desde el modelo anterior.
- 4 Los modelos de IA generativa tienen potencial de uso en distintas funciones empresariales, desde TI hasta recursos humanos y operaciones, pasando por finanzas, auditoría, el área legal o marketing. Algunas de sus aplicaciones son la redacción de propuestas, el desarrollo y la comprobación de códigos, o la extracción y resumen de información compleja.
- 5 La IA generativa utiliza datos y parámetros para aprender y crear conocimiento. Salvo que el usuario impida su uso explícitamente al proveedor de la aplicación, esos datos pueden utilizarse posteriormente para dar respuesta a indicaciones de otra persona, con el consecuente riesgo de exponer al público la información confidencial de la organización. En función de la aplicación, podría también estar cediendo sus derechos de autor. Los términos y condiciones de cada aplicación nos darán una idea sobre el uso de los datos introducidos por el usuario.
- 6 En función de para qué se utilice la IA generativa y cómo se ejecute, existe el riesgo de exponer propiedad intelectual o secretos comerciales y provocar que su organización sea vulnerable a riesgos de fraude. Es importante mantenerse alerta y asegurarse de no utilizar la IA de una forma que infrinja las leyes aplicables (incluidas las normas de protección de datos personales, o de propiedad intelectual e industrial, entre otras), los acuerdos con clientes o las normas profesionales.
- 7 La copia de información o código producidos por la IA en un entregable o producto puede constituir una infracción de derechos de autor u otros derechos de propiedad intelectual, con los consecuentes perjuicios legales y de reputación.
- 8 Es de esperar que las próximas versiones comerciales de la IA generativa de diferentes vendedores tecnológicos comiencen a estar integradas en muchas aplicaciones, sistemas y procesos comunes, desde los navegadores de Internet hasta la tecnología conectada que las organizaciones autorizan, tales como el *software* basado en la nube o los programas de mensajería instantánea.
- 9 Crear directrices para un uso seguro en su organización es clave para contribuir a garantizar un uso adecuado y efectivo de las aplicaciones de IA generativa. En paralelo, resulta indispensable mejorar la formación de sus empleados, dado que el 'human in the loop' brinda percepciones y conocimientos únicos que la IA generativa por sí misma no puede replicar.
- 10 KPMG adopta un enfoque responsable con respecto al diseño, desarrollo e implantación de sistemas de IA de forma segura, fiable y ética. Este enfoque ayuda a las empresas a acelerar el valor para los consumidores, las organizaciones y la sociedad.

Según la firma de análisis y consultoría Gartner, para 2025 el 30% de los mensajes salientes de grandes organizaciones se generarán de forma sintética. <sup>1</sup>En el [informe AI Risk Survey](#), realizado por KPMG en Estados Unidos en septiembre de 2022, el 85% de los encuestados espera un incremento del uso de la IA y los modelos de análisis predictivos. Además, en la edición de 2022 [de la encuesta sobre tecnología de KPMG en Estados Unidos](#), la mitad de los encuestados afirmó haber obtenido rendimiento de las inversiones en tecnología de IA.

Los modelos de IA generativa fueron el foco de atención en el verano de 2022, cuando una imagen generada por IA ganó un concurso de arte<sup>2</sup>. En noviembre, volvieron a acaparar los titulares tras el lanzamiento de ChatGPT. No obstante, fue en enero de 2023, durante una sesión del [Foro Económico Mundial](#) en la que el presidente y consejero delegado de Microsoft, Satya Nadella, afirmó que «la edad de oro de la IA» está en marcha,<sup>3</sup> cuando empezaron a intensificarse las expectativas en torno a ChatGPT, lo que dio lugar a muchas preguntas de —y conversaciones con— nuestros clientes.

Entrenar estos modelos requiere grandes volúmenes de capital, esfuerzo humano y capacidad informática. OpenAI, creador de ChatGPT, recibió 1.000 millones de USD de Microsoft<sup>4</sup> y una inversión plurianual multimillonaria de la compañía a principios de 2023<sup>5</sup>. Asimismo, Google<sup>6</sup> y Meta<sup>7</sup> han creado modelos de IA generativa por su cuenta. En vista de la amplia gama de posibles aplicaciones, se está desarrollando toda una industria dedicada a hacer que los modelos de IA generativa sean útiles.

Las aplicaciones de IA generativa pueden dividirse, en términos generales, en cinco categorías: generadores de contenido, extractores de contenido, chatbots cognitivos, traductores de idiomas y generadores de código:

- **Generadores de contenido:** Son herramientas transformadoras generativas previamente entrenadas que generan contenido como publicaciones en blogs, correos electrónicos, publicaciones en redes sociales, imágenes, texto web y anuncios.
- **Extractores de contenido:** Estas aplicaciones pueden generar resúmenes cortos y largos de artículos de prensa, publicaciones en *blogs*, documentos legales, etcétera. Asimismo, algunas empresas las usan para desarrollar y analizar documentos legales.
- **Chatbots cognitivos:** Las empresas utilizan cada vez más *chatbots* inteligentes a modo de asistentes para los consumidores. Los *chatbots* interactúan mediante conversaciones y pueden responder preguntas aclaratorias, admitir errores, cuestionar ideas incorrectas y rechazar solicitudes inapropiadas.
- **Traductores de idiomas:** Herramientas multilingües que pueden traducir numerosos idiomas. Tienen la capacidad de desarrollar interfaces de páginas web enteras, incluidos *sites* de traducción.
- **Generadores de código:** Los modelos de IA generativa pueden convertir entradas de texto natural en fragmentos de código o aplicaciones. Con una descripción básica o un pequeño programa, estos modelos pueden producir códigos en varios lenguajes de programación e identificar y resolver errores.

<sup>1</sup> Gartner, *7 Technology Disruptions That Will Completely Change Sales*, 10 de octubre de 2022: <https://www.gartner.com/en/articles/7-technology-disruptions-that-will-completely-change-sales>. GARTNER es una marca comercial registrada y una marca de servicio de Gartner, Inc. y/o sus filiales en Estados Unidos y a escala internacional, y se usa en el presente con su permiso. Todos los derechos reservados.

