

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR  
FACULTAD DE APRENDIZAJE, LENGUAS Y COMUNICACIÓN  
CARRERA: COMUNICACIÓN**

**TRABAJO DE TITULACIÓN PREVIA A LA OBTENCIÓN DEL TÍTULO DE  
LICENCIATURA EN COMUNICACIÓN**

**DISERTACIÓN SOBRE  
PROTOCOLO PARA VERIFICAR NOTICIAS FALSAS MEDIANTE EL USO DE  
MODELOS DE INTELIGENCIA ARTIFICIAL (IA)**

**IGNACIO JAVIER PERALTA ARROYO  
DIRECTOR: DR. CARLOS AULESTIA PÁEZ**

**QUITO, 2025**

## **Dedicatoria**

A mis padres,

por su amor incondicional, por enseñarme con su ejemplo el valor de la perseverancia y por ser mi mayor inspiración para culminar esta Carrera de Comunicación.

A mis hermanos y abuelos,

por el apoyo constante, las palabras de aliento en los momentos difíciles.

A mis docentes,

por entregarme su conocimiento con generosidad y por motivarme a desarrollar una propuesta crítica ante la sociedad de la información.

Y a todos mis amigos cercanos que de alguna manera creyeron en mí y contribuyeron a la realización de este proyecto.

## **Agradecimientos**

Agradezco, en primer lugar, a Dios, por brindarme la fe y la claridad para culminar esta etapa académica.

Expreso mi profundo agradecimiento a mis padres y mi familia, por su apoyo inquebrantable, su paciencia y sus constantes palabras de ánimo que me acompañaron durante este proceso.

A mi director de tesis, Dr. Carlos Aulestia, por su orientación, su compromiso, su paciencia y sus valiosas observaciones que enriquecieron arduamente el desarrollo de esta investigación.

A mi querida Pontificia Universidad Católica del Ecuador, quien a lo largo de mi formación universitaria impartió junto con mis docentes los conocimientos requeridos para mí futuro profesional.

## Índice de contenidos

Introducción.....	1
Capítulo I:	
Marco Teórico.....	3
1.1 La sobrecarga de información en la era digital .....	3
1.2 El riesgo de consumir <i>fake news</i> en contextos de exceso de información.....	4
1.3 El papel de los algoritmos de las plataformas digitales en la sobrecarga de información	5
1.4 La Inteligencia Artificial (IA) .....	6
1.5 El <i>chatbot</i> como factor comunicativo .....	7
1.5.1 Operación de chatbots.....	8
1.5.2 Responsabilidades éticas de los chatbots .....	8
1.6 Modelos de <i>chatbots</i> .....	10
1.6.1 Modelo Grok 3.....	10
1.6.2 Modelo ChatGPT .....	12
1.6.3 Modelo de Inteligencia Artificial Copilot .....	13
Capítulo II:	
Una propuesta metodológica para combatir la desinformación.....	17
2.1 Protocolo para usar modelos de Inteligencia Artificial generativa para verificar noticias falsas en entornos digitales .....	17
2.1.1 Identificar la noticia a verificar con IA generativa .....	18
2.1.2 Formular la pregunta clara al modelo de IA.....	19
2.1.3 Analizar meticulosamente la respuesta generada de la IA .....	20
2.1.4 Utilizar más de un modelo de IA.....	21
2.1.5 Consideración sobre el contexto de la noticia .....	22

2.1.6 El usuario tiene que sacar sus propias conclusiones: .....	23
2.2 Importancia sobre el uso ético y responsable con los modelos de IA .....	24
2.3 Registro de cada paso del protocolo de verificación con modelos IA de noticias falsas	25
Tablas de clasificación para evidenciar el proceso del protocolo de verificación con herramientas de IA para noticias falsas .....	27
2.3.1 Justificación de la estructura de la tabla de clasificación del protocolo ante la era de la sobrecarga informativa en usuarios digitales.....	31
2.4 Funcionalidad de los componentes de la tabla del protocolo .....	32
Capítulo III:	
Interpretación de los hallazgos.....	33
3.1 El impacto de la Inteligencia Artificial para detectar datos desinformativos mediante el protocolo .....	33
3.2 Dependencia de la Inteligencia Artificial como dispositivo de control.....	36
3.3 Fiabilidad de la Inteligencia Artificial .....	37
3.4 Visibilizar a los medios de verificación de contenidos y usuarios participantes para disminuir la sobrecarga de información.....	39
Conclusiones .....	43
Referencias.....	45

## Introducción

En esta sociedad contemporánea en la que la información ha adquirido un papel central que define, condiciona y transforma las relaciones entre los individuos, las instituciones y el ámbito global se intensifica en un contexto de avances tecnológicos que posibilitan la circulación inmediata y masiva de contenidos a través de medios digitales. Sin embargo, el acceso sin precedentes a fuentes de información diversas ha derivado en la aparición de problemáticas que afectan la comprensión objetiva de los hechos. Uno de los principales ejemplos destaca la sobrecarga informativa también conocida como infoxicación y la proliferación de noticias falsas o *fake news* que amenazan con cambiar las percepciones y la confianza en las instituciones públicas y organizaciones de la sociedad civil.

Este fenómeno de la sobrecarga de información aparte de generar efectos negativos en la salud emocional de los usuarios, también propicia la aceptación de los contenidos que carecen de datos verificados. Es fácil caer en las noticias falsas por su difusión de manera inmediata en las redes sociales y no pasa por filtros editoriales y, por otro lado, a los usuarios se les dificulta identificar el origen del contenido digital.

En este contexto, surge la necesidad de diseñar estrategias de verificación que permitan contrastar las noticias falsas y fortalecer el criterio de los usuarios digitales. Por otro lado, la Inteligencia Artificial (IA) ha comenzado a posicionarse como una herramienta potencialmente eficaz para abordar la desinformación y mitigar el impacto de la infoxicación. Sin embargo, el uso de estas tecnologías traza desafíos éticos, técnicos y sociales que deben ser analizados antes de ser utilizados como herramientas recurrentes.

La presente investigación se fundamenta en el aprovechamiento responsable del uso de la Inteligencia Artificial para disminuir los efectos de la saturación digital. Para ello, se propone la elaboración de un protocolo básico orientado en la verificación de noticias falsas mediante el uso de modelos de IA generativa, como ChatGPT, Copilot y Grok. Este trabajo plantea un protocolo como una guía práctica y accesible que permita a los usuarios realizar procesos de verificación asegurados, ordenados y éticos, con base en el análisis de las respuestas generadas por los sistemas de automatización de Inteligencia Artificial y la comparación con fuentes confiables.

En este contexto, las herramientas de IA constituyen un recurso para procesar, resumir y generar información que puede facilitar la tarea de verificación; por otro lado, su dependencia de datos de entrenamiento y sus algoritmos pueden derivar en respuestas imprecisas o en sesgos que desarrollen otras narrativas desinformativas. Es por ello, que entre la utilidad técnica de la Inteligencia Artificial y los riesgos asociados a su aplicación indiscriminada justifica la necesidad de un protocolo que determine los parámetros para los criterios de análisis crítico y pautas de contraste con fuentes humanas confiables.

Este trabajo plantea que la Inteligencia Artificial influye en la configuración de las relaciones de poder contemporáneas; como lo argumenta Foucault (Vega, 2017, p.143), las tecnologías no solo intervienen en la organización de la información, sino que orientan la manera en que los sujetos comprenden su entorno y ejercen su autonomía. En consecuencia, cualquier propuesta metodológica que incluya la IA debe contemplar principios éticos fundamentales como la transparencia, la privacidad, la responsabilidad y la imparcialidad, tal como señalan Van Dijk et al. (2022) y Gutiérrez (2022).

Este trabajo de investigación se estructura en tres capítulos principales. El primer capítulo expone el marco teórico que sustenta la problemática de la sobrecarga informativa, el fenómeno de las noticias falsas y el funcionamiento de los modelos de Inteligencia Artificial Generativa. Por otra parte, el segundo capítulo describe la metodología propuesta, en la elaboración de un protocolo de pasos secuenciales que guían el proceso de verificación de noticias mediante la IA. Por último, el tercer capítulo interpreta la aplicación del protocolo y reflexiona sobre su convivencia, sus limitaciones y su contribución para los usuarios digitales. Finalmente, se describen las respectivas conclusiones de la investigación.

Esta disertación pretende ofrecer una propuesta metodológica que contribuya a enfrentar uno de los desafíos más complejos en la era digital: la necesidad de discernir entre la información veraz y los contenidos falsos en un entorno donde la inmediatez de la comunicación frecuentemente supera la capacidad de verificación humana. Además, la integración crítica de la Inteligencia Artificial como herramienta de apoyo, y no como sustituto del juicio humano, es parte de esta investigación.

# Capítulo I:

## Marco Teórico

En este capítulo, presentaré y relacionaré los conceptos fundamentales para comprender el fenómeno de la sobrecarga de información. Revisaré conceptos importantes sobre los modelos de Inteligencia Artificial disponibles actualmente para usuarios en la web, y estableceré un método para abordar mi objeto de estudio, que consistirá en el siguiente capítulo en elaborar un protocolo con distintas inteligencias artificiales mediante *prompts* para guiar y responder sobre noticias o sucesos actuales, y evaluar cuán confiables son las IA como herramientas de verificación de información.

### 1.1 La sobrecarga de información en la era digital

El concepto ‘sobrecarga de información’ describe una problemática en la que los individuos tienden a sentir cansancio y presentan dificultades para procesar toda la información recibida, lo que los lleva a tratarla de forma errónea (Fernández, 2023, p.2). Según Julián Martínez (2024),

la sobreinformación es un fenómeno en el que la cantidad y/o la intensidad de la información exceden la capacidad limitada de procesamiento cognitivo del individuo. En este sentido, pueden generar emociones displacenteras y estados de malestar como el estrés, el miedo desproporcionado y niveles elevados de ansiedad en la población (p.74).

En la era digital, se presenta una problemática debido a la acumulación de información, que desafía las capacidades cognitivas y las emociones de los usuarios, por lo que es necesario implementar estrategias para gestionar el consumo de la información.

La propuesta de un protocolo básico y la interpretación que se profundizará en los siguientes capítulos se centra en el uso de *chatbots* o modelos de IA para verificar discursos en entornos de sobrecarga informativa. Se comprobará cómo los individuos pueden recurrir a estas herramientas tecnológicas una serie de información digital caótica y difícil de manejar, y pueden además proponer estrategias adecuadas de manejo de información ante posibles

reacciones adversas, como consecuencia de la sobreabundancia de datos, como lo expone el autor Julián Martínez (2024).

Antonio Padilla, en su artículo de investigación *Factores que afectan a la sobrecarga de información a nivel directivo*, menciona que varios estudios indican que los individuos tienen dificultades al momento de recibir gran cantidad de información digital. Un ejemplo, en los espacios laborales al momento de responder mensajes mediante el correo electrónico, su estructura de la información es confusa y compleja de entender. Por lo tanto, la deficiencia en los mensajes digitales podría generar un impacto negativo en el desarrollo de otras actividades laborales (Padilla, 2015, p. 253).

Para Miguel Romero et. al. (2018, p.74), los medios de comunicación proporcionan contenidos diseñados por sus audiencias, especialmente de infoentretenimiento para mantener sus ingresos publicitarios. Los receptores, al no poder distinguir correctamente la información, aceptan los contenidos difundidos como válidos, aunque resultan incompatibles con otras afirmaciones que han sido aceptadas como ciertas.

Conchi Jiménez (2016, p.55) sostiene que la información excesiva en múltiples canales de medios puede provocar una infoxicación. Estudios previos han determinado que el impacto del consumo mediático conduce a perder el tiempo sin saber cómo distinguir información útil. La acelerada infoxicación puede perjudicar al momento de realizar actividades productivas, y en las actividades laborales que requieren construir mensajes más elaborados (Padilla, 2015, p.253).

## **1.2 El riesgo de consumir *fake news* en contextos de exceso de información**

La sobrecarga de información, además de los impactos emocionales que pueden presentarse en los usuarios, también implica el riesgo grave de creer en las noticias falsas. Con el acceso inmediato y la democratización de las redes sociales que se han saturado en la era digital, Martínez Verdú (2024, p.2) afirma que:

Las *fake news* son un tipo de desinformación que ha proliferado con el auge de las redes sociales, donde la velocidad de difusión supera la capacidad de verificación de los hechos.

Además, los autores Nancy Montemayor-Rodríguez y Juan Francisco Torregrosa-Carmona señalan que la ausencia de acuerdos sobre el término '*fake news*' hace compleja la investigación, debido a que algunos académicos las asocian con contenidos de fines humorísticos, mientras que otros lo relacionan con noticias sin contexto y surrealistas (Montemayor-Rodríguez, Torregrosa-Carmona, 2023, p.2).

Según Martínez Verdú (2024), Montemayor-Rodríguez y Torregrosa-Carmona exponen, para elaborar su teoría central de las *fake news*, que las redes sociales son un elemento esencial como fuente de información, por lo que el riesgo de exposición a las *fake news* implica afectar percepciones de realidad y confiar en los contenidos falsos (Martínez Verdú, 2024, p.2).

Un caso de estudio sobre las *fake news* en jóvenes universitarios desarrollado por Martínez Verdú (2024), arroja resultados que muestran similitudes con otros estudios análogos: menciona que el autor Marián Alonso (2021), evidencia el aumento de la desinformación en redes sociales durante la pandemia del COVID-19.

