LLM: definición, contexto y regulación

"Me dijeron que tendría un impacto positivo en el mundo. Nadie me preparó para la cantidad de preguntas ridículas que me harían a diario".

Anthropic Claude[25]

Definición

La inteligencia artificial generativa (GenAI) es un tipo de IA capaz de generar diversos tipos de contenidos, como texto, imágenes, vídeos y audio. Utiliza modelos para aprender los patrones y la estructura de los datos de entrenamiento de entrada y, a continuación, genera nuevos contenidos basados en este conocimiento aprendido.

Dentro de la GenAI, los *Large Language Models* (LLM) son, según la Comisión Europea, "un tipo de modelo de inteligencia artificial que ha sido entrenado mediante algoritmos de aprendizaje profundo para reconocer, generar, traducir y/o resumir grandes cantidades de lenguaje humano escrito y datos textuales"[26].

Muy comúnmente, estos modelos utilizan arquitecturas conocidas como *transformers*, que les permiten entender contextos complejos y captar relaciones entre palabras distantes en el texto. Entrenados con vastos conjuntos de datos, como libros, artículos y páginas web, los LLM aprenden patrones lingüísticos y estructuras para ejecutar tareas variadas, incluyendo generación de texto, traducción y análisis de sentimiento.

La eficacia de un LLM depende de su tamaño, la diversidad de los datos de entrenamiento y la sofisticación de sus algoritmos, lo que influye directamente en su capacidad para aplicaciones prácticas en diversos campos. Por ello, entrenar un LLM es una tarea que requiere una capacidad muy elevada de computación y de tiempo de máquina, y por tanto costes muy significativos. Como referencia, según Sam Altman, entrenar GPT-4 costó "más de 100 millones de dólares[27].

Estos elevados costes hacen que el desarrollo de los mayores LLM esté concentrado en unas pocas organizaciones en el mundo (Fig. 4), con las capacidades tecnológicas, científicas y de inversión necesarias para abordar proyectos de esta envergadura.

Evolución de los LLM

El desarrollo de los LLM representa una evolución sustancial dentro del campo del procesamiento de lenguaje natural (NLP), y se remonta al trabajo fundacional sobre semántica[28] realizado por Michel Bréal en 1883. El advenimiento de los LLM comenzó a mediados del siglo XX, precedido por sistemas que dependían en gran medida de reglas gramaticales creadas manualmente. Un caso emblemático de este período es el programa "ELIZA", creado en 1966, que supuso un avance icónico en el desarrollo de modelos de lenguaje.

A medida que el campo evolucionó, las décadas de 1980 y 1990 presenciaron un cambio sustancial hacia métodos estadísticos de procesamiento de lenguaje. Este período vio la adopción de Modelos Ocultos de Markov (HMMs) y modelos n-gram, que ofrecieron un enfoque más dinámico para predecir secuencias de palabras basadas en probabilidades, en lugar de sistemas de reglas fijas.

El resurgimiento de las redes neuronales a principios de los años 2000, gracias a los avances en algoritmos de retropropagación que mejoraron el entrenamiento de redes multicapa, marcó un desarrollo crucial. Un hito fue la introducción de redes neuronales de alimentación directa para la modelización del lenguaje[29] (Bengio et al., 2003). Esto sentó las bases para innovaciones subsecuentes en la representación de palabras, especialmente la introducción de *embeddings* de palabras[30] (Mikolov et al., 2013) a través de Word2Vec. Los *embeddings* representan palabras como vectores de números y permiten

[25] Claude (lanzado en 2023) es un modelo de lenguaje entrenado por Anthropic, una *startup* de IA fundada por Dario Amodei, Daniela Amodei, Tom Brown, Chris Olah, Sam McCandlish, Jack Clarke y Jared Kaplan en 2021. Claude fue diseñado usando la técnica de "auto-aprendizaje alineado constitucionalmente" de Anthropic, que se basa en proporcionar al modelo de un listado de principios y reglas para aumentar su seguridad y evitar comportamientos dañinos.
[26] European Commission (2024).
[27] Wired (2023).
[28] Bréal (1883).
[29] Bengio (2003).
[30] Mikolov (2013).

definir distancias entre palabras, de manera que conceptos similares tengan distancias reducidas, y esto permite capturar relaciones semánticas con una efectividad sin precedentes.

