Newsletter trimestral

CÁTEDRA

# iDANAE

INTELIGENCIA · DATOS · ANÁLISIS · ESTRATEGIA

## 2T23

# Large Language Models: una nueva era en la inteligencia artificial

POLITÉCNICA

UNIVERSIDAD POLITÉCNICA DE MADRID

Management Solutions
Making things happen

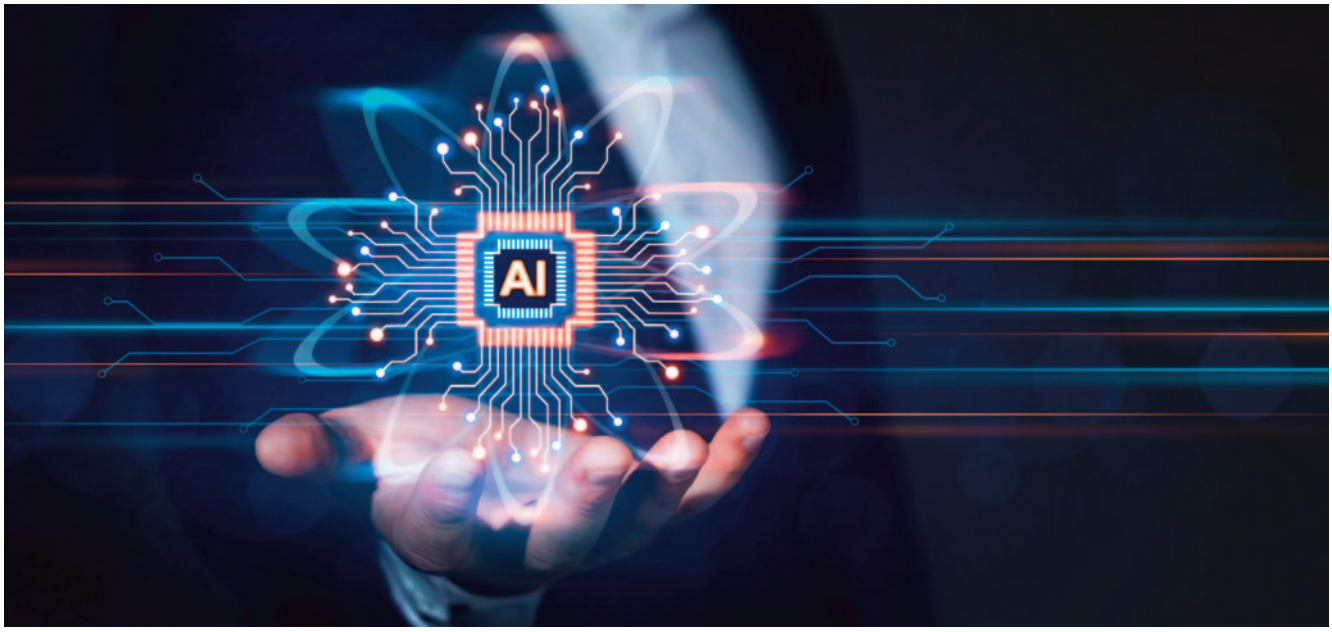
# Introducción

La rápida aparición de modelos grandes de lenguaje (LLM por sus siglas en inglés) ha captado la atención de profesionales de diversos sectores. Dotados de capacidades excepcionales para comprender, generar y procesar texto, estos modelos han demostrado ser activos valiosos para numerosas tareas asociadas al lenguaje, como la traducción automática, el análisis de sentimiento y la creación automática de contenidos, entre otras.

A medida que los modelos grandes de lenguaje siguen evolucionando, sus aplicaciones se vuelven cada vez más sofisticadas. Estos modelos agilizan los procesos y automatizan tareas que antes requerían de intervención humana, sobre todo en campos como la atención al cliente, el periodismo o la creación de contenidos. Por tanto, esta tecnología emergente abre numerosas oportunidades de negocio, impulsando la reducción de costes y optimizando la generación de ingresos.

Sin embargo, esta tecnología no está exenta de retos y riesgos. Aunque mejora sistemáticamente la eficiencia y reduce los gastos operativos, también suscita inquietudes en relación con la posible reducción de puestos de trabajo, protección de datos, sesgos, problemas medioambientales o el posible uso indebido de información confidencial, entre muchos otros.

El objetivo de esta newsletter es proporcionar una comprensión global de los LLM examinando sus fundamentos técnicos y sus avances recientes. Además, se analizan los retos que pueden plantear, tanto desde el punto de vista técnico como ético. Por último, se resumen algunas reflexiones finales sobre el estado actual y las perspectivas de los LLM en el panorama profesional.



# Definición y actualizaciones del sector

Un modelo grande de lenguaje (LLM) es un modelo de aprendizaje automático entrenado con grandes cantidades de datos de texto para comprender y generar un lenguaje natural similar al humano. Mediante técnicas de aprendizaje profundo, estos modelos se entrenan para captar patrones lingüísticos complejos. Tras entrenar grandes cantidades de parámetros, un LLM está capacitado para realizar tareas como responder preguntas, resumir textos, parafrasear, traducir o revisar la ortografía y la gramática.