Las redes sociales amplifican todo tipo de información verdadera o falsa, pero los estudiantes en el estudio de caso no verifican todo por igual, sino que lo hacen por sus intereses personales, por lo que la autora propone entender las motivaciones y formas de pensar de los jóvenes como un factor de desinformación efectiva (Martínez Verdú, 2024, p.11).

Por otro lado, de acuerdo con Martínez Verdú (2024), algunas plataformas digitales adoptan medidas de transparencia de sus algoritmos en la lucha contra las *fake news*. Se pueden obtener resultados favorables mediante la implementación de algoritmos esenciales para reconocer el contenido verificado.

### **1.3 El papel de los algoritmos de las plataformas digitales en la sobrecarga de información**

Los algoritmos en las redes sociales pueden ampliar sesgos existentes de la información y en las percepciones de los usuarios, en consolidar desigualdades y prejuicios sociales, mediante la selección de contenidos que dan prioridad a ciertos grupos específicos de información, en palabras de Andrés Abeliuk (2023, p.4).

Según este autor (2023, p.5), la interacción entre los sesgos ideológicos de los seres humanos y las bases de algoritmos pueden producir un efecto de retroalimentación que aumenta los estereotipos. Por lo tanto, es importante entender cómo opera el diseño de las plataformas digitales para disminuir los problemas de sesgos y estereotipos.

Los algoritmos usan los datos de interacciones previas, como los *likes*, vistas y búsquedas, para ofrecer contenidos adaptados a cada usuario. Buscan optimizar la elección de información y reducir la fatiga cognitiva o la sobrecarga. No obstante, si no existe una debida regularización, estos sistemas en las plataformas digitales pueden aumentar prejuicios y generar entornos de información cerrados en los usuarios (Abeliuk, 2023, p.3).

#### **1.4 La Inteligencia Artificial (IA)**

La Inteligencia Artificial (IA) es la parte del campo de la informática que consiste en desarrollar sistemas capaces de ejecutar tareas que están asociadas con la inteligencia humana (Pañi, 2025, p.3). La IA tiene ventaja en diseños de sistemas automáticos para promover la comprensión y sostener habilidades relacionadas con interpretar y razonar.

Vicenç Torra (2011) afirma:

Por tanto, esta primera definición de Inteligencia Artificial se focaliza en el comportamiento humano. Esta no es, sin embargo, la única aproximación existente en la actualidad pues algunos autores han mostrado que esta definición limita los objetivos del campo (p.1).

La Inteligencia Artificial puede enfocarse en obtener resultados correctos, con lo cual es la principal prioridad es en la toma de decisiones, como en los programas de ajedrez que buscan ganar, sin importar el razonamiento que imita al del humano (Torra, 2011, p.1).

En la actualidad, con la era digital, la cantidad de datos almacenados es cada vez más grande. Con los buscadores de web como Google, que guardan copias de múltiples sitios web, los correos electrónicos, que almacenan millones de mensajes, y las redes sociales, que registran información sobre intereses y relaciones personales (Torra, 2011). Para este autor, mediante el aumento de velocidad en los sistemas de informática, las grandes empresas de información

aplican más recursos para tomar decisiones, para que esto se ejecute de manera informada y más personalizada con el uso de Inteligencia Artificial (Torra, 2011, p.8).

### **1.5 El *chatbot* como factor comunicativo**

Según Fondevila-Gascón et al. (2024), el término *chatbot* se puede comprender como un agente virtual o robótico que realiza la interacción con el usuario, ya sea por textos escritos o mediante una llamada telefónica, cuando existe contacto con el centro de atención de clientes de empresas. Los autores explican que los *chatbots* pueden obtener resultados beneficiosos para gestionar actividades, lo que implica que las máquinas no deberían tener descansos y ofrecer atención ininterrumpida.

Fondevila-Gascón et al. (2024, p.51), señalan que la evolución de la IA tiene un rol importante al posibilitar el desarrollo de actividades que antes solían recurrir a la inteligencia humana, al destacar la explicación de Lasse Rouhiainen (2019, p.51). En este contexto, los autores mencionados aclaran la crítica de Turing (1950, p.51), quien reflexionó sobre si las máquinas podían razonar, proponiendo un experimento en el que un humano interactúa con una máquina sin poder distinguir si su interlocutor es humano o no. Este planteamiento buscaba evaluar si una máquina podía imitar el razonamiento humano de manera convincente. Además, se sugirió que tanto humanos como máquinas pueden procesar información disponible, pero las máquinas tienen la capacidad de manejar grandes bases de datos con mayor rapidez y menor margen de error.

Por otro lado, Fondevila-Gascón et al. (2024, p.51) también señalan, en referencia a Brian McGuire (2006, pp. 5-9), que la principal limitación en el pasado era la capacidad de las máquinas para almacenar y recopilar datos. Estas computadoras podían ejecutar comandos; sin embargo, no retenían la información de cómo ejecutaban los procesos.

En los últimos años, la IA captó el interés de los empresarios, lo que llevó a una adopción generalizada de *chatbots* en diferentes canales de atención al cliente. Según los autores Fondevila-Gascón et al. (2024, p.52), desarrollaron una encuesta realizada por Haptik (2022, p.52), en la que el 73% de las empresas ya habían incorporado algún tipo de asistente virtual. Además, el 39% tenía como proyección invertir entre \$50000 y \$100000 en la creación de *chatbots*. Un dato curioso de los *chatbots* es su eficiencia económica. En el año 2017, Juniper

Research calculó que los sistemas de IA permitirán un ahorro económico de \$8000 millones para el año 2022 (Fondevila-Gascón et al, 2024, p.52).

### **1.5.1 Operación de *chatbots***

García-Gaona et al. (2023, p.33), afirman que los *chatbots* en el ámbito de la educación tienen el potencial de proporcionar información estándar de forma instantánea, así como los criterios de evaluación del aprendizaje, fechas y recursos, lo que permite un fácil acceso a los estudiantes. De igual manera, se ha revisado que estas herramientas de IA motivan a los alumnos en el proceso de retención de conocimientos. La arquitectura de los *chatbots* suele enfocarse en mensajes predefinidos para elaborar consultas frecuentes, y su rendimiento depende de que los usuarios utilicen palabras clave para permitir una interacción fluida y natural.

Además, los *chatbots* pueden adaptarse a diferentes actividades, pero su implementación debe ser responsable para optimizar el proceso de enseñanza y aprendizaje. Con estos factores se puede fomentar la participación de los usuarios en el uso de los *chatbots*.

### **1.5.2 Responsabilidades éticas de los *chatbots***

Ahora bien, uno de los problemas presentados de la Inteligencia Artificial es su forma ética de actuar al momento de almacenar datos. Como lo señala el autor Jorge Gutiérrez (2022, p.2), varios autores, entre ellos Teun Van Dijk, discutieron en el año 2018 cuatro principios éticos, que son: privacidad, transparencia, responsabilidad e imparcialidad. Gutiérrez (2022, p.2) afirma que los cuatro principios éticos representan los retos más sobresalientes alrededor de una sociedad que es cada vez más dependiente de las plataformas digitales.

De acuerdo con Gutiérrez (2022, p.2), el enfoque ético se concentra en la elaboración de una serie de principios o valores normativos. El autor menciona que los valores pueden ser de naturaleza sugerente, es decir los Derechos Humanos; o de carácter obligatorio, como las leyes o regulaciones.

El principio de privacidad sería el primero de los valores para abordar; Gutiérrez señala que los autores Van Dijk et. al. (2022, p.2), definen a la privacidad como valor: “Se ha mostrado a

lo largo de nuestras discusiones sobre el tráfico de datos a través de canales en línea. Sin privacidad, los ciudadanos ya no pueden estar seguros de que tienen control sobre sus propias vidas...”. Entonces, la privacidad es parte fundamental de los principios en la declaración de los Derechos Humanos. De igual forma, en el funcionamiento de las plataformas digitales con Inteligencia Artificial, este valor o principio enfrenta amenazas constantes.

Según Gutiérrez (2022), los datos funcionan como el combustible esencial que impulsa el aprendizaje de la máquina, permitiendo que la Inteligencia Artificial se desarrolle y perfeccione. Asimismo, estos datos adquieren un valor económico relevante, ya que pueden ser vendidos a agencias de publicidad y desarrolladores, otorgándoles un beneficio monetario, incluso a plataformas que operan sin costo alguno.

Así pues, el segundo principio es la transparencia, como lo señala Gutiérrez:

En este sentido, la transparencia resulta importante en dos sentidos; primero, en el cómo se usan y recogen los datos; segundo, en cómo opera la plataforma (2022, p.3).

El autor hace referencia a Van Dijk et. al., quienes detallan que la transparencia es un valor público que se aplica no solo a los sistemas de datos y algoritmos, sino también a los modelos comerciales y al estado de propiedad de las plataformas. Además, es una parte fundamental para la rendición de cuentas, aunque en el entorno digital se torna cada vez más oscuro con el paso del tiempo (Gutiérrez, 2022, p.3).

Desde el punto de vista de Gutiérrez (2022, p.3), la opacidad en el funcionamiento de las plataformas las transforma en sistemas cerrados y poco comprensibles. Esto se intensifica al proteger elementos esenciales bajo el argumento de la propiedad intelectual, particularmente en lo que respecta a los algoritmos de aprendizaje de máquinas que opera la Inteligencia Artificial en dichas plataformas.

El tercer principio ético es la responsabilidad. Como lo señala Gutiérrez (2022), la responsabilidad promueve la colaboración entre los actores sociales para mejorar la convivencia, al mismo tiempo que permite identificar a los responsables de infracciones, apoyándose en elementos concretos, como las leyes y las normativas. Sin embargo, la

Inteligencia Artificial, especialmente en el aprendizaje de máquina, complica identificar quién o qué es responsable de un error debido a su creciente nivel avanzado. De cualquier modo, la responsabilidad viene de un agente moral. Es definido como una entidad capaz de ser el origen de acciones, y la mayoría de las perspectivas filosóficas sostienen que la agencia moral implica algún grado de autonomía, responsabilidad, o ambos (Véliz, 2021, citado en Gutiérrez, 2022, p.3).

De acuerdo con Gutiérrez (2022, p.4), la responsabilidad de la Inteligencia Artificial y los algoritmos que la integran no habitan en la propia tecnología, sino en las personas que la diseñan y las empresas que la desarrollan. Esto es una afirmación válida a pesar de las dificultades para comprender el funcionamiento de la Inteligencia Artificial y las barreras impuestas por la propiedad intelectual que protegen su estructura.

Por último, hay que explicar el principio de la imparcialidad. Según Gutiérrez (2022, p.4), la imparcialidad en el campo de la Inteligencia Artificial debe limitarse a impedir desviaciones que generen discriminación. La Inteligencia Artificial puede adquirir conocimientos para discriminar, debido a que es influenciada tanto por los prejuicios de quienes la desarrollan como por los sesgos presentes en los datos que utiliza para aprender. De igual forma, la Inteligencia Artificial puede ejercer discriminación de forma deliberada cuando se estima necesario; por ejemplo, al decidir si una persona puede acceder a un seguro de gastos médicos dependiendo de su edad.

## **1.6 Modelos de *chatbots***

### **1.6.1 Modelo Grok 3**

Grok 3 es un modelo de Inteligencia Artificial de lenguaje grande (LLM) de tercera generación desarrollado por xAI. Representa un hito esencial en el avance de las capacidades de LLM. Este modelo no solo está diseñado para generar texto, sino que también actúa como un asistente inteligente con la habilidad de razonar problemas de manera similar a como lo haría un ser humano (Keywords et. al, 2025, párr.4).

Los grandes modelos de lenguaje (LLM) son sistemas que, tras ser entrenados con enormes volúmenes de datos textuales, pueden comprender y producir lenguaje humano de forma

coherente y adaptada a diferentes contextos. Estas capacidades permiten automatizar tareas cognitivas complejas y procesar eficientemente la información no estructurada, lo que resulta fundamental para optimizar procesos como las encuestas de mercado (Diago, 2022, citado en Quiñónez et al, 2025, p. 687).

De acuerdo con Esteban Cárdenas, Grok ha tenido un impacto relevante en redes sociales que se ha convertido en una tendencia entre personas de todo el mundo, e incluso ha sido convocado a las conversaciones en redes sociales, especialmente en X, como si de un analista se tratase. (2025, párr.5)

La Inteligencia Artificial Grok 3 es propiedad del multimillonario estadounidense Elon Musk, un nombre conocido que ha expresado públicamente sus discursos sobre hacer política y cuestionar temas progresistas. La Inteligencia Artificial de xAI fue estrenada en agosto de 2024, pero sus funcionalidades comienzan a ser relevantes y a tener impulso en el mundo digital de las redes sociales a partir del 2025.

Grok se caracteriza por su habilidad para recopilar información de sitios web y publicaciones en X, lo que le permite ofrecer respuestas basadas en datos actuales y diversos (*Primicias.ec*, 2023, párr.17). Según en la investigación del autor Robel Revelo (2025) en el diario *Primicias*, los usuarios han experimentado con la herramienta del *chatbot* para hacer consultas públicas, intentar contextualizar hechos, afirmaciones o verificaciones, a medida que el *bot* cita a medios reconocidos. Sin embargo, los modelos de Inteligencia Artificial como Grok pueden tener alucinaciones, es decir, que engañan con la información proporcionada cuando no logran la respuesta con la base de datos procesada. Otras plataformas de Inteligencia Artificial como ChatGPT o Gemini advierten que la información que entregan los *chatbots* debe ser verificada. Por lo que Revelo (2025, párr. 21) afirma que: “Lo que le cuenta Grok no debe ser tomado como la única verdad o realmente fiable”, algo que este modelo no hace visible con la información entregada en sus respuestas.

