


"Prototipo inteligente para analizar el estado del arte en investigaciones científicas usando ChatGPT"

Christian Ovalle¹

Departamento de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad Privada del Norte, Lima, Perú

Información del artículo	Resumen (10 PT)
<p>Historia del artículo:</p> <p>Received month dd, yyyy Revised month dd, yyyy Accepted month dd, yyyy</p> <hr/> <p>Palabras clave:</p> <p>Inteligencia artificial IA Generativa GPT-4o Validación de contenido Ética en IA</p>	<p>La inteligencia artificial generativa, representada por modelos como GPT-4o, ha revolucionado la producción de contenido técnico y científico. Este estudio explora los avances recientes y los desafíos asociados con la generación automatizada de textos de alta precisión y coherencia. Se discute la evaluación de la calidad del contenido generado, resaltando la necesidad de discernir entre resultados de IA y contribuciones humanas en contextos críticos como la investigación científica. Además, se examinan las innovaciones tecnológicas, como el Mixture of Experts y el switch transformer, diseñadas para mejorar la fiabilidad y relevancia del texto generado. Este análisis subraya la importancia de desarrollar métodos robustos de validación y verificación para garantizar la exactitud del conocimiento producido por IA. En conclusión, se destaca el potencial transformador de la IA generativa en la ciencia, junto con las consideraciones éticas y técnicas fundamentales para su implementación efectiva y responsable.</p> <p><i>This is an open access article under the CC BY-SA license.</i></p> 

Corresponding Author:

Christian Ovalle

Department of Electrical Engineering, Faculty of Engineering, Universidad Privada del Norte
Lima, Perú

Email: denis.ovalle@upn.pe

1. INTRODUCCIÓN

En el contexto actual, el contenido generado por inteligencia artificial (IA) a menudo carece de precisión y calidad. Por ejemplo, Según una escala de 1 a 5, las respuestas generadas por GPT-3 muestra métricas que indican diversos puntajes para Habilidades básicas con 4.17, Habilidades científicas de nivel bachillerato con 3.76, nivel de maestría con 3.35, nivel de doctorado con 3.43 y habilidades más allá del conocimiento científico con 2.11[1], y donde las decisiones tomadas por estas mismas son, en esencia, regularmente aceptables en un entorno controlado [2], se pueden observar distintos patrones que podrían ser catalogados como desfavorables en distintas áreas como en la medicina y ciencias, específicamente en el área de conocimientos que involucren literatura científica [3]. Asimismo, se ha buscado calibrar este aspecto con técnicas creativas de monitoreo y evaluación [4]. No obstante, en paralelo se han desarrollado asistentes que sirven como apoyo para intentar minimizar el impacto negativo que algunas IA generativas puedan naturalmente provocar [5]. A pesar de ello, podemos incluir, bajo este rol de asistente/copiloto en el área de principalmente medicina dan datos principalmente precisos, aunque se sigan dando estas falencias [6]. Cabe destacar, que, actualmente existen algunas IA generativas muy preferibles unas antes otras, y que algunos expertos manifiestan su preferencia hacia las que proporcionan comprensibilidad[7]. Inclusive, en otros sectores como el periodismo, se han aplicado chatbots que han sido alimentados con una muestra muy limitada,

lo que, a corto plazo, puede parecer un problema, determinando en inexactitud el contexto en el cual se planea construir una conversación, teniendo como data un total de 160 noticias, 20 de cada uno de los diarios nacionales, [8]. Sin embargo, el hecho de generar contenido a través de IA generativas es en esencia una práctica que reduce el tiempo de trabajo realizado por una persona común y corriente, esto no se presenta como un problema, sin embargo, para algunos expertos es tedioso determinar si un texto es generado por IA o por humanos. [9]. Como consecuencia, debido a la reciente ola de tecnologías emergentes relacionadas a la inteligencia artificial, se tiene un campo o espectro muy prematuro para determinar conclusiones favorables en campos como la educación, se encontró que, en promedio, GPT-3 tuvo una puntuación de 9,7 en fluidez y 10,0 en patrones tonales, en comparación con los promedios de 6,45 y 7,01 de los estudiantes en esos aspectos respectivamente[10]. En otro orden de ideas, existen estudios donde se proponen soluciones de clasificación de problemas legales empleando modelos GPT-3, con el propósito de obtener un mejor score o superar experimentos anteriores manteniendo el marco de desarrollo y el contexto [11], paralelamente, en el marco de la literatura científica, se busca automatizar y aplicar inteligencia artificial con el propósito de obtener nuevas perspectivas frente a anteriores experimentos [12] . Algo semejante ocurre con la necesidad de emplear sistemas inteligentes en varios aspectos trae consigo el descubrimiento de nuevas necesidades que teóricamente podrían solucionar muchos problemas repetitivos y complejos [13] , Sobre todo, reconocer que actualmente muchas organizaciones, negocios, instituciones, etc. Están adoptando o creando tecnologías estrechamente relacionadas a la inteligencia artificial que además de beneficiar en reducción de tiempo y costo, también solucionan problemas en áreas ajenas en las que comúnmente se aplicarían, como lo son en la literatura científica [14] , Aunque en otro sentido la necesidad de emplear tecnologías emergentes como ChatGPT radica en que existe una alta demanda o interés en solucionar, esto podría ser llamado como una oportunidad para emplearlos, sin embargo, la precisión o capacidad de modelos de GPT pueden no ser realmente precisas y trae como resultado discrepancias o confusión [15].