Los orígenes de los LLM se remontan a la introducción de los transformadores, que surgieron como una arquitectura de red neuronal innovadora. Un transformador es un tipo de arquitectura de red neuronal especialmente adecuado para manejar datos secuenciales, como el texto. Este importante avance se presentó en el artículo de investigación titulado "Attention Is All You Need" [1] escrito por un equipo de investigadores de Google. Posteriormente, el campo de los LLM ha experimentado un crecimiento y un avance exponenciales. A continuación, se presentan algunos de los modelos más avanzados en el campo de los LLM que ejemplifican los últimos progresos y avances en investigación y tecnología:

- **GPT-3.5 de OpenAI [2].** Los investigadores desarrollaron e implementaron el concepto de pre-entrenamiento generativo [3]. A pesar de la abundancia de grandes corpus de texto sin etiquetar, los datos etiquetados para el aprendizaje de tareas específicas de procesamiento del lenguaje natural (PLN) son escasos, lo que dificulta el rendimiento adecuado de los modelos entrenados. Se pueden conseguir mejoras significativas en estas tareas mediante el preentrenamiento generativo de un modelo lingüístico en un corpus diverso de texto sin etiquetar, seguido de un ajuste discriminativo en cada tarea específica.

- **GPT-4 de OpenAI [4].** GPT-4 es un extenso modelo multimodal que procesa tanto imágenes como texto para producir resultados de texto. No se ha revelado información precisa sobre la arquitectura y el entrenamiento del modelo por motivos de competencia y seguridad. GPT-4 supera a sus predecesores en las pruebas de referencia estándar, lo que demuestra avances considerables en la comprensión de las intenciones del usuario y la garantía de las funciones de seguridad. Además, el modelo alcanza un rendimiento equivalente al humano en numerosas pruebas, como la puntuación dentro del 10% superior en un examen simulado del Colegio de Abogados de Estados Unidos.

- **LLaMA de Meta [5].** LLaMA (Large Language Model Meta AI) es una colección de modelos lingüísticos con 7B (mil millones) a 65B parámetros. LLaMA 33B y 65B se entrenaron con 1,4B de tokens (es decir, una unidad de texto que el modelo procesa, puede ser un solo carácter de una palabra), mientras que el modelo más pequeño, LLaMA 7B, se entrenó con un billón de tokens. Utilizaron exclusivamente conjuntos de datos disponibles públicamente, sin depender de datos patentados o restringidos. El equipo de Meta AI afirma que los modelos más pequeños entrenados con más tokens son más fáciles de reentrenar y ajustar para aplicaciones de productos específicos. LLaMA-13B superó a GPT-3 siendo 10 veces más pequeño.

- **Bloom [6]** es un modelo de 176B parámetros desarrollado en colaboración por más de 1.000 investigadores de IA de varias empresas, entre ellas Hugging Face, Microsoft y NVIDIA. Dado que la mayoría de los LLM son creados por organizaciones con muchos recursos y a menudo permanecen inaccesibles para el público, los autores se propusieron democratizar esta potente tecnología publicando Bloom como modelo de código abierto.

- **PaLM de Google [7].** Pathways Language Model (PaLM) es un modelo de lenguaje basado en transformadores con 540B parámetros. PaLM fue entrenado utilizando Pathways, un nuevo sistema de ML para un entrenamiento eficiente. El modelo demuestra las ventajas de la escalabilidad en el aprendizaje de pocos disparos (few-shot learning), logrando resultados de vanguardia en cientos de pruebas de referencia de comprensión y generación de lenguaje. PaLM supera a los modelos más avanzados en tareas de razonamiento de varios pasos y supera el rendimiento humano medio en la prueba de referencia BIG-bench. Recientemente, Google ha desarrollado una versión mejorada de su modelo, PaLM 2.

- **LaMDA de Google [8].** Language Models for Dialog Applications (LaMDA) es un conjunto de modelos basados en transformadores con hasta 137B parametros y preentrenado sobre 1.56T palabras de datos públicos. LaMDA es capaz de generar texto a partir de texto, images, o datos estructurados como datos de entrada. El chatbot de Google, Bard, está basado en este modelo.

- **Claude de Anthropic [9].** Se trata de un LLM preentrenado de 52B parámetros que utiliza un modelo de aprendizaje por refuerzo, en lugar de humanos, para generar la clasificación inicial de modelos ajustados. Este LLM es el elegido por AWS como modelo base en su servicio Bedrock.