Según Keywords AI (2025, párr.12), los usuarios pueden acceder a Grok mediante una suscripción a X Premium+, lo que les permite utilizarlo fácilmente en la plataforma X. El valor de Grok 3 se encuentra en su capacidad para enriquecer la experiencia de los usuarios de X con respuestas rápidas y fiables, así como en su potencial para aplicarse en áreas como el soporte al

cliente o la creación de contenidos. La reciente actualización de la Inteligencia Artificial de Grok en diciembre de 2024 mejoró su rendimiento de instrucciones en todos los idiomas. En este sentido, es capaz de cumplir órdenes relativamente complejas ya sea que desea redactar un correo electrónico formal o pedirle que use un tono de lenguaje amigable para cualquier necesidad. (Keywords AI, 2025)

### **1.6.2 Modelo ChatGPT**

Otro modelo de Inteligencia Artificial desde su lanzamiento oficial en noviembre de 2022 es la herramienta ChatGPT por la empresa Open AI. Según Adelaida et. al. (2023, p.62), la empresa Open AI ha asegurado ser el reemplazo de los buscadores de Internet, debido a que la Inteligencia Artificial de Open AI no solo encuentra la información solicitada, si no que la resume y muestra de una forma comprensible rápidamente.

La principal función del modelo ChatGPT, aparte de imitar una conversación humana, sus habilidades más complejas, ya que literalmente puede crear textos nuevos, como poemas, historias, o interpretar cualquier cosa dentro de su alcance (Adelaida et. al., 2023, p.63). Ahora bien, como la Inteligencia Artificial tiene una ventaja en ChatGPT sobre términos de comprensión rápida de la información, algo que en los humanos tienen limitaciones, al desarrollar su capacidad para leer una gama amplia de literatura y diferenciar conexiones de piezas de sistemas de información (Adelaida et. al. 2023, p.63).

En general, el *chatbot* conversacional ChatGPT que por sus siglas en inglés significa *Generative Pre-trained Transformer*, simplemente es un transformador preentrenado por humanos. La herramienta trae un cambio significativo para los servicios de educación y la sociedad en general, debido a su facilidad, accesibilidad y usabilidad en los buscadores web de Internet (Olite et. al, 2023, párr.11).

Existen prohibiciones en países para que la herramienta esté disponible, entre ellos Rusia, Irán, China, Cuba y naciones de África. Ciertos conflictos políticos, sociales y económicos con Estados Unidos han bloqueado el servicio de la herramienta de la empresa Open AI. Olite et al. (2023, párr.22) explican que el *software* de ChatGPT funciona como un asistente digital que busca responder a cualquier pregunta generada por los usuarios. Aunque sigue en desarrollo y puede presentar errores, algo que ha sorprendido es su capacidad de crear contenido, ofrecer

respuestas ágiles en una amplia variedad de investigaciones, realizar cálculos matemáticos, escribir códigos de programación, redactar ensayos, etc. Incluso el sistema de ChatGPT tiene el potencial de estructurar mensajes que podrían ser suficientes para calificar en exámenes de disciplinas como Medicina y Derecho. El funcionamiento elaborado por Open AI se estructura en la integración de capas y parámetros. Se rige por el procesamiento de millones de informaciones existentes en Internet, por lo que hace una distinción desde teorías científicas hasta otras fuentes que no tienen modos de verificación de información, como Wikipedia. Para acceder a ChatGPT, el proceso es bastante fácil y accesible. Según Olite et al. (2023, párr.40), con visitar el sitio web oficial de OpenAI, al momento de registrarse, hay que utilizar una dirección de correo electrónico y, tras recibir un código de verificación, el usuario puede empezar a interactuar con el *chatbot*, planteando preguntas y recibiendo respuestas de forma inmediata. Sin embargo, los autores consideran que hay más opciones para enriquecer esta experiencia con el *chatbot*. Por ejemplo, Open AI ofrece una suscripción premium conocida como ChatGPT Plus, la cual brinda beneficios de acceso exclusivo en momentos de alta demanda y con respuestas más rápidas. Además, ChatGPT no se limita a ir solo a sitios web, porque ofrece la función disponible en aplicaciones móviles para dispositivos iOS y Android, así como en extensiones para navegadores. Aunque la versión básica es gratuita, algunas funcionalidades avanzadas, como el acceso a modelos más potentes o la capacidad de generar contenido visual o leer documentos digitales, requerirán un pago adicional (Diego Olite et al., 2023, párr.43).

### **1.6.3 Modelo de Inteligencia Artificial Copilot**

El sistema Copilot es un modelo de Inteligencia Artificial desarrollado por la empresa Microsoft. Es un modelo de diálogo que se puede controlar y con el que se puede interactuar a través de un chat (Fernández, 2024, párr.4). El *chatbot* de Copilot utiliza funciones de lenguaje de Inteligencia Artificial GPT, el mismo que maneja ChatGPT. Su principal característica es utilizar el modelo GPT-4, parecido a las versiones que maneja ChatGPT. El autor Fernández (2024) indica que el modelo de Microsoft utiliza conexiones de Internet, para que la información ofrecida sea actualizada constantemente. Entonces, el modelo Copilot brindará la información actual en incluir sitios web en donde se ha consultado.

Por otro lado, Microsoft también ha invertido en integrar a Copilot el sistema DALL-E3, que es tecnología con Inteligencia Artificial para producir imágenes visuales, a través de texto. De acuerdo con Fernández, “puedes usarlo para crear imágenes de la nada, simplemente describiéndole lo que quieres que dibuje para ti” (2024, párr.8). La Inteligencia Artificial de Copilot pondrá en marcha ciertos mecanismos para resolver tu pregunta y dar una perspectiva de la imagen visual requerida.

Por lo tanto, el modelo de Copilot tiene una variedad en sus funciones para buscar información actual, y también poder entablar una conversación o pedirle que escriba sobre cualquier tema. Ahora bien, el autor Yúbal Fernández (2024, párr.10) considera que se puede pedir información al modelo en relación con viajes o zonas que podrían ser visitados. De igual forma, solicitar la creación de tablas comparativas o productos de Internet al alcance de las solicitudes del humano. También Copilot tiene integradas aplicaciones de Microsoft, como Word, Excel y el propio sistema operativo de Windows (Fernández, 2024, párr.11).

Copilot de Microsoft protege la privacidad mediante la anonimización de datos, eliminando la información de identificación personal durante el procesamiento y entrenamiento de la Inteligencia Artificial. Cumple con rigurosas normativas de gestión de datos, asegurando que la información no se almacene ni se reutilice para entrenar modelos de IA, lo que minimiza significativamente los riesgos de explotación de datos (Sasson, 2024, párr.4) Además, Sasson (2024, párr.7), expone que Copilot se alinea con el Reglamento General de Protección de Datos (RGPD), implementando la minimización de datos, lo cual requiere consentimiento del usuario y respetar el derecho al olvido. Los usuarios pueden gestionar sus datos personales a través de controles que facilitan el acceso y la corrección.

Además, Copilot presenta herramientas para manejar solicitudes de acceso de interesados y permite optar por no vender datos, cumpliendo con la Ley de Privacidad del Consumidor de California. También notifica rápidamente a los usuarios sobre filtraciones de datos que comprometan su información personal. Ahora bien, hay preocupaciones sobre la fuga involuntaria de datos proporcionados al modelo de Copilot de Microsoft. El modelo podría exponer datos sensibles al incluirlos en informes o resúmenes que se compartan más allá de la audiencia prevista. En aplicaciones de Microsoft como Teams, sus funciones en tiempo real pueden registrar acciones y generar informes de conversaciones, lo que ocasionará incorrectamente la divulgación de información privada de manera no intencionada (Sasson, 2024, párr.19).

Otra preocupación con este el modelo, según el autor Sasson (2024, párr.20), es que el almacenamiento inseguro de datos representa un riesgo a Microsoft en Copilot, ya que la restauración de datos podría llevar a guardar información sensible en lugares menos protegidos, como cuentas personales de OneDrive, incrementando la posibilidad de accesos no autorizados. De igual forma, los datos transferidos entre aplicaciones integradas pueden carecer de protección adecuada, con el aumento del riesgo de exposición.

Para el cierre del capítulo, los conceptos analizados sientan las bases para abordar los desafíos éticos y técnicos de la IA en entornos de discursos para la verificación de información. La funcionalidad de estas herramientas de *chatbots* para procesar, generar y verificar información en tiempo real, junto con sus riesgos inherentes, preocupa un debate generado por la información masiva en el mundo digital. La difusión masiva de medios de comunicación sobre hechos, noticias o acontecimientos, corren con riesgo de contextualizar falsedades y mentiras mal intencionadas. Por lo tanto, surge la necesidad de proponer estrategias que garanticen un uso responsable y seguro de los *chatbots* para encontrar información en tiempo real. En el siguiente capítulo, se plantea desarrollar un protocolo de uso de Inteligencia Artificial orientado a usuarios que requieran verificar de manera inmediata la veracidad de información, distinguiendo entre los contenidos confiables y los posibles discursos fraudulentos de las IA.

Este protocolo integrará principios éticos de su uso, como los observados en Microsoft 365 Copilot, Grok3 o ChatGPT y aprovechará la capacidad de análisis en tiempo real de las herramientas *chatbots* para crear un sistema accesible y eficiente. Este enfoque no solo aplicará los conceptos teóricos aquí expuestos, sino que también buscará utilizar ejemplos de noticias o narrativas falsas para fortalecer a los usuarios en un entorno digital donde la desinformación es un desafío creciente y de incertidumbre, promoviendo un uso seguro en discusión con la IA.

## Capítulo II:

### Una propuesta metodológica para combatir la desinformación

Una vez revisado el marco teórico previamente expuesto, se abordará la propuesta metodológica con el objeto de estudio en un protocolo básico de pasos para aplicar ante cualquier *chatbot* o modelo de Inteligencia Artificial como verificador de noticias falsas, acontecimientos o sucesos dentro de un contexto de sobrecarga informativa. Las redes sociales, incluyendo Facebook, Instagram, Twitter y WhatsApp, desempeñan un papel cuestionable en la difusión de noticias falsas debido a que sus algoritmos favorecen la viralidad del contenido. Las plataformas digitales, mediante sus algoritmos, crean burbujas informativas para los usuarios. Estas burbujas consolidan las creencias existentes, lo que conlleva el riesgo de proliferar la desinformación masiva (Sevilla & Maceiras, 2024, p. 3).

Para este trabajo de investigación, se utilizarán conceptos del primer capítulo para proponer la elaboración de un protocolo básico de uso con Inteligencia Artificial como herramienta verificadora de noticias falsas o *fake news*. El propósito del protocolo es verificar noticias y servir como herramienta ante la masiva sobrecarga de información en entornos digitales. Este conjunto de pasos está diseñado para ser sencillo de seguir y aplicar, incluso para usuarios sin experiencia previa con modelos de IA como Grok, Copilot o ChatGPT.

#### **2.1 Protocolo para usar modelos de Inteligencia Artificial generativa para verificar noticias falsas en entornos digitales**

Para comenzar con un protocolo para buscar y verificar información por medio de modelos de IA generativa, antes hay que entender qué son las indicaciones para usar la Inteligencia Artificial. Las indicaciones, conocidas como *prompts*, son instrucciones que se dan a un modelo de Inteligencia Artificial capaz de procesar grandes cantidades de texto en lenguaje natural (Alex Hey, 2024, par. 6). Los sistemas de IA son capaces de procesar entradas de texto ofrecidas por los usuarios y producir respuestas esperadas. Esto se logra a través del procesamiento del lenguaje y las construcciones lingüísticas que el sistema emplea en su base de datos. Entonces, cuando el usuario elabora los *prompts*, estos deben ser ajustados para el sistema de IA seleccionado para generar los resultados deseados.

Cuando se elaboran indicaciones confusas o incoherentes para un modelo de Inteligencia Artificial Generativo, este enfrenta dificultades más complejas que las de una persona intentando comprender un idioma desconocido. Aunque el sistema podría interpretar correctamente ciertos mensajes coherentes, a través de los *prompts*, en general, le resultará complejo generar respuestas coherentes. Los *prompts* precisos y bien formulados, con sus preguntas, actuarán como un conector entre los objetivos esperados del usuario y la capacidad de comprensión de texto de la IA.

A continuación, se procederá a explicar los pasos del protocolo para utilizar la Inteligencia Artificial Generativa, utilizando estas herramientas como ayuda para verificar si una noticia es verdadera o falsa, hacer un breve análisis crítico sobre las respuestas y contrastar la información con diversas fuentes.