En primer lugar, el estudio 'Analysing the impact of ChatGPT in research' evalúa cuatro herramientas de detección de contenido generado por ChatGPT utilizando métricas para medir la fluidez y coherencia del texto generado. La perplejidad, que se define como la medida de la dificultad de comprensión del texto, muestra que un texto con baja perplejidad es más coherente y comprensible. Por otro lado, se considera la intermitencia, que refleja la variabilidad en la calidad del texto generado. Una baja intermitencia indica un texto más consistente y de alta calidad. Además, se comparan las tasas de acierto de estas herramientas mediante matrices de confusión en conjuntos de datos de referencia, obteniendo tasas superiores al 85% en la detección de contenido generado por ChatGPT [16]. Por consiguiente, Los grandes modelos de lenguaje, a pesar de su éxito en tareas de procesamiento del lenguaje natural, carecen intrínsecamente de la capacidad para entender el significado del lenguaje [17]. Para abordar esta limitación , se ha introducido el Mixture of Experts (MoE) en deep learning, Este enfoque como tal asigna diferentes parámetros para cada entrada, generando modelos dispersamente activados, en paralelo, surge el switch transformer, que simplifica el MoE y propone modelos mejorados con reducción en costos de procesamiento y de comunicación [18]. Estas innovaciones están contribuyendo a mejorar la capacidad de comprensión de los modelos de lenguaje y a abordar ciertos desafíos que aparecen en su implementación. Del mismo modo, es importante destacar la importancia del desarrollo de la IA en la actualidad, abarcando una amplia sección de áreas, desde la detección de enfermedades hasta la accesibilidad de información en determinados idiomas lo que resulta beneficioso, además, es crucial resaltar que, durante los últimos seis años ,el potencial de la IA ha transformado significativamente la manera en la que se organiza y accede a la información , marcando una nueva etapa en la evolución tecnológica . Por otro lado, se presenta a “Bard”, un servicio experimental de IA conversacional que aspira a integrar el conocimiento del mundo con la inteligencia y la creatividad de los modelos de lenguaje[19] Esta iniciativa representa un paso adelante en la exploración de las capacidades de la IA en la generación de contenido y la interacción conversacional avanzada. Como contrapartida, ChatGPT ha alcanzado un éxito cultural en menos de dos meses, principalmente debido a su capacidad para generar texto basado en indicaciones escritas. El éxito significativamente rápido plantea preguntas sobre como la tecnológica , logra influencia en la percepción publica de IA y como podría afectar la forma en que enseña e investiga. Aunque ChatGPT utiliza aprendizaje por refuerzo a partir de retroalimentación humana, ocasionalmente produce respuestas incorrectas o sin sentido. Lo que plantea preocupaciones sobre cómo la tecnología puede impactar negativamente a la calidad de la información en la educación e investigación científica. La evaluación automática de respuestas enfocadas en texto y en la educación superior, enfrenta desafíos en la identificación de respuestas incoherentes o incorrectas, lo que destaca la necesidad de reforzar estos problemas. Se observa que la evaluación automática de respuestas basadas en texto en la educación superior plantea desafíos significativos, especialmente en la identificación de respuestas incoherentes o incorrectas[4].Además, el artículo de Nature hace mención que los resúmenes escritos por ChatGPT pueden engañar a los científicos , lo que subraya la importancia de garantizar la calidad y la veracidad de la información generada por estos sistemas[20]. Por lo tanto , esto es especialmente relevante

en la escritura de ensayos y la creación de artículos académicos, donde la precisión y la fiabilidad son fundamentales.[21]. Por lo general, los investigadores enfrentan dificultades para distinguir entre resúmenes generados por IA y originales. Esta situación plantea desafíos importantes en la confiabilidad y credibilidad de la información generada por la IA, lo que resalta la necesidad de desarrollar métodos efectivos para verificar la autenticidad del texto generado[20].

En esta investigación, se llevó a cabo una exploración innovadora para evaluar la capacidad de GPT-4 mediante un proceso de retroalimentación continua. El objetivo era determinar el estado del arte de un artículo científico. Esta solución presupone la capacidad de GPT-4 de estar en el mismo nivel o superior al estado actual mediante la aplicación de técnicas de retroalimentación y almacenamiento en memoria [22]. El algoritmo de retroalimentación utilizará técnicas como la programación dinámica y la memorización para comprender cómo GPT-4 genera el estado del arte de una investigación. Esto implica procesar diferentes estados del arte y utilizarlos como nuevas entradas [23]. La implementación de este proceso, combinada con una interfaz amigable para el usuario final, simplificará el análisis exhaustivo necesario para generar el estado del arte de una investigación científica, al tiempo que reducirá la carga de trabajo manual para los investigadores. Además, esta solución aborda los desafíos únicos enfrentados por diversos investigadores al proporcionar un enfoque unificado y automatizado para la evaluación del estado del arte. Al ofrecer una interfaz intuitiva y colaborativa para el usuario final, esta implementación no solo simplifica el proceso de análisis, sino que también promueve la colaboración entre investigadores al facilitar el intercambio de información. Asimismo, sienta las bases para futuras investigaciones y desarrollos en el campo de la inteligencia artificial. En síntesis, nuestra innovación viene a cubrir la brecha metodológica que implica generar el estado del arte de manera tradicional a una manera automática, para ello entendemos que nuestro modelo de retroalimentación basado en almacenamiento en memoria ayudará a generar un estado del arte totalmente coherente y conciso [24].

