

Módulo 4. Uso de ChatGPT como complemento para el análisis y tratamiento de datos en RStudio



☰ 4. Uso de ChatGPT como complemento para el análisis y tratamiento de datos en RStudio

☰ Referencias

4. Uso de ChatGPT como complemento para el análisis y tratamiento de datos en RStudio

Prácticamente a diario vemos noticias, publicaciones en redes sociales, menciones en compañías tecnológicas sobre herramientas de inteligencia artificial (IA) y, desde el 2022, cuando se lanzó ChatGPT, gran parte de la atención se ha centrado sobre esta herramienta.

La mayoría de estas informaciones muestran cómo las herramientas de IA a disposición de los usuarios finales tienen gran aplicación y ventajas en distintos ámbitos, y que su nivel de sofisticación y funcionalidades mejora a un ritmo muy alto.

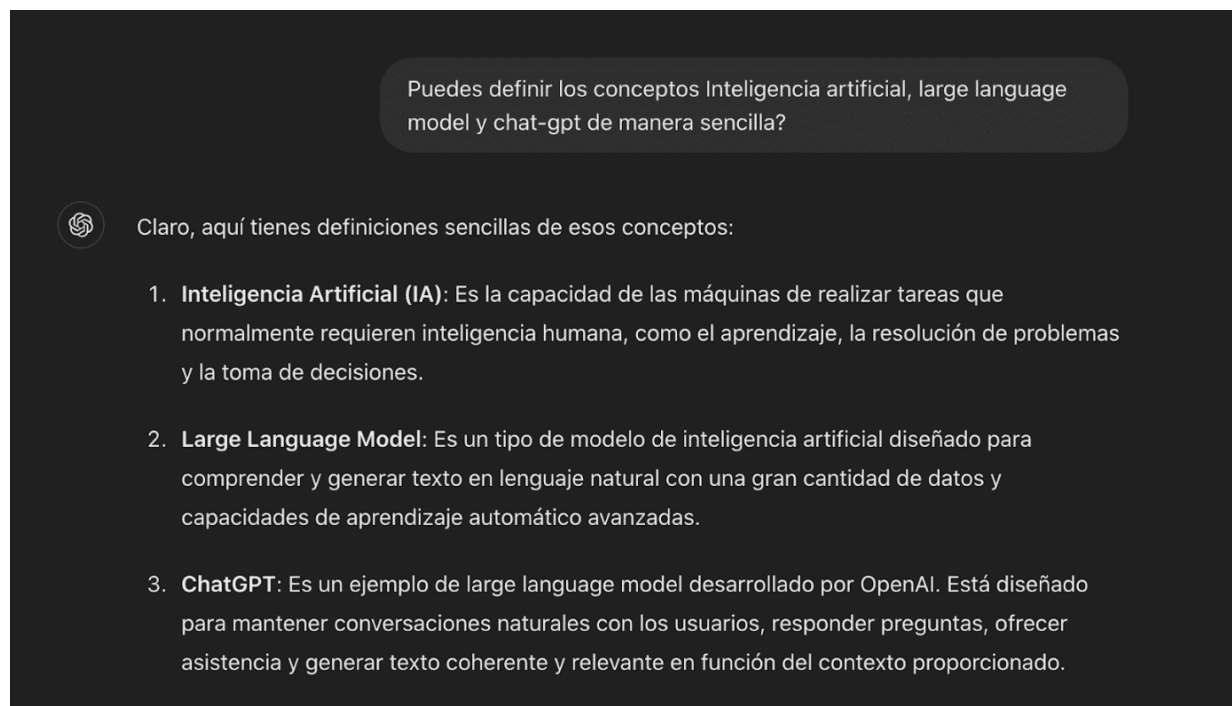
El ámbito del análisis de datos físicos no se queda fuera de esta evolución, y podemos utilizar herramientas como ChatGPT como complemento en los procesos que hemos ido destacando a lo largo del certificado.

ChatGPT (Chat *Generative Pre-trained Transformer*) es un sistema de inteligencia artificial basado en un modelo «*large language model*» (LLM). Estos modelos son entrenados a partir de una gran cantidad de

datos de tipo texto para que sean capaces de interpretar el lenguaje del usuario y generen respuestas como si se tratara de una conversación entre personas. Este tipo de LLM se basan en redes neuronales profundas, las cuales aprenden de los patrones en el texto y el significado del mismo (Almarie et al., 2023). Por lo tanto, ChatGPT es una interfaz que permite interactuar con el modelo creado, donde el usuario puede formular preguntas y la interfaz devolverá las respuestas como si fuera parte de una conversación.

Hemos mencionado ciertos conceptos que pueden ser nuevos para nuestro contexto, vamos a utilizar ChatGPT para aclarar los mismos.

Figura 1: Definición de Inteligencia Artificial, *Large Language Model* y ChatGPT



Puedes definir los conceptos Inteligencia artificial, large language model y chat-gpt de manera sencilla?

