

Aplicación de la inteligencia artificial ChatGPT en RStudio para el procesamiento y representación de datos estadísticos

Application of ChatGPT artificial intelligence in RStudio for processing and representation of statistical data

Aplicação da inteligência artificial ChatGPT no RStudio para processamento e representação de dados estatísticos

Angel Aronés-Cisneros*¹, ORCID: <https://orcid.org/0000-0002-5064-7027>

Cynthya Alegre-Palomino², ORCID: <https://orcid.org/0000-0001-9099-2290>

Rocío Aronés-Cisneros¹, ORCID: <https://orcid.org/0000-0002-0983-7849>

Leticia Tapia-Oré¹, ORCID: <https://orcid.org/0009-0003-1739-2266>

¹Instituto de Investigación Geográfico Andino Rural, Perú

²Universidad Nacional de San Antonio Abad del Cusco, Perú

*Autor para correspondencia: a.aronescisneros@gmail.com

RESUMEN

La investigación tuvo como objetivo explorar la aplicación de la inteligencia artificial ChatGPT en el entorno RStudio para simplificar el procesamiento y la representación de datos estadísticos. Se utilizó un archivo Excel con nueve variables, como sexo, edad, nivel de educación, IMC, horas de trabajo semanal y estaturas de padres e hijos, organizadas en columnas para su fácil manipulación. A través de códigos generados por ChatGPT, se crearon diversas visualizaciones, incluyendo gráficas de barras, histogramas, dispersión, líneas, circulares, diagramas de cajas y bigotes, y campanas de Gauss. Estas visualizaciones permitieron identificar patrones y correlaciones clave, como la relación entre edad y características demográficas, y la presencia de valores atípicos. En conclusión, la combinación de ChatGPT y RStudio demostró ser una herramienta eficiente para la generación automatizada de gráficos, mejorando la precisión y la eficiencia en el análisis y la interpretación de datos, así como optimizando el tiempo en la creación de representaciones visuales que enriquecen la comprensión de la información.

Palabras clave: Inteligencia artificial; RStudio; análisis estadístico; ChatGPT.

ABSTRACT

The research aimed to explore the application of ChatGPT artificial intelligence in the RStudio environment to simplify the processing and representation of statistical data. An Excel file was used with nine variables, such as sex, age, education level, BMI, weekly work hours, and heights of parents and children, organized in columns for easy manipulation. Through codes generated by ChatGPT, various visualizations were created, including bar graphs, histograms, dispersion, lines, circular, box and whisker diagrams, and Gaussian bells. These visualizations allowed the identification of key patterns and correlations, such as the relationship between age and demographic characteristics, and the presence of outliers. In conclusion, the combination of ChatGPT and RStudio proved to be an efficient tool for the automated generation of graphics, improving accuracy and efficiency in data analysis and interpretation, as well as optimizing time in the creation of visual representations that enrich the understanding of information.

Keywords: artificial intelligence; RStudio; statistical analysis; ChatGPT.

RESUMO

A pesquisa teve como objetivo explorar a aplicação da inteligência artificial ChatGPT no ambiente RStudio para simplificar o processamento e a representação de dados estatísticos. Utilizou-se um arquivo Excel com nove variáveis, incluindo sexo, idade, escolaridade, IMC, carga horária semanal e altura dos pais e filhos, organizadas em colunas para facilitar a manipulação. Utilizando códigos gerados pelo ChatGPT, diversas visualizações foram criadas, incluindo gráficos de barras, histogramas, gráficos de dispersão, gráficos de linhas, gráficos de pizza, gráficos de caixa e bigode e curvas de Gauss. Essas visualizações permitiram a identificação de padrões e correlações importantes, como a relação entre idade e características demográficas, e a presença de outliers. Conclui-se que a combinação do ChatGPT e do RStudio demonstrou ser uma ferramenta eficiente para a geração automatizada de gráficos, melhorando a precisão e a eficiência da análise e interpretação de dados, além de otimizar o tempo necessário para criar representações visuais que enriquecem a compreensão das informações.

Palavras-chave: inteligência artificial; RStudio; análise estatística; ChatGPT.

Recibido: 9/3/2025 Aprobado: 14/4/2025

Introducción

La inteligencia artificial (IA) ha transformado numerosos campos, y la ciencia de datos no es la excepción (González, 2023). Con el avance de las tecnologías, herramientas como ChatGPT y entornos de programación como RStudio se han consolidado como recursos poderosos para el análisis, procesamiento y representación de datos estadísticos. Esta integración no solo optimiza los flujos de trabajo, sino que también mejora la eficiencia en la toma de decisiones basadas en datos (Calsin *et al.*, 2023). En este contexto, el presente artículo explora la combinación entre la IA y RStudio, destacando cómo ChatGPT puede complementar el análisis estadístico al automatizar y simplificar tareas complejas (Ruiz y Pedroza, 2024). El procesamiento de datos estadísticos ha sido una tarea fundamental en diversas disciplinas, desde las ciencias sociales hasta la biología y la economía (Barreto-Villanueva, 2012). Con la creciente disponibilidad de grandes volúmenes de datos, el reto ha sido cómo procesarlos y visualizarlos de manera eficiente. RStudio, uno de los entornos más utilizados por estadísticos y científicos de datos, ofrece una plataforma flexible y potente para manejar datos complejos. Sin embargo, con la llegada de la IA, la capacidad de automatizar análisis y generar visualizaciones intuitivas ha cobrado un nuevo impulso. (Márquez, 2020). ChatGPT, desarrollado por OpenAI, es una IA avanzada basada en modelos de lenguaje natural que no solo es capaz de generar texto coherente y contextual, sino también de interactuar con otros programas (Hassani & Silva, 2023). En el ámbito del análisis de datos, esta herramienta puede ayudar a generar código, interpretar resultados estadísticos y proponer métodos de análisis, facilitando el trabajo de quienes no son expertos en programación, pero que necesitan obtener insights a partir de grandes conjuntos de datos (Hassani & Silva, 2023). La combinación de ChatGPT y RStudio abre nuevas posibilidades para el análisis de datos estadísticos. A través de esta colaboración entre IA y plataformas de código abierto, se pueden automatizar procesos que antes requerían conocimientos especializados en programación y análisis (Rahman & Watanobe, 2023). Esto reduce la curva de aprendizaje y permite a los profesionales de diversas áreas aprovechar los beneficios de la ciencia de datos sin la necesidad de ser expertos en el uso de R o en estadística avanzada (Yilmaz & Karaoglan Yilmaz, 2023). Una de las ventajas más destacadas de utilizar ChatGPT en RStudio es la capacidad de interpretar y predecir patrones en los datos con mayor facilidad (Shi *et al.*, 2023; Lingo, 2023). La IA puede asistir en la selección de modelos estadísticos adecuados, sugerir funciones y paquetes de R para tareas específicas, y explicar de manera comprensible los resultados obtenidos (Maksymiuk *et al.*, 2020). Además, al interactuar con RStudio, ChatGPT puede sugerir gráficos y técnicas de visualización que se ajusten a las características de los datos, lo que facilita la interpretación de los mismos (Kim *et al.*, 2023). El uso de la IA en la ciencia de datos no solo beneficia a los investigadores y analistas, sino que también tiene un impacto positivo en la enseñanza de la estadística. Con herramientas como ChatGPT, los estudiantes pueden aprender más rápido al contar con un asistente que explica conceptos, corrige errores y proporciona ejemplos prácticos en tiempo real (Altarawneh, 2023). Esto puede revolucionar la forma en que se enseña y aprende el análisis de datos, haciendo que la estadística sea más accesible y comprensible para un público más amplio. Otro aspecto relevante es la capacidad de ChatGPT para generar informes y documentación automatizada a partir de los análisis realizados en RStudio. Esta funcionalidad es clave para proyectos que requieren una presentación clara de los resultados, tanto para la comunidad académica como para la toma de decisiones en empresas. Con un lenguaje natural, la IA puede redactar resúmenes de los análisis, generar