## Revisión técnica

El rendimiento de un LLM puede variar en función de su arquitectura o tipo, y cada uno de ellos está diseñado para sobresalir en determinadas tareas. Por ejemplo, los modelos codificadores (denominados encoder, un tipo de red neuronal que transforma la información de una representación a otra con el fin de comprimirla) son más competentes en la clasificación de frases, el reconocimiento de entidades con nombre y la respuesta extractiva a preguntas. Por el contrario, los modelos decodificadores (*decoder*, red que ejecuta la tarea inversa del *encoder*) demuestran un rendimiento superior en la generación de textos. Los modelos codificador-decodificador, por su parte, están óptimamente diseñados para tareas como el resumen, la traducción y la respuesta generativa a preguntas.

Los transformadores están el núcleo de los LLM dentro del estado del arte. Los transformadores se basan en tres componentes clave:

- **Mecanismo de autoatención:** permite al modelo sopesar la importancia de las distintas partes de la secuencia de entrada para una posición de salida específica. En otras palabras, ayuda al modelo a centrarse en los tokens de entrada relevantes a la hora de generar tokens de salida. Para ello se utilizan puntuaciones de atención calculadas a partir de las incrustaciones de los tokens de entrada, que luego se utilizan para calcular una suma ponderada de las representaciones de entrada. El mecanismo de autoatención puede aplicarse varias veces en paralelo, lo que da lugar a una atención *multi-head*.

- **Codificación posicional:** la arquitectura de transformador no tiene un procesamiento secuencial inherente, como las redes neuronales recurrentes (RNN); en su lugar, incorpora la información de posición de los tokens en la secuencia de entrada de una forma novedosa. La codificación posicional se añade a las incrustaciones (embedding) de entrada para proporcionar esta información. Los transformadores utilizan un esquema de codificación posicional inteligente, en el que cada posición/índice se asigna a un vector. Por lo tanto, la salida de la capa de codificación posicional es una matriz, donde cada fila de la matriz representa un objeto codificado de la secuencia sumado con su información posicional. Esto puede hacerse utilizando funciones sinusoidales o incrustaciones que se aprenden, lo que permite al modelo comprender las posiciones relativas de los tokens en la secuencia.

- **Normalización de capas:** los transformadores están formados por múltiples capas de redes neuronales de autoatención y feed-forward. La normalización de capas se utiliza para estabilizar el proceso de entrenamiento, normalizando las activaciones de cada capa antes de pasarlas a la siguiente. Esto ayuda a mitigar el problema del desvanecimiento o explosión del gradiente, que puede producirse durante el entrenamiento de redes neuronales profundas.

El rendimiento de los LLM se suele medir utilizando diversas tareas de referencia y conjuntos de datos que evalúan la capacidad del modelo para comprender y generar lenguaje natural. El rendimiento en estas tareas se mide utilizando diferentes métricas de evaluación que cuantifican la precisión, fluidez y relevancia de los modelos.

Algunas métricas de evaluación habituales son

- Perplejidad: medida de la precisión con la que un modelo lingüístico predice una secuencia determinada de palabras.

- BLEU (Bilingual Evaluation Understudy): métrica utilizada habitualmente para evaluar tareas de traducción automática. Mide la similitud entre la traducción generada por el modelo y la traducción humana de referencia.

- Precisión, recall y puntuación F1: se utilizan para evaluar tareas como el reconocimiento de entidades con nombre y la respuesta a preguntas. Una precisión baja indica un elevado número de falsos positivos, y un recall bajo indica un elevado número de falsos negativos. La puntuación F1 refleja el equilibrio entre la precisión y el recall.

- ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [10]: conjunto de métricas utilizadas para evaluar el resumen automático de textos. Mide el solapamiento entre el resumen generado y el resumen humano de referencia en términos de n-grams, secuencias de palabras u otras unidades lingüísticas.

- Tasa de error en carácter/palabra/frase (CER, WER, SER): estas métricas también pueden añadirse al análisis del rendimiento en términos de tasa de error en carácter, palabra o frase (aunque se trata de métricas utilizadas habitualmente en tareas de transcripción automática).

Existen varios conjuntos de datos de referencia y tablas de clasificación que facilitan la comparación de los LLM en función de su rendimiento en diversas tareas:

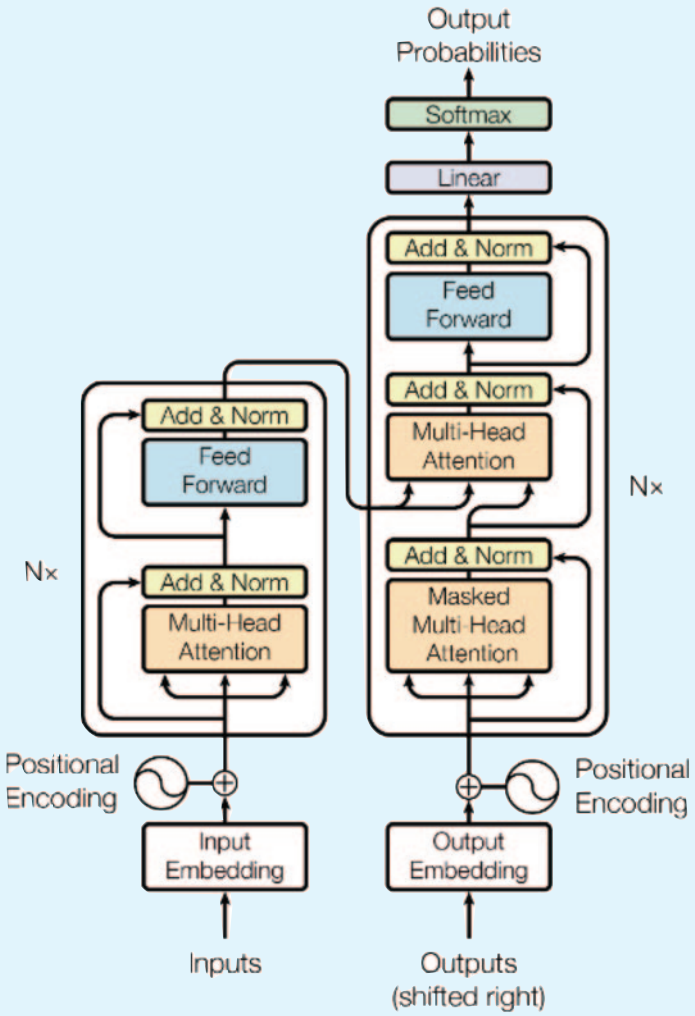
- GLUE [11] (General Language Understanding Evaluation): evaluación comparativa de nueve tareas de procesamiento de lenguaje natural, incluidos el análisis de sentimiento, la inferencia del lenguaje natural y la detección de paráfrasis, cuyo objetivo es evaluar la comprensión general de los modelos lingüísticos.

- SuperGLUE [12]: se trata de una ampliación de la prueba GLUE que introduce tareas más complejas y conjuntos de datos más variados para evaluar mejor el rendimiento de los modelos.

- SQuAD [13] (Stanford Question Answering Dataset): referencia para evaluar la comprensión lectora y las capacidades de respuesta a preguntas de los modelos lingüísticos. Consiste en un conjunto de datos que contiene preguntas y respuestas basadas en artículos de Wikipedia.

- LAMBADA (Language Modeling Broadened to Account for Discourse Aspects) [14]: tarea de modelización lingüística diseñada para evaluar la capacidad de los modelos para predecir la palabra final de una frase, lo que requiere la comprensión de dependencias de largo alcance y del contexto.
- BIG-bench [15]: este conjunto de datos consta actualmente de 204 tareas, aportadas por 444 autores de 132 instituciones. Los temas de las tareas son variados y abarcan problemas de lingüística, matemáticas, razonamiento de sentido común, biología, física, prejuicios sociales, desarrollo de software y otros campos. BIG-bench se centra en tareas que se considera que superan las capacidades de los modelos lingüísticos actuales.

*Figura 1. Arquitectura original del transformador [1]*



# Aplicaciones prácticas

Los LLM han demostrado una notable capacidad para procesar y generar textos similares a los humanos en diversos ámbitos. Estos modelos de IA versátiles pueden realizar una amplia gama de tareas, transformando la forma en que el ser humano interactúa con los datos textuales. La siguiente lista ofrece una visión general de algunas de las tareas clave que pueden realizar los LLM, destacando su potencial para transformar numerosas aplicaciones tanto en el ámbito empresarial como en el de la investigación:

- **Generación de textos:** creación de un texto coherente y contextualmente pertinente a partir de una indicación o entrada dada.
- **Análisis del sentimiento:** identificación del sentimiento o emoción (por ejemplo, positivo, negativo, neutro) expresado en un texto.
- **Reconocimiento de entidades con nombre:** identificación y clasificación de entidades como nombres, organizaciones, lugares y fechas dentro de un texto.
- **Resumen de textos:** generación de un resumen conciso de un documento o fragmento de texto más largo conservando su significado esencial.
- **Respuesta a preguntas:** proporcionar respuestas precisas y pertinentes a las consultas de los usuarios a partir de un contexto o una fuente de conocimiento determinados.
- **Traducción automática:** conversión de texto de un idioma a otro conservando su significado y contexto.
- **Clasificación de textos:** clasificación de un texto en clases o temas predefinidos en función de su contenido.
- **Extracción de palabras clave:** identificación y extracción de palabras clave y frases significativas de un texto dado.
- **Parafraseo:** reescribir un texto con otras palabras, pero conservando su significado original.
- **Conversadores Chatbot:** conversaciones naturales y contextualizadas con los usuarios.
- **Búsqueda semántica:** mejora de los resultados de búsqueda mediante la comprensión del significado y la intención de una consulta, en lugar de basarse únicamente en la coincidencia de palabras clave.

- **Filtrado de contenidos:** identificación y filtrado de contenidos inapropiados, ofensivos o irrelevantes.

- **Generación de textos en lenguaje natural a partir de datos estructurados:** creación de informes escritos a partir de datos comerciales o financieros.

- **Tareas de codificación:** generación de código, traducción entre distintos lenguajes de programación, documentación y revisión del código.