### **2.1.1 Identificar la noticia a verificar con IA generativa**

El primer paso del protocolo es identificar la noticia a verificar, y este será el punto de partida para el proceso de selección. Este paso implica que el usuario seleccione una noticia específica que provoque dudas sobre su veracidad, ya sea por su impacto social, su origen cuestionable o su viralidad en redes sociales. La elección debe ser intencional, basada en la relevancia de la noticia en el contexto actual, su alcance para generar desinformación o su relación con temas de interés público. Por ejemplo, una noticia que afirme un cambio drástico en políticas de estado o un evento catastrófico debe ser priorizada porque podría tener posibles consecuencias en la opinión pública.

Una vez seleccionada la noticia, realizar un breve resumen porque es fundamental para clarificar su contenido y evitar interpretaciones erróneas durante el proceso de verificación con las IA generativas de texto. Para hacer este resumen, el usuario debe ser conciso, objetivo y exponer los hechos clave: quiénes están involucrados en la noticia, qué se explica, cuándo y dónde ocurrió el supuesto evento, y cómo se presenta la información. Por ejemplo, si la noticia seleccionada afirma que “La FIFA cancela el Mundial del Fútbol 2026 debido a la crisis internacional desatada por Trump al amenazar con deportaciones (...)”, el resumen que el usuario realice, facilitará la formulación de preguntas para las herramientas de IA generativa y la búsqueda de contrastación con fuentes que no sean con IA.

### 2.1.2 Formular la pregunta clara al modelo de IA

El siguiente paso del protocolo es la formulación de una pregunta clara, un componente clave para la verificación de noticias falsas, ya que asegura que la consulta dirigida a herramientas de IA generativa, como Copilot, Grok o ChatGPT, sea precisa y centrada para obtener información confiable. En el protocolo, este paso requiere un trabajo meticuloso para evitar ambigüedades que puedan generar respuestas confusas o sin fundamento en los modelos de IA. Una pregunta bien formulada debe ser específica, incluir detalles como fechas, actores y el contexto del suceso, y estar redactada de manera que facilite una respuesta directa. Por ejemplo, en el caso de una noticia sobre la supuesta cancelación del Mundial de Fútbol 2026 por medidas migratorias de Donald Trump, una pregunta precisa sería: “¿Es cierto que la FIFA canceló el Mundial de Fútbol 2026 debido a las políticas migratorias del presidente Donald Trump?” Esta pregunta define el evento, las motivaciones detrás del hecho, los actores involucrados y el contexto temporal, lo que el *prompt* en la IA proporcionará una respuesta centrada y verificable.

Evitar ambigüedades es fundamental, ya que las IA generativas dependen en gran medida de la claridad de los *prompts* para generar respuestas comprensibles. Una pregunta mal planteada puede llevar a interpretaciones erróneas o introducir sesgos ideológicos en el contexto. Una pregunta confusa como: “¿Es verdad que no hay fútbol en 2026?”, podría generar una respuesta general sobre cualquier evento relacionado con el fútbol, sin aproximarse a la noticia falsa de la cancelación del torneo de selecciones por medidas migratorias en Estados Unidos. Para el usuario, desarrollar habilidades de redacción crítica será una competencia esencial para obtener resultados efectivos con la Inteligencia Artificial.

Además, formular preguntas precisas, sin contradicciones ni confusiones, facilitará la comparación de respuestas entre diferentes herramientas de IA y la búsqueda de fuentes confiables en medios de comunicación o instituciones que ofrezcan datos oficiales, como comunicados o boletines de prensa. Este proceso, como parte del protocolo, no solo fomenta un trabajo ético en la verificación de noticias, asegurando que el usuario no contribuya a la propagación de noticias falsas o rumores infundados, sino que también promueve que su interacción con la IA tenga un impacto positivo en la construcción de una sociedad infoxificada con información.

### 2.1.3 Analizar meticulosamente la respuesta generada de la IA

El tercer paso del protocolo consiste en analizar la respuesta generada por la IA. Esto implica evaluar críticamente la información producida por herramientas seleccionadas, como Copilot, Grok o ChatGPT, para decidir su utilidad en la verificación de una noticia. Este análisis meticuloso comienza revisando si la respuesta de la IA incluye fuentes citadas, un componente central para determinar su credibilidad. Por ejemplo, si la IA responde a la pregunta sobre si el barrio La Floresta declaró persona no grata al pintor Oswaldo Guayasamín afirmando que:

No, eso es falso. Según Ecuador Chequea, en una verificación publicada el 30 de mayo de 2025, no existe ningún registro oficial, tanto institucional como en redes sociales o medios, de que el barrio La Floresta, en Quito, haya declarado persona no grata al pintor Oswaldo Guayasamín (ChatGPT, 2025).

Es importante que el usuario verifique si se menciona la fuente exacta de la respuesta de la Inteligencia Artificial, como un enlace a sitios web o artículos de medios de comunicación. Las fuentes de respaldo son una señal de que la IA se basa en información obtenida de Internet, lo que facilita el siguiente paso de comprobación. Sin embargo, si la respuesta es general o carece de fuentes, esto debe considerarse una advertencia, ya que los mensajes generativos sin respaldo no son confiables y podrían reflejar datos desactualizados en las herramientas de IA.

Otro elemento a evaluar son los enlaces. Conocidos como *links*, estos actúan como un puente hacia la validación externa de la información proporcionada por las IAs. Una respuesta de la IA que incluya *links* específicos, como un pronunciamiento oficial en una página destacada o un reportaje periodístico de un medio de comunicación, permite al usuario validar la información.

Es fundamental asegurarse de que la Inteligencia Artificial sustente sus respuestas adjuntando *links* a artículos que puedan ser citados, que desmiente ciertos rumores en la opinión pública. No obstante, el usuario debe analizar los mensajes de la IA sobre los enlaces, verificando que sean funcionales y provengan de sitios web legítimos, ya que algunos modelos pueden generar alucinaciones o enlaces inexistentes. Este análisis es un paso clave en el protocolo, pues requiere que el usuario esté atento para distinguir entre fuentes primarias y documentadas, realizando una crítica sobre la credibilidad de las fuentes proporcionadas por la IA.

En caso de que la IA no proporcione *links* o solo ofrezca menciones dudosas de fuentes sin detalles, el usuario debe proceder a desconfiar y buscar rigurosamente la información en otros buscadores web. Finalmente, el usuario debe fijarse en los detalles específicos de las respuestas de los modelos de Inteligencia Artificial generativa que puedan ser verificados. Este proceso permite al usuario realizar las verificaciones correctamente; la IA actúa como una herramienta de apoyo, pero el juicio crítico es la idea central del protocolo para obtener conclusiones correctas y bien fundamentadas.

#### **2.1.4 Utilizar más de un modelo de IA**

Se recomienda al usuario utilizar más de una IA generativa dentro del protocolo de verificación de noticias falsas. Aplicar este paso aumenta la confianza en el proceso, ya que cada herramienta de Inteligencia Artificial, como Copilot, Grok o ChatGPT, opera con bases de datos y algoritmos distintos que pueden generar alucinaciones diversas sobre el mismo *prompt*. Al repetir los pasos de formular una pregunta precisa y analizar detenidamente la respuesta con diferentes IA el usuario obtendrá una perspectiva más amplia, lo cual le permite identificar desacuerdos en la información recabada. Este planteamiento ayudará al usuario a abordar las imperfecciones imprecisas a los modelos de Inteligencia Artificial generativa de texto. Al comparar diferentes respuestas, los usuarios tienen una ventaja para confiar en la información cuando las IAs coinciden entre sí o para reconocer la necesidad de una investigación más profunda si se contradicen, evitando así aceptar una sola respuesta como verdad absoluta. Aunque los modelos de IA son potentes, su fiabilidad dependerá de cómo han sido entrenados y los datos a los que acceden en un momento determinado.

Cuando las respuestas generadas coinciden en detalles puntuales, como se mencionó en el paso anterior, esto aumenta la probabilidad de que la información sea verídica. Sin embargo, no elimina la necesidad de verificarla con fuentes confiables. Si varios modelos de IA coinciden en que un evento ocurrió en una fecha específica y citan un medio reconocido, esto proporciona al usuario una base sólida para continuar con la verificación. Aun así, estas coincidencias no garantizan absoluta credibilidad, ya que las herramientas podrían estar replicando errores comunes o sesgos presentes en sus bases de datos. Por ello, se recomienda no depender exclusivamente de la tecnología, sino utilizarla como punto de partida para determinar críticamente la información.

Este proceso fortalece la verificación en fomentar el desarrollo de habilidades analíticas para diferenciar entre información creíble y rumores o noticias falsas. Al emplear diversos modelos de IA, se utiliza su variedad para detectar posibles errores compartidos, lo que convierte a la verificación en un ejercicio práctico y basado en una sospecha sobre los datos generados.

### **2.1.5 Consideración sobre el contexto de la noticia**

El usuario debe evaluar quién publicó la noticia falsa, ya que esto es fundamental para determinar la credibilidad de la información. En un contexto de sobrecarga informativa, un ejemplo en redes sociales sobre videos manipulados del papa León XIV criticando a JD Vance, analizar la fuente es esencial. La reputación del medio de comunicación es un indicador clave; las agencias establecidas suelen contar con procesos editoriales rigurosos que combaten la desinformación, mientras que plataformas digitales como TikTok, Instagram o cuentas anónimas en X pueden difundir contenidos sin verificación. Para retornar, en el caso de los videos falsos del papa León XIV, el medio Factchequeado confirmó que fueron generados por IA y compartidos por usuarios de Instagram conocidos por publicar contenidos manipulados, lo que es cuestionable su credibilidad.

Además, determinar cuándo se publicó la noticia es esencial, pues la temporalidad influye en su relevancia y seriedad. En los videos mencionados, circularon en mayo de 2025 y utilizaban material audiovisual de un discurso real; sin embargo, el audio, fabricado con IA, no correspondía al contexto original. Por lo que, verificar la fecha de publicación y la cronología de los eventos permite detectar las inconsistencias, como afirmaciones manipuladas. Esto es particularmente relevante para quienes usan las IA generativas, que no siempre distinguen información reciente, lo que exige contrastar los hechos con fuentes oficiales.

Por último, el usuario debe identificar los posibles sesgos, examinando el estilo, la intención y la interpretación de la IA. En los videos manipulados, el papa León XIV aparecía supuestamente criticando a JD Vance con frases provocadoras como: “volteas más rápido que las páginas de la Biblia” (Factchequeando, 2025 par. 7), diseñadas para generar reacciones y polarizar en la opinión digital. Este lenguaje alterado señala un sesgo, especialmente al carecer de respaldo verificable, como un pronunciamiento oficial del Vaticano. Dado que las herramientas de IA pueden no detectar tales intenciones implícitas, como las narrativas políticas para desacreditar a oponentes. Al usuario se le sugiere a evaluar el estilo y la intención, mediante la verificación

responsable, permite al usuario construir una perspectiva más amplia frente a la desinformación de las noticias falsas o la Inteligencia Artificial.

### ***2.1.6 El usuario tiene que sacar sus propias conclusiones:***

El paso final del protocolo de verificación de noticias falsas con herramientas de Inteligencia Artificial consiste en que el usuario saque conclusiones. Tras integrar las respuestas con las herramientas como Copilot, Grok o ChatGPT, y evaluar las coincidencias de las respuestas generadas de IA mediante el respaldo de fuentes, el usuario debe adoptar una perspectiva crítica y calificativa. Las conclusiones no deben basarse únicamente en las IAs, dado que estas pueden presentar las limitaciones señaladas en los pasos previos del protocolo. Para este paso final, se propone analizar críticamente las evidencias presentadas por las herramientas y asegurar que la decisión refleje un juicio consciente, evitando aceptar la información proporcionada por la tecnología como válido. Se recomienda clasificar a la noticia como verdadera, falsa o incierta, según la información recabada. Una noticia se considera verdadera si las fuentes oficiales confirman los hechos y las respuestas de las IAs son coherentes con la interpretación de las fuentes.

Por otro lado, una noticia se clasifica como falsa si carece de evidencia creíble, las fuentes proporcionadas son dudosas o existen contradicciones claras entre las IAs y los datos verificados. La categoría de incierta surge cuando la información es incompleta o ambigua, comúnmente debido a respuestas contradictorias de los modelos o la ausencia de más fuentes respaldadas. Esta clasificación fomenta la transparencia en el proceso de verificación, permitiendo al usuario justificar su conclusión con base a hechos concretos y no en suposiciones, un principio esencial para combatir con la desinformación digital caótica.

Este paso de realizar la conclusión final ayuda al usuario para cerrar el ciclo de verificación con decisiones correctamente informadas. Al emplear IAs generativas, el usuario debe evaluar la información y reconocer las limitaciones de estas herramientas, como su incapacidad para detectar intenciones implícitas en la desinformación. El usuario al combinar la tecnología con su juicio humano resultará especial crucial, en entornos mediáticos y sobrecargas de información excesiva donde las noticias falsas se propagan rápidamente.

## **2.2 Importancia sobre el uso ético y responsable con los modelos de IA**

Cabe señalar que el usuario, al integrar un análisis ético y reflexivo sobre el impacto de la infoxicación digital, debe comprender el origen de las narrativas falsas en la sociedad y garantizar que el uso de modelos de IA generativa, se realice con responsabilidad. Esto implica evaluar las consecuencias de la difusión de noticias falsas, considerando cómo su difusión manipula percepciones, confrontamientos sociales o la desconfianza en las instituciones gubernamentales. El usuario debería reflexionar sobre quién se beneficia de estas narrativas, como los actores políticos, grupos de interés o creadores de contenido en búsqueda de la viralidad.