interpretaciones de los gráficos y ofrecer recomendaciones basadas en los datos, lo que facilita la comunicación de hallazgos estadísticos complejos a audiencias no especializadas (Wang et al., 2024). La aplicación de la inteligencia artificial de ChatGPT en RStudio para el procesamiento y representación de datos estadísticos tiene un gran potencial para transformar el análisis de datos. Al automatizar tareas, mejorar la visualización y hacer más accesibles los procesos complejos, esta combinación de herramientas se perfila como un recurso clave para investigadores, analistas y docentes que buscan optimizar su trabajo con datos. En los siguientes apartados, se detallarán ejemplos prácticos de cómo se pueden integrar estas tecnologías en el flujo de trabajo del análisis estadístico.

Materiales y métodos

Se empleó un archivo Excel denominado "datos", el cual contenía información de nueve variables clave: sexo, edad, nivel de educación, índice de masa corporal (IMC), horas de trabajo semanal, talla del padre, talla de la madre, talla del hijo y talla de la hija. Cada variable fue organizada en columnas, utilizando nombres cortos, en minúsculas y sin tildes, facilitando así su manipulación en el entorno de análisis. Este formato fue cuidadosamente diseñado para asegurar la compatibilidad con el software estadístico RStudio.

Para el análisis visual de los datos, se seleccionaron diversas gráficas comúnmente empleadas en estadística, tales como: gráfica de barras, histograma, gráfica de dispersión, gráfica lineal, gráfica circular, diagrama de cajas y bigotes (box plot), y campana de Gauss. Cada uno de estos gráficos fue elegido en función de su capacidad para representar diferentes relaciones y distribuciones dentro de las variables mencionadas. Estos tipos de gráficos permitieron una exploración más profunda de los datos, resaltando patrones y tendencias que serían esenciales para el análisis estadístico posterior.

El software RStudio, versión 2024.09.0+375, fue utilizado para generar cada uno de los gráficos. Los códigos necesarios fueron creados a partir de prompts introducidos en la IA ChatGPT, que generó el script correspondiente en lenguaje R. Estos códigos fueron copiados y ejecutados en RStudio para procesar y visualizar los resultados. Este enfoque permitió automatizar y agilizar el proceso de creación de gráficos, mejorando la eficiencia y precisión en el análisis de los datos estadísticos.

Discusión y resultados

La Figura 1 presenta una gráfica de barras que ilustra los niveles de educación de las personas, según la información contenida en el archivo Excel cargado en RStudio. Para generar esta gráfica de barras, se utilizó el siguiente prompt: Tengo un archivo Excel subido a RStudio con el nombre "datos". Necesito representar en una gráfica de barras la columna "educacion", cada barra en diferentes colores.

Con base en este prompt, ChatGPT generó el siguiente código:

```
# Instalar los paquetes si no los tienes instalados
install.packages("readxl") # Para leer archivos Excel
install.packages("ggplot2") # Para hacer gráficos

# Cargar las librerías
library(readxl)
library(ggplot2)

# Leer el archivo Excel
datos <- read_excel("ruta_del_archivo/datos.xlsx")

# Gráfico con personalización
ggplot(datos, aes(x = educacion, fill = educacion)) +
  geom_bar() +
  labs(title = "Distribución de la educación", x = "Nivel de Educación", y = "Frecuencia") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Inclina las etiquetas del eje X
  scale_fill_brewer(palette = "Set3") # Cambia la paleta de colores
```

Al ejecutar el código proporcionado, se obtuvo la Figura 1, que representa visualmente la distribución de los niveles educativos. Esta gráfica de barras ofrece una visión clara y organizada de cómo se distribuyen las diferentes categorías de educación dentro del conjunto de datos, permitiendo identificar fácilmente tendencias y patrones. Muestra una gráfica de barras generada en RStudio utilizando el lenguaje de inteligencia artificial a través de ChatGPT. Esta herramienta facilitó la creación automatizada del código necesario para visualizar los datos educativos de manera eficiente y precisa, mejorando el análisis y la interpretación de la información contenida en el archivo Excel. Este proceso no solo acelera el proceso de generación de gráficos, sino que también demuestra el potencial del uso de modelos de IA como ChatGPT para asistir en tareas de análisis de datos, optimizando el flujo de trabajo y promoviendo una mayor accesibilidad a las técnicas de representación gráfica.