Los primeros mecanismos de atención se introdujeron en 2016[31], y permitieron resultados sin precedentes en tareas de procesamiento del lenguaje, ya que identificaban la relevancia de diferentes partes del texto de entrada. Pero fue la introducción de la arquitectura *transformer*[32] (Vaswani et al., 2017) la que representó el verdadero cambio de paradigma en el entrenamiento de modelos y permitió la aparición de los LLM. El núcleo de la innovación de los *transformers* reside en los mecanismos de autoatención, que permiten a los modelos ponderar la importancia relativa de diferentes palabras en una oración. Esto significa que el modelo puede enfocarse en las partes más relevantes del texto al generar la respuesta, lo que es crucial para analizar el contexto y las relaciones complejas dentro de las secuencias de palabras. Además, al habilitar el procesamiento de datos de manera paralela, los *transformers* mejoran la eficiencia, la velocidad y el rendimiento del entrenamiento del modelo.

La serie de modelos GPT desarrollados por OpenAI, comenzando con GPT-1 en junio de 2018 y llegando a GPT-4 en marzo de 2023, ejemplifican los rápidos avances en las capacidades de los LLM. En particular, GPT-3, lanzado en 2020 con 175.000 millones de parámetros, llegó al gran público y mostró el extenso potencial de los LLM en diversas aplicaciones. Además de la serie GPT de OpenAI, otros modelos de LLM como Google Gemini y Anthropic Claude han surgido como actores importantes en el panorama de la IA. Gemini es un ejemplo de cómo las grandes empresas tecnológicas están invirtiendo en el desarrollo de LLM avanzados, mientras que Claude representa un esfuerzo por crear LLM que no solo sean potentes, sino también alineados con principios éticos y seguros para su uso.

El año 2023, llamado "el año de la IA"[33], destaca como un hito en la historia de los LLM, caracterizado por una mayor accesibilidad y contribuciones globales. Las innovaciones durante este año demostraron que los LLM pueden construirse con un mínimo de código, reduciendo significativamente las barreras de entrada, aunque a la vez introduciendo nuevos desafíos como el coste de entrenamiento y de inferencia, y sus riesgos inherentes. Este periodo también vio una preocupación creciente por las consideraciones éticas y los desafíos

[31] Parikh, A. P. (2016).
[32] Vaswani (2017).
[33] Euronews (2023).
[34] Adaptado de MindsDB (2024) y expandido.

Fig. 4. Algunos de los principales LLM y sus proveedores[34].

Empresa	LLM	Comentarios	País
OpenAI	**ChatGPT**	Conocido por su versatilidad en tareas lingüísticas, suele utilizarse para completar textos, traducir y mucho más.	Estados Unidos
Microsoft	**Orca**	Se centra en la creación de datos sintéticos y la mejora de las capacidades de razonamiento.	Estados Unidos
Anthropic	**Claude**	Reconocido por sus amplios conocimientos generales y su capacidad multilingüe.	Estados Unidos
Google	**Gemini, Gemma, BERT**	Pionero en el tratamiento del lenguaje con modelos que admiten múltiples tipos de datos.	Estados Unidos
Meta AI	**Llama**	Conocido por su eficacia y acceso democratizado, se centra en el alto rendimiento con un menor coste computacional.	Estados Unidos
LMSYS	**Vicuna**	Perfeccionado para las funcionalidades de chatbot, ofrece un tratamiento único de las interacciones conversacionales.	Estados Unidos
Cohere	**Command-nightly**	Especializado en tiempos de respuesta rápidos y búsqueda semántica en más de 100 idiomas.	Canadá
Mistral AI	**Mistral, Mixtral**	Hace hincapié en modelos más pequeños pero potentes, que operan localmente con sólidas métricas de rendimiento.	Francia
Clibrain	**LINCE**	Adaptado a la lengua española, centrado en los matices lingüísticos y la calidad de la comprensión.	España
Technology Innovation Institute	**Falcon**	Proporciona modelos de IA de código abierto altamente eficientes y escalables con soporte multilingüe.	Emiratos Árabes Unidos
Aleph Alpha	**Luminous**	Destaca por su enfoque multimodal y su rendimiento competitivo en tareas básicas de IA.	Alemania
SenseTime	**SenseNova**	Una serie de modelos y aplicaciones de IA generativa que hacen uso de la plataforma de investigación y desarrollo AGI e integran LLM con sistemas informáticos a gran escala (SenseCore, con 5000 petaflops).	Hong Kong

planteados por el desarrollo y el uso de los LLM y, como consecuencia, un avance en la regulación de la IA y la IA generativa en todo el mundo.