A medida que los LLM avanzan y se vuelven más sofisticados, sus aplicaciones prácticas se extienden a diversos sectores, agilizando procesos y automatizando tareas que antes requerían la intervención humana. La siguiente lista destaca una serie de casos de uso empresarial en los que los LLM pueden emplearse eficazmente, demostrando su potencial para mejorar la eficiencia, ahorrar recursos e impulsar el crecimiento en diversos sectores:

**1. Automatización del servicio de atención al cliente**

- Chatbots: los chatbots inteligentes podrían entender las consultas de los usuarios y ofrecer respuestas precisas, mejorando la calidad y la eficiencia del servicio al cliente.

- Asistentes virtuales: los asistentes virtuales potenciados por IA pueden encargarse de tareas rutinarias como concertar citas, responder a preguntas frecuentes y ofrecer asistencia general, liberando a los agentes humanos para que se ocupen de tareas más complejas.

- Gestión de tickets del servicio de asistencia al cliente: analizar los tickets entrantes, categorizarlos en función de la urgencia y el tema, y dirigirlos al equipo o agente adecuado para una resolución más rápida.

**2. Creación de contenidos**

- Material de marketing: generación de textos persuasivos y atractivos para campañas publicitarias, boletines electrónicos y material promocional, ayudando a las empresas a llegar a su público objetivo con mayor eficacia.

- Contenidos para redes sociales: creación de publicaciones atractivas en las redes sociales que lleguen al público objetivo de una empresa y aumenten la notoriedad de la marca y la participación.

- Descripciones de productos: elaboración de descripciones de productos precisas y persuasivas que muestren las ventajas y características de los productos, mejorando las tasas de conversión y las ventas.

**3. Estudios de mercado**

- Análisis del sentimiento: análisis de las opiniones de los clientes, las publicaciones en redes sociales y otros contenidos en línea para medir el sentimiento del público hacia una marca o un producto, lo que ayuda a las empresas a tomar decisiones basadas en datos y mejorar la satisfacción del cliente.

- Supervisión de las redes sociales: las herramientas basadas en IA pueden rastrear las menciones a la marca, las tendencias del sector y las conversaciones relevantes en las plataformas de medios sociales, lo que permite a las empresas identificar oportunidades y amenazas en tiempo real.

- Análisis de la competencia: extracción de información valiosa de los sitios web, materiales de marketing y contenidos en línea de los competidores, lo que ayuda a las empresas a comprender su panorama competitivo y desarrollar estrategias eficaces para destacar.

**4. Sistemas de personalización y recomendación**

- Contenido personalizado: al analizar el comportamiento, las preferencias y el historial de navegación de los usuarios, los LLM pueden generar contenidos personalizados que se adapten a cada usuario, lo que aumenta el compromiso y la satisfacción de los usuarios.

- Publicidad dirigida: creación de mensajes de marketing a medida para segmentos de audiencia específicos, mejorando la relevancia de los anuncios y el retorno de la inversión.

**5. Legal y cumplimiento regulatorio**

- Análisis de contratos: revisión de contratos legales, identificación de términos clave, cláusulas y riesgos potenciales, ahorro de tiempo y reducción de la probabilidad de descuidos costosos.

- Supervisión del cumplimiento normativo: los LLM pueden ayudar a las organizaciones a seguir cumpliendo la normativa supervisando los cambios regulatorios y proporcionando información sobre cómo las nuevas normas o actualizaciones pueden afectar a sus operaciones empresariales, reduciendo el riesgo de sanciones por incumplimiento.

**6. Sanidad e investigación médica**

- Análisis de historiales clínicos: análisis de historiales clínicos electrónicos, extracción de información relevante e identificación de patrones que podrían ayudar a los profesionales sanitarios a tomar decisiones mejor informadas y mejorar los resultados de los pacientes.

**7. Recursos Humanos**

- Selección de currículos: escanear currículos, identificar a los candidatos más relevantes en función de sus habilidades, experiencia y otros factores, agilizar el proceso de contratación y reducir los sesgos de contratación.

- Búsqueda de empleo: emparejar a los demandantes de empleo con oportunidades laborales adecuadas analizando sus currículos y preferencias laborales, y comparándolos con los puestos disponibles, mejorando así la experiencia global de búsqueda de empleo.

**8. Finanzas**

- Análisis de noticias financieras: seguimiento y análisis de la actualidad financiera, identificando los acontecimientos, tendencias y sentimientos del mercado que pueden influir en las decisiones de inversión y el rendimiento de la cartera.

- Estrategias de trading basadas en el sentimiento: los LLM pueden utilizarse para desarrollar estrategias de trading basadas en el análisis del sentimiento de las noticias financieras, las redes sociales y otras fuentes, mejorando potencialmente el rendimiento de las inversiones.

**9. Educación y formación**

- Asistente de aprendizaje: explicar conceptos complejos en términos sencillos, dar respuesta a preguntas frecuentes y ofrecer recursos de aprendizaje personalizados en plataformas de educación en línea.