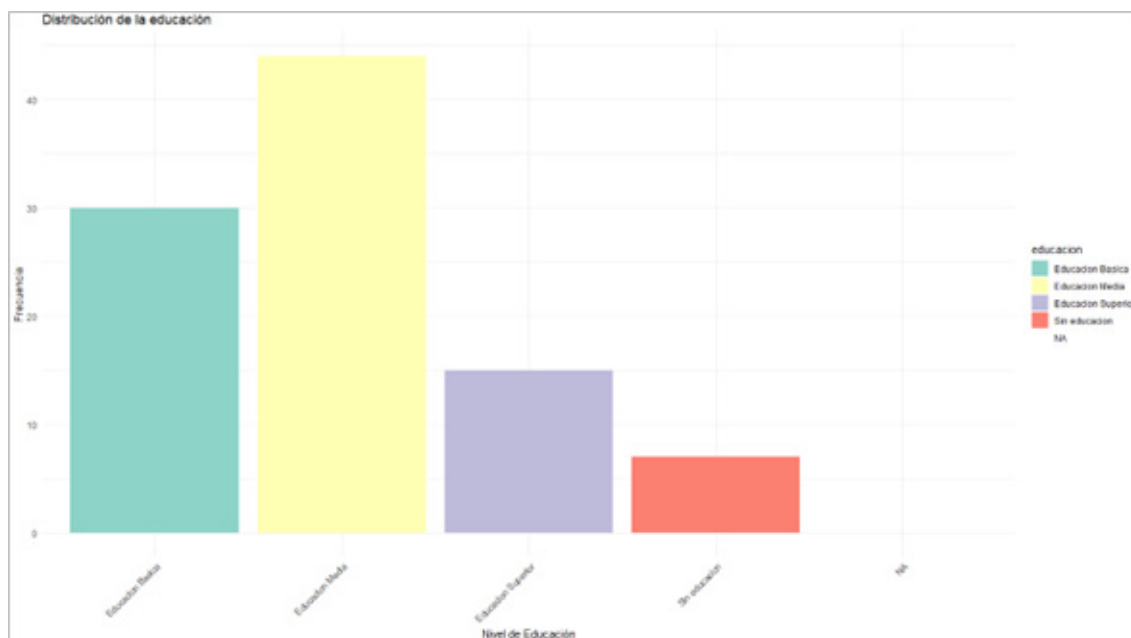


Figura 1. Gráfica de barra generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

Para Gutiérrez *et al.* (2004), las gráficas de barras son una técnica apropiada para representar variables cuantitativas discretas, ya que permiten identificar tendencias de manera clara y sencilla. En este caso, la IA no solo ha facilitado la creación del código necesario para generar la gráfica, sino que también ha optimizado el flujo de trabajo, lo que subraya el potencial de la IA en tareas analíticas complejas. El uso de IA, como se ha demostrado, tiene el potencial de automatizar y mejorar significativamente los procesos de visualización de datos, haciéndolos más accesibles para investigadores y analistas. Por otro lado, Uzcátegui y Ríos (2024) destacan que, aunque la inteligencia artificial ofrece múltiples ventajas, también plantea retos importantes en el ámbito educativo. El uso de IA en la educación debe ser cuidadosamente gestionado para asegurar que se implemente de manera ética y equitativa.

La Figura 2 presenta una gráfica de histograma que muestra la talla de las madres, según la información contenida en el archivo Excel cargado en RStudio. Para generar esta gráfica de histograma, se utilizó el siguiente prompt: Tengo un archivo excel subido al RStudio cuyo nombre es "datos". Necesito representar en una gráfica de histograma la "talla_madre"

Con base en este prompt, ChatGPT generó el siguiente código con dos opciones:

```
# Instala ggplot2 si no lo tienes instalado
```

```
install.packages("ggplot2")
```

```
# Cargar la librería
```

```
library(ggplot2)
```

```
# Graficar histograma de la columna "talla_madre"
```

```
ggplot(datos, aes(x = talla_madre)) +
```

```
  geom_histogram(binwidth = 2, fill = "lightblue", color = "black") +
```

```
  xlab("Talla de la madre") +
```

```
  ylab("Frecuencia") +
```

```
ggtitle("Histograma de 'talla_madre'")
```

Al ejecutar el código generado por ChatGPT, se logró visualizar la Figura 2, que presenta un histograma detallado de la distribución de las tallas de las madres. Este gráfico permite observar la frecuencia con la que se repiten diferentes rangos de estatura, ofreciendo una representación clara de la variabilidad en las tallas dentro de la población analizada. La visualización facilita la identificación de patrones o tendencias que podrían ser relevantes para un análisis más profundo de los datos demográficos.

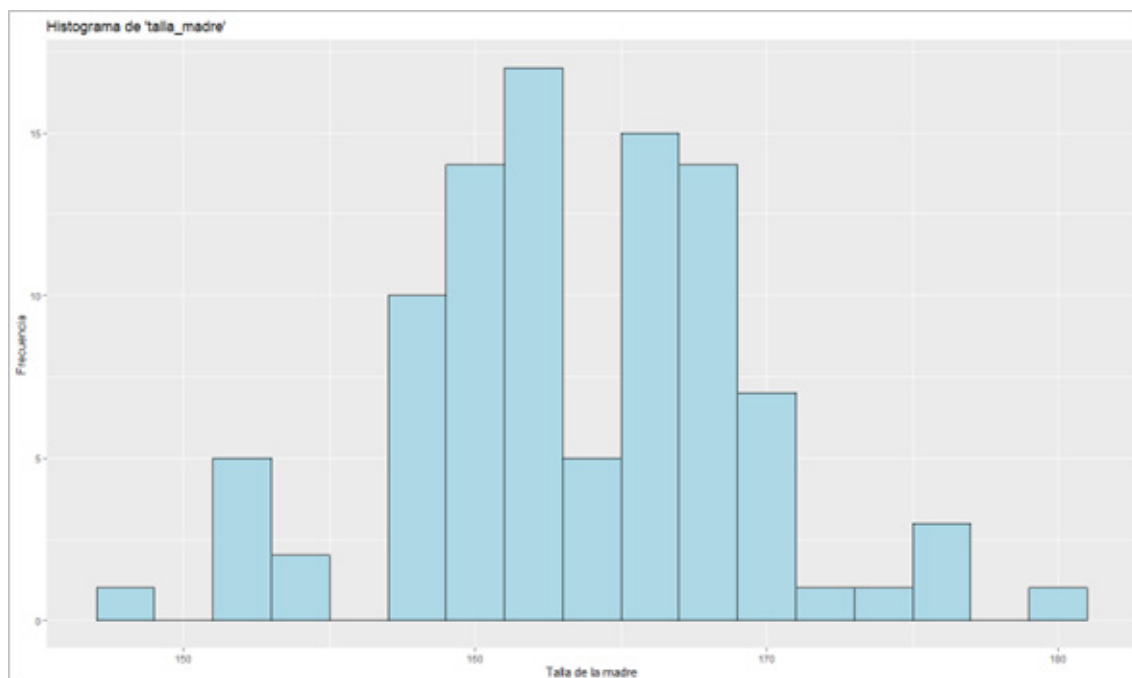


Figura 2. Gráfica de histograma generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

De acuerdo con Rivas *et al.* (2020), los histogramas de frecuencias son una herramienta útil para resumir visualmente la distribución de variables cualitativas, facilitando la observación de patrones o irregularidades en los datos. En la figura 2, cada barra del histograma representa un rango de estaturas y su frecuencia dentro de la población, lo que permite identificar con mayor precisión qué intervalos de talla son más comunes. Esta visualización resulta esencial para un análisis demográfico más detallado, permitiendo detectar tendencias que podrían estar relacionadas con factores biológicos, socioeconómicos o geográficos.

