



LLM e Inteligencia Artificial Generativa

La irrupción de ChatGPT dentro de la categoría de Inteligencia Artificial Generativa basada en LLM y su habilidad para responder a cualquier tipo de pregunta abre una nueva vía en el desarrollo de chatbots para la atención de clientes y usuarios.

Este artículo explica los conceptos fundamentales de esta nueva tecnología que algunos expertos sitúan a nivel disruptivo a la altura de la aparición de Internet.

¿Que es LLM e Inteligencia Artificial Generativa?

Los [Modelos de Lenguaje de Gran Escala](#)(LLM) son un tipo de inteligencia artificial (IA) que ha revolucionado la forma en que las máquinas comprenden y generan texto de manera natural.

Un ejemplo destacado de LLM es GPT (Generative Pre-trained Transformer) desarrollado por [OpenAI](#). Estos modelos, como GPT, son entrenados en grandes conjuntos de datos para aprender patrones y contextos del lenguaje natural. El «pre-entrenamiento» se refiere a la fase inicial de aprendizaje antes de afinar el modelo para tareas específicas.

GPT utiliza capas de atención y aprendizaje no supervisado para captar patrones y relaciones en el lenguaje, lo que le permite realizar tareas diversas, como generación de texto, traducción automática, respuestas a preguntas y más.

La Inteligencia Artificial Generativa, impulsada por LLM, se refiere a sistemas de inteligencia artificial capaces de generar contenido de manera autónoma, como textos, imágenes y música. Estos sistemas no se limitan a responder preguntas o seguir comandos específicos, sino que pueden crear contenido original y contextualmente relevante.

En resumen, un LLM como GPT es un modelo de inteligencia artificial generativa que puede entender y producir texto de manera coherente y contextual, siendo versátil para diversas aplicaciones en procesamiento de lenguaje natural.

Cómo cambia las reglas del juego a nivel empresarial

La incorporación de LLM y la IA generativa en el entorno empresarial ha transformado radicalmente la forma en que las organizaciones interactúan con sus clientes, gestionan la información y toman decisiones. Algunas de las formas en que esta tecnología ha cambiado las reglas del juego son,

Atención al cliente y chatbots avanzados

Los chatbots impulsados por LLM pueden mantener conversaciones más naturales y útiles con los clientes. Pueden responder preguntas, brindar asesoramiento y resolver problemas con una precisión que antes sólo era posible con la intervención humana.

Esto mejora la satisfacción del cliente y reduce la carga de trabajo del personal de atención al cliente.

Generación de contenido automatizado

Las organizaciones pueden utilizar la IA generativa para crear contenido de marketing, artículos de blog y descripciones de productos.

También pueden ser usados para extraer información de documentos ya sea a modo de resumen o de datos importantes que deben ser tenidos en cuenta.

Esto acelera la producción de contenido y puede adaptarse a audiencias específicas, lo que ahorra tiempo y recursos.

Automatización de procesos empresariales

Las organizaciones pueden automatizar una amplia gama de procesos, desde la gestión de datos hasta la programación de reuniones, utilizando IA generativa. Esto aumenta la eficiencia y reduce los errores humanos.



Análisis de datos avanzado

LLM y la IA generativa pueden analizar grandes conjuntos de datos y extraer información valiosa, tanto para la toma de decisiones como para la optimización de procedimientos.

Esto es especialmente útil en áreas como la inteligencia de negocios y la predicción de tendencias.

Generación de contenido multilingüe y globalización

La capacidad de LLM para trabajar en varios idiomas y culturas permite a las empresas expandirse globalmente y llegar a una audiencia más diversa.

Conclusiones

En resumen, LLM y la IA generativa han abierto un abanico de posibilidades, mejorando la eficiencia, la personalización y la calidad de las interacciones con los clientes, y ayudando a las empresas a tomar decisiones más informadas y creativas. Esto les permite mantenerse competitivas y adaptarse a un entorno empresarial en constante evolución.

Pero la implantación de este tipo de proyectos no es trivial y, como todo, tiene sus ventajas e inconvenientes que comentamos en nuestro artículo [Chatbots con Plataformas NLP vs. Modelos de Lenguaje GPT](#).



Chatbots con Plataformas NLP vs. Modelos de Lenguaje GPT

Actualmente, las principales alternativas para desarrollar chatbots empresariales que utilicen Inteligencia Artificial (IA) son el uso de Plataformas NLP y la generación mediante modelos de lenguaje GPT (Generative Pre-trained Transformer).

En este artículo realizamos una comparativa entre estas dos alternativas, analizando las ventajas e inconvenientes de cada una de ellas.

Plataformas NLP vs. Modelos de Lenguaje GPT

Un ejemplo de Plataforma NLP es [SapiensXBot](#) y son, fundamentalmente, arquitecturas de procesamiento y comprensión del lenguaje natural (NLP y NLU, por sus siglas en inglés, respectivamente).