2. METODOLOGÍA

El propósito metodológico para mejorar la calidad del estado del arte generado por medio de ChatGPT se detallan a continuación. Nuestra propuesta está compuesta por un flujo de procesos que, inicialmente, cuenta con un input (carga de archivos) donde el sistema web recibe los archivos, posteriormente, recibimos el documento y lo transformamos a texto con el fin de almacenar cada archivo en una variable que posteriormente será usada para hacer funcionar el algoritmo, una vez tengamos el contenido de los archivos a texto procederemos a generar el estado del arte de los artículos de investigación, estos textos serán evaluados para determinar si los documentos cumplen con las características necesarias para generar el estado del arte con GPT-4 [25]. Inicialmente se tienen 7 aspectos fundamentales encontrados en otros autores para la evaluación del estado del arte [26]. Acto seguido se vuelven a generar el estado del arte a través de GPT-4 para luego realizar una comparativa y determinar si el primer estado del arte generado con GPT-4 cumple con los 7 aspectos previamente configurados[27]. En caso que la API de GPT-4 devuelva un resultado parcial, inconcluso, genérico o en su defecto que no contenga información realmente considerada, será reevaluado por segunda vez y, si en todo caso, el resultado vuelve ser genérico y sin ningún tipo de aporte de valor, será informado hacia el usuario que el artículo que ha subido no contiene información relevante para determinar el estado del arte[28]. A continuación, se describen las etapas involucradas

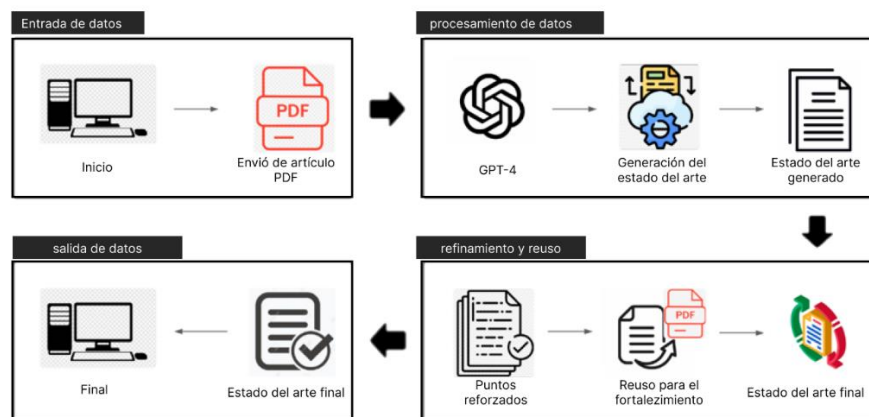


Figura 1. Modelo de generación del estado del arte

2.1 Envío del artículo PDF:

En esta etapa, el proceso de importación y limpieza de documentos PDF enviados desde el frontend se realiza en varias etapas para asegurar que los datos estén en un formato adecuado y libre de errores antes de ser procesados. Primero, se hace una carga los documentos PDF a través de la interfaz web, donde se almacenan temporalmente en el servidor. A continuación, estos documentos PDF se convierten a texto utilizando bibliotecas como pdfminer. El texto extraído se almacena en variables adecuadas y se somete a un proceso de normalización. Posteriormente, el texto se envía como parte del prompt que será enviado a la API de GPT-4.

```

1  import os
2  from pdfminer.high_level import extract_text
3  from docx import Document
4  import re
5  import nltk
6  from spellchecker import SpellChecker
7
8
9  def receive_files(upload_dir):
10     files = [os.path.join(upload_dir, f) for f in os.listdir(upload_dir) if os.path.isfile(os.path.join(upload_dir, f))]
11     return files
12
13  def convert_to_text(file_path):
14     text = ""
15     if file_path.endswith('.pdf'):
16         text = extract_text(file_path)
17     elif file_path.endswith('.docx'):
18         doc = Document(file_path)
19         for para in doc.paragraphs:
20             text += para.text + '\n'
21     return text
22
23  def clean_text(text):
24
25     text = re.sub(r'[\w\s]', '', text)
26     text = text.lower().strip()
27     spell = SpellChecker()
28     corrected_text = " ".join([spell.correction(word) for word in text.split()])
29     return corrected_text
30

```

Fig 2. Importando archivos enviados por el frontend

2.2 Estado del arte con GPT

En la fase inicial, se utiliza el modelo GPT-4 para solicitar el estado del arte en el área de estudio específica, aprovechando su capacidad para generar contenido basado en una amplia gama de datos y conocimientos previos. Una vez obtenida esta información, se almacena en la memoria del sistema para su posterior consulta y referencia. Este paso es crucial, ya que el estado del arte ofrece una visión panorámica de los avances más recientes en el campo de estudio, estableciendo un marco de referencia sólido para el desarrollo y la investigación futura. El proceso incluye la introducción de datos relevantes, la generación de contenido por GPT-4, la verificación y ajuste del contenido, y su almacenamiento eficiente para futuras etapas de investigación.

```

51
52     MODEL="gpt-4o"
53
54     completion = client.chat.completions.create(
55         model=MODEL,
56         messages=[
57             {"role": "system", "content":
58             | "Eres un investigador científico en ingeniería informática, debes cuidar tu
59             | {"role": "user", "content": "Genera el estado del arte de la siguiente infor
60
61         ]
62     )
63     app.logger.info(completion.choices[0].message.content)
64
65     return jsonify({"text_responses": text_responses,
66                    | "completion": completion.choices[0].message.content
67                    | }), 200
68

```

Fig 3. Estado del arte generado con el modelo GPT-4o

2.3 Estado del arte generado

En esta etapa, se solicita a ChatGPT que elabore un estado del arte detallado y contextualizado, analizando exhaustivamente las investigaciones y desarrollos recientes. Esto incluye identificar tendencias, avances y áreas clave de interés, así como evaluar métodos, enfoques y perspectivas teóricas relevantes. La integración del contexto histórico y las implicaciones prácticas proporciona una comprensión más profunda y crítica de los avances en el campo. El proceso implica la generación de contenido detallado por GPT-4, que captura no solo hallazgos y resultados, sino también los métodos y teorías que sustentan las investigaciones, estableciendo una base sólida para evaluar críticamente los desarrollos recientes.