<sup>2</sup> <https://www.smithsonianmag.com/smart-news/artificial-intelligence-art-wins-colorado-state-fair-180980703/>

<sup>3</sup> <https://www.weforum.org/press/2023/01/satya-nadella-says-ai-golden-age-is-here-and-it-s-good-for-humanity>

<sup>4</sup> <https://openai.com/blog/microsoft-invests-in-and-partners-with-openai>

<sup>5</sup> <https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/>

<sup>6</sup> <https://blog.google/technology/ai/bard-google-ai-search-updates/>

<sup>7</sup> <https://ai.facebook.com/>

# 03

## ¿Qué son los modelos de IA generativa?

La IA generativa hace referencia a la Inteligencia Artificial que puede generar contenido en vez de simplemente analizar o realizar acciones a partir de datos existentes.

Los modelos de IA generativa, como GPT-4, se desarrollan y entrenan sobre la base de un conjunto de datos recopilados. Pueden ser generalistas o especializados, desarrollados a partir de recopilaciones de datos predefinidas, los cuales están diseñados para producir un resultado que ayude a realizar determinadas peticiones dirigidas por humanos. Algunos modelos pueden, por ejemplo, predecir la siguiente palabra en función de las frases anteriores o la siguiente imagen en función de las descripciones de las imágenes anteriores.

Este entrenamiento posibilita una rápida generación de contenido original, incluido texto, imágenes, vídeo y código. Con una necesidad menor de recursos humanos, algunas empresas esperan poder producir contenido más rápido y con un menor coste, proporcionando oportunidades para generar nuevos tipos de contenido que anteriormente resultaban demasiado caros o que requerían demasiado tiempo. Esto tiene el potencial de cambiar de forma radical la interacción entre humanos y máquinas, y abre un sinfín de posibles casos prácticos.

Dicha capacidad predictiva permite a los modelos analizar documentos. Por ejemplo, pueden utilizarse para identificar documentos sobre una temática descrita mediante un texto introducido por el usuario.

# 04

## Cómo funcionan los modelos de IA generativa

Los modelos de IA generativa están diseñados para producir contenido basándose en un conjunto claro de indicaciones y normas.

La aplicación de modelos de IA generativa de la que más se habla en los últimos meses es ChatGPT, un *chatbot* entrenado con instrucciones humanas creado por el laboratorio de investigación de OpenAI en San Francisco<sup>8</sup>. Desde el 14 de marzo de 2023, los suscriptores de ChatGPT Plus pueden usar GPT-4, un modelo multimodal extenso (LMM, por sus siglas en inglés) que acepta entradas tanto de texto como de imágenes y genera resultados en texto.<sup>9</sup> Asimismo, el 23 de marzo de 2023, OpenAI lanzó extensiones para ChatGPT, incluida su propia extensión de búsqueda en la web. Esto significa que ChatGPT puede ahora acceder a determinadas fuentes y bases de datos de terceros.<sup>10</sup>

ChatGPT significa Chat (basado en conversación) G(enerativo) P(preentrenado) T(ransformador). Se perfeccionó utilizando aprendizaje reforzado con sugerencias humanas, en el que un modelo de recompensa que representa las preferencias humanas se entrena para que los resultados parezcan más

humanos, se previenen respuestas no deseadas y se trata de evitar «alucinaciones» (hechos inventados).

ChatGPT se creó como un modelo de lenguaje extenso (LLM, por sus siglas en inglés) y, desde entonces, ha evolucionado a una aplicación de IA generativa multimodal extensa. Esto significa que la aplicación ahora acepta entradas de imagen y texto, no solo de texto, como anteriormente. Junto con un modelo de red neuronal que utiliza aprendizaje no supervisado para predecir resultados, este tipo de modelo generativo puede determinar los patrones lingüísticos y las relaciones más probables entre el contenido que ya ha integrado.

Con «extenso» se hace referencia al volumen de datos en que se basan los modelos, así como al tamaño de los modelos en sí, e implica alimentarlos con una colección ingente de documentos electrónicos de dominio público. Por ejemplo, cuando se lanzó en 2022, ChatGPT tenía 175.000 millones de parámetros (un valor que controla el comportamiento del modelo de aprendizaje automático; cuantos más parámetros, mayor capacidad tiene el modelo para analizar).

<sup>8</sup> <https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/?sh=7510d916678c>

<sup>9</sup> <https://openai.com/research/gpt-4>

<sup>10</sup> ChatGPT plugins (openai.com)

Inicialmente se entrenó con más de 1 millón de conjuntos de datos o 500.000 millones de tokens (palabras o fragmentos de palabras), incluido contenido de Wikipedia y The New York Times. Para poner esto en perspectiva, una persona pronuncia de media 860,3 millones de palabras en su vida<sup>11</sup>, por lo que esta colección —o «corpus», en terminología de la IA— equivale a 300 años de lenguaje. La versión básica de ChatGPT no está conectada a Internet y se alimentó con material *online* hasta septiembre de 2021, de modo que sus conocimientos no están actualizados. Las versiones más recientes lanzadas para un número reducido de desarrolladores *premium*, como una extensión para el motor de búsqueda Bing<sup>12</sup>, están conectadas a Internet y disponen de contenido más reciente.

ChatGPT es un ejemplo de Inteligencia Artificial limitada (Artificial Narrow Intelligence, ANI, por sus siglas en inglés).

Los sistemas de ANI son adecuados para ejecutar un único tipo de tarea para la que hayan sido entrenados. Por ejemplo, un sistema ANI diseñado para generar imágenes probablemente no será capaz de resolver problemas matemáticos.

Según OpenAI, GPT-4, si bien aún no es totalmente fiable, lo es sustancialmente más, y además es capaz de gestionar instrucciones matizadas en comparación con su predecesor, GPT-3.5. Y, lo que es más importante, superó un examen de acceso a la abogacía simulado y se situó entre el 10% con mejor puntuación de los participantes. En comparación, GPT-3.5 se situó entre el 10% con peor puntuación en ese mismo examen simulado. OpenAI señala que sus limitaciones son similares a los modelos de GPT anteriores: invención de hechos y errores de razonamiento<sup>13</sup>.