Estas plataformas están diseñadas para comprender y procesar el lenguaje humano, identificando las intenciones detrás de las consultas de los usuarios y extrayendo entidades relevantes. Para ello es necesario entrenarlas con una gran cantidad de ejemplos para que sean capaces de identificar que es lo que se les está pidiendo.

Por su parte, la aparición de modelos de lenguaje GPT (como [ChatGPT](#), [Llama2](#) u otros open source que se pueden descargar desde [Hugging Face](#)), que destacan por su capacidad para comprender y generar texto de manera coherente y contextual, ha supuesto un cambio radical en la forma en que se desarrollan los chatbots tal como hemos comentado en [LLM e Inteligencia Artificial Generativa](#).

Uso de modelos de lenguaje GPT

Ventajas:

Comprensión del lenguaje natural

Los LLM están entrenados con cientos de miles de millones de parámetros. Esto le permite entender y generar texto en contextos complejos sin necesidad de un entrenamiento específico para situaciones concretas.

Capacidad de aprendizaje

Continuamente están apareciendo modelos nuevos mejorados que podemos incorporar con pocos cambios a nuestros proyectos.

Multi-idioma

Los modelos más potentes (a partir de decenas de miles de millones de parámetros 13B y superiores) están entrenados en varios idiomas por lo que son capaces de comprender la pregunta independientemente del idioma en que se ha realizado y responder en el mismo idioma.

Escalabilidad

La escalabilidad viene determinada por el uso de una API pública para realizar las preguntas. Esto nos permite utilizar los modelos sin necesidad de tener que proveer una infraestructura compleja y cara que necesitará de ordenadores con varias GPUs.

Desventajas

Potenciales errores

Uno de los principales problemas de los modelos LLM son las llamadas «Alucinaciones». Como su propio nombre indica, la IA Generativa, se caracteriza por generar respuestas imaginativas a las preguntas.

Esta capacidad de inventarse la respuesta si no la conoce es muy peligrosa en caso de que necesitamos respuestas concretas y fiables en entornos de consulta de documentación interna o servicio de atención a clientes.



No pueden responder cualquier pregunta

Como hemos comentado en varias ocasiones, los modelos GPT están entrenados con miles de millones de parámetros y millones de documentos públicos (wikis, foros, libros,...) por lo que son capaces de responder a un número ilimitado de preguntas.

Sin embargo, las diferentes empresas y organizaciones disponen de una gran cantidad de información interna que son parte de su «know-how» y que no está disponible públicamente.

Toda esta información no se ha podido (ni seguramente se debe) utilizar en los entrenamientos, por lo que los modelos GPT no podrán responder a preguntas sobre estos temas.

Además, el entrenamiento de estos modelos es muy costoso en recursos, por lo que las actualizaciones, aunque constantes, no son frecuentes. Esto hace que no sepa responder a preguntas sobre novedades aparecidas después de su entrenamiento.

Costos de implementación

Existen dos planteamientos a la hora de implantar chatbots con modelos GPT, el pago por uso vía API y la utilización de modelos privados en infraestructura propia.

En el primero de los casos, el coste de inicio del proyecto es bajo, pero tiene un crecimiento lineal «infinito» en función de su uso.

Por el contrario, el uso de modelos privados requiere de una infraestructura costosa inicialmente debido a los requerimientos de GPU pero con un crecimiento logarítmico.

Uso de plataformas NLP

Ventajas

Flexibilidad y control total

Ofrecen un alto grado de flexibilidad y control sobre el desarrollo de chatbots. Por un lado, los desarrolladores pueden definir flujos de conversación personalizados y lógica empresarial compleja mediante la definición de historias.

Pero además, una vez [identificada la intención](#), es posible llamar a APIs, ejecutar programas o consultar bases de datos para ofrecer una respuesta totalmente personalizada a las demandas del usuario.

Integración con plataformas externas

Se integra fácilmente con otros sistemas y plataformas, tanto a nivel de interfaz de usuario (canales) como a bases de conocimiento.

Por lo que respecta a los canales de entrada pueden ser la propia página web de la organización como redes sociales (Facebook, Telegram,...) u otros dispositivos (Alexa, Chat rooms, ...)

Para obtener el conocimiento y generar la respuesta se pueden ejecutar aplicaciones, hacer llamadas a servicios webs o consultar bases de datos

Cumplimiento de normativas

El hecho de que la respuesta sea definida o generada mediante una programación previa, permiten un mayor control sobre la gestión de datos y la privacidad, lo que facilita el cumplimiento de regulaciones como el RGPD.

Desventajas:

Mayor complejidad de desarrollo

Dado que son una plataformas de desarrollo más técnica, puede requerir más tiempo y habilidades de desarrollo en comparación con herramientas de chatbot basadas en modelos GPT.



Desafíos en NLU

Aunque ofrecen capacidades de procesamiento de lenguaje natural (NLU), la calidad de las predicciones de NLU puede variar y puede requerir una cantidad significativa de datos y ajustes para lograr un alto rendimiento.