2.4 Desglose del estado del arte – grupo de archivos

En esta fase del proceso, se desglosa el estado del arte en siete subtemas fundamentales, lo que permite abordar exhaustivamente los aspectos más relevantes del campo de estudio. Esta división facilita la comprensión y análisis detallado de la situación actual, estableciendo una estructura organizada que guía la revisión bibliográfica. Se identifican los subtemas clave, se solicita a GPT-4 que genere un estado del arte específico para cada uno, y la información resultante se almacena y organiza adecuadamente. Esto

```

1  import openai
2  openai.api_key = 'apikey'
3  def generate_subtopic_state_of_the_art(subtopic):
4      prompt = f"Genera un estado del arte detallado sobre el subtema: {subtopic}."
5      response = openai.Completion.create(
6          engine="gpt-4",
7          prompt=prompt,
8          max_tokens=1500
9      )
10     subtopic_state_of_the_art = response.choices[0].text.strip()
11     return subtopic_state_of_the_art
12 def store_in_memory(content, filename):
13     with open(filename, 'w') as file:
14         file.write(content)
15
16 def process_subtopic_state_of_the_art(subtopics):
17     all_subtopics_content = {}
18     for subtopic in subtopics:
19         state_of_the_art = generate_subtopic_state_of_the_art(subtopic)
20         filename = f'{subtopic.replace(" ", "_")}_state_of_the_art.txt'
21         store_in_memory(state_of_the_art, filename)
22         all_subtopics_content[subtopic] = state_of_the_art
23     return all_subtopics_content

```

Fig 4. Dividir la información del archivo en subtemas

permite una fácil consulta y referencia, facilitando el análisis comparativo y la identificación de tendencias, enfoques y hallazgos relevantes para enriquecer y fundamentar el marco teórico de la investigación.

2.5 Reutilización para fortalecer el estado del arte

En esta etapa del proceso, se utiliza el estado del arte previamente desarrollado como punto de partida. Luego, se lleva a cabo una segunda evaluación que incluye tanto el estado del arte original como el que se generará en esta fase. El objetivo es determinar cuál de los dos conjuntos de información proporciona un contexto más amplio y profundo sobre el tema en cuestión. Este enfoque dual permite identificar áreas de convergencia y discrepancia, así como posibles lagunas en el conocimiento. Esta comparación crítica enriquece la comprensión del tema y guía futuras investigaciones y decisiones en el campo correspondiente.

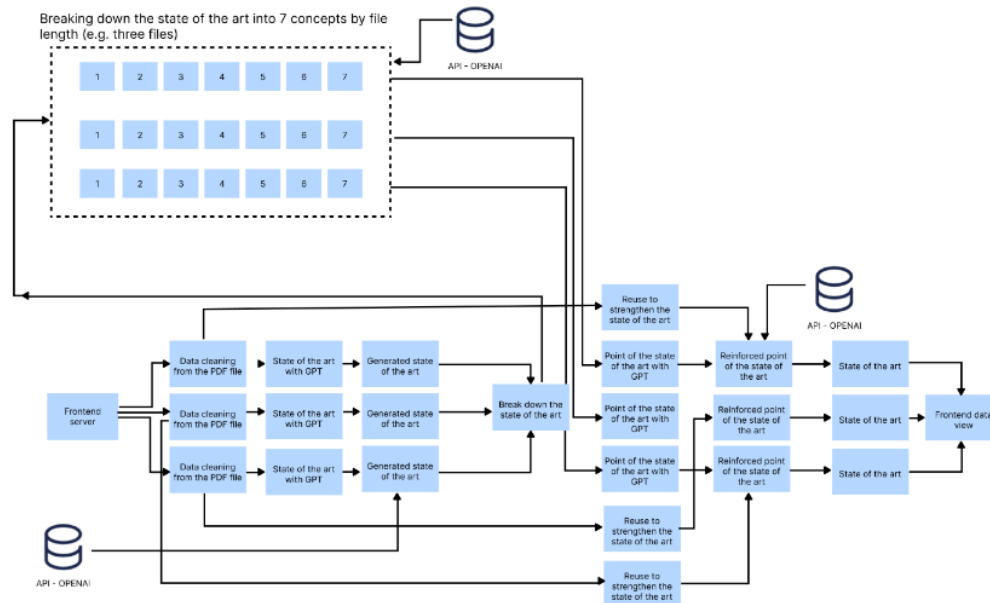


Figura 5. Arquitectura del sistema de generación del estado del arte

2.6 Punto del estado del arte con GPT

En esta etapa, se solicita a ChatGPT que elabore el estado del arte, aprovechando el estado del arte previamente generado, con un enfoque en la inclusión de contexto. Esto implica analizar exhaustivamente las investigaciones y desarrollos recientes, identificando tendencias, avances y áreas clave de interés. Se busca no solo capturar los hallazgos y resultados, sino también los métodos, enfoques y perspectivas teóricas relevantes. La inclusión de datos contextuales como el contexto histórico y las implicaciones prácticas establece una base sólida para comprender y evaluar críticamente los avances en el campo.

2.7 Punto del estado del arte reforzado

Durante esta fase, nos adentramos en la exploración exhaustiva del estado del arte en el campo relevante, fortaleciendo nuestro conocimiento mediante el empleo de un algoritmo específico. Esta metodología nos permite profundizar aún más en el panorama actual, identificando y analizando las últimas tendencias, investigaciones y avances en el área de interés. Al integrar este análisis con la aplicación estratégica de un algoritmo especializado, obtenemos una perspectiva más robusta y enriquecida del estado actual del conocimiento, lo que nos capacita para avanzar de manera más informada y precisa en nuestro trabajo.