Claro, aquí tienes definiciones sencillas de esos conceptos:

1. **Inteligencia Artificial (IA):** Es la capacidad de las máquinas de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la resolución de problemas y la toma de decisiones.
2. **Large Language Model:** Es un tipo de modelo de inteligencia artificial diseñado para comprender y generar texto en lenguaje natural con una gran cantidad de datos y capacidades de aprendizaje automático avanzadas.
3. **ChatGPT:** Es un ejemplo de large language model desarrollado por OpenAI. Está diseñado para mantener conversaciones naturales con los usuarios, responder preguntas, ofrecer asistencia y generar texto coherente y relevante en función del contexto proporcionado.

Las principales aplicaciones del uso de la herramienta son proporcionar definiciones, resumir contenido, etc. Pero debemos tener en cuenta que tiene ciertas limitaciones. Como cualquier modelo de *machine learning* o inteligencia artificial, está entrenado con unos datos específicos, por lo tanto, las características de los datos que han sido proporcionados para el modelo condicionarán las respuestas. Esto tiene implicaciones claras en el *output* que recibimos, por ejemplo, en las primeras versiones gratuitas y disponibles de ChatGPT, se informó que el modelo se había entrenado con datos hasta una fecha específica, por lo tanto, no sería una herramienta útil para responder preguntas sobre situaciones actuales o recientes, ya que el modelo no dispone de esa información. Hay que recordar que esta herramienta no es un motor de búsqueda como Google, por lo que tampoco proporcionara *links* o páginas de acceso a otros recursos.

Para obtener el mejor *output* posible, al tratarse de un modelo basado en el lenguaje, es necesario ser cuidadosos en cómo pedimos la información a ChatGPT. La forma en que estructuramos nuestra pregunta o petición condicionarán el detalle y la precisión de la respuesta. Las preguntas que realizamos a ChatGPT son conocidas como «prompt».

- Incluir contexto antes de la pregunta:
 - o quién es la supuesta persona que está realizando la pregunta (un estudiante, un entrenador, etc.) y a quién va dirigida la respuesta;
 - o cuál es el objetivo de la pregunta.
- Qué información quieres obtener:
 - o ser lo más específicos posible;
 - o si hay cierta información que no es necesaria, especificarlo.
- Especificar la longitud de la respuesta.
- Pedir ejemplos.

Ejemplo de un prompt «malo»:

- gráfico 3D RStudio saque de tenis.

Ejemplo de un prompt «bueno»:

- ¿Podrías describir el proceso de creación de un gráfico 3D en RStudio para visualizar la trayectoria del brazo de un jugador de tenis durante el servicio? Estoy interesado en conocer distintas posibilidades, librerías que sean útiles y código como ejemplo.

Una de las particularidades de ChatGPT es que podemos realizar seguimiento de las preguntas que hagamos, es decir, si queremos que extienda su respuesta en alguno de los puntos podemos pedir que detalle más ese punto específico, que utilice otros ejemplos, etc.

Estos puntos únicamente describen cuál debería ser la forma o el formato de plantear las preguntas, pero el punto más importante, como hemos destacado múltiples veces durante el curso, es la elección de la pregunta en sí. Debemos, primeramente, definir aquella inquietud específica en nuestro contexto para la cual buscamos respuesta y luego, determinar cómo ChatGPT puede ayudarnos durante el proceso.

El enlace para acceder a la herramienta es:

OpenAI. (2021). ChatGPT. <https://www.openai.com/chatgpt/>

ES POR ESE MOTIVO QUE DEBEMOS SEGUIR UNA
SERIE DE CONSIDERACIONES QUE DESARROLLAMOS
A CONTINUACIÓN.

USO DEL CHATGPT DURANTE EL PROCESO DE
ANÁLISIS DE DATOS.

En el material de video, veremos ejemplos sobre cómo utilizar la herramienta para realizar consultas que podamos tener durante nuestros proyectos de análisis de datos físicos. Durante el certificado, se han destacado muchos tipos de análisis y funcionalidades de RStudio para intentar mostrar todo su potencial. RStudio es una herramienta con una alta curva de aprendizaje, ya que, al inicio, requiere una comprensión correcta de la estructura que puede resultar compleja si se trata del primer lenguaje de programación con el que trabajamos. ChatGPT puede ayudarnos a resolver dudas durante el proceso, como veremos en los siguientes ejemplos.

- Comprender código: podemos pedir a ChatGPT que describa cuál es cada una de las partes del código utilizado y qué objetivo consigue/persigue.
- Escribir código: eliminar valores ausentes, añadir filtros a columnas, importar datos, etc., pueden ser necesidades que tenemos en nuestra fuente de datos antes de seguir con los procesos de análisis. Como ayuda, podemos describirlas a ChatGPT para que proponga el código específico para poder cumplir esas necesidades.