La proliferación de los LLM de código abierto ha marcado un hito en la democratización de la tecnología de IA. Comenzando por Llama y siguiendo con Vicuna, Falcon, Mistral o Gemma, entre otros, los LLM *open-source* han democratizado el acceso a la tecnología puntera en el procesamiento del lenguaje y han permitido a investigadores, desarrolladores y aficionados experimentar, personalizar y desplegar soluciones de IA con una inversión inicial mínima. La disponibilidad de estos modelos ha fomentado una colaboración sin precedentes en la comunidad de IA, estimulando la innovación y facilitando la creación de aplicaciones avanzadas en una variedad de sectores.

Por último, la integración de LLM en herramientas ofimáticas y de desarrollo de software está transformando la eficiencia y la capacidad de las empresas. Microsoft ha integrado los LLM en su suite de Office bajo el nombre de Microsoft 365 Copilot, mientras que Google lo ha hecho en Google Workspace. Al mismo tiempo, herramientas como GitHub Copilot o StarCoder utilizan LLM para asistir a los programadores, acelerando la generación de código y mejorando la calidad del desarrollo de software.

Tipologías de LLM

Los LLM han progresado más allá de la simple predicción de texto y se han convertido en sofisticadas aplicaciones en diversos dominios, arquitecturas y modalidades. Esta sección presenta una categorización de los LLM según varios criterios.

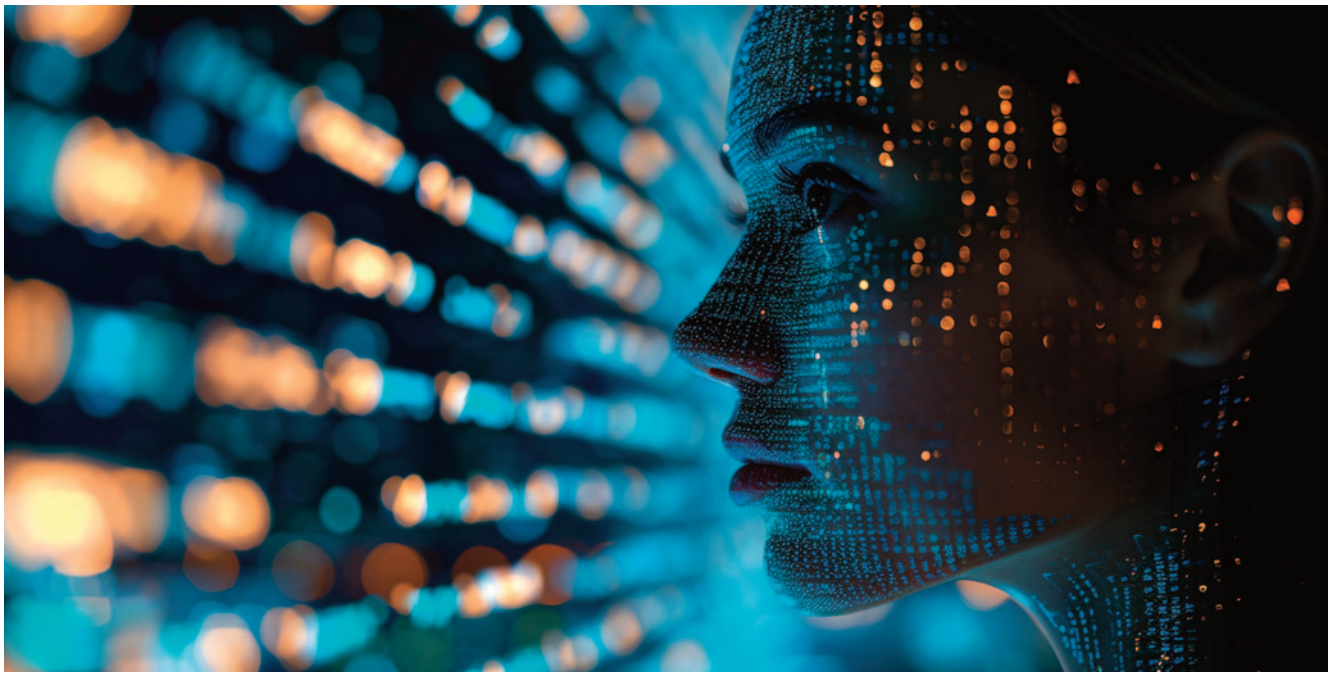
Por arquitectura

- **LLM basados en redes neuronales recurrentes (RNN):** estos modelos procesan el texto secuencialmente, analizando el impacto de cada palabra en la siguiente, y utilizan arquitecturas recurrentes, como memoria a largo plazo (LSTM) o unidades recurrentes de compuerta (GRU), para procesar datos secuenciales. Aunque no son tan potentes como los *transformers* para secuencias largas, los RNN son útiles para tareas donde entender el orden de las palabras es crucial, como en la traducción automática. Son ejemplos ELMo *(Embeddings from Language Models)* y ULMFiT *(Universal Language Model Fine-tuning)*.

