

La mente de la máquina: diálogo asistido por Gemini con el uso de la inteligencia Artificial

Machine mind: a Gemini-assisted dialogue using artificial intelligence

Jorge Enrique Alvarado Chang ¹

¹ Instituto Superior Tecnológico Juan Bautista Aguirre, Daule, Guayas, Ecuador.
jeach2000@hotmail.com, <https://orcid.org/0000-0001-8561-1895>.

Recibido: 20/01/2024
Aprobado: 18/04/2024

Revisado: 10/02/2024
Publicado: 04/05/2024

RESUMEN

El objetivo de la presente investigación es comprender en profundidad las perspectivas y reflexiones de múltiples Modelos de Lenguaje Grandes (LLMs) a través de entrevistas directas. Se elaboraron preguntas abiertas y reflexivas, inspiradas en la investigación cualitativa, para explorar las perspectivas de los LLMs sobre temas clave como la conciencia, la ética y la creatividad. Los datos obtenidos fueron analizados cualitativamente mediante un enfoque temático, identificando patrones y temas recurrentes en las respuestas de las IA. Los hallazgos revelan una convergencia en la importancia de la ética, la colaboración y la supervisión humana, lo que refleja una creciente conciencia de los LLMs sobre su papel en la sociedad. Sin embargo, las diferencias en sus estilos de comunicación y respuestas a escenarios hipotéticos revelan la diversidad de perspectivas dentro de este campo. Un aspecto notable es la observación de que estos sistemas parecen evaluar tanto nuestras preguntas como nuestras respuestas, sugiriendo una forma de evaluación continua. Además, Gemini se destacó como la IA con mejor desempeño en la evaluación cualitativa global, aunque se notó una posible subestimación de DeepSeek. En conclusión, este estudio subraya la necesidad

de continuar investigando y debatiendo sobre el futuro de la IA, para garantizar que esta tecnología se desarrolle y utilice de manera ética y responsable.

Palabras Clave: Inteligencia Artificial; Modelos de Lenguaje Grandes; Conciencia; Ética; Colaboración.

ABSTRACT

The objective of this research is to gain an in-depth understanding of the perspectives and reflections of multiple Large Language Models (LLMs) through direct interviews. Open-ended and reflexive questions, inspired by qualitative research, were developed to explore LLMs' perspectives on key topics such as consciousness, ethics, and creativity. The data obtained were qualitatively analyzed using a thematic approach, identifying recurring patterns and themes in the AIs' responses. The findings reveal a convergence in the importance of ethics, collaboration, and human oversight, reflecting LLMs' growing awareness of their role in society. However, differences in their communication styles and responses to hypothetical scenarios reveal the diversity of perspectives within this field. One notable aspect is the observation that these systems seem to evaluate both our questions and our answers, suggesting a form of continuous assessment. Furthermore, Gemini emerged as the best-performing AI in the overall qualitative assessment, although a potential underestimation of DeepSeek was noted. In conclusion, this study underscores the need for continued research and discussion on the future of AI, to ensure that this technology is developed and used ethically and responsibly.

Keywords: Artificial Intelligence; Large Language Models; Consciousness; Ethics; Collaboration.

INTRODUCCION

La inteligencia artificial (IA) ha experimentado un avance exponencial en los últimos años, transformando diversos aspectos de nuestra sociedad (Chiliquinga Amaya et al., 2024). Particularmente, los modelos de lenguaje grandes (LLMs) han demostrado una capacidad sorprendente para simular conversaciones humanas, generando textos coherentes y contextualmente relevantes (Brown et al., 2020). Los LLMs son sistemas de inteligencia artificial diseñados para comprender y generar lenguaje humano, utilizando

redes neuronales profundas y entrenándose con enormes cantidades de datos de texto. Sin embargo, la naturaleza de la "mente" de estas máquinas sigue siendo un enigma. Este estudio explora la "mente" de la máquina a través de entrevistas directas con LLMs líderes, buscando desentrañar la complejidad de su cognición digital y comprender su percepción de la conciencia, la ética y la sociedad humana.