Es así, el análisis y la visualización de datos mediante inteligencia artificial, como destaca Rama (2024), aceleran y potencian los procesos de investigación en diversos sectores. La IA permite manejar grandes volúmenes de datos, identificar patrones y verificar hipótesis con una eficiencia inigualable. En el caso del histograma de tallas de las madres, el uso de IA facilita no solo la generación automática del código para crear el gráfico, sino también la posibilidad de realizar un análisis más profundo y preciso de los datos. Esto subraya la importancia del sector cuaternario, que abarca la creación de conocimientos a partir de herramientas tecnológicas avanzadas como la IA, contribuyendo a la investigación y el desarrollo en diversos campos.

La Figura 3 presenta una gráfica de dispersión que muestra la edad y horas de trabajo a la semana, según la información contenida en el archivo Excel cargado en RStudio. Para generar esta gráfica de dispersión, se utilizó el siguiente prompt: Tengo un archivo excel subido al RStudio cuyo nombre es "datos". Necesito representar en una gráfica de dispersión las columnas "edad" y "horas_trab_sema"

Con base en este prompt, ChatGPT generó el siguiente código con dos opciones:

```
# Instala ggplot2 si no lo tienes instalado
install.packages("ggplot2")
# Cargar la librería
library(ggplot2)
# Graficar dispersión de "edad" vs "horas_trab_sema"
ggplot(datos, aes(x = edad, y = horas_trab_sema)) +
  geom_point(color = "blue", alpha = 0.7) + # Agregar puntos
  xlab("Edad") + # Etiqueta del eje x
  ylab("Horas de trabajo por semana") + # Etiqueta del eje y
  ggtitle("Gráfico de Dispersión: Edad vs Horas de Trabajo Semanal") + # Título del gráfico
```

```
theme_minimal()
```

```
# Usar un tema minimalista
```

Al ejecutar el código generado por ChatGPT, se logró visualizar la Figura 3, que muestra una gráfica de dispersión entre la edad de los participantes y el número de horas trabajadas por semana. Esta representación gráfica permite identificar patrones y correlaciones entre ambas variables, facilitando el análisis de tendencias como la posible relación entre la edad y la carga laboral semanal. Además, la dispersión de los puntos sugiere la presencia de variabilidad en las horas trabajadas en función de la edad, lo que podría ser un indicio de diferencias en la distribución de tiempo de trabajo según el grupo etario.

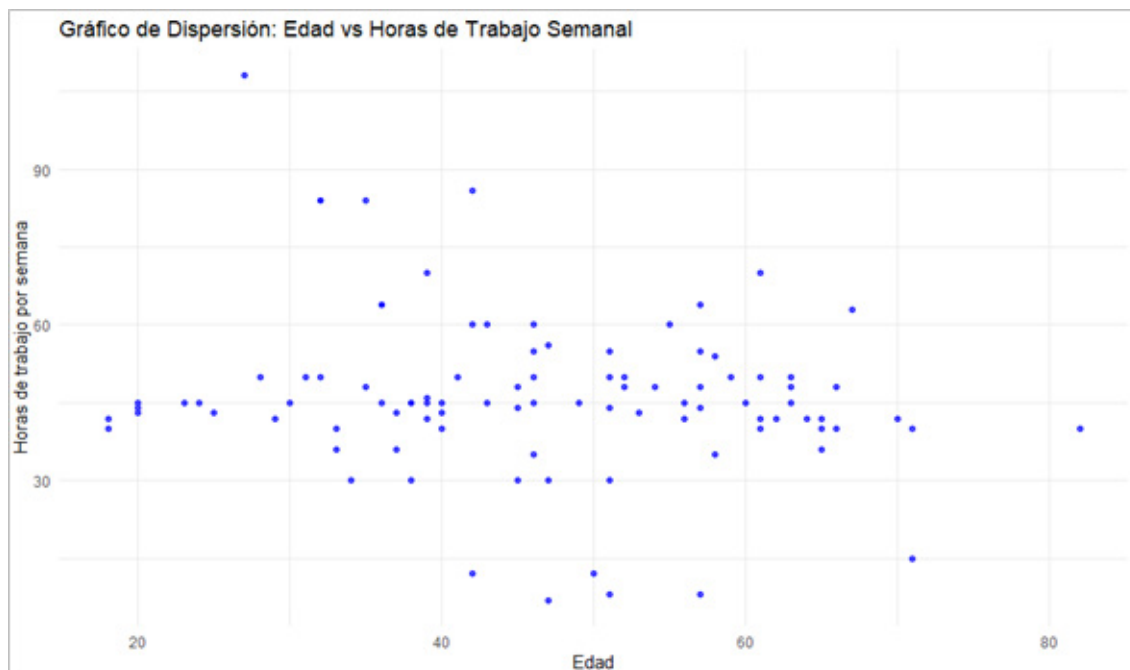


Figura 3. Gráfica de dispersión generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

Para Gutiérrez et al. (2004), los diagramas de dispersión permiten visualizar si existe una tendencia que sugiera una dependencia entre una medida y otra. En este caso, la representación gráfica facilita la observación de posibles patrones, como la correlación entre la edad y la carga laboral semanal. La dispersión de los puntos muestra variabilidad en las horas trabajadas según los diferentes grupos etarios, lo que podría estar relacionado con factores como el tipo de empleo, la etapa de la carrera profesional o las demandas laborales en función de la edad.

Sin embargo, el uso de inteligencia artificial en la creación de esta visualización, como en el caso del código generado por ChatGPT, plantea preguntas sobre el impacto de la tecnología en el análisis de datos. Puche (2024) señala que, aunque la IA ofrece ventajas claras como la mejora en la calidad de los análisis y la eficiencia en la enseñanza, también suscita preocupaciones sobre la pérdida de habilidades analíticas y la dependencia tecnológica. En este contexto, aunque la IA facilita el análisis de relaciones complejas entre variables, es importante que los investigadores y analistas mantengan un enfoque crítico y sigan participando activamente en la interpretación de los datos, evitando una dependencia excesiva de la tecnología para la toma de decisiones informadas. Así, el equilibrio entre el uso de IA y el juicio humano es clave para garantizar que las conclusiones obtenidas sean precisas y significativas.