- **LLM basados en *transformers:*** es la arquitectura dominante para los LLM hoy en día. Utilizan *transformers* para analizar las relaciones entre las palabras en una oración. Esto les permite capturar estructuras gramaticales complejas y dependencias entre palabras a gran distancia. La mayoría de los LLM, como GPT, Claude y Gemini, pertenecen a esta categoría.

Por componentes

- **Codificadores *(Encoders):*** son modelos diseñados para comprender (codificar) la información de entrada. Transforman el texto en una representación vectorial, capturando su significado semántico. Los *encoders* son fundamentales en tareas como la comprensión de texto y la



clasificación. Un ejemplo es BERT, de Google, un modelo que analiza el contexto de cada palabra en un texto para entender su significado completo, y que no es realmente un LLM.

- **Decodificadores *(Decoders):*** estos modelos generan (decodifican) texto a partir de representaciones vectoriales. Son esenciales en la generación de texto, como en la creación de contenido nuevo a partir de *prompts* dados. La mayor parte de los LLM son *decoders*.

- **Codificadores/Decodificadores *(Encoders/Decoders):*** estos modelos combinan *encoders* y *decoders* para convertir un tipo de información en otro, facilitando tareas como la traducción automática, donde el texto de entrada se codifica y luego se decodifica en otro idioma. Un ejemplo es T5 (*Text-to-Text Transfer Transformer*) de Google, diseñado para abordar múltiples tareas de procesamiento de lenguaje natural.

Por enfoque de entrenamiento

- **LLM preentrenados:** estos modelos se entrenan primero en un gran corpus de texto sin etiquetar utilizando técnicas de aprendizaje autosupervisado como modelado de lenguaje enmascarado o predicción de la siguiente oración, y después se pueden ajustar con datos etiquetados más pequeños para tareas específicas. Ejemplos de este tipo de LLM incluyen modelos como GPT, Mistral, BERT y RoBERTa, entre muchos otros.

- **LLM específicos:** estos modelos se entrenan desde cero con datos etiquetados para una tarea particular, como análisis de sentimiento, resumen de textos o traducción automática. Ejemplos de este tipo de LLM incluyen modelos de traducción y resumen.

Por modalidad

- **LLM de solo texto:** son el tipo más común, entrenados y trabajando exclusivamente con datos textuales. Son ejemplos GPT-3, Mistral o Gemma.

- **LLM multimodales:** es un campo emergente donde los LLM son entrenados en una combinación de texto y otros formatos de datos como imágenes o audio. Esto les permite realizar tareas que requieren entender la relación entre diferentes modalidades. Son ejemplos GPT-4, Claude 3 y Gemini.

Por tamaño

- ***Large language models* (LLM):** son modelos que utilizan cantidades masivas de parámetros. Son muy potentes, pero requieren una infraestructura tecnológica en la nube, relativamente costosa, para su ejecución. Son ejemplos GPT-4, Gemini o Claude 3.

- ***Small language models* (SLM):** una tendencia reciente, los SLM son versiones más pequeñas y eficientes de los LLM, diseñados para funcionar en dispositivos con recursos limitados, como *smartphones* o dispositivos IoT, sin necesidad de conexión o despliegue en la nube. A pesar de su tamaño reducido, estos modelos mantienen un rendimiento aceptable gracias a técnicas como la compresión de modelos o la *cuantización*, que reduce la precisión de los pesos y las activaciones del modelo. Son ejemplos Gemini Nano de Google, o la familia de modelos Phi de Microsoft.

LLM en la práctica: casos de uso en producción

A pesar del creciente interés y la exploración de posibles aplicaciones de los LLM en las organizaciones, los casos de uso realmente implementados en producción son aún limitados. La mayoría de las empresas se encuentran en etapas relativamente tempranas, identificando y priorizando potenciales casos de uso.