- Creación de contenidos educativos: elaboración de materiales de estudio, exámenes prácticos y otros recursos de aprendizaje.

**10. Automatización de procesos empresariales**

- Automatización de informes: generación de informes escritos a partir de datos estructurados, como ventas, métricas de rendimiento o datos financieros.

- Flujo de trabajo del correo electrónico: categorizar y responder a los correos electrónicos rutinarios, permitiendo a los trabajadores centrarse en tareas más importantes.

## *Un ejemplo de uso: mejorar la extracción de datos de documentos no estructurados*

En este ejemplo se muestra un uso potencial de los LLM en una aplicación real en el mundo empresarial, ejemplificando cómo pueden aprovecharse estos modelos avanzados de IA para resolver problemas complejos y mejorar los procesos existentes.

En relación con el proceso de contratación, las grandes empresas suelen recibir cientos, o incluso miles, de currículos diariamente, que deben ser procesados para la toma de decisiones en materia de contratación. Tradicionalmente, estos documentos se someten a una revisión inicial por parte de una persona que filtra a los candidatos en función de su formación o experiencia. A continuación, los currículos seleccionados reciben una evaluación más profunda como parte del proceso de selección en curso.

Para realizar esta tarea de forma más eficiente, las empresas piden a los candidatos que introduzcan sus datos en formularios en línea o que carguen el CV en un formato específico predefinido; sin embargo, este planteamiento podría disuadir a los candidatos de realizar el trabajo, lo que podría hacer que las organizaciones pasaran por alto a personas muy cualificadas y con talento. Para solucionar este punto, una

alternativa es solicitar solo el CV y, a continuación, utilizar modelos para extraer la información sobre los distintos criterios.

Para realizar esta tarea, el siguiente sistema de IA seguiría estos pasos:

- Convierte el archivo inicial del currículum (PDF, word, imagen) en un archivo de imagen unificado.
- Ejecuta e inspecciona la imagen a través de un sistema de IA de reconocimiento óptico de caracteres (OCR). Este tipo de tecnología de IA transforma las imágenes en texto.
- Utiliza el texto extraído del currículum por el OCR para inferir información relevante sobre el candidato.

Para obtener la información en forma de texto tras aplicar el OCR, es necesario identificar algunos campos en la fase de extracción, utilizando técnicas como el reconocimiento de identidades. La información específica, como el nombre del candidato (normalmente situado en la parte superior del currículum), el número de teléfono o la dirección de correo electrónico, puede extraerse fácilmente utilizando expresiones regulares (una expresión regular es un patrón de búsqueda utilizado para cotejar y manipular texto, que permite realizar tareas como la comparación de patrones, la validación, la manipulación de texto y el análisis sintáctico).

Sin embargo, en campos como la educación, la experiencia y las competencias, la estrategia anterior no se puede aplicar, ya que los datos de texto suelen ser no estructurados y no siguen ningún patrón específico, como en el caso de un correo electrónico. Por lo tanto, no se pueden utilizar expresiones regulares. Dado que los LLM permiten la extracción automática de información a partir de texto, podrían utilizarse para resolver este problema, ya sea resumiendo segmentos de texto o extrayendo directamente la información en el formato deseado. Por ejemplo, si en un CV la información sobre la educación viene dada por el texto "Estudié en la Universidad de Barcelona, donde obtuve la licenciatura en matemáticas", tras aplicar el LLM los datos estructurados de salida podrían ser: "Universidad de Barcelona" y "Licenciado en Matemáticas".

Se podría pensar que la búsqueda de coincidencias exactas en el texto sería adecuada; sin embargo, la lista de universidades y titulaciones disponibles evoluciona constantemente. Además, este método solo sería eficaz para currículos y coincidencias en un único idioma. Recorrer la lista de todas las universidades disponibles, en todos los idiomas, exigiría una potencia de cálculo considerable, lo que daría lugar a una experiencia de usuario poco manejable. Por tanto, la integración de un LLM en este proceso aumentaría la eficacia y mejoraría la experiencia del candidato en el proceso de selección.

# Limitaciones y retos futuros

Los LLM, aunque potentes y capaces de generar resultados impresionantes, tienen algunas limitaciones. Algunas de las principales limitaciones son:

*Cuestiones relacionadas con la salida generada:*

- **Sesgos:** los LLM se entrenan con grandes cantidades de datos de texto de Internet, que pueden incluir contenido sesgado, ofensivo o controvertido. En consecuencia, estos modelos pueden aprender inadvertidamente y perpetuar estos sesgos en el texto que generan.
- **Sensibilidad a la formulación de las preguntas:** en ocasiones, el rendimiento de los LLM puede ser sensible a la forma en que se formulan las preguntas o los mensajes. Esto significa que una ligera reformulación puede producir resultados diferentes y potencialmente más precisos.
- **Alucinaciones:** se dice que un modelo "alucina" cuando genera información o detalles que no estaban presentes en los datos de entrada o en el contexto proporcionado.
- **Dificultades con la verificación de la información:** los LLM pueden generar respuestas basadas en patrones lingüísticos y no en hechos verificados, lo que puede conducir a la propagación de información incorrecta o falsa.
- **Requiere supervisión:** los LLM, a pesar de ser autónomos hasta cierto punto, siguen requiriendo una considerable supervisión y ajuste humanos para garantizar la precisión y evitar resultados no deseados.