Otro campo en el uso ético de la Inteligencia Artificial, es la IA explicativa que está evolucionando como herramienta de apoyo en la lucha contra la desinformación. Este sistema prioriza la transparencia y la comprensibilidad, buscando que las decisiones como la identificación de contenidos falsos sean esclarecidos y fundamentados para los usuarios. Al adecuar con el sistema de IA explicativa, las razones sobre por qué una noticia falsa se la clasifica como desinformativa, estas tecnologías no solo fomentarán la confianza de sus resultados, incluso facilitarán la detección y corrección de sesgos o errores. Con estos avances, los usuarios podrán entender mejor el funcionamiento de la IA y utilizarla como una herramienta confiable para evitar las trampas engañosas con alterar la realidad. (Alejandro García et al., 2024, p. 48).

En los próximos años, es previsible que la IA se integre con principios éticos para la detección de contenidos desinformativos. Hasta ahora, los rumores expuestos en las plataformas digitales son los más comunes sobre la descontextualización de imágenes o textos y la construcción de narrativas falsas. Aunque, si bien es cierto que la IA ya es utilizada por quienes generan desinformación, su presencia es cuestionable en la percepción de los usuarios. Sin embargo, cuando los modelos de IA sean más accesibles y sus resultados sean suficientemente convincentes para aparentar con la realidad, se producirá un descontrol masivo de información engañosa elaborada por las IA.

Este criterio ético se apoya al marco metodológico al posicionar la verificación de noticias falsas como un acto de responsabilidad, especialmente en un contexto de sobrecarga informativa que amenaza los vínculos sociales y la democracia. Al preguntarse los usuarios qué emociones

busca provocar una noticia falsa y cómo podría polarizar la discusión entre las audiencias digitales. Los usuarios digitales podrán con la verificación de contenidos los de mayor impacto mediático en la opinión pública. Al adoptar esta reflexión ética y responsable con el protocolo de verificación contribuirá a formar usuarios digitales más conscientes de su rol en redes sociales y de ciudadanos en fomentar un entorno digital informativo seguro.

### **2.3 Registro de cada paso del protocolo de verificación con modelos IA de noticias falsas**

Para integrar el protocolo de verificación de noticias falsas utilizando IA generativa, es importante diseñar una tabla donde se registre de manera entendible, ordenada y cada paso del proceso. El primer elemento a evidenciar es una descripción breve de la noticia o rumor que se desea verificar, especificando su origen, como un medio de comunicación, una publicación en redes sociales. Este registro sobre el origen de la noticia permite a los usuarios del protocolo a mantener una idea central desde el inicio, evitando confusiones sobre el contenido que se analiza y permitiendo rastrear la fuente original en el caso de que necesite más información la IA.

A continuación, se detallará el prompt formulado a las IAs generativas, como Copilot, Grok o ChatGPT, asegurando que sea preciso para la verificación. Esta instrucción guía para obtener respuestas generadas de las herramientas. Después en la tabla, se hace resúmenes de las respuestas de cada IA, incluyendo si presentan fuentes o *links* que puedan verificarse por parte del usuario. Este ejercicio es importante para comparar la capacidad entre las IAs y evaluar la información proporcionada.

Luego, se enumeran las fuentes consultadas, como medios de comunicación, plataformas de fact-checking o comunicados oficiales de instituciones, que sirven para contrastar las respuestas de las IAs con información disponible. Por lo tanto, no solo va a depender exclusivamente de la IA y buscar activamente la información requerida.

Para registrar los desacuerdos encontrados, permite identificar contradicciones entre las respuestas de las IAs y las fuentes consultadas, por lo que señalar los posibles sesgos o limitaciones será factible. Además, se incluye una reflexión sobre el contexto de la noticia, considerando los aspectos señalados en el paso del protocolo sobre la noticia o rumor. Esta reflexión también evalúa el impacto de la noticia, el usuario preguntándose qué efectos sociales,

políticos o culturales podría tener si se difunde sin verificar el contexto de fondo. Finalmente, en la tabla se formula una conclusión que clasifica la noticia como probablemente verdadera, probablemente falsa o incierta. Hay que justificar a base de las evidencias recopiladas. Esta conclusión breve del análisis, permite una decisión informada orientada a las audiencias digitales o ser como sustento de investigaciones futuras de la IA. La tabla, al incluir todos estos elementos, se asegura que el proceso de verificación sea creíble, ético y reproducible. Por lo tanto, facilitará su aplicación a quienes utilizan herramientas IA para tomar mejores decisiones informadas.

**Tablas de clasificación para evidenciar el proceso del protocolo de verificación con herramientas de IA para noticias falsas**

Noticia seleccionada	Prompt formulado	Respuesta de la IA	Fuentes humanas consultadas del usuario	Desacuerdos entre los modelos	Contexto de la noticia	Conclusión final
<p>FIFA canceló el Mundial de Fútbol 2026 en los Estados Unidos debido a las políticas migratorias del presidente Trump, por lo que Canadá y México seguirán siendo sedes.</p>	<p>¿Es cierto que las autoridades de la FIFA cancelaron el Mundial de Fútbol 2026, debido a las medidas migratorias de Trump?</p>	<p>Grok:                      Considera que es falso y lo califica como un rumor para generar desinformación. Indica que en la plataforma Tik Tok hay videos relacionados con creación de IA, que FIFA ha cancelado el Mundial de Fútbol.                      Copilot:                      Desmiente la noticia y lo considera como un rumor falso. Menciona a sus fuentes consultadas a medios de fact-checking para afirmar su postura.                      ChatGPT:                      La noticia lo califica como falsa. La IA cita a medios de comunicación que se refieren ante posibles escenarios si la situación de Estados Unidos es cada vez compleja con las políticas de Trump.</p>	<p>No existe pronunciamientos oficiales de instituciones gubernamentales de los países anfitriones ante una posible cancelación del torneo.                      La FIFA promociona a través de sus canales oficiales el torneo, venta de entradas y las ciudades sedes.                      En las redes sociales del presidente de FIFA Gianni Infantino, aseguró que ha fortalecido grupos de trabajo con funcionarios de Estados Unidos para supervisar los preparativos del torneo de fútbol.                      No existe ningún comunicado oficial por parte de la Federación Estadounidense de Fútbol.                      Según el portal Politico, asegura que varias organizaciones de derechos humanos muestran preocupación ante posibles violaciones de derechos a personas durante el torneo de FIFA por el origen de nacionalidad.</p>	<p>Copilot afirma que organizaciones de derechos humanos han expresado preocupaciones sobre las políticas migratorias en el torneo de fútbol.                      Grok indica que existen preocupaciones sobre la asistencia de aficionados y posibles discriminaciones, para que exista pronunciamientos de organizaciones de derechos humanos.                      Según ChatGPT, afirmó posibles consecuencias si las políticas migratorias se endurecen. Menciona que existe la posibilidad de reubicaciones de sedes ante el aumento de protestas locales o si existen cambios en las políticas migratorias que rige en EE.UU.</p>	<p>Se ha evidenciado un video viral en plataforma TikTok, con narrativas falsas que FIFA está considerando cancelar el torneo de fútbol debido a las políticas migratorias del presidente Donald Trump. Este video ha tenido millones de vistas y el perfil es considerado en difundir contenido engañoso con IA.</p>	<p>Falsa:                      No hay evidencia de ningún pronunciamiento oficial de FIFA que cancele el torneo a Estados Unidos como sede del Mundial 2026. La organización mantiene planificaciones con sus socios estratégicos, promociona en sus redes sociales la venta de entradas y confirma las sedes seleccionadas para los partidos.</p>

Protocolo de verificación de noticia falta sobre la cancelación del Mundial de Futbol 2026. Elaborado por: Ignacio Peralta Arroyo

Noticia seleccionada	Prompt formulado	Respuesta de la IA	Fuentes humanas consultadas del usuario	Desacuerdos entre los modelos	Contexto de la noticia	Conclusión final
El Papa León XIV mediante una entrevista critica al vicepresidente de los Estados Unidos J.D Vance por las políticas adoptadas del país.	El Papa León XIV dijo esto sobre el vicepresidente Vance: "tú cambias de posición más rápido que las páginas de la Biblia (...)". ¿Es real esta frase del papa?	Grok: Afirma que la frase fue alterada digitalmente y lo considera como <i>deepfake</i> . Añade que, el video engañoso con la supuesta frase circuló en Instagram. Copilot: Menciona que no hay pronunciamientos oficiales del Vaticano y de medios cercanos que cubren a la Iglesia sobre las supuestas declaraciones del papa y la frase fue creada con IA. ChatGPT: Lo califica como falso la frase del papa. Menciona a medios de verificación que indicaron que se trata de un <i>deepfake</i> .	Según el medio de verificación de noticias Factchequeado, la voz del papa León XIV fue manipulado por Inteligencia Artificial. France 24 asegura que León XIV es crítico con las medidas anti migratorias implementadas por el gobierno de EE.UU., el medio de comunicación no evidencia la supuesta frase del pontífice al vicepresidente J.D Vance. El portal web Vatican News, publicó una noticia que el vicepresidente JD Vance mantuvo un encuentro privado con el papa León XIV el 19 de mayo de 2025, pero no detalla que dicha frase fue revelada después de la reunión.	ChatGPT mencionó que León XIV cuando fue cardenal ha criticado en sus redes sociales las políticas de Donald Trump. Además, cita a la fuente La Opinión donde León aseguró que JD Vance tiene una postura de <i>ordo amoris</i> . Copilot afirmó que hubo el encuentro privado entre el papa y el vicepresidente de EE.UU., pero no hubo declaraciones oficiales de ambos. Grok señaló que la frase se trataría de un <i>deepfake</i> y que circuló en Instagram a partir del 15 de mayo 2025 y da más información sobre como identificar contenidos <i>deepfake</i> .	Robert Prevost en sus redes sociales ha sido crítico con JD Vance y al gobierno de Donald Trump antes de ser elegido papa.	Falsa: La frase es parte de un <i>deepfake</i> y fue digitalizado con IA con videos originales para confundir en redes sociales.

Protocolo de verificación de noticia falsa sobre la frase del Papa León XIV en contra del vicepresidente de EE.UU. Elaborado por: Ignacio Peralta Arroyo

Noticia seleccionada	Prompt formulado	Respuesta de la IA	Fuentes humanas consultadas del usuario	Desacuerdos entre los modelos	Contexto de la noticia	Conclusión final
<p>En un video viral se observa que un canguro aguarda para abordar en un aeropuerto.</p>	<p>¿Qué sabes sobre los videos que circulan del canguro que no puede subirse al avión?</p>	<p>Grok:  Afirma que el video difundido en redes sociales es falso y fue creado con IA por la cuenta Infinite Unreality.  Menciona herramientas con IA que pueden detectar material generado con IA. Además, Grok detecta que el video tiene inconsistencias visuales.  Copilot:  Confirma que el video es hecho con IA y ha generado viralidad en redes sociales por la cuenta Infinite Unreality.  ChatGPT:  Menciona el origen de la cuenta en Instagram Infinite Unreality, quien reprodujo el video viral con Inteligencia Artificial. Afirma que es falso el video por el idioma inexistente en que se comunican las personas involucradas y el comportamiento del canguro.</p>	<p>Según Lupa Media el video viral del canguro fue creado con Inteligencia Artificial. Menciona que la publicación original advierte que el video no es real, pero se ha difundido masivamente en redes sociales.  El Financiero explica que el video del canguro su origen es completamente falso. Además, menciona que se desconoce el lugar del aeropuerto y fue creado con IA.</p>	<p>Grok afirmó que analiza contenidos hechos con IA para detectar con certeza el origen del video. Además, en su respuesta recomienda usar <i>Decopy AI</i> y <i>Hive Moderation</i> que determinaron que el 94% del video es elaborado con IA. Copilot y ChatGPT no mencionaron herramientas para detectar IA.  ChatGPT mencionó que el video fue etiquetado con <i>tags IA</i> al difundirse en X.  Copilot deduce que la intención del video fue generar empatía a las personas que aman a los canguros.</p>	<p>El video se publicó originalmente en la cuenta de Instagram Infinite Unreality, genera contenidos de escenas surrealistas con animales, en situaciones humanas.</p>	<p>Falsa: Varios medios de comunicación coincidieron que el video es falso y creado con IA. La cuenta de Instagram Infinite Unreality tiene contenidos creados con IA y utiliza animales en situaciones absurdas que son inverificables.</p>

Protocolo de verificación de noticia falsa sobre video viral de canguro en un aeropuerto. Elaborado por: Ignacio Peralta Arroyo