Conclusiones

El análisis comparativo entre las Plataformas NLP y los Modelos de Lenguaje GPT para la creación de chatbots nos muestra que la principal ventaja de los sistemas chatbot basados en modelos GPT es su facilidad para entender y generar respuestas en lenguaje natural sin un gran esfuerzo gracias a que los modelos han sido entrenados previamente con un gran volumen de datos. Por el contrario, este mismo hecho, provoca su principal desventaja: las alucinaciones.

Por su parte, la construcción de chatbots con Plataformas NLP requieren de un gran esfuerzo en entrenamiento para la comprensión de cualquier tipo de pregunta que se pueda realizar, pero una vez comprendida la intención del usuario la respuesta es más fiable, personalizada y segura.



Ventajas de utilizar IA en el desarrollo de Chatbots

Estamos acostumbrados a escuchar que determinada aplicación o servicio ha creado o dispone de un chatbot que permite ofrecer un servicio al cliente. La mayoría de estos bots están basados en una serie de flujos activados por botones. Este tipo de chatbots no aprovechan las ventajas de utilizar Inteligencia Artificial (IA) lo cual hace que la experiencia del usuario pueda ser frustrante y no difiera mucho de un asistente o de un menú.

Tampoco debemos confundir un Chatbot que utiliza la Inteligencia Artificial con un buscador que se limita a comparar expresiones idénticas y mostrar los resultados.

SapiensXBot utiliza la Inteligencia Artificial (IA) tanto para la [identificación de intenciones](#) como para la [gestión de la conversación](#).

Ventajas de utilizar inteligencia artificial en la Identificación de intenciones

Respecto de la identificación de intenciones, SapiensXBot utiliza la Inteligencia artificial para crear un modelo que permita interpretar la frase recibida por parte del usuario/ciudadano.

Con este modelo, se evalúa cada una de las frases del ciudadano y se detecta la intención con un porcentaje de confianza.

En concreto, el núcleo de SapiensXBot dispone de una Unidad de procesamiento de Lenguaje Natural (NLU). El NLU utiliza técnicas de Machine Learning conocidas como Procesamiento de Lenguaje Natural (NLP).

Por su parte, el NLP permite identificar los diferentes elementos de la frase con el objetivo de detectar patrones que permitan detectar frases similares a las utilizadas como ejemplo.

La ventaja de utilizar IA en la identificación de intenciones, es que permite detectar variaciones significativas respecto de las frases utilizadas para entrenar el modelo.

Por ejemplo, la utilización de sinónimos en verbos o en sustantivos, el cambio de orden de los elementos de la frase (sujeto, verbo y predicado) e, incluso, texto con errores tipográficos o faltas de ortografía.

Puedes leer más sobre este tema en nuestros artículos sobre [la identificación de intenciones](#) y [la detección de entidades](#).

Ventajas de utilizar inteligencia artificial en la Gestión de la conversación

Una vez que se reconoce la intención del usuario, es posible que el chatbot necesite información adicional para proporcionar la respuesta.

Por ejemplo, si tenemos un chatbot para concertar una cita, necesitaremos saber la fecha en la que se quiere concertar, el tema y posiblemente la ubicación para ofrecer una alternativa cercana.

En función de la información proporcionada en la pregunta, el chatbot deberá preguntar por el resto de parámetros necesarios. En otras palabras, si el ciudadano ha pedido «Quiero concertar una cita», el chatbot deberá preguntar por

- el día,
- el tema y
- la ubicación

pero si la pregunta es «Quiero concertar una cita el jueves» el chatbot únicamente deberá preguntar por

- la ubicación y
- el tema



Más de una conversación simultánea

Los chatbots que NO utilizan IA tienen que tener toda esta casuística programada de antemano y no permiten introducir nuevos temas de conversación hasta que no se finaliza el flujo actual.

Dicho con otras palabras, si el ciudadano, antes de acabar de concretar la cita desea saber que documentación debe aportar, no lo podrá hacer hasta que acabe el trámite de concertación de la cita.

Por consiguiente, respecto de la gestión de la conversación, la ventaja de utilizar inteligencia artificial viene dada porque podemos construir modelos de conversación básicos que serán lanzados en función de la frase detectada de forma dinámica.

Por ejemplo, si a mitad del trámite de concertación de la cita, el ciudadano pregunta «que documentación debo llevar?», SapiensXBot es capaz de responder correctamente a esa pregunta y continuar solicitando la información necesaria para acabar de concretar la cita.

Puedes leer más de este tema en nuestro artículo sobre la [gestión de la conversación](#).

Conclusiones

La utilización de Inteligencia Artificial para la creación de Chatbots, facilita tanto

- la **definición del propio chatbot**, dado que requiere un número limitado de ejemplos para construir el modelo de identificación de intenciones y puede detectar variaciones sobre dichos ejemplos, como
- la **gestión de las conversaciones**, evitando tener en cuenta todo tipo de excepciones a la hora de construir un flujo de secuencias.

En definitiva permite poder mantener conversaciones en un lenguaje próximo al usuario/ciudadano, de manera que éste no pueda distinguir si al otro lado del chat hay una persona o un robot.