*Cuestiones relacionadas con la arquitectura:*

- **Falta de comprensión:** los modelos lingüísticos pueden generar textos que parezcan coherentes y adecuados al contexto, pero es posible que no posean una comprensión profunda del contenido. En consecuencia, el texto generado puede ser incorrecto o carecer de sentido.
- **Incapacidad para manejar datos multimodales:** aunque los últimos avances han empezado a acortar distancias, los LLM están diseñados principalmente para datos de texto y pueden tener dificultades para manejar datos multimodales, como imágenes o audio, sin modificaciones adicionales de la arquitectura.
- **Ausencia de memoria a largo plazo:** los LLM pueden no mantener un seguimiento consistente de interacciones pasadas, lo que puede limitar su capacidad para mantener conversaciones coherentes y contextuales a largo plazo.

*Cuestiones computacionales:*

- **Tamaño del modelo y requisitos computacionales:** los LLM más avanzados tienen miles de millones de parámetros, lo que conlleva importantes requisitos computacionales tanto para el entrenamiento como para la inferencia. Esto hace que consuman muchos recursos y sean menos accesibles para usuarios con una capacidad de cálculo limitada.
- **Impacto medioambiental:** el consumo de energía asociado al entrenamiento y al funcionamiento de los LLM contribuye a su huella de carbono, lo que suscita preocupación por su impacto medioambiental.

*Cuestiones éticas:*

- **Preocupación por la seguridad y el uso indebido:** las capacidades de los LLM pueden utilizarse indebidamente para generar contenidos nocivos, como desinformación o material ofensivo, lo que suscita dudas sobre su uso ético y su posible regulación.

Otra cuestión preocupante en relación con los LLM es su rápido avance y su impacto impredecible en el mercado laboral mundial. Estos modelos abren la posibilidad de automatizar en muy poco tiempo muchas tareas que actualmente realizan las personas, lo que convierte la reasignación de recursos humanos en un reto importante.

OpenAI ha publicado recientemente "GPTs are GPTs: An Early Look at the Labour Market Impact Potential of Large Language Models" [16], un documento de trabajo en el que los autores investigan las posibles implicaciones de los LLM en el mercado laboral estadounidense. Evaluaron las ocupaciones en función de su alineación con las capacidades de los LLM, integrando tanto la experiencia humana como las clasificaciones GPT-4. Descubrieron que alrededor del 80% de la mano de obra estadounidense podría ver afectado al menos el 10% de sus tareas laborales por la introducción de los LLM, mientras que aproximadamente el 19% de los trabajadores podría ver afectado al menos el 50% de sus tareas.

Los efectos previstos se extienden a todos los niveles salariales, y los empleos con mayores ingresos se enfrentan potencialmente a una mayor exposición a las capacidades de los LLM. En particular, estas repercusiones no se limitan a los sectores con mayor crecimiento reciente de la productividad. Los investigadores descubrieron que el 15% de todas las tareas de los trabajadores en EE.UU. podrían realizarse mucho más rápido manteniendo el mismo nivel de calidad. Cuando se incorpora software y herramientas creados sobre LLM, esta proporción aumenta hasta entre el 47% y el 56% de todas las tareas. Este hallazgo sugiere que el software basado en LLM desempeñará un papel importante en la amplificación de las repercusiones económicas de los modelos subyacentes.

Por último, existe una creciente preocupación por la protección de datos. Italia se convirtió en el primer país occidental en prohibir ChatGPT [17]: La Autoridad de Protección de Datos italiana ordenó a OpenAI que detuviera temporalmente el procesamiento de los datos de los usuarios italianos debido a una investigación en curso sobre una posible violación de la estricta normativa europea en materia de privacidad. A partir del 30 de abril de 2023, ChatGPT volvió a ser accesible en Italia.

Las empresas propietarias de LLM deben cumplir el Reglamento General de Protección de Datos (RGPD) si desean ofrecer sus servicios en países europeos. Uno de los principales retos a los que se enfrentan estas empresas es el cumplimiento del artículo 17, relativo al derecho al olvido. Según el RGPD, las personas tienen derecho a solicitar la eliminación de su información personal de los registros de una organización.

Los sistemas de IA, como las redes neuronales, no olvidan como los humanos. En lugar de ello, la red ajusta sus ponderaciones para adaptarse mejor a los nuevos datos, lo que produce resultados diferentes para entradas idénticas. Este proceso no constituye un olvido en el sentido tradicional, sino que la red da prioridad a los últimos datos que adquiere. Toda la información permanece en el sistema cuando se reajustan los pesos de la red.