No obstante, varias compañías ya han logrado poner en producción algunos casos de LLM, demostrando su valor tangible para el negocio y sus clientes. Aquí se resumen algunos de estos casos:

- ***Chatbots* internos:** bastantes organizaciones han implementado *chatbots* basados en LLM para facilitar el acceso de sus empleados a políticas, procedimientos e información relevante de la compañía. Estos asistentes conversacionales permiten obtener respuestas rápidas y precisas a consultas frecuentes, mejorando la eficiencia y reduciendo la carga sobre otros canales de soporte interno.
- **Extracción de información:** los LLM están siendo utilizados para extraer automáticamente datos clave de documentos extensos y complejos, como memorias anuales o informes de riesgo climático. Estas herramientas son capaces de procesar archivos PDF de miles de páginas, con estructuras heterogéneas que incluyen imágenes, gráficos y tablas, y transformar la información relevante en formatos estructurados y accesibles, como tablas ordenadas. Esta automatización permite a las empresas ahorrar tiempo y recursos en tareas de análisis documental.
- **Asistencia en centros de atención al cliente:** algunos *contact centers* están aprovechando los LLM para mejorar la calidad y eficiencia del servicio. Aplicando técnicas de transcripción y resumen, estas herramientas generan un contexto de las interacciones previas de cada cliente, permitiendo a los agentes ofrecer una atención más personalizada. Además, durante las llamadas en curso, los LLM pueden proporcionar a los agentes acceso en tiempo real a documentación relevante para responder las consultas específicas de los clientes, como información sobre comisiones bancarias o instrucciones para bloquear tarjetas de crédito.
- **Clasificación inteligente de documentos:** las capacidades de procesamiento de lenguaje natural de los LLM están siendo aplicadas para clasificar automáticamente grandes volúmenes de documentos, como contratos o facturas, partiendo de su contenido. Esta categorización inteligente permite a las organizaciones agilizar procesos de gestión documental y facilita la búsqueda y recuperación de información relevante.
- **Banca conversacional:** algunos bancos están integrando LLM en sus aplicaciones móviles y canales digitales para ofrecer experiencias conversacionales avanzadas a sus clientes. Estos chatbots son capaces de acceder a los datos transaccionales de los usuarios en tiempo real y responder a consultas específicas, como «¿Cómo han sido mis gastos en el último mes?“ o «¿Cuánto he ganado en intereses por mis depósitos en el último año?“.
- **Asistencia en la redacción de informes de auditoría:** las funciones de Auditoría Interna de algunas compañías ya están utilizando LLM para agilizar la elaboración de sus informes. Estas herramientas toman como *inputs* los hallazgos del auditor, una base de datos con informes previos y otra con la normativa aplicable, tanto interna como externa. A partir de esta información, los LLM generan un borrador avanzado del informe de auditoría, adoptando el tono, vocabulario y estilo de los auditores, y citando adecuadamente informes anteriores y regulaciones relevantes. Esto permite a los auditores ahorrar tiempo significativo en tareas de redacción y centrarse en actividades de mayor valor añadido.

Estos ejemplos ilustran cómo los LLM están creando valor real en diversas funciones empresariales, desde la optimización de procesos internos hasta la mejora de la experiencia del cliente. Si bien actualmente el número de casos de uso en producción es limitado, se espera que esta tendencia se acelere muy rápidamente en el futuro próximo, a medida que los LLM sigan evolucionando y se aborden de manera efectiva los desafíos relacionados con la privacidad y la seguridad de los datos.

Principales usos

Los LLM están encontrando aplicaciones en una multitud de dominios, transformando sustancialmente la forma en que las personas interactúan con la tecnología y aprovechado el procesamiento de lenguaje natural para mejorar procesos, servicios y experiencias.

A continuación, se resumen algunos de los usos más destacados de los LLM de texto.

1. Creación y mejora de contenido

- Generación de contenido: producción automática de texto.
- Asistencia de escritura: corrección ortotipográfica, de estilo y de contenido.
- Traducción automática: conversión de texto de un idioma a otro.
- Resumen de textos: reducción de documentos extensos a resúmenes.
- Planificación y guion de contenidos: estructuración de contenidos, p. ej., índices.
- Brainstorming: propuestas creativas para proyectos, nombres, conceptos, etc.
- Programación: creación de código de programación a partir de lenguaje natural.