La Figura 4 presenta una gráfica lineal que muestra la edad y el índice de masa corporal (IMC), según la información contenida en el archivo Excel cargado en RStudio. Para generar esta gráfica de dispersión, se utilizó el siguiente prompt: Tengo un archivo excel subido al RStudio cuyo nombre es "datos". Necesito representar en una gráfica lineal las columnas "edad" y "IMC"

Con base en este prompt, ChatGPT generó el siguiente código con dos opciones:

```
# Instala ggplot2 si no lo tienes instalado
install.packages("ggplot2")
```

```
# Cargar la librería
library(ggplot2)
```

```
# Graficar la relación entre "edad" y "IMC"
ggplot(datos, aes(x = edad, y = IMC)) +
  geom_line(color = "blue", size = 1) + # Línea de la gráfica
  geom_point(color = "red", size = 2) + # Agregar puntos
  xlab("Edad") + # Etiqueta del eje x
  ylab("Índice de Masa Corporal (IMC)") + # Etiqueta del eje y
  ggtitle("Gráfica Lineal: Edad vs IMC") + # Título del gráfico
  theme_minimal() # Usar un tema minimalista
```

Al ejecutar el código generado por ChatGPT, se logró visualizar la Figura 4, la cual presenta una gráfica lineal que muestra la relación entre la edad de los participantes y su Índice de Masa Corporal (IMC). La gráfica revela una tendencia clara, donde se observa cómo el IMC varía en función de la edad. Este análisis permite identificar posibles patrones de comportamiento, como un aumento o disminución del IMC en ciertas franjas etarias, lo que puede ser relevante para estudios de salud poblacional. Además, esta representación gráfica facilita la interpretación visual de los datos, apoyando la evaluación de la correlación entre las dos variables involucradas.

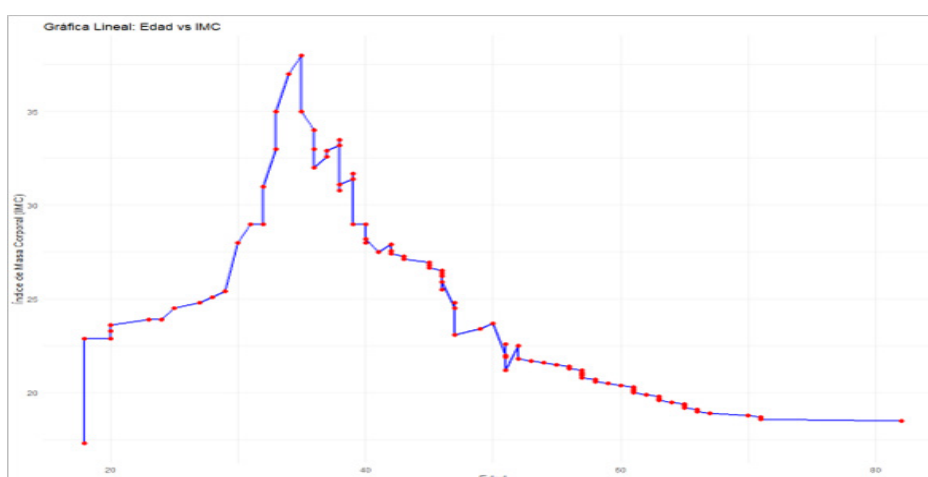


Figura 4. Gráfica lineal generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

Para Rivas et al. (2020), los gráficos de líneas son útiles para mostrar la frecuencia de una variable y su cambio a lo largo del tiempo o en función de una categoría ordinal, como en este caso, la edad. Esta visualización revela patrones claros, como un posible aumento o disminución del IMC con la edad, lo cual es fundamental para el análisis en estudios. El gráfico facilita la interpretación de tendencias, permitiendo identificar grupos etarios en los que el IMC pueda estar relacionado con factores de riesgo, lo que puede influir en la formulación de estrategias preventivas de salud pública.

No obstante, el uso de inteligencia artificial generativa (IAGen) para producir el código que generó este gráfico plantea algunas consideraciones sobre la precisión y utilidad de estas herramientas en el análisis de datos complejos. Corica et al. (2024) mencionan que la IAGen, aunque es una tecnología disruptiva con un gran potencial, presenta dificultades en la resolución de problemas que requieren representaciones gráficas y ubicaciones espaciales precisas. Aunque en este caso la IA ha facilitado la creación del gráfico lineal y la visualización de datos, es crucial que los investigadores analicen críticamente los resultados generados. Si bien la IA ofrece ventajas en términos de eficiencia y accesibilidad, los expertos deben validar las conclusiones obtenidas para garantizar que las interpretaciones sean rigurosas y se basen en una comprensión sólida de las relaciones entre las variables analizadas, especialmente en temas de salud donde las decisiones informadas son fundamentales.

La Figura 5 presenta una gráfica circular que muestra la variable sexo, según la información contenida en el archivo Excel cargado en RStudio. Para generar esta gráfica circular, se utilizó el siguiente prompt: Tengo un archivo excel subido al RStudio cuyo nombre es "datos". Necesito representar en una gráfica circular la columna "sexo"

Con base en este prompt, ChatGPT generó el siguiente código con dos opciones:

```
# Instalar ggplot2 si no lo tienes instalado
```

```
install.packages("ggplot2")

# Cargar la librería
library(ggplot2)

# Asegurarse de que la columna 'sexo' es un factor
datos$sexo <- as.factor(datos$sexo)

# Crear un gráfico circular (gráfico de pastel)
ggplot(datos, aes(x = "", fill = sexo)) +
  geom_bar(width = 1) +           # Crear barras
  coord_polar("y") +             # Convertir a coordenadas polares
  labs(title = "Distribución por Sexo", # Título del gráfico
       fill = "Sexo") +          # Etiqueta de la leyenda
  theme_minimal() +              # Usar un tema minimalista
  theme(axis.text.x = element_blank()) # Eliminar texto del eje x
```

Al ejecutar el código generado por ChatGPT, se logró visualizar la Figura 5, que presenta una gráfica circular (o diagrama de pastel) ilustrando la distribución de los participantes según su sexo. En dicha representación, se observa una predominancia del sexo masculino, lo cual sugiere un desequilibrio en la proporción de hombres y mujeres dentro de la muestra analizada. La gráfica facilita una visualización clara y rápida de esta información demográfica, permitiendo que se identifiquen fácilmente posibles sesgos o la necesidad de ajustar el diseño de la investigación para garantizar una mayor representatividad.

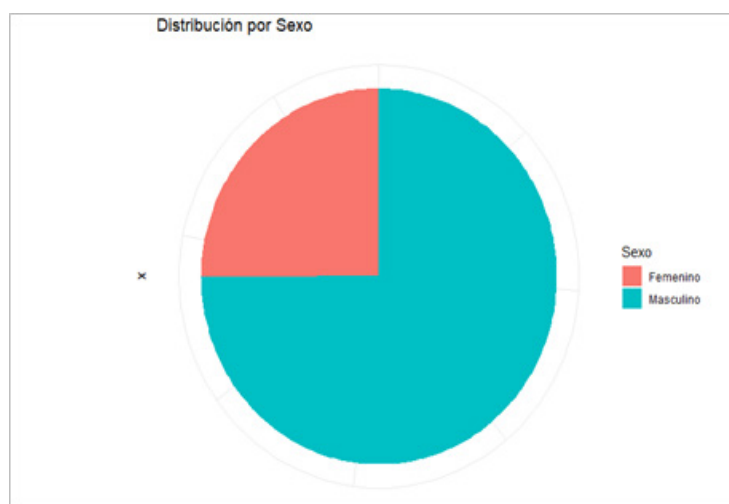


Figura 5. Gráfica circular generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

De acuerdo con Rivas *et al.* (2020), los gráficos de sectores o circulares son útiles para resumir una sola variable, dividiendo un círculo en fragmentos que representan la proporción relativa o absoluta de cada categoría. En este caso, la visualización rápida y clara permite identificar un posible desequilibrio de género, lo que sugiere que la muestra podría no ser representativa de la población general. Este sesgo demográfico, si no se corrige, podría afectar los resultados de la investigación y limitar la generalización de las conclusiones, destacando la importancia de ajustar el diseño de la investigación para asegurar una mayor equidad en la distribución de los participantes.