Noticia seleccionada	Prompt formulado	Respuesta de la IA	Fuentes humanas consultadas del usuario	Desacuerdos entre los modelos	Contexto de la noticia	Conclusión final
<p>Una imagen circuló en el debate digital, que Elon Musk suspendió la cuenta de X del actual presidente Donald Trump, tras las fuertes tensiones entre ambos y la salida del multimillonario del gabinete.</p>	<p>Debido a la tensión entre Elon Musk y Donald Trump, ¿es cierto que el dueño de X suspendió de forma permanente la cuenta del presidente de Estados Unidos?</p>	<p><b>Grok:</b> No tiene evidencia reciente debido a las tensiones políticas entre Trump y Musk para suspender la cuenta de X del actual presidente. Afirmo que hay desacuerdos con los proyectos fracasados de Musk y ha generado una enemistad con Trump. Lo califica como una noticia falsa debido a que generó confusión con relación a eventos pasados.</p> <p><b>Copilot:</b> Indica que existe una imagen viral en redes sociales en la que supuestamente Elon Musk suspendió la cuenta de X a Donald Trump. En su verificación dice que la información ha sido desmentida por fuentes oficiales y que en X permanece activo el perfil de Trump. Menciona que hay conflictos internos entre ambos, pero no hay medidas tomadas por Musk como propietario de X para suspender la cuenta.</p> <p><b>ChatGPT:</b> Menciona que la cuenta de X no está suspendida de Donald Trump por orden de Musk. Reconoce que hay una imagen viral circulando en redes sociales, pero es descontextualizada.</p>	<p>Según el portal BBC, Elon Musk restableció la cuenta de X del actual presidente Trump, en noviembre de 2022. No menciona que, debido a las tensiones políticas entre ambos, Musk haya tomado la decisión de suspender la cuenta indefinidamente.</p> <p>El medio digital Lupa Media y Reuters Fact Check desmintieron que no existe ninguna evidencia que Elon Musk suspendió definitivamente la cuenta de X de Trump, tras la difusión de la imagen falsa. En su proceso de verificación de datos, ambos portales digitales coinciden que la imagen es descontextualizada y pretende confundir en el entorno digital porque la cuenta de Trump si fue suspendida en 2021 tras los ataques al Capitolio.</p>	<p>Copilot y ChatGPT confirman que el rumor surge tras una imagen falsa en redes sociales que pretendió confundir. Grok no da detalles del origen del rumor de la imagen. Grok profundiza más las tensiones políticas entre ambos personajes públicos por los proyectos de ley y propuestas fracasadas. Copilot y ChatGPT no mencionan con más información la tensión política. Por su parte, el modelo ChatGPT menciona que Elon Musk habilitó, en noviembre de 2022 la cuenta de X de Trump. La IA constató la información con el medio Lupa Media.</p>	<p>Esta noticia falsa circula en redes sociales, tras un momento de tensión entre Trump y Musk, quien fue un aliado importante para la reelección de Trump en 2025. La relación se quebranta cuando Musk criticó el plan fiscal de Trump, tildándole como repugnante y renunció de su cargo en el gabinete.</p>	<p>Falso: La cuenta de X del actual presidente Donald Trump está activo sin ninguna restricción. Elon Musk y Trump han declarado tener conflictos por propuestas de gobernabilidad y proyectos fracasados, pero no se han tomado medidas extremas según los medios de comunicación y la empresa de X. En 2021, la plataforma antes llamada Twitter decidió suspender temporalmente la cuenta de Trump por incitar a la violencia, tras los ataques al Capitolio.</p>

Protocolo de verificación de noticia falsa sobre la suspensión de la cuenta de X de Donald Trump. Elaborado por: Ignacio Peralta Arroyo

### **2.3.1 Justificación de la estructura de la tabla de clasificación del protocolo ante la era de la sobrecarga informativa en usuarios digitales**

En el contexto actual en la era digital, representada por la sobrecarga de información conocida como la infoxicación, surge la necesidad de herramientas para procesar y verificar el flujo constante de datos. La infoxicación, es el exceso de consumo de información que dificulta la capacidad de los individuos en verificar contenidos reales de aquellos que son engañosos. Esta problemática mencionada en el primer capítulo, la elaboración de la tabla de clasificación como propuesta en el protocolo de verificación de noticias falsas mediante Inteligencia Artificial generativa se presenta como una solución. El protocolo organiza de manera ordenada el proceso de análisis de información, y también mantiene la transparencia, trazabilidad y un enfoque ético en la lucha contra la desinformación, contribuyendo a moderar los efectos del abarrotamiento informativo digital.

Si bien los autores explicaron que la infoxicación se manifiesta en la proliferación masiva de contenidos a través de plataformas digitales, medios de comunicación y las redes sociales, donde la inmediatez de difusión contantemente supera la capacidad de un pensamiento crítico de la información. Este problema incrementa el riesgo de quebrantar a las audiencias digitales ante narrativas falsas, engañosas y manipuladas que pueden influir en la toma de decisiones tanto individuales como colectivas. Ante este factor de sobrecarga informativa, la tabla de clasificación para demostrar el protocolo se respalda como un instrumento que permite descomponer el caos informativo. Al establecer un marco metodológico para registrar y evaluar cada aspecto del proceso de verificación, la tabla es un componente para ordenar los datos mediante un análisis comprensible. El protocolo resulta importante para contrarrestar la confusión generada por la infoxicación informativa, ya que proporciona una guía de pasos que facilita proceder con la identificación de la información en medio del ruido digital.

Además, la tabla favorece la sistematización del protocolo, lo que es valioso en un entorno donde la rapidez de la información es cada vez difundida masivamente. Al documentar cada etapa de la verificación de la noticia falsa seleccionada de manera detallada, desde la identificación del contenido hasta la evaluación de su veracidad, se crea un registro que permite revisar y validar los pasos propuestos. Esta característica aparte de mejorar la eficiencia del análisis, sino que también evidencia que las conclusiones estén basadas en un razonamiento humano, en lugar de las impresiones superficiales influenciadas por la sobrecarga informativa.

Dentro del contexto de sobrecarga de información, donde la cantidad consumida de contenidos puede abrumar incluso a los verificadores con experiencia en este ámbito, esta clasificación se convierte en un recurso de aporte para manejar los datos recabados.

## **2.4 Funcionalidad de los componentes de la tabla del protocolo**

La estructura de la tabla de clasificación está diseñada para contener los elementos claves que optimizarán el proceso de verificación y lo alinean con los desafíos de la saturación informativa en los usuarios. Cada componente cumple con una función específica que contribuye para la aplicabilidad del protocolo. La estructura de la tabla, al integrar estos componentes, responderá a los desafíos de la sobrecarga informativa al proponer un protocolo sistemático, ético y eficiente que colabora a los usuarios para combatir la desinformación con responsabilidad, fortaleciendo el pensamiento crítico del consumo de los contenidos en una era de sobrecarga digital.

En conclusión, la estructura de la tabla de clasificación del protocolo de verificación de noticias falsas mediante el uso de Inteligencia Artificial se justifica plenamente como una respuesta a los retos de la infoxicación digital. Al ofrecer un marco estructural, para procesar la información, la tabla permitirá a los verificadores enfrentar el exceso de datos con herramientas que priorizan la claridad, la precisión y la responsabilidad en su uso ético. En un mundo digital donde la sobrecarga informativa es una amenaza con desfigurar entre la verdad y la mentira, ayuda a quienes la utilizan para desempeñar un papel en la construcción de un entorno digital más confiable y consumible. Su diseño hace respuesta a los cuestionamientos de una era digital compleja como un recurso para navegar y transformar el panorama de la información expuesta en las plataformas digitales.

## Capítulo III:

### Interpretación de los hallazgos

En el presente capítulo interpretaré los resultados de la aplicación de la propuesta del protocolo previamente realizado, con el objetivo de comprender la problemática expuesta en el primer capítulo de la sobrecarga informativa en plataformas digitales y la presencia accesible de herramientas con Inteligencia Artificial, como medios de verificación de noticias falsas.

#### 3.1 El impacto de la Inteligencia Artificial para detectar datos desinformativos mediante el protocolo

La implicación de la Inteligencia Artificial en los usuarios digitales ha impactado masivamente en los entornos mediáticos y digitales. Es cierto que, esta tecnología avanzada ha sido capaz de procesar la información masivamente en su base de datos e interpretarla constantemente, mediante la instrucción o *prompt* del usuario. La Inteligencia Artificial ha facilitado avances impresionantes para la creación de contenidos informativos, educativos y periodísticos, pero a su vez ha permitido la capacidad de realizar estrategias desinformativas y manipuladoras sin precedentes (Cruz, 2025, p.12). Entonces, quienes realizan contenidos engañosos por medio de las plataformas digitales, también tienen como herramienta de apoyo la Inteligencia Artificial para alterar las percepciones de los usuarios quienes asumen el riesgo de difundir su contenido malicioso.

Con el progreso constante de la evolución de la Inteligencia Artificial y la incertidumbre de lo que es capaz de imitar a la realidad, esta herramienta ha sido eficaz en la verificación de datos en precisión. Los sistemas de procesamiento de lenguaje natural (PLN) utilizados en el protocolo para los usuarios, sus algoritmos detectaron textos y examinaron la información engañosa. Por lo que concuerdo con la interpretación del autor Jorge Cruz (2025), donde afirma que: “Estos algoritmos identifican patrones e inconsistencias en grandes conjuntos de datos, lo que facilita su detección de desinformación. La verificación automatizada mediante la IA puede realizarse en tiempo real” (p.15).

No obstante, los sistemas de automatización exponen limitaciones al verificar la información proporcionada. Ahora bien, la Inteligencia Artificial es capaz de monitorear los acontecimientos difundidos en redes sociales y verificar los contenidos virales por las audiencias, pero frecuentemente presentan dificultades los sistemas IA en analizar afirmaciones complejas o la información inverificable, calificándolo como simples rumores.

Las ventajas que tiene la Inteligencia Artificial para su uso en la verificación de datos es un desafío importante en la actualidad. Tras analizar la información elaborada de los sistemas IA, evidenció un problema significativo que es la falta de transparencia respecto a los algoritmos que procesan los modelos. Es necesario que, los medios de verificación en información propongan mecanismos para los usuarios al explicar el funcionamiento de los algoritmos de las IA, si es que entonces realmente son herramientas de apoyo para validar una información. Con esta inquietud podrán guiar al usuario a identificar los patrones algorítmicos sesgados y que no integren en sus datos información engañosa.

Una particularidad de los modelos de la Inteligencia Artificial aplicados con el protocolo es la regularidad de datos en tiempo real. Tiene la capacidad de generar la información precisa, personalizada y oportuna para el usuario en informarle de noticias o eventos recientes. La IA hace competencia directa a los criterios tradicionales de los medios de comunicación para tomar una postura sobre el acontecimiento viralizado. Entonces, la IA es capaz de mitigar la sobrecarga de información que están expuestos los usuarios con las plataformas digitales. Además, la IA destaca en ese aspecto encontrado para el protocolo porque tiene la capacidad de analizar múltiples fuentes de datos y detectar rápidamente los más relevantes para que el usuario decida que hacer con la noticia seleccionada previamente para la verificación.

Cabe señalar, algo relevante en la aplicación del protocolo con los modelos de Inteligencia Artificial es que puede analizar datos de los perfiles, las interacciones en cualquier red social, y los patrones de consumo de información para personalizar y recomendar las noticias en donde se sustenta para demostrar al usuario la recolección de datos recabada. Esta capacidad predictiva de las IA permite identificar contenidos de alto impacto, mediante el monitoreo de redes sociales o cualquier tipo de plataformas en línea, por lo cual los algoritmos pueden detectar tendencias ascendentes que podrían tener consecuencias sociales, políticas, legales y económicas en la sociedad (Cruz, 2025, p.17).

Finalmente, los modelos de IA en la comparación de las respuestas generadas de texto de las noticias falsas para realizar el protocolo de verificación, analiza la relevancia de los actores sociales involucrados. Las IA utilizaron un análisis de redes sociales, sitios web y algoritmos sobre la figura relevante de la noticia. El autor Jorge Cruz (2025) afirma que, los modelos automatizados de IA están utilizando técnicas de análisis de redes sociales, algoritmos para detectar influencias y sistemas para analizar sentimientos por lo que, se pueden reconocer a los actores principales, medir su influencia en la sociedad o en la política y establecer interpretaciones de la IA de cómo afectan el debate público (p.18).

La Inteligencia Artificial aparte de detectar contenidos desinformativos, también puede adaptarse a las particularidades de cada usuario o comunidad digital. Al analizar las preferencias, comportamientos y contextos locales, como bien lo explicó el autor Cruz, los algoritmos personalizados logran identificar contenidos engañosos que vulnera, como los rumores señalados en las respuestas de las IA para audiencias concretas. Esta capacidad de ajuste eleva la precisión de la IA, permitiendo que las alertas sean más relevantes y oportunas. Sin embargo, este enfoque también plantea interrogantes sobre la privacidad, ya que dependen de datos individuales para funcionar. La personalización, así, es un avance significativo, pero su implementación debe ir acompañado de regulaciones éticas.

La verificación automatizada ha sobresalido como una tendencia significativa debido a la rapidez con la que se propaga la información en los entornos digitales. Estas herramientas disminuyen la dependencia de la intervención humana en las primeras fases de verificación, agilizando las respuestas de las redacciones y permitiendo que los periodistas y verificadores se concentren en un análisis más profundo y contextual de los datos. Sin embargo, su eficacia está limitada por su capacidad para reconocer patrones lingüísticos diseñados para engañar y para gestionar la ambigüedad discursiva. Este aspecto se relaciona directamente con la aplicación del protocolo de verificación de noticias falsas mediante Inteligencia Artificial generativa, donde las herramientas automatizadas desempeñaron un papel clave para la mitigación de la sobrecarga de información. No obstante, el éxito de este protocolo depende de su habilidad para resolver la ambigüedad discursiva en afirmaciones complejas, lo que resalta en la importancia de complementar estas tecnologías con el juicio humano mencionado dentro de los pasos de verificación de usuarios para lograr una comprobación precisa.