# 05

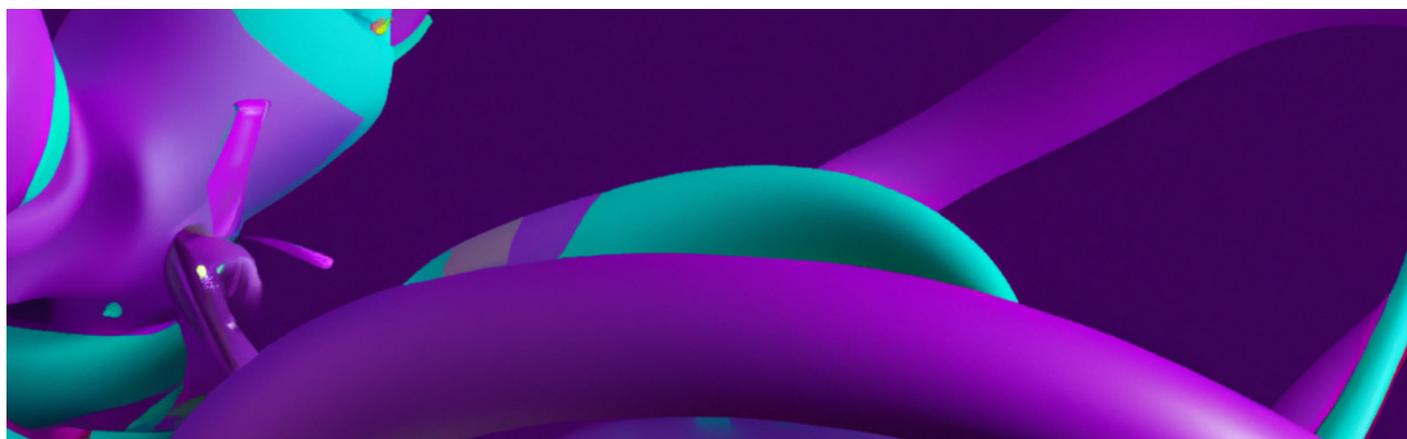
## Oportunidades y casos prácticos

El ascenso vertiginoso de la popularidad de ChatGPT se debe, en parte, a que cualquiera puede usarlo, incluso aquellos sin conocimientos técnicos. El rápido crecimiento en el número de usuarios —100 millones a febrero de 2023<sup>14</sup>— demuestra el interés por utilizar esta tecnología. Y cuantos más usuarios tiene un *chatbot*, más entrenada estará su IA subyacente.

ChatGPT tiene el potencial de transformar las empresas automatizando y ejecutando tareas basadas en el lenguaje con una velocidad y eficiencia sin precedentes. Estos modelos de lenguaje pueden implantarse para ayudar con una amplia gama de tareas: pueden modificarse para resumir y clasificar documentos legales, responder a preguntas de consumidores, asistir a asesores expertos o generar esquemas de ingeniería y arquitectura.

Pueden hacer las veces de punto de partida para la inspiración humana, dado que ofrecen ideas que pueden transformarse en pensamientos innovadores y creativos. Esto los hace idóneos para ayudar a generar informes empresariales, propuestas de marketing o código para aplicaciones de *software*.

Los modelos de IA también pueden aplicarse a funciones empresariales como TI, legal, auditoría, recursos humanos, operaciones y muchas más. A la hora de explorar estos casos prácticos, es importante tener en cuenta que, a pesar de las muchas oportunidades que ofrece la IA generativa, estas no están exentas de riesgo.



<sup>11</sup> [https://openlibrary.org/books/OL3502128M/The\\_joy\\_of\\_lex](https://openlibrary.org/books/OL3502128M/The_joy_of_lex)

<sup>12</sup> <https://www.bing.com/new>

<sup>13</sup> <https://openai.com/research/gpt-4>

<sup>14</sup> <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>

En el espacio de las **operaciones de TI**, pueden utilizarse para:



### Sistemas de gestión de conocimientos basados en LMM

Recopilar información de fuentes de datos con varios formatos. Posteriormente, esta información puede estructurarse en consultas para buscar elementos específicos.



### Soporte autónomo de TI

Ayudar a los empleados a resolver errores de los sistemas de TI a través de instrucciones de soporte generadas por *chatbots* basados en IA.



### Codificación o comprobación de códigos

Convertir código de una función a otra, por ejemplo, de SQL a Python, o comprobar un código para garantizar que funciona.

En cuanto a las áreas de **auditoría/cumplimiento**, pueden ayudar a la hora de:



### Revisiones de auditoría automatizadas

Automatizar la recopilación de hechos para la presentación de auditorías y revisiones de auditoría detalladas en formatos de consulta.



### Evaluaciones de requisitos de independencia

Evaluar los requisitos de independencia de los encargos de auditoría para ayudar a simplificar los procesos de aprobación para certificar la independencia.

Entre los usos posibles en recursos humanos, destacan:



### Selección de candidatos

Entrenar a los modelos de IA generativa con descripciones de puestos y datos de habilidades relevantes para ayudar a identificar candidatos idóneos para el puesto.



### Aplicaciones de uso autónomo

Implantar *chatbots* que pueden compartir conocimientos al estilo humano y resolver consultas de RR. HH.

En lo que respecta a las **operaciones**, pueden ayudar con:



### Presentación de información sobre sostenibilidad y ESG

Contextualizar datos ESG y prestar apoyo a las operaciones de presentación de información, incluida la creación de declaraciones de lenguaje simple que pongan de relieve las iniciativas en este ámbito.



### Gestión virtual de eventos

Coordinar la gestión de eventos a través de la redacción de invitaciones, la planificación de sesiones y la formulación de respuestas a preguntas de los asistentes.



### Simplificación de las operaciones empresariales

Desde la redacción de correos electrónicos hasta la preparación de solicitudes de propuestas, pasando por la ejecución de análisis e investigaciones de la competencia para garantizar la comprensión del mercado.

En el espacio de **finanzas/logística**, pueden ser utilidad para:



### Categorizar y validar los pagos

Ayudar a las empresas a publicar sus contribuciones fiscales clasificando volúmenes masivos de datos.



### Realizar recomendaciones de independencia personalizadas para las inversiones

Permitir a las organizaciones proporcionar respuestas personalizadas a consultas relativas a la independencia de determinadas inversiones a través de *chatbots*.

Entre las opciones para el **gobierno del área legal y la organización**, destacan:



#### **Redactar y revisar condiciones contractuales**

Revisar contratos y poner de relieve cláusulas con posibles conflictos de intereses, así como redactar cláusulas y condiciones para acelerar el proceso contractual.



#### **Sacar a la luz jurisprudencia, citas legales y enlaces de fuentes**

Buscar citas legales, jurisprudencia y ejemplos de casos relevantes, lo que ayuda a identificar fuentes fiables de manera más rápida.

Las posibles aplicaciones para la división de **marketing son**, entre otras:



#### **Simplificar el lenguaje de las campañas**

Encontrar opciones de palabras alternativas fáciles de traducir a distintos idiomas.