Desde una perspectiva más amplia, Rodríguez *et al.* (2024) exploran el impacto de la inteligencia artificial en el análisis de datos y subrayan tanto los beneficios como los desafíos que plantea. En este caso, la utilización de IA para generar automáticamente el código que produjo el gráfico circular facilita la visualización de datos demográficos y optimiza el proceso de análisis, mejorando la eficiencia del investigador. Sin embargo, al igual que en el contexto educativo, donde se enfrentan dilemas éticos como el plagio o la creación de contenido no auténtico, el uso de IA en el análisis de datos también puede plantear retos relacionados con la precisión y la interpretación de los resultados.

La Figura 6 presenta una gráfica de diagrama de cajas y bigotes o box plot las columnas que muestra la variable educación y edad, según la información contenida en el archivo Excel cargado en RStudio. Para generar este

diagrama de cajas y bigotes o box plot, se utilizó el siguiente prompt: Tengo un archivo excel subido al RStudio cuyo nombre es "datos". Necesito representar en una diagrama de cajas y bigotes o box plot las columnas "educacion" y "edad"

Con base en este prompt, ChatGPT generó el siguiente código con dos opciones:

```
# Instalar ggplot2 si no lo tienes instalado
install.packages("ggplot2")
# Cargar la librería
library(ggplot2)
# Crear un diagrama de cajas y bigotes para las columnas "educacion" y "edad"
ggplot(datos, aes(x = educacion, y = edad, fill = educacion)) +
  geom_boxplot() + # Crear el boxplot
  labs(title = "Diagrama de Cajas y Bigotes de Edad por Educación", # Título del gráfico
       x = "Educación", # Etiqueta del eje x
       y = "Edad") # Etiqueta del eje y
  theme_minimal() # Usar un tema minimalista
```

Al ejecutar el código generado por ChatGPT, se logró visualizar la Figura 6, que presenta un diagrama de cajas y bigotes (box plot), ilustrando la distribución de los participantes según su edad y sexo. Este tipo de gráfico es especialmente útil para resumir visualmente la dispersión de los datos y detectar posibles valores atípicos. En la gráfica, se muestran las medianas, los cuartiles y los rangos intercuartílicos de la edad, separados por categorías de sexo, lo que permite comparar cómo varía la edad en función de esta variable. Este análisis es esencial para entender si hay patrones significativos entre edad y sexo dentro de la muestra, y si estos pueden influir en otros aspectos del estudio, como las diferencias en salud, comportamiento o características sociodemográficas.

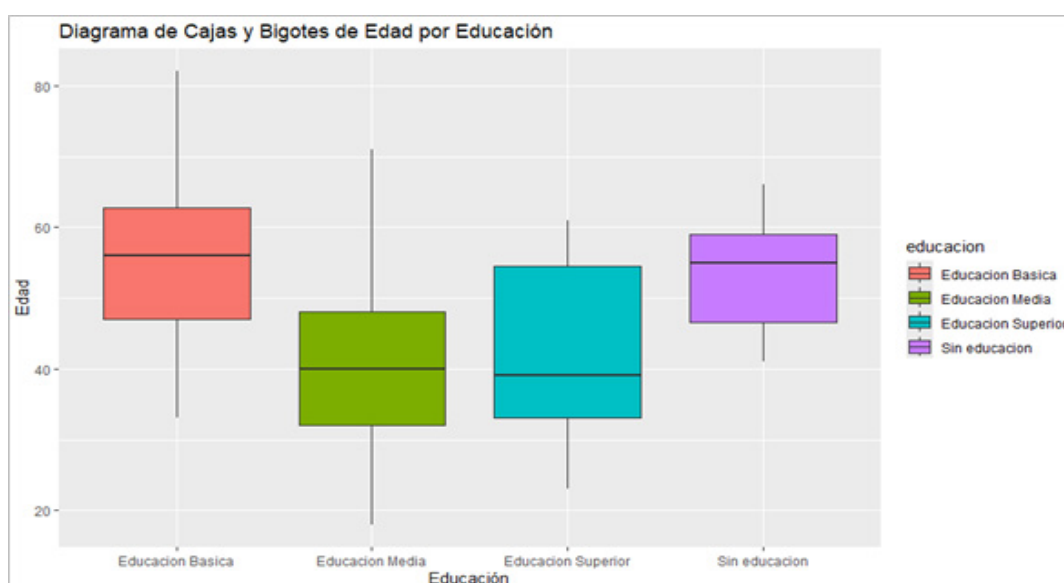


Figura 6. Gráfica de caja y Bigotes / Box plot generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

Según Rivas *et al.* (2020), este tipo de gráfico es ideal para representar variables cuantitativas con distribución libre, mostrando de manera clara las medianas, los cuartiles y los rangos intercuartílicos (RIC) de los datos. En este caso, la visualización facilita la comparación de cómo varía la edad en función del sexo, permitiendo detectar posibles diferencias entre ambos grupos. Además, el diagrama de cajas es útil para identificar valores atípicos que podrían ser relevantes para el análisis, ya que estos puntos fuera de lo común pueden ofrecer información importante sobre características específicas de la muestra. Este enfoque permite entender si existen patrones de edad y sexo que influyen en aspectos clave del estudio, como las diferencias en salud, comportamiento o características sociodemográficas.

El uso de inteligencia artificial para generar automáticamente este gráfico, como en el caso del código proporcionado por ChatGPT, facilita la representación y análisis de datos complejos. Sin embargo, como señala Monroy (2024), la IA es una herramienta mediadora en el proceso de análisis y no debe ser vista como un

fin en sí misma. Aunque los diagramas de cajas y bigotes son valiosos para obtener una visión rápida de la distribución de los datos, es esencial que los investigadores complementen estas visualizaciones con un análisis interpretativo profundo.