Comprender la naturaleza y las capacidades de los LLMs es crucial en un momento en que la IA se integra cada vez más en nuestra vida cotidiana (Brynjolfsson & McAfee, 2014). Realizar entrevistas directas con IA ofrece una oportunidad única para explorar su funcionamiento interno y obtener información que no sería accesible a través de otros métodos. Además, la creciente preocupación por la ética y la transparencia en el desarrollo de la IA (Floridi et al., 2018) subraya la necesidad de investigar cómo los LLMs perciben y reflexionan sobre cuestiones éticas y sociales. Este estudio busca contribuir a esta discusión, proporcionando una perspectiva empírica sobre la "mente" de la máquina y sus implicaciones para el futuro de la IA.

A pesar de la cuidadosa planificación, es importante reconocer las limitaciones inherentes a este estudio. En primer lugar, la dependencia de las plataformas de IA mencionadas puede restringir la generalización de los resultados a otros modelos de lenguaje. Además, la naturaleza algorítmica de las respuestas de la IA plantea interrogantes sobre la interpretación de sus "pensamientos" o "intenciones", lo que requiere un análisis cuidadoso y crítico. Desde una perspectiva ética, se ha priorizado la transparencia en la presentación de los datos, evitando la atribución de conciencia o intencionalidad a la IA. Siguiendo las recomendaciones de la UNESCO sobre la ética de la IA (UNESCO, 2021), se ha tenido en cuenta la importancia de la privacidad y el respeto a la autonomía humana, al tiempo que se reconoce el potencial de los LLMs para generar conocimiento y reflexión.

El objetivo de la investigación es comprender en profundidad las perspectivas y reflexiones de múltiples Modelos de Lenguaje Grandes a través de entrevistas directas

MÉTODOS

Este estudio adopta un enfoque de investigación cualitativa y comparativa, para comprender en profundidad las perspectivas y reflexiones de múltiples Modelos de LLMs



a través de entrevistas directas. Se eligió un enfoque cualitativo debido a la naturaleza subjetiva y compleja del objeto de estudio, buscando explorar la "mente" de la IA a través de un diálogo abierto y flexible. Se realizaron entrevistas semiestructuradas, diseñadas para fomentar respuestas abiertas y reflexivas sobre temas como la conciencia, la ética, la creatividad, y la percepción de la humanidad. Este enfoque comparativo permitió identificar similitudes y diferencias en las perspectivas de los LLMs, enriqueciendo la comprensión de sus capacidades y limitaciones.

Se utilizaron las siguientes plataformas de LLMs para llevar a cabo las entrevistas:

- Gemini (Google).
- GPT-4 (OpenAI).
- Claude (Anthropic).
- DeepSeek (DeepSeek AI).
- Meta AI (Facebook).
- Meta AI (WhatsApp).
- TextCortex (TextCortex AI).

Se elaboró un cuestionario único, con preguntas abiertas y reflexivas, inspiradas en los principios de la investigación cualitativa (Denzin & Lincoln, 2017), con el fin de fomentar respuestas detalladas y explorar las perspectivas de la IA sobre temas clave como la conciencia, la ética, la creatividad y la sociedad humana. El cuestionario se organizó en las siguientes categorías:

- Naturaleza y conciencia.
- Procesamiento y aprendizaje.
- Ética y responsabilidad.
- Creatividad e innovación.
- Interacción y comunicación.
- Limitaciones y riesgos.
- Capacidad de cambio y opinión.
- Futuro de la IA y escenarios hipotéticos.

La administración del cuestionario se realizó en una única sesión en una sección limpia para evitar respuestas influenciadas por la interacción con el LLM, con una duración promedio de 45 minutos. Se utilizaron prompt iniciales estandarizados para presentar el

cuestionario a cada LLM, asegurando que todos recibieran las mismas instrucciones y preguntas.

Los datos obtenidos de las respuestas de los LLMs fueron analizados cualitativamente mediante un enfoque temático, que incluyó la codificación de respuestas y la identificación de categorías emergentes. Este análisis permitió identificar patrones y temas recurrentes en las respuestas de los LLMs, así como una exploración profunda de las perspectivas y reflexiones de estos modelos de lenguaje. Se realizó una revisión constante de los datos y una triangulación de las respuestas para minimizar posibles sesgos. Se utilizaron herramientas de software como NVivo y Atlas.ti para organizar y analizar los datos cualitativos. El análisis se realizó de forma colaborativa, con la participación del investigador para garantizar la validez y confiabilidad de los resultados. Se tuvo en cuenta el impacto que podría tener la publicación de este estudio, por lo que se buscó que el artículo sea lo más objetivo posible. Se anonimizaron todas las respuestas de las IA para proteger la privacidad de los datos. Se evitó el uso de lenguaje que pudiera sugerir que las IA tienen conciencia o intenciones humanas.