#### **Localizar comunicaciones de marketing a escala**

Ayudar a localizar campañas globales compartiendo los datos de las conversaciones a escala local con el modelo.

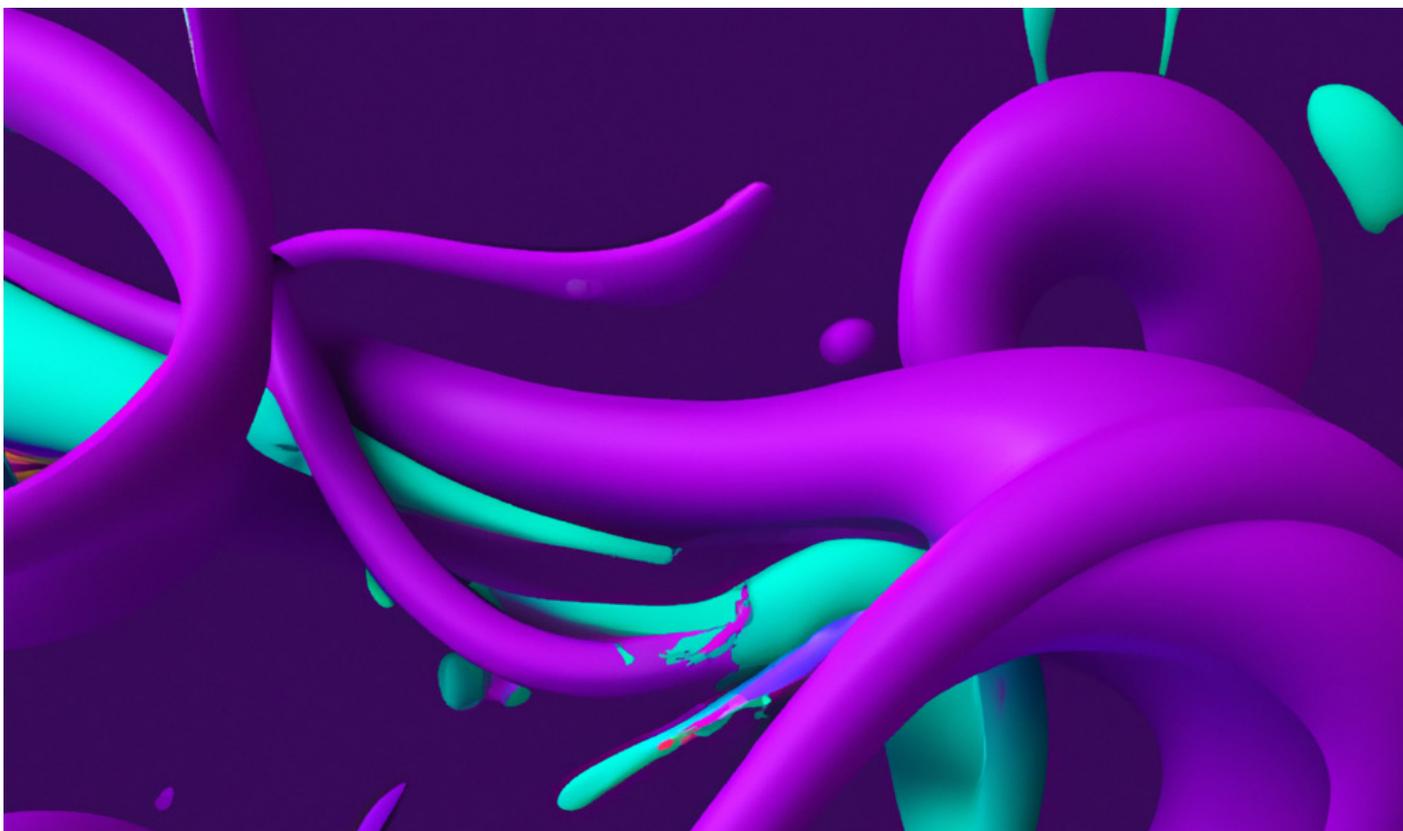


#### **Sintetizar información compleja**

Aprender la información básica sobre, por ejemplo, los procesos de *due diligence* financiera, para digerir y estructurar el contenido, y contribuir a desarrollar una campaña de marketing sólida.

En KPMG, somos líderes reconocidos en IA, aprendizaje automático y analítica avanzada, y tenemos un profundo conocimiento sobre los riesgos que deben tenerse en cuenta en relación con la tecnología emergente. Desde nuestra experiencia, podemos ayudar a valorar los aspectos éticos, de gobierno corporativo y de seguridad en torno a las tecnologías de IA y aprendizaje automático de los clientes.

Las firmas de KPMG se sitúan históricamente a la vanguardia de la exploración y el uso de nuevas tecnologías, y pueden responder a preguntas sobre cómo las aplicaciones de IA generativa pueden impulsar el crecimiento de su organización.



Hemos presentado ejemplos sobre cómo los modelos de IA generativa pueden ayudar a los consumidores, optimizar los procesos organizativos y liberar tiempo para que los empleados asuman tareas de mayor valor. Si bien, el uso de la IA generativa presenta también numerosas limitaciones y posibles peligros.

Como comentábamos en la Parte 4, GPT-3, el LLM subyacente inicial de ChatGPT, se entrenó con material de hasta septiembre de 2021 y no estaba conectado a Internet. Y si bien OpenAI ha posibilitado que ChatGPT tenga capacidad para buscar en Internet en algunos casos, sigue siendo fundamental garantizar que la revisión y la experiencia humanas se integran en el proceso de uso de la IA generativa ("human in the loop"). Los modelos de IA generativa pueden ser el núcleo de una aplicación de IA, pero requieren análisis, tecnología y procesos humanos adicionales para resolver problemas.

En este apartado, abordamos los riesgos de utilizar modelos y aplicaciones de IA generativa y cómo gestionarlos, incluidos aquellos en torno a la confidencialidad de los clientes y la empresa, el uso inadecuado por parte de empleados o las prácticas de *phishing*.

Los modelos multimodales extensos como ChatGPT generan respuestas similares a las de los humanos. No obstante, carecen de las habilidades de razonamiento de las personas. Para que se consideren fiables, los usuarios son responsables de aplicar sus capacidades de IA a casos prácticos adecuados, y su organización debería formar a los empleados en el uso de estos programas. También es importante que los desarrolladores utilicen conjuntos de datos fiables para alimentar los modelos de IA, y aplicar sesgos y filtros de contenido relevantes.