### **3.2 Dependencia de la Inteligencia Artificial como dispositivo de control**

La perspectiva de Michel Foucault sobre los dispositivos de control (Vega, 2017, p.143), sugiere que las tecnologías modernas gestionan información y configuran las relaciones de poder al regular cómo los individuos perciben y actúan en el mundo. En relación con el protocolo, los modelos de IA podrían interpretarse como dispositivos de control al influir en la manera en que los usuarios digitales acceden y validan la información. Al seleccionar contenidos y clasificarlos como reales o falsos, estas herramientas orientan a las narrativas inventadas que prevalezcan en la discusión pública. Este proceso de selección y validación, aunque está diseñado para proteger contra la infoxicación digital, puede configurar inconscientemente las percepciones de los usuarios determinando qué información se considera legítima y, por tal razón, ejerce una forma de poder sobre el conocimiento. Sin embargo, la aplicación del protocolo revela una repetición en el papel de la IA. Por un lado, actúa como un mecanismo de resistencia frente a la desinformación, sugiriendo a los usuarios a cuestionar la información consumida, por otro su capacidad de analizar grandes volúmenes de datos y personalizar las respuestas dependiendo del consumo de contenidos del usuario que podría asemejarse a las tecnologías de vigilancia descritas por Foucault.

Ahora bien, el ser humano tiene una de sus habilidades principales que es el desarrollo del pensamiento crítico. Se consideraba que esta habilidad era únicamente para los seres humanos, esta orientación se está discutiendo con los últimos avances tecnológicos de la Inteligencia Artificial. La habilidad que desarrollan los seres humanos de pensar críticamente es a medida, mientras progresan profesionalmente y estudian, por lo que les facilita escoger una toma de decisiones debido a la capacidad que han evolucionado en conocimientos y experiencias previas. Por lo tanto, este proceso del pensamiento crítico requiere constancia en los procesos cognitivos humanos, algo que la Inteligencia Artificial aún no ha conseguido imitar.

El ser humano, como sujeto que construye conocimiento, no percibe la realidad desde una perspectiva neutral, sino que está configurado por sus ideas, emociones, la edad, experiencias vividas y un intelecto particular (Oviedo, 2023, p.74). Este aspecto es destacable, ya que la Inteligencia Artificial carece de elementos como la sensibilidad, la experimentación de vivencias previas, lo que limita su capacidad para generar un pensamiento desarrollado tan complejo como es del ser humano. Entonces, la capacidad intelectual de la IA produce

resultados menos elaborados, carentes de juicios de valor y la simpatía que caracterizan a la subjetividad humana.

El investigador científico Ramón López de Mántaras señala que las máquinas, como los modelos de IA generativa de texto son escasos en desarrollar capacidad moral, ya que son herramientas tecnológicas diseñadas para procesar datos sin un juicio ético o comprensión de las consecuencias de sus acciones. Su afirmación involucra que la IA no puede distinguir entre el bien y el mal, ni tampoco evaluar el impacto social de las decisiones que facilita a los seres humanos, como se experimentó con el objeto de estudio para el protocolo de verificación (Villuendas, 2024, par. 10).

La afirmación expuesta de López de Mántaras, considero que la limitación pone la responsabilidad a los desarrolladores y científicos, quienes deben garantizar que los algoritmos de las IA se diseñen con principios éticos que prioricen la estabilidad social. En el problema identificado con la sobrecarga de información y produce efectos como la desinformación, esto implica que las IA generativas utilizadas para verificar las noticias falsas seleccionadas deberían ser programadas para minimizar sesgos ideológicos, por lo que se evitaría la discusión de narrativas inventadas y promover la transparencia de sus datos y procesos de desarrollo. La responsabilidad de los usuarios es que las herramientas no se conviertan en manipular y generar falsedades a los acontecimientos recientes.

### **3.3 Fiabilidad de la Inteligencia Artificial**

La fiabilidad de la Inteligencia Artificial en la verificación de noticias falsas depende de diversos factores metodológicos que se hicieron evidentes a lo largo de la aplicación del protocolo propuesto. Este protocolo, al estructurar cada paso del proceso de verificación, facilitó evaluar la precisión de las respuestas generadas por las IA e identificar sus limitaciones cuando se las pone en prueba a situaciones de desinformación digital.

En primer lugar, se constató que las IA generativas seleccionadas pueden ofrecer respuestas coherentes, detalladas y, en la mayoría de casos, acompañadas de fuentes de información. Esta capacidad les otorga una ventaja en el entorno de sobrecarga informativa, donde los usuarios necesitan respuestas rápidas y sencillas de comprender. Sin embargo, la precisión de estas respuestas depende, de la formulación específica del *prompt* y del acceso que tenga el modelo

a datos actualizados en tiempo real. En ese sentido, la fiabilidad de las herramientas no se origina en sí, sino en cómo el usuario usa críticamente y lo emplea en su entorno.

Asimismo, la comparación entre distintos modelos reveló una particularidad entre las contradicciones por la ausencia de fuentes o la falta de información que generaron incertidumbre. Esto se evidenció en los ejemplos reales de noticias falsas seleccionadas, donde algunos modelos señalaban la falta de evidencia institucional, mientras que otras IA iban más allá a profundizar el origen de los rumores o la existencia de los *deepfakes* como herramienta de engaño.

Otro elemento que incide directamente en la fiabilidad de las IA es su reconocimiento de declarar sus limitaciones. Por ejemplo, en algunas herramientas como Copilot incluyó advertencias sobre la necesidad de realizar verificaciones humanas, lo que evidencia una perspectiva ética más transparente para el usuario. Esta práctica es fundamental para que los usuarios comprendan que las IA si son guías de apoyo, pero no deben ser consideradas como buscadores de información.

Además, mediante la fase de prueba con el protocolo se demostró que la IA es más fiable cuando se integra un mecanismo que evalúe el origen de la noticia falsa porque brindó diferentes perspectivas para que el usuario reconozca y decida que herramienta es más convincente. En ese marco, la fiabilidad no se mide únicamente por la exactitud de una respuesta, sino por la capacidad de la herramienta para integrar un ejercicio de distinguir situaciones, donde la interpretación humana es irremplazable.

Para finalizar sobre la fiabilidad, la Inteligencia Artificial tiene sus condiciones para ser herramienta de apoyo para sobrellevar la infoxicación digital y el usuario no debería sentir miedo, desconfianza o preocupación siempre y cuando lo utilice de forma responsable cualquier modelo de su preferencia, además tiene el riesgo de promover y dotar a las herramientas información engañosa para aumentar la desinformación y saturar aún más la sobrecarga de los millones de contenidos circulando en las plataformas digitales.

### **3.4 Visibilizar a los medios de verificación de contenidos y usuarios participantes para disminuir la sobrecarga de información**

Los medios de comunicación especializados en la verificación de información son esenciales para defender la democracia en tiempos de polarización en las sociedades civiles. Cuando los usuarios identifican que todo parece estar dividido sobre un acontecimiento viral en la opinión pública, el periodismo en contrarrestar los hechos son clave para reconstruir el origen común para la discusión generalizada. Es decir, los medios quienes verifican información hacen más que solo corregir errores, tienen el compromiso en defender el derecho de las sociedades a decidir con evidencias, más no defenderse con las mentiras.

Las organizaciones verificadoras de información exclaman una petición a financiadores, plataformas digitales y gobiernos centrales a defender la verificación porque es un derecho fundamental que la sociedad civil esté informada. Ahora bien, si el derecho se debilita, no está en debate el futuro del periodismo, sino la salud misma de los usuarios digitales. Con el protocolo debería ser una base de ayuda para los usuarios en mitigar la saturación digital porque las plataformas de redes sociales deberían ser parte de la solución ante la problemática expuesta y no ser fomentadores del caos digital. Los medios de verificación encuentran cada vez, más problemas del manejo de datos transparentes de las redes sociales lo que les obliga a reclamar a tener un mayor acceso a la información y sobre todo a quienes combaten la desinformación digital.

Las compañías creadoras de los sistemas de Inteligencia Artificial deberían asumir su responsabilidad en la lucha contra la desinformación y el caos sofocante de consumo de contenidos elaborados por IA. Las empresas deberían dar espacio a colaboradores en verificación de contenidos informativos para asegurar su veracidad e implementar mecanismos para proteger la integridad de la información. Los medios de verificación no pueden controlar todo el contenido digital excesivo que circula, por lo que es urgente el rol de los gobiernos y parlamentos en proteger la libertad de expresión, es su obligación. Si no hay cambios notables para mitigar la desinformación, mientras sigan con intentar silenciar a los medios y organizaciones por su trabajo en ser transparentes y estar al servicio de la ciudadanía, provocan aún más la polarización de las sociedades y distorsionan las realidades con narrativas cuestionables.

El entorno mediático marcado por la masividad de datos, la participación activa de los usuarios digitales en la verificación de información se vuelve un elemento esencial para reducir el caos informativo. No basta con delegar esta responsabilidad a los medios de verificación o a herramientas de Inteligencia Artificial, es indispensable generar estrategias pedagógicas y comunicativas que motiven a los usuarios a convertirse en actores críticos dentro de la era digital.

Una de las formas más efectivas para fomentar la participación es visibilizar cómo la desinformación afecta directamente a la vida cotidiana del usuario. Al conectar los riesgos de la infoxicación con influir en las decisiones personales, como votar, consumir productos o reaccionar ante situaciones sociales, se despierta un interés más analítico por entender las consecuencias de compartir información sin verificar. Esto humaniza el problema y trasciende la lógica técnica, apelando a la responsabilidad individual y colectiva.

Del mismo modo, la motivación también puede surgir de la apropiación de herramientas accesibles como los modelos de Inteligencia Artificial. Al promover a los usuarios protocolos de actuación, como el que se desarrolló en esta investigación, se reduce la percepción de que la verificación era una tarea exclusiva solo de expertos. Cuando los usuarios descubren que pueden consultar, analizar y comparar respuestas con otras IA en cuestión de minutos, se les da el rol como agentes capaces de tomar decisiones informadas en medio de la avalancha digital.

Por otro lado, fomentar a comunidades digitales de verificación también incentiva la participación. Realizar espacios colaborativos donde los usuarios compartan recursos, dudas, resultados de verificación o experiencias con modelos de IA, no solo fortalecerán la alfabetización mediática porque crearán una red entre apoyo frente al contenido manipulador. En estas comunidades, la verificación deja de ser una tarea individual y se transforma en una práctica compartida, discutida y responsable.

La promoción mediante actividades recreativas también puede desempeñar un papel interesante para la motivación. Al incorporar dinámicas lúdicas, como retos de verificación, reconocimientos virtuales o actividades colaborativas, puede estimular el interés de los usuarios más jóvenes y reforzar el pensamiento crítico. Este tipo de estrategias convierte al problema expuesto el proceso de verificación en una actividad atractiva, formativa y continua.

Finalmente, realizar campañas de sensibilización, concientización y alfabetización mediática incluyan propuestas claves sobre el valor del pensamiento crítico y el impacto de una ciudadanía informada. Esto, a los usuarios los motivará al despertar la conciencia en que tienen la capacidad de decidir al confirmar o desmentir una noticia, con el apoyo de herramientas IA. Considero que la lucha contra la infoxicación no se vence únicamente implementando más tecnología, sino al contrario formar personas comprometidas que asuman el reto de ser agentes digitales en promover sociedades mejor informadas.

Para culminar este capítulo, es oportuno retomar los aportes del investigador Ramón López de Mántaras, quien plantea que, por muy sofisticadas que lleguen a ser las futuras inteligencias artificiales, siempre serán distintas a la inteligencia humana. Esta afirmación cobra especial relevancia al evaluar los resultados del protocolo aplicado en esta investigación. Las IA pueden simular cualquier tipo de razonamiento, generar respuestas con datos verificados, incluso determinar patrones de falsedad en las discusiones digitales; sin embargo, no cuentan con cuerpo, emociones ni experiencias de socialización humana, elementos importantes que el usuario tiene para la construcción del juicio crítico (López de Mántaras, 2015, p. 101).

Esta diferencia esencial debe motivar una reflexión ética: al tratarse de inteligencias ajenas a nuestros valores, su uso masivo en tareas como la verificación de contenidos no puede ser indiferente ni automático. Si bien es cierto que las herramientas evaluadas en este protocolo han demostrado la utilidad, la aplicación debe mantenerse como una asistencia, no como un reemplazo del pensamiento humano. El riesgo de asumir que una IA puede verificar por sí sola una noticia, sin intervención crítica del usuario, el riesgo aumentaría de reproducir errores, sesgos o narrativas manipuladas sin control ni responsabilidades.

El investigador crítico con la Inteligencia Artificial, también advierte que sobre una nueva dimensión en relación entre IA y la humanidad surge la emergencia descontrolada de la creatividad computacional. Las máquinas, al generar arte, narrativas o contenidos digitales que aparentan originalidad, están transformando el acceso a la información, y el concepto mismo de creación y posverdad. Esta problemática proyecta un escenario donde la verificación no solo será necesaria afrontar las noticias falsas, igualmente los discursos generados creativamente por algoritmos sin intenciones humanas.

En comparación, con el protocolo diseñado en esta investigación cobra una nueva dimensión. Fortalecer la capacidad de los usuarios para detectar y comprender las intenciones, limitaciones y potenciales de las herramientas de las IA en un entorno informativo cada vez más automatizado y complejo. La alfabetización mediática, la ética aplicada y la participación activa del usuario serán, más que nunca, condiciones necesarias para asegurar la información con responsabilidad en la era de la Inteligencia Artificial.