Debemos ser conscientes de que este contenido se ha incluido en el último curso, ya que es importante en el proceso de aprendizaje de una herramienta como RStudio comprender desde cero lo que se consigue con cada línea de código, experimentar con las distintas funciones y buscar soluciones propias para conseguir las transformaciones o modificaciones de datos que creemos,

así como comprender de dónde provienen los errores que vamos realizando a lo largo de la elaboración de código y cómo corregirlos. Si utilizamos la información que proporciona ChatGPT sin comprender lo que sucede, podemos dejar pasar errores que condicionen los resultados finales y limitar nuestra capacidad de aprendizaje de la propia herramienta, lo que condicionará un desarrollo más extenso de la misma en un futuro.

Si tenemos más experiencia con RStudio, ChatGPT sigue siendo un buen complemento para los proyectos en los que trabajemos. Como siempre, en el trabajo de un Sport Scientist, nuestro objetivo es automatizar los procesos de nuestro día a día, para que la información que se utiliza en la toma de decisiones esté disponible lo más rápido posible. Veremos algunos ejemplos a continuación.

- Eficiencia en el código: podemos utilizar ChatGPT para obtener métodos alternativos al código que hemos desarrollado, especificando que busquemos una manera en la que el código sea ejecutado lo más rápido posible, por ejemplo. Cada librería o función que utilizamos en R es más o menos eficiente según cómo utiliza los datos y qué procesos lleva a cabo y hemos visto en cursos anteriores, que múltiples opciones son válidas para conseguir el objetivo en varios de los casos.
- Funciones: si queremos repetir un mismo proceso múltiples veces, las funciones son de gran utilidad. En el caso de que tuviéramos un código escrito y necesitáramos transformarlo en una función, ChatGPT podría responder a esa necesidad.
- Automatización: para las dudas que han surgido durante el certificado sobre cómo leer múltiples archivos a la vez o producir la misma visualización para distintos jugadores/tareas, podemos utilizar ChatGPT para obtener indicaciones sobre cómo conseguir una mayor automatización.

Como Sport Scientists, existe la necesidad de tener un rango de aptitudes y de herramientas, el análisis de datos y RStudio es una más de ellas. Sin embargo, como hemos visto dentro del análisis de datos, existen varias ramas o partes

del proceso y es extremadamente difícil dominar cada una de ellas debido a su rápida expansión, novedades constantes y necesidad de mantenerse al día con el resto de tecnología. ChatGPT puede servir también como material para aumentar nuestro dominio y exploración en ámbitos avanzados, así como dar un paso más en partes del proceso, como en los siguientes ejemplos.

- Visualización y comunicación de resultados: librerías como ggplot y plotly tienen grandes capacidades de customización, con la ayuda de ChatGPT podemos conseguir visualizaciones más avanzadas.
- Modelos estadísticos: preguntas sobre la comprensión de los modelos, ejemplo que podamos ejecutar en nuestra sesión de RStudio para experimentar con los resultados. También describir resultados y ayudar en la comprensión de ellos.

Con la versión de suscripción de ChatGPT es posible obtener una clave para vincular la cuenta a un *package* específico de RStudio que permite utilizar la herramienta dentro de la sesión de R. De la misma manera que en el módulo anterior se creó esa clave/*token* para autorizar el acceso de Rstudio a Google Sheets, utilizaríamos una clave proporcionada por ChatGPT para utilizarlo. En el caso de que esto sea una aplicación que pueda ser interesante para determinados contextos, tendríamos esta posibilidad a nuestro alcance.

Hay que destacar de nuevo que es importante conocer las limitaciones, no es una herramienta que reemplace la labor del Sport Scientist o el analista de datos, ya que la parte fundamental es la elaboración de la pregunta y el desarrollo del plan de trabajo hasta responderla. Además, como profesionales tratando con jugadores, conocemos el contexto y las particularidades que puedan estar afectando a los datos, estas serán importantes y podrían condicionar el análisis. Las decisiones que se tienen que tomar irán a cargo del Sport Scientist, por lo que ChatGPT no puede reemplazar esta labor. De la misma manera, cuando seleccionemos el método con el que comuniquemos los resultados, nosotros conocemos nuestra audiencia, sabemos qué información y en qué formato será recibida de mejor forma entre el *staff*, jugadores o directiva, según su nivel de comprensión, detalle de la

información, etc. Esta decisión, por lo tanto, también deberá ser tomada por el Sport Scientist.

Por último, debemos recordar que se trata de una herramienta basada en un modelo, y como vimos en módulos anteriores, todos los modelos tienen su error correspondiente. Es importante conocer lo que estamos buscando responder y ser críticos con las respuestas que obtenemos, es decir, asegurarnos que el resultado que estamos obteniendo es el correcto y deseado.

CONTINUAR

Referencias

Almarie, B.; Teixeira, P. E. P.; Pacheco-Barrios, K.; Rossetti, C. A.; Fregni, F. (2023). The Use of Large Language Models in Science: Opportunities and Challenges. *Principles and Practice of Clinical Research*, 9(1), 1–4. <https://doi.org/10.21801/ppcrj.2023.91>

OpenAI. (2021). ChatGPT. <https://www.openai.com/chatgpt/>

CONTINUAR