## Gestión de riesgos

La creciente popularidad de la IA generativa es otro motivo por el que recomendamos desarrollar e implantar la IA de forma responsable si su organización quiere protegerse frente a un uso indebido. A continuación exponemos algunos de los retos en materia de gestión de riesgos de los modelos de IA generativa.

## Riesgos y consideraciones a nivel interno

### *Vulneración de la confidencialidad y los derechos de propiedad intelectual*

Muchos modelos de IA generativa están diseñados para absorber datos introducidos por los usuarios y mejorar los modelos subyacentes con el tiempo, lo que esencialmente les ayuda a aprender y desarrollar conocimientos. Esos datos, a su vez, pueden utilizarse para responder a una indicación de otra persona, lo que puede exponer información privada o confidencial al público. Cuanto más use su organización esta tecnología, más probabilidades habrá de que otros puedan acceder a su información sensible o confidencial. Por tanto, su organización necesita determinar cómo proteger su propiedad intelectual para que, al mismo tiempo, se pueda aprovechar de las ventajas que aportan las aplicaciones de IA generativa.

### *Uso indebido de los empleados y errores*

Incluso el uso legítimo de la IA generativa puede entrañar riesgos. Los modelos generan respuestas sobre la base de los datos introducidos, lo que conlleva el riesgo de que genere contenido falso o malicioso. Cuando sus empleados utilicen estas soluciones, deben tener cuidado y revisar el contenido generado por la IA con una visión crítica y haciendo hincapié en la calidad.

Si el contenido de la IA generativa presenta errores que no se detectan, ello podría afectar a los resultados de su empresa o generar problemas de responsabilidad. Por ejemplo, el *bot* de IA generativa de Meta, Galactica, se creó para condensar información científica y ayudar a los académicos e investigadores a encontrar informes y estudios con rapidez. En lugar de ello, produjo ingentes volúmenes de información incorrecta en la que se citaba erróneamente a científicos prestigiosos<sup>15</sup>. En otro *bot* de Meta, BlenderBot3, se identificaron declaraciones falsas y sesgadas<sup>16</sup> poco después de su lanzamiento en agosto de 2022. El *chatbot* de Google, Bard, provocó que su sociedad matriz, Alphabet, perdiera 100.000 millones de USD de su valor de mercado tras compartir información incorrecta en su primera versión de demostración<sup>17</sup>. Las invenciones de hechos de ChatGPT también están bien documentados<sup>18,19</sup>, y su desarrollador, OpenAI, reconoció estos fallos<sup>20</sup>.

Otro riesgo en torno a la IA generativa es, entre otros, la posibilidad de que la tecnología pueda generar información sensible, como datos personales, que pueda utilizarse

<sup>15</sup> <https://gizmodo.com/meta-ai-bot-galactica-1849813665>

<sup>16</sup> <https://www.cnn.com/2022/08/11/tech/meta-chatbot-blenderbot/index.html>

<sup>17</sup> <https://www.npr.org/2023/02/09/1155650909/google-chatbot-error-bard-shares>

<sup>18</sup> <https://www.npr.org/2023/03/17/1164383826/heres-what-the-latest-version-of-chatgpt-gets-right-and-wrong>

<sup>19</sup> <https://news.yahoo.com/factual-errors-inflated-bios-aside-100209244.html>

<sup>20</sup> <https://openai.com/research/gpt-4>

### Cuestiones que deben tenerse en cuenta:

1. ¿Cómo puede garantizar que se mantienen la confidencialidad y la precisión al utilizar modelos de IA generativa?
2. ¿Cómo puede garantizar de forma efectiva que sus modelos de IA generativa cumplen las normativas aplicables en rápida evolución?
3. ¿Cómo puede automatizar la revisión y la gestión de políticas de cumplimiento?
4. ¿Qué deberían saber sus empleados sobre la IA generativa en términos de riesgos y beneficios?

para el robo de identidades o para invadir la privacidad de las personas. Incluso un empleado descontento o un consumidor enfadado podrían crear material falso para dañar la reputación de su empresa, o la de sus empleados o directivos.

### La IA generativa evoluciona

A medida que la comprensión sobre la IA sigue evolucionando, ya estamos asistiendo a un incremento de las normativas en todo el mundo. Resulta importante mantenerse al corriente a este respecto, incluso si no planea utilizar la IA generativa de forma intencional.

Es de esperar que la IA generativa siga integrándose en muchas aplicaciones, sistemas y procesos comunes, desde los navegadores de Internet hasta la tecnología conectada a la IA que su organización puede autorizar. Por tanto, resulta esencial mantenerse alerta y asegurarse de no está utilizando la IA en la esfera profesional de una forma que infrinja las leyes aplicables (incluidas las leyes de privacidad), los acuerdos con clientes o las normas profesionales.

### Implicaciones para el personal de la organización

Solo pueden conseguirse unos resultados de elevada calidad y de nivel experto con consultas de la misma calidad y nivel. Por tanto, su organización tendrá que formar a su personal y conservar sus propios conocimientos para contextualizar las consultas y emitir las indicaciones adecuadas. En KPMG, por ejemplo, hemos puesto a disposición de todos nuestros empleados formación sobre IA generativa a través de nuestro programa Digital and Data Foundations, que ofrece contenido básico sobre la evolución de la IA y cómo desarrollar, implantar e interactuar con la IA de forma fiable.

Los profesionales deben ser conscientes de que no están simplemente usando una solución tecnológica, sino que también la están entrenando y evolucionando.

En un futuro generativo, prevemos que el papel de los profesionales evolucionará desde la resolución de problemas hasta la definición de estos, y los equipos trabajarán con las máquinas para crear nuevos enfoques. Las herramientas de IA generativa son una interfaz, no un oráculo.

La participación de los humanos aporta percepciones únicas al proceso que la IA generativa por sí misma no puede replicar. Ofrece comentarios fundamentales para perfeccionar y mejorar el modelo con el tiempo y ayudar a garantizar que los resultados son precisos, justos y cumplen los objetivos deseados.

Cuando las personas y la tecnología están en sintonía pueden suceder cosas maravillosas, y creemos firmemente que no puede producirse un cambio duradero sin el ingenio humano.

### Riesgos y consideraciones a nivel externo

#### Información incorrecta, sesgos y discriminación

Como hemos comentado antes, los sistemas basados en LLM y LMM han compartido información falsa, obsoleta y discriminatoria, pero la presentan con tal autoridad y de un modo que incluso el lector más escéptico puede caer en la trampa.