2.8 Estado del arte

Durante esta fase inicial, nos enfocamos en la preparación exhaustiva del estado del arte, un proceso crucial para establecer una base sólida y bien informada. Este estado del arte, meticulosamente elaborado, es luego formateado en JSON para su óptima utilización y consumo por parte del front-end de nuestra aplicación. Este proceso de formateo se realiza con el propósito específico de garantizar una integración sin problemas con las tecnologías que estamos utilizando, incluyendo Axios y React.js. De esta manera, aseguramos que la información es entregada de manera eficiente y efectiva, facilitando su manipulación y presentación dentro de nuestra interfaz de usuario desarrollada en React.js. Este enfoque nos permite no solo mantener la coherencia en la presentación de datos, sino también optimizar el rendimiento y la experiencia del usuario.

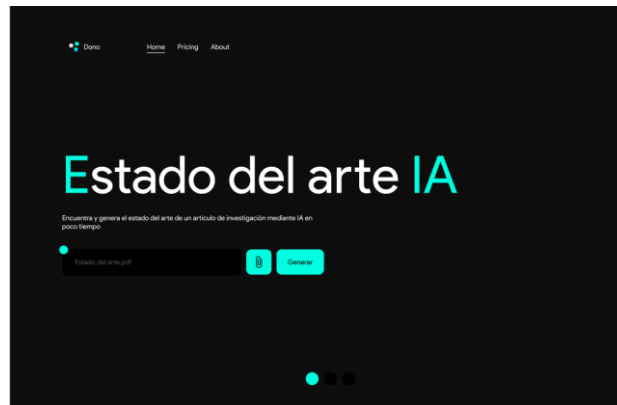


Fig 6. Concepto de la interfaz de usuario realizada en figma

2.9 Estado del arte en la respuesta del servidor Backend

Durante esta fase, se configura la información de vanguardia contenida en la respuesta generada por la API de OpenAI. Esta información se formatea en JSON (Javascript object notation) y se envía al servidor frontend. Es fundamental destacar que, en ciertos escenarios, la API puede no responder adecuadamente, lo que podría resultar en un error interno. Dicho potencial fallo requiere una estrategia robusta de manejo de errores por parte del servidor, que debe estar preparado para gestionar tales situaciones y proporcionar notificaciones claras al usuario final, sin embargo esto es un escenario de fallo.

2.10 Recepción y visualización de la información

Inmersos en la etapa de visualización de información, los usuarios podrán acceder al estado del arte generado a partir del artículo de investigación científica mediante una interfaz intuitiva y fácil de entender. Solo se mostrará la información relevante contenida en el paquete de datos JSON proporcionado por el servidor, lo que permite a los usuarios enfocarse únicamente en los datos que les conciernen. La interfaz está diseñada para presentar los datos de manera clara y concisa, destacando los puntos más importantes y facilitando la comprensión rápida de la información crítica. Este enfoque no solo mejora la experiencia del usuario, sino que también asegura que la información se presente de manera eficiente y efectiva, eliminando cualquier ruido innecesario.

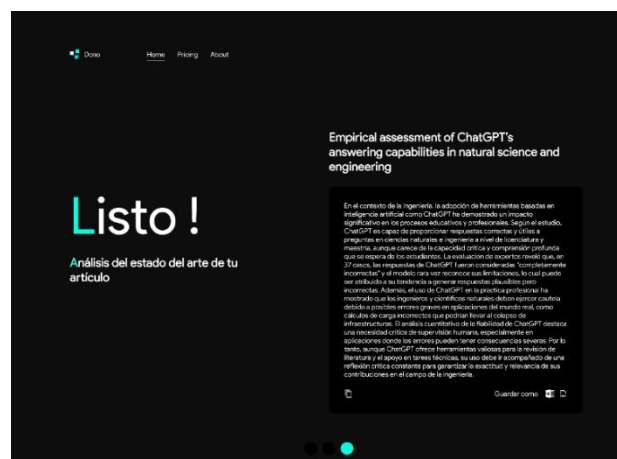


Fig 7. Concepto de la interfaz de usuario resultado realizado con figma

3. RESULTADOS Y DISCUSIONES

En la evaluación de la calidad del estado del arte generado por GPT-4, se llegó a un total de 20 documentos PDF convertidos a texto. Los estados del arte iniciales fueron evaluados según siete aspectos fundamentales: claridad y coherencia, relevancia del contenido, cobertura del tema, precisión de la información, originalidad del análisis, estructura y organización, y concisión. Los resultados iniciales

mostraron un cumplimiento promedio del 79% en estos aspectos. Posteriormente, se generaron nuevos estados del arte para 15 documentos que no cumplieron con los criterios establecidos, resultando en una mejora promedio del 10%. La entropía del texto original y del estado del arte generado fue de 6.5 bits y 5.8 bits, respectivamente, indicando que se mantuvo una buena cantidad de información relevante tras la síntesis. Sin embargo, para este experimento se consideraron dos alternativas como la entropía de datos y la evaluación de expertos.

3.1 Entropía de datos

La entropía de datos, es una medida cuantitativa de la incertidumbre o diversidad de información, y en la presente investigación se justifica como una herramienta importante en la evaluación de la calidad de resúmenes generados, particularmente en el contexto de aplicaciones como la generación de estados del arte con GPT-4.