Todas estas limitaciones y retos plantean la necesidad de que las empresas definan e integren un marco y unas políticas adecuadas para el uso de los LLM en su trabajo diario, incluyendo la incorporación de estos elementos a la gobernanza, la creación de políticas específicas, el establecimiento de usos de los LLM aceptados internamente, la formación de los empleados y la garantía del uso correcto de las herramientas.

# Conclusiones

Los grandes modelos lingüísticos (LLM) son modelos basados en el aprendizaje profundo capaces de procesar grandes cantidades de texto. Pueden realizar diversas tareas, como el resumen de textos, el análisis de sentimientos y la traducción automática. El rápido desarrollo y popularización de estos modelos ha abierto la puerta a numerosas aplicaciones empresariales, sobre todo en la automatización de tareas que actualmente realizan las personas.

La generación de contenidos para redes sociales, el análisis de tendencias de mercado, la selección de currículos para recursos humanos y la automatización del servicio de atención al cliente mediante chatbots representan solo una parte de las aplicaciones potenciales de los LLM. El alcance de estas aplicaciones es enorme y abarca numerosos sectores empresariales en los que los LLM pueden aportar un valor significativo.

Las numerosas ventajas y aplicaciones de estos modelos también van acompañadas de notables limitaciones, como el sesgo de los modelos y las repercusiones medioambientales ligadas a su uso. Además, la integración de estos modelos en nuestra sociedad presenta retos importantes. Entre ellos destacan los efectos imprevisibles sobre el mercado laboral, las preocupaciones asociadas a la protección de los datos de los usuarios o su impacto medioambiental.

Por todo ello, la incorporación de LLM en los procesos empresariales debe realizarse asegurando un uso eficaz, ético y seguro. Esto implica identificar los usos que la organización está dispuesta a acometer a través de esta tecnología, garantizar la confidencialidad de la información de la compañía (evitando fugas de información a servidores de terceros y anonimizando la información), y asegurar que las salidas de estos modelos se contrastan con fuentes fiables antes de integrarlas en procesos de decisión o de gestión.

---

Nota: esta publicación se ha elaborado con la ayuda de varias herramientas de inteligencia artificial (IA). Estas herramientas se utilizaron para diversas tareas, como buscar información, recopilar y organizar datos y generar resúmenes. En cualquier caso, el material final de esta publicación fue escrito por una persona y no por una IA.

# Bibliografía

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is All You Need", in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017.

[2] L. Floridi and M. Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences", Minds and Machines, vol. 30, no. 4, pp. 681-694, 2020.

[3] A. Radford, K. Narasimhan, T. Salimans and I. Sutskever, "Improving Language Understanding by Generative Pre-Training," 2019.

[4] OpenAI, "GPT-4 Technical Report", 2023.

[5] Meta AI, "LLaMA: Open and Efficient Foundation Language Models," 2023.

[6] BigScience Workshop, "BLOOM: A 176B-Parameter Open-Access MultilingualLanguage Model", 2022.

[7] Google Research, "PaLM: Scaling Language Modeling with Pathways," 2022.

[8] Google, "LaMDA: Language Models for Dialog Applications", 2022.

[9] Anthropic, "Constitutional AI: Harmlessness from AI Feedback," 2022.

[10] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", in Text Summarization Branches Out, Barcelona, Association for Computational Linguistics, 2004, pp. 74-81.

[11] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy and S. Bowman, "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding", in Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2018.

[12] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy and S. R. Bowman, "SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems", in NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019.

[13] [Online]. Available: https://rajpurkar.github.io/SQuAD-explorer/.

[14] D. Paperno, G. Kruszewski, A. Lazaridou, N. Q. Pham, R. Bernardi, S. Pezzelle, M. Baroni, G. Boleda and R. Fernandez, "The LAMBADA dataset: Word prediction requiring a broad discourse context", in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016.

[15] Google, [Online]. Available: https://github.com/google/BIG-bench.

[16] T. Eloundou, S. Manning, P. Mishkin and D. Rock, "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models", 2023.

[17] CNBC, [Online]. Available: https://www.cnbc.com/2023/04/04/italy-has-banned-chatgpt-heres-what-other-countries-are-doing.html. [Accessed 25 April 2023].

## *Autores*

Ernestina Menasalvas (UPM)
Víctor Robles (UPM)
Manuel Ángel Guzmán (Management Solutions)
Sergio Ruiz Bonilla (Management Solutions)

La Universidad Politécnica de Madrid es una Entidad de Derecho Público de carácter multisectorial y pluridisciplinar, que desarrolla actividades de docencia, investigación y desarrollo científico y tecnológico.

**www.upm.es**

Management Solutions es una firma internacional de consultoría, centrada en el asesoramiento de negocio, finanzas, riesgos, organización, tecnología y procesos, que opera en más de 50 países y con un equipo de más de 3.300 profesionales que trabajan para más de 1.500 clientes en el mundo.

**www.managementsolutions.com**

**Para más información visita**

# blogs.upm.es/catedra-idanae/