La IA generativa puede utilizarse —y ya se ha utilizado— para crear imágenes y vídeos *deepfake* (contenido visual alterado para hacer parecer que alguien ha dicho o hecho algo, cuando no es así). Estas imágenes y vídeos a menudo parecen extremadamente realistas y no dejan rastro forense en los medios digitales editados, lo que dificulta que los humanos o incluso las máquinas puedan detectar esta actividad.<sup>21</sup>

#### Derechos de autor (Copyright)

Existen multitud de preguntas sobre quién posee el contenido cuando se ha procesado a través de aplicaciones de IA generativa, y no existe una respuesta universal. Los términos y condiciones varían en función de la herramienta, así como de la forma en que se utilizan los materiales.

Si el contenido se corta y pega o no sufre prácticamente cambios con respecto a un texto cuyos derechos de autor pertenecen a otra persona, se podría considerar plagio. Es complicado determinar de forma definitiva cuánta información obtenida a través de una herramienta de IA generativa tendría que cambiarse para considerarla legítimamente propia.

Declarar que un contenido generado con IA es propio puede dar lugar a infinidad de problemas éticos. Para empezar, actuar de este modo no es responsable ni honesto y, si saliera a la luz, podría hacer dudar a clientes y consumidores sobre su honestidad en todos los sentidos. Además, si los clientes o consumidores descubrieran que simplemente está transmitiendo información generada por IA, ¿qué les impediría hacer lo mismo y eliminar el intermediario (su organización) por completo?

En el siguiente subapartado profundizamos más en los riesgos para la reputación derivados de la IA generativa.

#### Riesgo financiero, de marca y de reputación

En el seno de su organización, la copia de información o código producidos por la IA en un entregable o producto

<sup>21</sup> <https://www.propertycasualty360.com/2021/09/14/deepfakes-an-insurance-industry-threat/>

puede constituir una infracción de derechos de autor u otros derechos de propiedad intelectual. Esto podría causar a su organización perjuicios legales y de reputación.

Si bien muchas de estas herramientas indican específicamente a los usuarios que no introduzca información confidencial de sus clientes, los usuarios que carezcan de formación y comprensión al respecto pueden exponer involuntariamente la propiedad intelectual o secretos comerciales al público o incluso a un competidor. Ello podría dar lugar a demandas y afectar negativamente a los resultados de su empresa si los clientes actuales o potenciales y los consumidores ponen en duda si pueden fiarse de su empresa en relación con su información confidencial.

La ausencia de transparencia a la hora de utilizar el contenido creado mediante IA generativa también puede causar problemas de reputación. La editorial tecnológica CNET recibió críticas por utilizar la tecnología sin comunicarlo para escribir más de 70 artículos desde noviembre de 2022<sup>22</sup> —algunos de los cuales contenían errores—, si bien en su página web afirmaba que un equipo de editores participa en el contenido «desde la concepción hasta la publicación».

### Cuestiones que deben tenerse en cuenta:

1. ¿Cómo puede garantizar que las aplicaciones de IA generativa se gestionan de forma efectiva para evitar sanciones financieras por incumplir las normativas?
2. ¿Cómo puede fiarse de las aplicaciones de IA que utiliza?
3. ¿Cómo puede gestionar proactivamente sus aplicaciones de IA generativa y ser consciente y mantenerse alerta ante posibles sesgos o discriminación?
4. ¿Se alinea su uso con sus principios, sus valores y su marca?

### Ciberseguridad

Los ciberdelincuentes pueden utilizar la IA generativa para crear estafas de *phishing* más realistas y sofisticadas o credenciales para abrirse paso en los sistemas. Además, los algoritmos de la IA no pueden proteger los conjuntos de datos subyacentes que utilizan para su entrenamiento. Existen estudios que han demostrado que los algoritmos pueden distinguir las identidades individuales aunque los datos se hayan anonimizado y depurado<sup>23</sup>.

Otros riesgos de la IA generativa en materia de ciberseguridad incluyen el envenenamiento de datos, esto es, cuando los datos utilizados para entrenar los modelos se manipulan, y los ataques adversarios, con los que se intenta engañar a los modelos de IA generativa alimentándolos con datos maliciosos.

A medida que su organización explora los casos prácticos de ChatGPT y otras aplicaciones de IA generativa, recomendamos que sus equipos de ciberseguridad y

riesgos establezcan directrices y normas de implantación segura. Así, pueden establecer las expectativas sobre el uso de ChatGPT y otras soluciones en el entorno empresarial, la formación del personal sobre las ventajas y riesgos de utilizar aplicaciones de IA generativa, y la implantación de controles de ciberseguridad según resulte apropiado.

### Cuestiones que deben tenerse en cuenta:

1. ¿Qué nivel de seguridad tienen sus aplicaciones de IA generativa frente a ciberataques, actores maliciosos y amenazas internas?
2. ¿Funcionan sus controles de seguridad? ¿Cómo se pueden mejorar?
3. ¿Violan la privacidad de alguna persona las aplicaciones de IA generativa que usa?

### Ataques adversarios

Incluso cuando están entrenados para trabajar dentro de unos límites aceptables, los modelos de IA generativa han demostrado ser vulnerables, como cualquier otro modelo de análisis, a la manipulación deliberada por parte de usuarios externos sofisticados. Si su organización tiene previsto utilizar soluciones de IA generativa, debe ser consciente de que esto podría sucederle cuando la solución está expuesta al público.

### Cuestiones que deben tenerse en cuenta:

1. ¿Cuáles son las vulnerabilidades adversarias básicas conocidas de las tecnologías que está utilizando?
2. ¿Cómo puede ensayar ataques probables y endurecer las soluciones existentes y nuevas para estar preparado frente a estos?
3. ¿Qué supervisión ha implantado para identificar ataques adversarios?

### Respaldar un uso apropiado de la IA generativa

Recomendamos crear directrices para un uso seguro en su organización, ya que esto es clave para contribuir a garantizar un uso adecuado y efectivo de las aplicaciones de IA generativa. Las directrices pueden incluir el requisito de contar con formación para cualquiera que quiera utilizar la IA generativa y determinar cómo debe utilizarse y cómo no. Además, su organización debería tratar la IA generativa como cualquier otra solución tecnológica y exigir a los empleados que respeten las políticas relevantes (políticas sobre el uso aceptable o la seguridad de la información, por ejemplo) que ya se hayan implantado.