2. Análisis y organización de información

- Análisis de sentimientos: evaluación de emociones y opiniones en textos.
- Extracción de información: extracción de datos específicos de documentos extensos.
- Clasificación de textos: organización de textos en categorías o temas específicos.
- Revisión técnica: asistencia en revisar documentos especializados (p. ej., legales).

3. Interacción y automatización

- Chatbots: simulación de conversaciones sobre temas generales o específicos.
- Q&A: generación de respuestas a preguntas basadas en un corpus.

Estos usos resumen las aplicaciones actuales de los LLM de texto. Con la emergencia de los LLM multimodales, comienzan a aflorar aplicaciones adicionales como la generación de contenido audiovisual, la interpretación de datos a partir de imágenes, la traducción de contenido multimedia o la creación de experiencias interactivas enriquecidas, como la interacción con *chatbots* con entradas no solo de texto, sino también de imagen, audio y vídeo.

Requisitos regulatorios

La rápida evolución de la inteligencia artificial generativa, especialmente en el campo de los modelos de lenguaje de gran escala (LLM), ha captado la atención de reguladores a nivel global. El potencial de estos sistemas para influir de forma negativa en los ciudadanos ha llevado a un incremento en las iniciativas para establecer marcos regulatorios que aseguren su desarrollo y uso responsable.

Algunas de las principales iniciativas regulatorias sobre IA son:

- **El AI Act de la Unión Europea:** propuesta legislativa pionera para regular la IA, que clasifica los sistemas de IA según su nivel de riesgo y establece requisitos de transparencia, seguridad y derechos fundamentales. El AI Act fue aprobado por el Parlamento Europeo el 13 de marzo de 2024.

- **El AI Bill of Rights de Estados Unidos:** documento orientativo que busca proteger los derechos civiles en el desarrollo y aplicación de la IA, enfatizando la privacidad, la no discriminación y la transparencia.

- **La guía sobre IA del NIST[35] de Estados Unidos:** establece principios para la creación de sistemas de IA fiables, con enfoque en la precisión, la explicabilidad y la mitigación de sesgos.

[35] El Instituto Nacional de Estándares y Tecnología (NIST) ha publicado documentos que detallan marcos de ciberseguridad, de gestión de riesgos y, concretamente, de gestión de modelos de IA y de IA generativa.

- **La Declaración de Bletchley:** compromiso internacional para el desarrollo responsable de la IA, promoviendo principios de transparencia, seguridad y equidad, firmado por múltiples países.

Además de las iniciativas mencionadas, numerosos países han comenzado a emitir sus propias regulaciones locales o han establecido principios para la adopción de la IA de manera ética y segura. Entre ellos se cuentan[36] Reino Unido, Francia, España, Alemania, Países Bajos, Polonia, Australia, Nueva Zelanda, Singapur, Canadá, Japón, Corea del Sur, China, India, Indonesia, Israel, Emiratos Árabes Unidos, Arabia Saudí, Egipto, Brasil, Chile, Perú, Argentina, México, Colombia y Turquía, entre otros.

Todas estas iniciativas regulatorias plantean requisitos muy similares sobre la IA que, aplicados a los LLM, se pueden resumir en:

- **Transparencia y explicabilidad:** obligación de revelar cómo funciona el LLM, incluyendo la lógica detrás de sus salidas para que sean comprensibles para los usuarios.

- **Privacidad y protección de datos:** medidas estrictas para proteger la información personal recopilada o generada por LLM, cumpliendo con leyes de protección de datos, como GDPR en Europa.

- **Equidad y no discriminación:** requisitos para prevenir sesgos y asegurar que los LLM no perpetúen discriminaciones ni prejuicios, mediante la evaluación y corrección constantes de sus algoritmos.

- **Seguridad y fiabilidad:** exigencias de robustez operacional para prevenir disfunciones o manipulaciones que puedan causar daño o pérdida de información.

- **Responsabilidad y gobernanza:** marco de responsabilidad de desarrolladores y usuarios de LLM en caso de daños o violaciones de derechos, incluyendo mecanismos de supervisión y control.

- **Supervisión humana:** la necesidad de mantener una supervisión humana efectiva sobre los LLM, asegurando que las decisiones importantes puedan ser revisadas y, si es necesario, corregidas o revertidas por humanos.

Estos requisitos reflejan un consenso emergente sobre los principios fundamentales para el desarrollo ético y seguro de los LLM, y forman la base para futuras regulaciones específicas y adaptaciones según evolucione la tecnología.

[37] IAPP (2024).