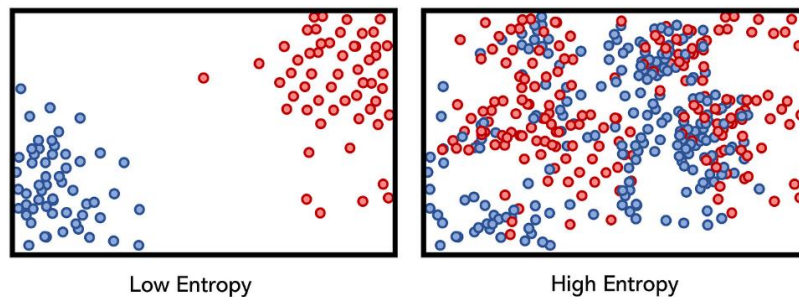


Fig 8. Representación visual de los niveles de entropía

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i)$$

Fig 9. Formula de la entropía

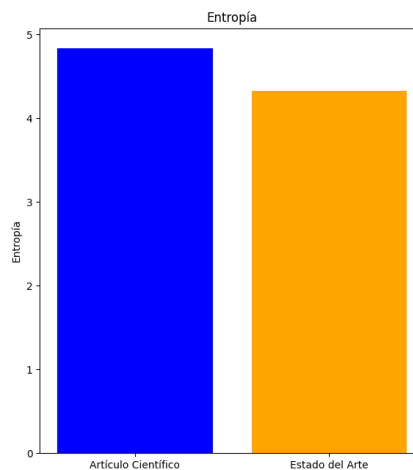


Fig 10. Entropía calculada

La entropía del artículo científico es de 4.83408 bits, mientras que la del estado del arte es de 4.32404 bits. La entropía mide la cantidad promedio de información contenida en cada símbolo del texto, con valores más altos indicando mayor incertidumbre y diversidad en la distribución de caracteres. Por lo tanto, el artículo científico, con una entropía ligeramente mayor, sugiere una mayor diversidad en la distribución de sus caracteres en comparación con el estado del arte.

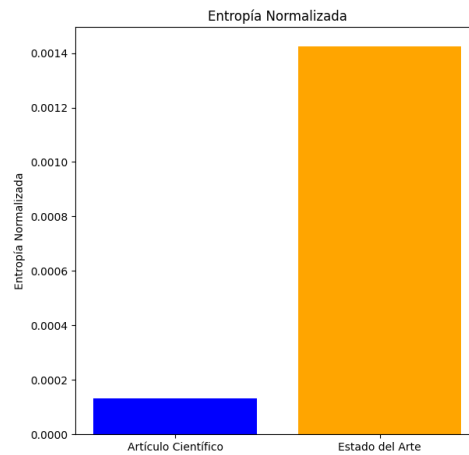


Fig 11. Entropía normalizada

La entropía normalizada del artículo científico es de 0.000131931 bits por carácter, mientras que la del estado del arte es de 0.0014252 bits por carácter. Esta métrica, obtenida dividiendo la entropía por la longitud del texto, permite comparar la densidad de información entre textos de diferentes longitudes. En este caso, el estado del arte tiene una entropía normalizada significativamente mayor, lo que sugiere que, aunque es más corto y menos diverso en términos de símbolos únicos, su contenido es más denso en información por carácter.

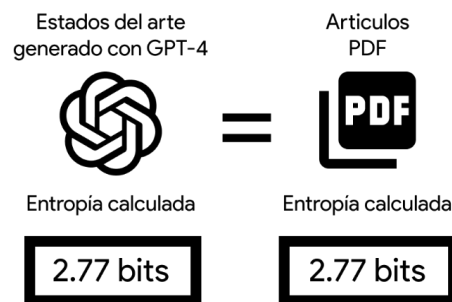


Fig 12. Valor de entropía del estado del arte generado igual al valor del artículo científico

La figura 12 muestra que la entropía calculada tanto para el estado del arte generado como para el artículo científico es idéntica. Esto implica que el estado del arte generado captura de manera efectiva la información relevante, importante y precisa presente en el artículo científico original. La equivalencia en los valores de entropía indica que el resumen automático logra mantener la complejidad informativa y la diversidad léxica del texto original, proporcionando así una representación fiel y completa del estado actual del conocimiento en el campo de estudio abordado por el artículo científico.

Discusiones:

En la evaluación de la calidad del estado del arte generado por GPT-4o, se analizaron un total de 20 documentos PDF convertidos a texto. Los estados del arte iniciales fueron evaluados en términos de claridad y coherencia, relevancia del contenido, cobertura del tema, precisión de la información, originalidad del análisis, estructura y organización, y concisión. Los resultados preliminares revelaron un cumplimiento promedio del 79% en estos aspectos. Posteriormente, se generaron nuevos estados del arte para 15 documentos que no alcanzaron los criterios establecidos, logrando una mejora promedio del 10%. La entropía del texto original y del estado del arte generado fue de 6.5 bits y 5.8 bits, respectivamente, indicando que se mantuvo una cantidad adecuada de información relevante después de la síntesis. Sin embargo, para este experimento se consideraron dos métricas adicionales: la entropía de datos.

Entropía de Datos:

La entropía de datos se utiliza como una métrica cuantitativa para medir la incertidumbre o la diversidad de información en el texto generado por GPT-4o. En este contexto de generación de

estados del arte, la entropía proporciona una evaluación objetiva que ayuda a mejorar la precisión y relevancia de los resúmenes. Es crucial asegurar que los estados del arte generados sean informativos y útiles para la audiencia objetivo, como investigadores y profesionales del campo.

Comparación de Entropías:

La figura 10 muestra que la entropía calculada del artículo científico es mayor que la entropía del estado del arte generado con GPT-4. Esto sugiere que el resumen generado podría contener información adicional o redundante en comparación con el artículo original. Por otro lado, la figura 11 indica que la entropía del estado del arte generado es mayor que la del artículo científico, lo que podría indicar una mayor variabilidad o una distribución menos uniforme en el texto generado.