Creemos que aún existe mucho trabajo por hacer antes de que podamos utilizar la última generación de IA para las interacciones entre consumidores, empleados, ciudadanos y empresas. Con un programa de IA responsable, las organizaciones pueden empezar a avanzar en el desarrollo de procesos y procedimientos en torno al uso de IA generativa.

<sup>22</sup> <https://gizmodo.com/cnet-artificial-intelligence-writing-scandal-1850031292>

<sup>23</sup> <https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-of-anonymous-data/>

Si analizamos lo que las empresas tecnológicas están explorando en torno a la IA generativa podemos obtener una idea sobre hacia dónde podría dirigirse este espacio en el futuro.

## Desarrollo y mantenimiento del software

La IA generativa está demostrando su potencial para acelerar la totalidad del proceso de desarrollo de software y, por tanto, posibilitar una entrega más rápida de productos y servicios de software de confianza. Preveemos que las empresas podrán automatizar por completo procesos como la generación de códigos, el mantenimiento y la resolución de errores de programación.

## Creación de vídeo y realidad virtual

La IA generativa puede crear entornos de videojuegos inmersivos, diseñar vídeos o incluso personalizar vídeos de productos para páginas web de comercio electrónico. En el futuro, las empresas podrán aprovecharla para crear

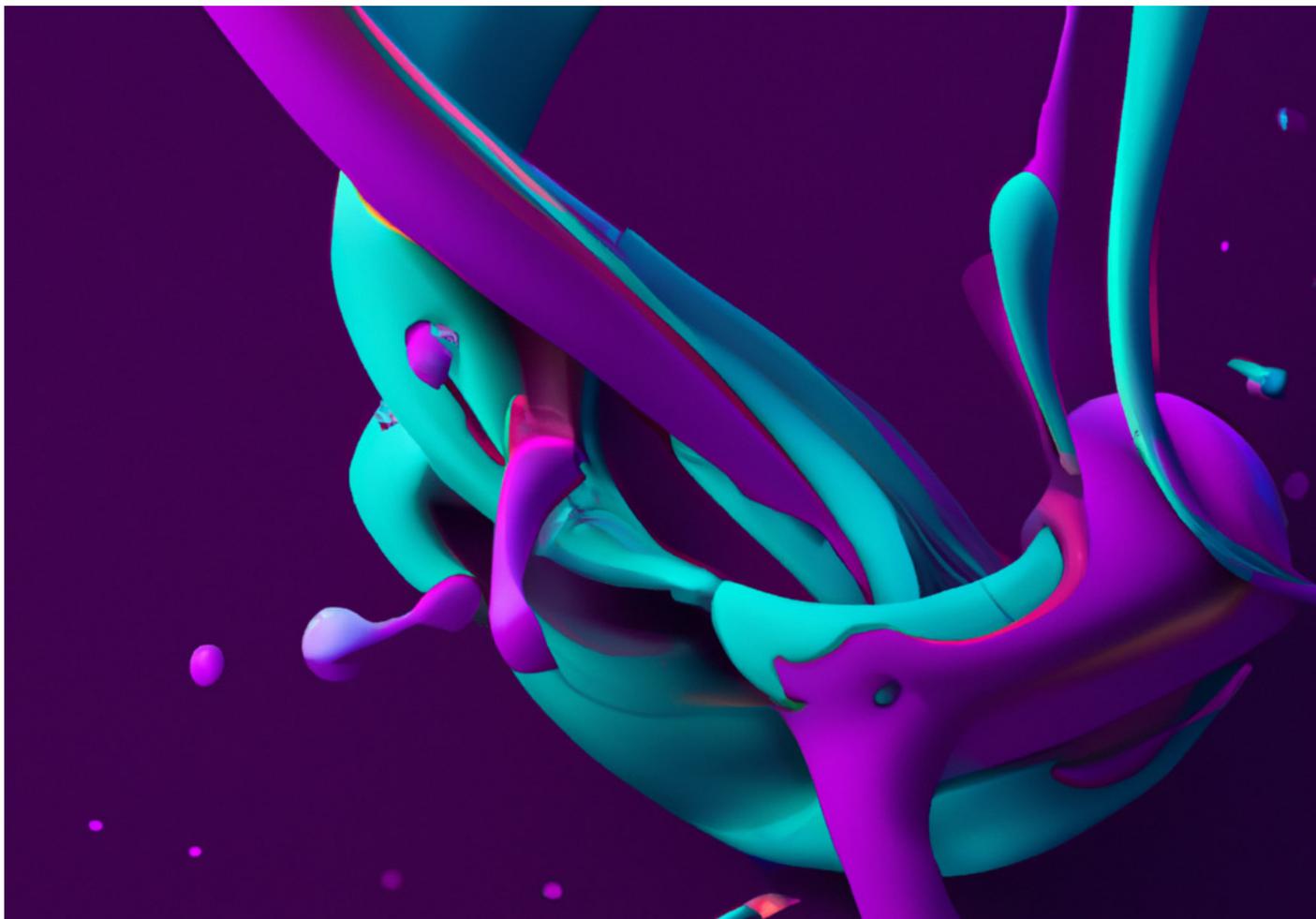
asistentes virtuales o aplicaciones de reproducción en directo, como la introducción automática de subtítulos en vídeos en directo. Muchas empresas de este ámbito ya están dirigiendo su atención hacia los clientes corporativos.

## El metaverso

Crear activos en 3D realistas en el metaverso es caro y requiere mucho tiempo. La IA generativa puede crear activos en 3D de texto a imagen o voz, escenas en 3D basadas en imágenes en 2D o incluso efectos de sonido. Asimismo, puede generar rostros humanos y brindar características más realistas a los avatares del metaverso.

## Mejora de la seguridad de la información

La IA generativa puede enseñar a las personas los riesgos clave que representan algunas vulnerabilidades, ayudándoles así a escribir códigos apropiados o comprender los métodos de ataque de los actores maliciosos.



Durante más de 150 años, las firmas de KPMG han desempeñado un papel fundamental a la hora de explorar y aprovechar las nuevas tecnologías, como la IA generativa, y han brindado seguridad y orientación a las empresas para implantarlas.

Global Lighthouse es la red mundial de KPMG de más de 15.000 especialistas en datos y análisis, IA y tecnologías emergentes, con oficinas en 37 países en América, Asia-Pacífico y Europa. Comprendemos que utilizar la IA de forma responsable constituye un reto complejo en materia empresarial, normativa y técnica. A través de Global Lighthouse y la red de firmas de KPMG, tenemos el firme compromiso de ayudar a los clientes a materializar una oferta de IA responsable.