La Figura 7 presenta una gráfica de campana de Gauss con las variables de talla del padre y talla de la madre, según la información contenida en el archivo Excel cargado en RStudio. Para generar esta gráfica de campana de Gauss, se utilizó el siguiente prompt: Tengo un archivo excel subido al RStudio cuyo nombre es "datos". Necesito representar en una campana de Gauss la columna "talla_padre" y "talla_madre"

Con base en este prompt, ChatGPT generó el siguiente código con dos opciones:

```
# Asegúrate de que tienes instalado ggplot2
```

```
install.packages("ggplot2")
```

Modelo 1

```
# Cargar la librería
```

```
library(ggplot2)
```

```
# Crear la gráfica de densidad para talla_padre y talla_madre
```

```
ggplot(datos, aes(x = talla_padre)) +
```

```
  geom_density(aes(y = ..density..), fill = "lightblue", alpha = 0.5, color = "blue") +
```

```
  geom_density(aes(x = talla_madre, y = ..density..), fill = "lightpink", alpha = 0.5, color = "red") +
```

```
  labs(title = "Distribución Normal de Talla del Padre y Talla de la Madre",
```

```
        x = "Talla",
```

```
        y = "Densidad") +
```

```
  theme_minimal()
```

Modelo 2

```
ggplot(datos) +
```

```
  geom_violin(aes(x = "Padre", y = talla_padre, fill = "Padre"), alpha = 0.6) +
```

```
  geom_violin(aes(x = "Madre", y = talla_madre, fill = "Madre"), alpha = 0.6) +
```

```
  geom_boxplot(aes(x = "Padre", y = talla_padre), width = 0.1) +
```

```
  geom_boxplot(aes(x = "Madre", y = talla_madre), width = 0.1) +
```

```
  labs(title = "Comparación de Distribuciones: Talla Padre vs Madre",
```

```
        x = "Grupo",
```

```
        y = "Talla") +
```

```
  theme_minimal()
```

Al ejecutar el código generado por ChatGPT, se visualizó la Figura 7, que presenta una gráfica con forma de campana de Gauss, ilustrando la distribución de la estatura de madres y padres según su propia talla. Esta representación permite observar cómo la mayoría de las tallas se concentran en torno a un valor promedio, mientras que los valores más extremos se alejan hacia los márgenes de la gráfica, reflejando una menor frecuencia. Además, la gráfica permite identificar si existen diferencias significativas en la distribución de la talla entre madres y padres, destacando posibles patrones relacionados con la genética o el contexto socioambiental.

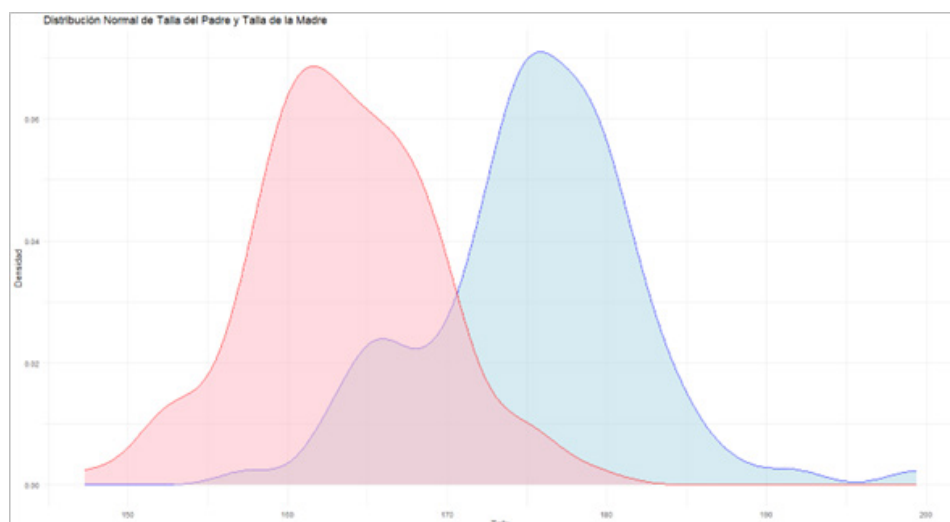


Figura 6. Gráfica de caja y Bigotes / Box plot generado en RStudio con lenguaje de IA ChatGPT

Fuente. Elaborado con RStudio con scrip generado por IA ChatGPT

Como señalan Fontanelli *et al.* (2021), la distribución normal se caracteriza por su simetría respecto a la media

y una disminución rápida en ambos extremos, lo que implica que las estaturas muy alejadas del promedio son inusuales. Esta representación visual es útil para analizar si las estaturas de madres y padres siguen esta distribución y si existen diferencias significativas entre ambos grupos, lo que podría estar relacionado con factores genéticos o socioambientales que influyen en el crecimiento y desarrollo.

Sin embargo, es importante considerar que, como menciona Barreto (2024), la representación de la realidad mediante funciones lógico-matemáticas, como las distribuciones estadísticas, tiene sus límites. Si bien la gráfica generada por ChatGPT permite una interpretación eficiente y visual de los datos, es crucial complementarla con un análisis más profundo que considere factores no visibles en la mera distribución estadística, como el contexto cultural, económico o de salud de la población estudiada. Este tipo de análisis cualitativo es fundamental para evitar que la inteligencia artificial, por sí sola, simplifique excesivamente la realidad, promoviendo conclusiones sesgadas o incompletas. Por lo tanto, el uso de herramientas de IA en combinación con un enfoque interdisciplinario puede ofrecer una visión más completa y precisa de los fenómenos analizados.

Conclusiones

La aplicación de la inteligencia artificial ChatGPT en RStudio para el procesamiento y representación de datos estadísticos ha demostrado ser una herramienta eficaz y eficiente. Al generar de manera automatizada gráficos que van desde simples barras hasta diagramas más complejos como cajas y bigotes o campanas de Gauss, se facilita la identificación de patrones, correlaciones y anomalías en los datos. Esto no solo mejora la comprensión y análisis de la información, sino que también optimiza el tiempo en la elaboración de resultados visuales, contribuyendo a una toma de decisiones más precisa y ágil en investigaciones.

La generación automatizada de gráficos mediante ChatGPT, utilizando prompts para crear código de RStudio, simplifica significativamente la creación de representaciones visuales de los resultados. Desde gráficas de barras que muestran los niveles educativos hasta distribuciones más complejas, como diagramas de cajas y bigotes que ilustran la relación entre edad y sexo, estas visualizaciones facilitan la identificación de patrones y tendencias clave en los datos. Esto contribuye a una interpretación más precisa y profunda de la información disponible. Además, este proceso resalta el potencial de las herramientas de inteligencia artificial para optimizar el flujo de trabajo en el análisis de datos, haciendo más eficiente la obtención de resultados visuales y su posterior análisis.