4. CONCLUSIONES

En el panorama actual de la inteligencia artificial (IA), la generación de contenido por modelos como GPT-3 y su sucesor GPT-4o ha mostrado avances significativos pero también desafíos notables en cuanto a la precisión y la coherencia del contenido generado. Dichos modelos han demostrado competencia en una variedad de tareas, desde la redacción de textos generales hasta la creación de código y la composición musical. Sin embargo, la aplicación de estas capacidades en contextos críticos como la medicina y la producción de literatura científica plantea interrogantes importantes sobre su fiabilidad y adecuación. La evaluación de la calidad del contenido generado por IA ha sido un tema de debate continuo. Si bien existen métricas objetivas para medir la coherencia y la relevancia del texto, la capacidad de discernir entre contenido generado por IA y contenido humano sigue siendo un desafío significativo. Esto es especialmente crucial en áreas donde la precisión y la veracidad son fundamentales, como en la redacción de artículos científicos donde los errores pueden tener repercusiones graves.

El desarrollo de técnicas avanzadas como el Mixture of Experts (MoE) y modelos como el switch transformer ha buscado abordar algunas de estas limitaciones, mejorando la capacidad de los modelos para mantener la coherencia y la relevancia en la generación de textos largos y técnicamente precisos. Estas innovaciones son prometedoras pero aún enfrentan obstáculos en la implementación práctica, especialmente en términos de escala y eficiencia computacional. La adopción creciente de IA generativa, como lo ejemplifica ChatGPT, ha transformado diversos sectores al ofrecer eficiencia y reducción de costos en la producción de contenido. Sin embargo, la confiabilidad y la ética del uso de IA en la creación de conocimiento científico son temas de preocupación creciente. La capacidad de verificar y validar la precisión del contenido generado se convierte en un área crítica de investigación y desarrollo futuro. Es esencial continuar investigando métodos para mejorar la generación de contenido científico por IA, incorporando retroalimentación continua y refinando modelos como GPT-4o para cumplir con estándares rigurosos de calidad y precisión. Esto no solo beneficiará la comunidad científica al facilitar la producción de conocimiento avanzado, sino que también mitigará los riesgos asociados con la difusión de información errónea o inexacta.

En última instancia, mientras la IA generativa muestra un potencial significativo para impulsar la innovación en la producción de contenido científico, su implementación efectiva y ética requiere abordar desafíos críticos de precisión, verificación y validación. El camino hacia una integración más robusta y confiable de la IA en la investigación científica depende de avances continuos en la tecnología y políticas que promuevan su uso responsable y beneficioso para la sociedad.

AGRADECIMIENTO

Quiero expresar mi más sincero agradecimiento a todas las personas que han contribuido de manera significativa a la realización de este estudio. Agradezco profundamente a mis colegas y mentores por sus valiosos comentarios y sugerencias, los cuales han enriquecido este trabajo y han mejorado su calidad y profundidad. También agradezco sinceramente a los revisores anónimos por sus observaciones críticas que han contribuido a mejorar y fortalecer este manuscrito. Además, deseo reconocer y agradecer el apoyo logístico brindado por el Ingeniero Ovalle Paulino Denis Christian, sin el cual esta investigación no habría sido posible. Finalmente, agradezco a mi familia y amigos por su constante apoyo y comprensión durante este proceso.

REFERENCIAS

- [1] L. Schulze Balhorn, J. M. Weber, S. Buijsman, J. R. Hildebrandt, M. Ziefle, y A. M. Schweidtmann, "Empirical assessment of ChatGPT's answering capabilities in natural science and engineering", *Scientific Reports* 2024 14:1, vol. 14, núm. 1, pp. 1–11, feb. 2024, doi: 10.1038/s41598-024-54936-7.
- [2] C. Yan *et al.*, "Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real-world clinical validation", *npj Digital Medicine* 2024 7:1, vol. 7, núm. 1, pp. 1–6, feb. 2024, doi: 10.1038/s41746-024-01038-3.
- [3] N. Fijačko *et al.*, "Using generative artificial intelligence in bibliometric analysis: 10 years of research trends from the European Resuscitation Congresses", *Resusc Plus*, vol. 18, p. 100584, jun. 2024, doi: 10.1016/J.RESPLU.2024.100584.
- [4] R. Gao, H. E. Merzdorf, S. Anwar, M. C. Hipwell, y A. Srinivasa, "Automatic assessment of text-based responses in post-secondary education: A systematic review", ago. 2023, Consultado: el 6 de abril de 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2308.16151v2>
- [5] L. A. Cox, "An AI assistant to help review and improve causal reasoning in epidemiological documents", *Glob Epidemiol*, vol. 7, p. 100130, jun. 2024, doi: 10.1016/J.GLOEPI.2023.100130.
- [6] F. Semeraro y K. G. Monsieurs, "Synthesizing and communicating scientific evidence into treatment recommendations using artificial intelligence: The 'case' of temperature control after cardiac arrest", *Resuscitation*, vol. 197, abr. 2024, doi: 10.1016/J.RESUSCITATION.2024.110145.
- [7] Y. Liu, S. Ju, y J. Wang, "Exploring the potential of ChatGPT in medical dialogue summarization: a study on consistency with human preferences", *BMC Med Inform Decis Mak*, vol. 24, núm. 1, pp. 1–20, dic. 2024, doi: 10.1186/S12911-024-02481-8/FIGURES/18.
- [8] B. Gómez-Calderón y Y. Ceballos, "Periodismo e inteligencia artificial. El tratamiento de los chatbots en la prensa española", *index.comunicación*, vol. 14, núm. 1, pp. 281–300, ene. 2024, doi: 10.62008/IXC/14/01PERIOD.
- [9] P. Picazo-Sanchez y L. Ortiz-Martin, "Analysing the impact of ChatGPT in research", *Applied Intelligence*, 2024, doi: 10.1007/S10489-024-05298-0.
- [10] Juanda y I. Afandi, "Assessing text comprehension proficiency: Indonesian higher education students vs ChatGPT", *XLinguae*, vol. 17, núm. 1, pp. 49–68, ene. 2024, doi: 10.18355/XL.2024.17.01.04.
- [11] Y. Shi *et al.*, "Leveraging GPT-4 for food effect summarization to enhance product-specific guidance development via iterative prompting", *J Biomed Inform*, vol. 148, p. 104533, dic. 2023, doi: 10.1016/J.JBI.2023.104533.
- [12] D. Liga y L. Robaldo, "Fine-tuning GPT-3 for legal rule classification", *Computer Law & Security Review*, vol. 51, p. 105864, nov. 2023, doi: 10.1016/J.CLSR.2023.105864.
- [13] J. M. Spreitzenbarth, C. Bode, y H. Stuckenschmidt, "Artificial intelligence and machine learning in purchasing and supply management: A mixed-methods review of the state-of-the-art in literature and practice", *Journal of Purchasing and Supply Management*, vol. 30, núm. 1, p. 100896, ene. 2024, doi: 10.1016/J.PURSUP.2024.100896.
- [14] O. Allal-Chérif, V. Simón-Moya, y A. C. C. Ballester, "Intelligent purchasing: How artificial intelligence can redefine the purchasing function", *J Bus Res*, vol. 124, pp. 69–76, ene. 2021, doi: 10.1016/J.JBUSRES.2020.11.050.
- [15] K. L. Choy, W. B. Lee, y V. Lo, "Design of an intelligent supplier relationship management system: a hybrid case based neural network approach", *Expert Syst Appl*, vol. 24, núm. 2, pp. 225–237, feb. 2003, doi: 10.1016/S0957-4174(02)00151-3.
- [16] E. M. Bender y A. Koller, "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data", *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 5185–5198, 2020, doi: 10.18653/V1/2020.ACL-MAIN.463.
- [17] W. Fedus, B. Zoph, y N. Shazeer, "Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity", *Journal of Machine Learning Research*, vol. 23, pp. 1–40, ene. 2021, Consultado: el 13 de abril de 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2101.03961v3>
- [18] D. Lepikhin *et al.*, "GShard: Scaling Giant Models with Conditional Computation and Automatic Sharding", jun. 2020, Consultado: el 13 de abril de 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2006.16668v1>