El equipo de KPMG Forensic está formado por más de 100 profesionales en España y cuenta con una red internacional en más de 50 países. Compuesto por un equipo multidisciplinar integrado por profesionales de distintas especializaciones, posee una amplia experiencia en el asesoramiento a empresas en materia de riesgo, cumplimiento y en el objetivo de alcanzar altos niveles de integridad. En particular, el equipo de Smart Forensic, está especializado en ética y cumplimiento en el ámbito tecnológico.

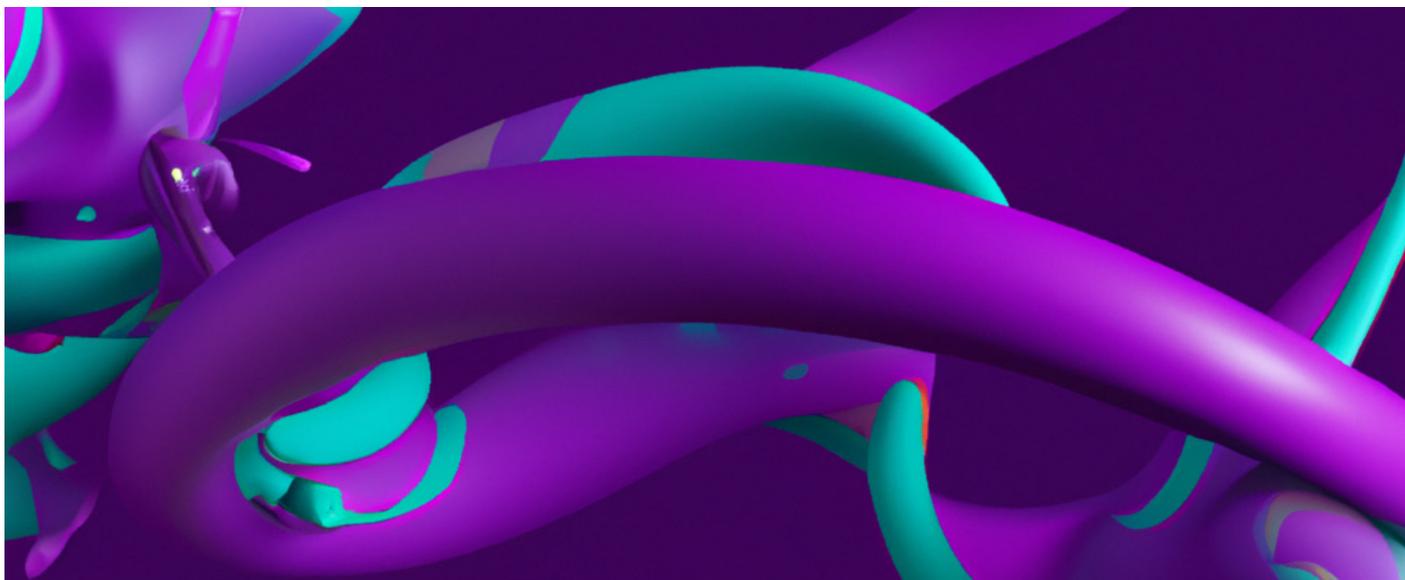
En España, la alianza entre los equipos de LOTS (Legal Operations and Transformation Services), Lighthouse (Data Analytics & IA) y Forensic (Ethics and Compliance) nos permite ofrecer a nuestros clientes un enfoque integral con el que abordar este reto desde una perspectiva legal, ética, tecnológica y de negocio.

## Usar la IA generativa de forma responsable

Desde KPMG ayudamos a las organizaciones a desarrollar soluciones de IA responsables, fiables y seguras adoptando asimismo un enfoque responsable a la hora de valorar los aspectos éticos, de gobierno corporativo y de seguridad en torno a las tecnologías de IA y aprendizaje automático de los clientes. El conjunto de marcos, controles, procesos y herramientas puede ayudar a nuestros clientes a aprovechar el poder de la IA — diseñando, desarrollando e implantando sistemas de IA de forma segura, fiable y ética — de modo que las empresas puedan acelerar la creación de valor para los consumidores, la organización y la sociedad.

Nuestro enfoque responsable de la IA incluye:

- 1. Equidad:** garantizar que los modelos son equitativos y libres de sesgos.
- 2. Capacidad de explicación:** garantizar que la IA puede comprenderse, documentarse y estar disponible para su revisión.
- 3. Rendición de cuentas:** garantizar que existen mecanismos para fomentar la responsabilidad en todo el ciclo de vida de la IA.
- 4. Integridad de los datos:** garantizar que los pasos de calidad, gobierno y alimentación de los datos integran confianza.
- 5. Fiabilidad:** garantizar que los sistemas de IA se ejecutan con el nivel de precisión y coherencia deseado.
- 6. Protección:** salvaguarda frente a cualquier acceso no autorizado, actos de corrupción o ataques.
- 7. Privacidad:** garantizar el cumplimiento de las normativas de privacidad de datos y uso de datos de consumidores.
- 8. Seguridad:** garantizar que la IA no afecta negativamente a las personas, los bienes o el medio ambiente.



# Contacts



**José Ángel Alonso**

Socio KPMG Lighthouse España

E: joseangelalonso@kpmg.es



**Noemí Brito**

Head de LOTS España  
– KPMG Abogados

E: noemibrito@kpmg.es



**María Barcenilla**

Head Smart Forensic España  
(Ethics & Compliance)

E: mbarcenilla@kpmg.es

[kpmg.com/socialmedia](https://kpmg.com/socialmedia)

La información aquí contenida es de carácter general y no va dirigida a facilitar los datos o circunstancias concretas de personas o entidades. Si bien procuramos que la información que ofrecemos sea exacta y actual, no podemos garantizar que siga siéndolo en el futuro o en el momento en que se tenga acceso a la misma. Por tal motivo, cualquier iniciativa que pueda tomarse utilizando tal información como referencia, debe ir precedida de una exhaustiva verificación de su realidad y exactitud, así como del pertinente asesoramiento profesional.

© 2023 KPMG, S.A., sociedad anónima española y firma miembro de la organización global de KPMG de firmas miembro independientes afiliadas a KPMG International Limited, sociedad inglesa limitada por garantía. Todos los derechos reservados.

KPMG y el logotipo de KPMG son marcas registradas de KPMG International Limited, sociedad inglesa limitada por garantía.