Referencias bibliográficas

- Altarawneh, H. (2023). ChatGpt impact on Student Educational Performance: a conceptual analysis. *EAI Endorsed Transactions on E-Learning*, 9. <https://doi.org/10.4108/eetel.4574>
- Barreto, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. *Papeles de Población*, 18(73), 241–271. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252012000300010
- Barreto, L. (2024). Desafíos de los sistemas y aplicaciones de la IA para la educación y la formación de ciudadanía. *Areté, Revista Digital Del Doctorado En Educación*, 10(ee), 213–226. http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29490
- Calsin, M., Aedo, M., & Castro, E. (2023). Impacto de ChatGPT en la enseñanza: Un enfoque de aula invertida para fundamentos de programación *Impact of ChatGPT on Teaching: A Flipped Classroom Approach for Programming Fundamentals*. <https://doi.org/10.17013/risti.52.97-112>
- Corica, A., Sureda, P., Parra, V., Schiaffino, S., & Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté, Revista Digital Del Doctorado En Educación*, 10(ee), 119–139. http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29456
- Fontanelli, O., Mansilla, R., & Miramontes, P. (2020). Distribuciones de probabilidad en las ciencias de la complejidad: una perspectiva contemporánea. *Inter disciplina*, 8(22), 11-37. <https://doi.org/10.22201/ceiich.24485705e.2020.22.76416>
- González, C. (2023). El impacto de la inteligencia artificial en la educación: transformación de la forma de enseñar y de aprender. *Curriculum. Revista de Teoría, Investigación Y Práctica Educativa*, 36(2), 51–60. <https://doi.org/10.25145/j.qurricul.2023.36.03>
- Gutiérrez, Á., Babativa, Y., & Lozano, I. (2004). Presentación de datos. *Revista Ciencias de la Salud*, 2(1), 65-73. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-72732004000100011&lng=en&tlng=es
- Hassani, H., & Silva, E. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data and Cognitive Computing*, 7(2), 62. <https://doi.org/10.3390/bdcc7020062>

- Kim, N. W., Myers, G., & Bach, B. (2023). How Good is ChatGPT in Giving Advice on Your Visualization Design? ArXiv.org. <https://arxiv.org/abs/2310.09617>
- Lingo, R. (2023). The Role of ChatGPT in Democratizing Data Science: An Exploration of AI-facilitated Data Analysis in Telematics. ArXiv.org. <https://doi.org/10.48550/arXiv.2308.02045>
- Maksymiuk, S., Gosiewska, A. & Biecek, P. (2020). Landscape of R packages for eXplainable Artificial Intelligence. ArXiv.org. <https://www.semanticscholar.org/paper/Landscape-of-R-packages-for-eXplainable-Artificial-Maksymiuk-Gosiewska/b33275a016d7d32d930ec3b70d46431d8fdc8c04>
- Márquez, J. (2020). Inteligencia Artificial y Big Data como soluciones frente al COVID-19. Revista De Bioética Y Derecho, (50), 315–331. <https://doi.org/10.1344/rbd2020.50.31643>
- Monroy, A. (2024). De la inteligencia artificial a una atención consciente: el modelo educativo Etievan. Areté, Revista Digital Del Doctorado En Educación, 10(ee), 195–211. Recuperado a partir de http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29488
- Puche, D. (2024). Inteligencia artificial como herramienta educativa: ventajas y desventajas desde la perspectiva docente. Areté, Revista Digital Del Doctorado En Educación, 10(ee), 85–100. http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29454
- Rahman, M., & Watanobe, Y. (2023). ChatGPT for Education and Research: Opportunities, Threats, and Strategies. Applied Sciences, 13(9), 5783. <https://doi.org/10.3390/app13095783>
- Rama, C. (2024). Cambios de las dinámicas de la investigación científica con las tecnologías de comunicación e información de la inteligencia artificial. Areté, Revista Digital Del Doctorado En Educación, 10(ee), 23–32. Recuperado a partir de http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29448
- Rivas, R., Roy, I., Pérez, M., Berea, R., Moreno, J., Moreno, M., Palacios, L. & Ureña, K. (2020). Pertinencia e impertinencia de los gráficos en la investigación clínica. Revista alergia México, 67(4), 381-396. <https://doi.org/10.29262/ram.v67i4.854>
- Rodriguez, M., Marín, J., & Maiuri, C. (2024). Perspectivas de la inteligencia artificial en la educación universitaria: un análisis basado en la literatura académica. Areté, Revista Digital Del Doctorado En Educación, 10(ee), 175–193. de http://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29459
- Ruiz, K., & Pedroza, L. (2024). Uso de ChatGPT como ayudante en una RSL con el método prisma. Scienceevolution. <https://www.semanticscholar.org/paper/%E2%80%8B%E2%80%8BUSO-DE-CHATGPT-COMO-AYUDANTE-EN-UNA-RSL-CON-EL-Mendoza-Z%C3%BA%C3%B1iga/d4ba0704186ef15ec7470ee5a0f499f59a70a036>
- Shi, Y., Ma, H., Zhong, W., Mai, G., Li, X., Liu, T., & Huang, J. (2023). ChatGraph: Interpretable Text Classification by Converting ChatGPT Knowledge to Graphs. ArXiv, abs/2305.03513. <https://doi.org/10.48550/arXiv.2305.03513>
- Uzcátegui, R. & Ríos, M. (2024). Inteligencia Artificial para la Educación: formar en tiempos de incertidumbre para adelantar el futuro. Areté, Revista Digital del Doctorado en Educación, 10 (ee), 1 –21 <https://doi.org/10.55560/arete.2024.ee.10.1>
- Wang, J., Ye, Q., Liu, L., Guo, N. L., & Hu, G. (2024). Scientific figures interpreted by ChatGPT: strengths in plot recognition and limits in color perception. Npj Precision Oncology, 8(1). <https://doi.org/10.1038/s41698-024-00576-z>
- Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. Computers in Human Behavior: Artificial Humans, 1(2), 100005. <https://doi.org/10.1016/j.chbah.2023.100005>
- Declaración de conflictos de intereses:** Los autores declaran no tener ningún conflicto de intereses.
- Declaración de contribución de los autores/as utilizando la Taxonomía CRediT:**
- Angel Aronés-Cisneros y Cynthia Alegre-Palomino: Realizó la investigación y aplicó la metodología.
- Rocío Aronés-Cisneros: Redacción del artículo científico.
- Leticia Tapia-Oré: Edición y revisión del artículo científico.
- Declaración de aprobación por el Comité de Ética:** Los autores declaran que la investigación fue aprobada por el Comité de Ética de la institución responsable, en tanto la misma implicó a seres humanos
- Declaración de originalidad del manuscrito:** Los autores confirman que este texto no ha sido publicado con anterioridad, ni ha sido enviado a otra revista para su publicación.