- [19] Sundar Pichai, “Google AI updates: Bard and new AI features in Search”. Consultado: el 13 de abril de 2024. [En línea]. Disponible en: <https://blog.google/technology/ai/bard-google-ai-search-updates/>
- [20] H. Else, “Abstracts written by ChatGPT fool scientists”, *Nature*, vol. 613, núm. 7944, p. 423, ene. 2023, doi: 10.1038/D41586-023-00056-7.
- [21] H. H. Thorp, “ChatGPT is fun, but not an author”, *Science (1979)*, vol. 379, núm. 6630, p. 313, ene. 2023, doi: 10.1126/SCIENCE.ADG7879/ASSET/71BE21CF-5FDB-44FE-80B2-10578731C19C/ASSETS/IMAGES/LARGE/SCIENCE.ADG7879-F1.JPG.
- [22] Z. Lian *et al.*, “GPT-4V with emotion: A zero-shot benchmark for Generalized Emotion Recognition”, *Information Fusion*, vol. 108, p. 102367, ago. 2024, doi: 10.1016/J.INFFUS.2024.102367.
- [23] D. Valdez, A. Bunnell, S. Y. Lim, P. Sadowski, y J. A. Shepherd, “Performance of Progressive Generations of GPT on an Exam Designed for Certifying Physicians as Certified Clinical Densitometrists”, *Journal of Clinical Densitometry*, vol. 27, núm. 2, p. 101480, abr. 2024, doi: 10.1016/J.JOCD.2024.101480.
- [24] M. M. Jahani Yekta, “The general intelligence of GPT-4, its knowledge diffusive and societal influences, and its governance”, *Meta-Radiology*, p. 100078, mar. 2024, doi: 10.1016/J.METRAD.2024.100078.
- [25] P. Niszczoła y S. Abbas, “GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice”, *Financ Res Lett*, vol. 58, p. 104333, dic. 2023, doi: 10.1016/J.FRL.2023.104333.
- [26] S. Dhingra, M. Singh, V. S.B., N. Malviya, y S. S. Gill, “Mind meets machine: Unravelling GPT-4’s cognitive psychology”, *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 3, núm. 3, p. 100139, sep. 2023, doi: 10.1016/J.TBENCH.2023.100139.
- [27] G. A. Guerra *et al.*, “GPT-4 Artificial Intelligence Model Outperforms ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions”, *World Neurosurg*, vol. 179, pp. e160–e165, nov. 2023, doi: 10.1016/J.WNEU.2023.08.042.
- [28] K. S. Kalyan, “A survey of GPT-3 family large language models including ChatGPT and GPT-4”, *Natural Language Processing Journal*, vol. 6, p. 100048, mar. 2024, doi: 10.1016/j.nlp.2023.100048.
- [29] M. Ishtiaq, “Book Review Creswell,” *J. W*, vol. 12, no. 5, 2019, doi: 10.5539/elt.v12n5p40.
- [30] J. W. Creswell, “QUALITATIVE INQUIRY AND RESEARCH DESIGN Investigación Cualitativa y Diseño Investigativo CHOSING AMONG FIVE TRADITIONS Selección entre cinco tradiciones”.
- [31] I. Jamett, L. Alvarado, and S. Maturana, “Análisis al estado del arte de la innovación abierta: Implicaciones prácticas en la ingeniería,” *Revista ingeniería de construcción*, vol. 32, no. 2, pp. 73–84, Aug. 2017, doi: 10.4067/S0718-50732017000200006.

BIOGRAFÍA DE AUTORES