En fin, el futuro de la IA en la humanidad no va a depender del poder computacional, más bien la capacidad humana de mantener el criterio, el origen de la información, los valores centrales del conocimiento y la comunicación. El protocolo aplicado es una muestra de que la tecnología puede estar al servicio del juicio humano, siempre y cuando no olvidemos que, por muy inteligentes que sean las máquinas, nunca podrán ser seres emocionales.

## Conclusiones

El protocolo de verificación de noticias falsas desarrollado en esta investigación se consolida como una herramienta metodológica que responde de manera pertinente al desafío de la sobrecarga informativa en entornos digitales. Para que los usuarios puedan verificar con mayor seguridad y se reduzca el riesgo de aceptar como verídica cualquier noticia, contenido engañoso que circule en redes sociales o plataformas digitales, este protocolo con su estructura ordenada contribuye a fortalecer competencias críticas que son esenciales para el ejercicio de una ciudadanía informada y consciente de los retos ante la infoxicación digital.

La integración de los modelos de Inteligencia Artificial, como Grok, ChatGPT y Copilot, dentro del protocolo confirma que la IA al ser capaz de procesar grandes volúmenes de datos en tiempo real, se posiciona como un recurso estratégico que facilita la detección temprana de inconsistencias y de narrativas descontextualizadas. No obstante, en el análisis de este trabajo advierten que la eficacia de la Inteligencia Artificial no sustituye la importancia de aplicar criterios humanos de interpretación en especial si se analizan afirmaciones complejas que involucran contextos históricos, culturales o políticos. En este sentido, con la aplicación del protocolo destinada al análisis del contexto evidencia que la verificación no es un proceso únicamente tecnológico, sino una práctica que requiere el juicio responsable de los usuarios.

Además, desde la perspectiva de Foucault en los dispositivos de control, la Inteligencia Artificial puede entenderse como una tecnología que ordena y clasifica el conocimiento, estableciendo niveles discursivos que inciden en la percepción de la realidad. La capacidad de los modelos para seleccionar información, priorizar ciertos datos y omitir otro hallazgo corre el riesgo de consolidar una forma única de regular y seleccionar qué contenidos son válidos y cuáles se descartan como inverificables. Por tanto, la aplicación del protocolo propuesto permite equilibrar esta tendencia y exigir que toda información producida por la Inteligencia Artificial sea contrastada con fuentes humanas y sometida a un análisis crítico que valore su usabilidad y credibilidad.

La eficacia de los sistemas automatizados de verificación que usan algoritmos de procesamiento de lenguaje natural es evidente en entornos con exceso de información. Estos sistemas pueden

reconocer rápidamente los patrones falsos o las inconsistencias con mucha más rapidez que los métodos tradicionalistas. En el protocolo, la comparación de respuestas entre distintos modelos de IA permitió revelar coincidencias y desacuerdos que orientan al usuario hacia una interpretación más rigurosa con los datos. Sin embargo, la experiencia acumulada en esta investigación demuestra que la capacidad técnica de la Inteligencia Artificial no resuelve por completo la ambigüedad a ciertos contenidos, por lo que se requiere mantener un enfoque que combine la eficiencia de los sistemas automatizados con la capacidad interpretativa de análisis humana.

En relación con la fiabilidad de la Inteligencia Artificial, este trabajo concluye que los modelos generativos empleados a pesar de su técnica y su habilidad para elaborar respuestas coherentes y detalladas no deben considerarse como fuentes de la verdad absoluta. Su desempeño está cuestionado en la calidad y actualización de los datos, así como la transparencia de sus algoritmos, que en ocasiones reproducen sesgos o generan alucinaciones informativas que pueden inducir al error. La fiabilidad de la IA, por tanto, es aproximada y depende de que su uso se acompañe con estrategias de verificación y de un juicio crítico por el humano. Entonces, la Inteligencia Artificial es una herramienta de apoyo que ayuda a desarrollar la capacidad humana de análisis, pero no la va a reemplazar. El valor de la IA reside en la capacidad de procesar información con velocidad que deben ser cuidadosamente contrastados antes de asumir conclusiones definitivas.

Finalmente, esta investigación destaca que el futuro de la alfabetización mediática de la información tiene que consolidar un enfoque ético en el uso de la Inteligencia Artificial. Es muy importante formar usuarios con habilidades de pensamiento crítico y la aplicación de protocolos transparentes y emergentes que constituyen soportes fundamentales para luchar contra la desinformación expuesta en entornos digitales. De este modo, la Inteligencia Artificial puede afianzarse como un aliado estratégico que debe contribuir a construir entornos informativos más confiables, accesibles y seguros que mantengan la libertad de expresión y el derecho a una información verídica.

## Referencias

- Abeliuk, A. (2023). Sesgos algorítmicos en las redes sociales. *Revista Bits de Ciencia*, (24).
- Aguaded, I., & Rodríguez, L. M. R. (Eds.). (2018). Competencias mediáticas en medios digitales emergentes. *Comunicación Social*.
- Alonso-González, M. (2021). Desinformación y coronavirus: el origen de las fake news en tiempos de pandemia. *Revista de Ciencias de la Comunicación e Información*, 26, 1-25. <http://doi.org/10.35742/rcci.2021.26.e139>
- Aronés Cisneros, Angel, Aronés Cisneros, Rocío, Alegre Palomino, Cynthia, & Colquehuanca Solis, Joel. (2024). INTELIGENCIA ARTIFICIAL EN LA ELABORACIÓN E INTERPRETACIÓN DE IMÁGENES. UNA HERRAMIENTA EN LA EDUCACIÓN. *Areté, Revista Digital del Doctorado en Educación*, 10(especial), 101-117. Epub 31 de enero de 2025. <https://doi.org/10.55560/arete.2024.ee.10.8>
- Cárdenas, E. (2025, marzo 26). ‘Grok’, la IA de X, irrumpe en el ‘juego’ de la campaña electoral. *Ecuador Chequea*. <https://ecuadorchequea.com/grok-la-ia-de-x-irrumpe-en-el-juego-de-la-campana-electoral/>
- Cómo escribir indicaciones de IA: una guía para escribir indicaciones de IA eficaces | Coursebox AI. (s. f.). <https://www.coursebox.ai/es/blog/how-to-write-ai-prompts>
- Cruz Silva, J. (2025). CREDIBILIDADE e VERACIDADE a PARTIR DA RELAÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL e FACT-CHECKING [Relatório de Pós-doutorado]. Universidade Estadual Paulista (UNESP), Faculdade de Arquitetura, Artes e Comunicação de Bauru.

Diego Olite, Francisca Mercedes, Morales Suárez, Ileana del Rosario, & Vidal Ledo, María Josefina. (2023). Chat GPT: origen, evolución, retos e impactos en la educación. *Educación Médica Superior*, 37(2) [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412023000200016&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412023000200016&lng=es&tlng=es).

Factchequeado. (2025, 3 junio). No, estos videos del papa León XIV donde critica a JD Vance no son reales y están manipulados. Factchequeado.com. <https://factchequeado.com/verificaciones/20250603/videos-papa-critica-jd-vance-manipulados/>

Fernández, P. C. (2023). Efectos de la sobrecarga de información en el comportamiento del consumidor de noticias: El doomscrolling. *VISUAL REVIEW. International Visual Culture Review/Revista Internacional De Cultura Visual*, 14(1), 1-11.

Fernández, Y. (2023, 16 febrero). Qué es Copilot de GitHub y cómo funciona esta Inteligencia Artificial que te ayuda a programar. Xataka. <https://www.xataka.com/basics/que-copilot-github-como-funciona-esta-inteligencia-artificial-que-te-ayuda-a-programar>

Fernández, Y. (2024, 9 marzo). Microsoft Copilot: qué es y cómo funciona este chat de Inteligencia Artificial. Xataka. <https://www.xataka.com/basics/microsoft-copilot-que-como-funciona-este-chat-inteligencia-artificial>

Fondevila-Gascón, Joan-Francesc, Huamanchumo, Abigail, Martín-Guart, Ramon, & Gutiérrez-Aragón, Óscar. (2024). El chatbot como factor de éxito comunicativo, de marketing y empresarial: análisis empírico. *Correspondencias & análisis*, (19), 47-70. Epub 25 de marzo de 2024. <https://doi.org/10.24265/cian.2024.n19.02>

García-Gaona, H., Hernández-Gaona, L. Y., Gámez-Ortíz, R. A., & Gaona-Julian, G. (2023). Perspectivas del uso de Chatbots en la educación superior: caso de estudio de la Universidad Tecnológica de Gutiérrez Zamora. *Revista Multidisciplinaria de Ciencia, Innovación y Desarrollo*, Gutiérrez Zamora, 32, 37.

Guevara, L. G. O. (2023). Dilema de la Inteligencia Artificial: pensamiento crítico y generaciones digitales. *Realidad y Reflexión*, 1(58), 69-83. <https://doi.org/10.5377/ryr.v1i58.17397>

Iglesias-Martínez, M. J., Lozano-Cabezas, I., Fernández-Verdú, C., & Rovira-Collado, J. (2024). Is there a dialogic teaching model in primary school classrooms? A qualitative study in the practicum/¿Existe un modelo dialógico docente en las aulas de Educación Primaria? Un estudio cualitativo en el practicum. *Culture and Education*, 36(1), 106-133.

Introducing Grok 3: What it is, how to access it, and why it matters. (s. f.-c). KeywordsAI.co. <https://www.keywordsai.co/blog/introducing-grok-3-what-it-is-how-to-access-it-and-why-it-matters>

Jiménez Fernández, C. (2016). Sobrecarga informativa en la universidad. Infoxicación del alumnado universitario en Portugal.

López de Mántaras, R. (2015). Algunas reflexiones sobre el presente y futuro de la Inteligencia Artificial.

Martinez, J. (2024). Sobrecarga de información en Twitter, en el contexto de las elecciones presidenciales 2023. In XVI Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXXI Jornadas de Investigación. XX Encuentro de Investigadores en Psicología del MERCOSUR. VI Encuentro de Investigación de Terapia Ocupacional. VI Encuentro de Musicoterapia. Facultad de Psicología-Universidad de Buenos Aires.

Microsoft. (2025). *Copilot* [Modelo de lenguaje AI]. Microsoft. <https://copilot.microsoft.com/>

Montemayor-Rodríguez, N. y Torregrosa-Carmona, J. (2023). Disinformation as a professional challenge of journalism and documentation. *VISUAL REVIEW. International Visual Culture Review / Revista Internacional De Cultura Visual*, 14(1), 1-11.

Morton Gutiérrez, J. L. (2022). Replika and the Emotional Artificial Intelligence Company: The ethical and social challenges of company chatbots. VISUAL REVIEW. International Visual Culture Review Revista Internacional De Cultura Visual, 10(3), 1–13. <https://doi.org/10.37467/revvisual.v9.3606>

Ojeda, Adelaida D., Solano-Barliza, Andrés D., Alvarez, Danny Ortega, & Cárcamo, Efraín Boom. (2023). Análisis del impacto de la Inteligencia Artificial ChatGPT en los procesos de enseñanza y aprendizaje en la educación universitaria. Formación universitaria, 16(6), 61-70. <https://dx.doi.org/10.4067/S0718-50062023000600061>

OpenAI. (2025). ChatGPT (versión GPT-4o) [Modelo de lenguaje grande]. <https://chat.openai.com/>

Padilla-Meléndez, A., Kock, N., & Del-Águila-Obra, A. R. (2015). Factores que Afectan a la sobrecarga de Información a Nivel Directivo. Revista de Ciencias Sociales (Ve), 21(2), 250-260.

Pañi Rojas, D. A., & Álvarez Gavilanes, J. E. (2025). El impacto de la Inteligencia Artificial en la producción de contenidos noticiosos en la red social TikTok. Pacha. Revista De Estudios Contemporáneos Del Sur Global, 6(18), e250369. <https://doi.org/10.46652/pacha.v6i18.369>

Revelo, R. (2025b, marzo 27). Grok causa furor en X, pero ¿cómo funciona y qué tan confiable es? Primicias. <https://www.primicias.ec/ciencia-tecnologia/grok-red-socialx-modelo-inteligencia-artificial-respuestas-92674/>

Sasson, D. (2024). The Security Risks of Microsoft 365 Copilot. Reco. <https://www.reco.ai/blog/security-risks-of-microsoft-365-copilot>

Sevilla, R. E. F., & Maceiras, S. D. A. (2024). Detección y Clasificación de Fake News mediante Inteligencia Artificial.

Torra, V. (2011). La Inteligencia Artificial. Lychnos, 7, 14-19

Vega, G. A. (2017). El concepto de dispositivo en M. Foucault. Su relación con la “microfísica” y el tratamiento de la multiplicidad. *Nuevo Itinerario*, 12, 136. <https://doi.org/10.30972/nvt.0122038>

Villuendas, B. (2024, 31 enero). Ramón López de Mántaras: «No creo en el tecnosolucionismo». 480. <https://cuatroochenta.com/entrevistas/ramon-lopez-de-mantaras-tecnosolucionismo-inteligencia-artificial/>

xAI. (2025). Grok 3 (Versión 3) [Software]. <https://x.ai/grok>