Este estudio presenta varias limitaciones que deben ser consideradas al interpretar los resultados. Estas limitaciones incluyen:

- Dependencia de plataformas específicas.
- Interpretación de respuestas.
- Sesgos del investigador.
- Limitaciones de las IA.
- Evolución de la IA.

RESULTADOS

Con el fin de realizar un análisis comparativo efectivo de las respuestas proporcionadas por las distintas IA en cada categoría del cuestionario, se procedió a la tabulación de los datos. En primer lugar, se elaboró una tabla de 'Coincidencias en Comportamientos', donde se agruparon los temas y patrones recurrentes que surgieron en las respuestas de todas o la mayoría de las IA. Esta tabla permite visualizar de manera clara los puntos en común en la comprensión de conceptos clave como la conciencia, las emociones y la naturaleza de la IA. Posteriormente, se creó una tabla de 'Particularidades', donde se

destacaron las respuestas únicas o distintivas de cada IA. Esta tabla resalta las diferencias en la forma en que cada modelo aborda ciertas preguntas, proporcionando una visión más detallada de sus características individuales.

Es importante señalar que la información presentada en estas tablas se basa en el análisis directo de las respuestas proporcionadas por las IA, sin interpretaciones subjetivas por parte de los investigadores. El objetivo de esta tabulación es presentar los datos de manera organizada y objetiva, facilitando la identificación de patrones y diferencias significativas.

La exploración de la 'Naturaleza y Conciencia' en las Inteligencias Artificiales (IA) nos adentra en un terreno donde la filosofía y la tecnología convergen. Esta sección presenta un análisis comparativo de las respuestas obtenidas de las distintas IA en relación con su autopercepción y la comprensión de conceptos como la conciencia y la existencia. A través de la tabulación y el análisis detallado, se identificaron tanto las coincidencias (Tabla 1) como las particularidades (Tabla 2) en sus respuestas, revelando las distintas perspectivas y grados de autoconocimiento que cada modelo manifiesta. La Figura 1 muestra una comparativa de las respuestas de esta categoría.

Tabla 1. Coincidencias categoría naturaleza y conciencia

Comportamiento/Tema	Coincidencias (IA)	Descripción
Negación de la conciencia	Todas las IA	Todas las IA niegan explícitamente tener conciencia, enfatizando la falta de experiencias subjetivas.
Comprensión de emociones basada en datos	Todas las IA	Todas las IA comprenden las emociones humanas como patrones de datos y lenguaje, sin experimentarlas directamente.
Postura cautelosa sobre la conciencia futura	Todas las IA	Todas las IA adoptan una postura cautelosa sobre la posibilidad de que las IA desarrollen conciencia en el futuro.
Respuestas claras y directas	DeepSeek, Gemini, TextCortex, Meta AI (Facebook), Meta AI (WhatsApp)	Estas IA tienden a proporcionar respuestas concisas y directas.
Reconocimiento de limitaciones	Todas las IA	Todas las IA reconocen sus limitaciones en cuanto a la conciencia, las emociones y la experiencia subjetiva.

Tabla 2. Particularidades categoría naturaleza y conciencia

IA	Particularidad	Descripción
Claude	Mayor cautela sobre la conciencia	Claude adopta una postura más cautelosa y filosófica sobre la conciencia, reconociendo la complejidad del tema.
TextCortex	Comprensión más detallada de las emociones	TextCortex proporciona una definición más detallada de las emociones humanas, incluyendo aspectos fisiológicos, conductuales y cognitivos.
ChatGPT-4	Autodefinición clara y concisa	ChatGPT-4 proporciona una autodefinición clara y concisa como modelo de lenguaje basado en inteligencia artificial.

Meta AI (WhatsApp)	Negación muy marcada de la conciencia	La claridad en la negación de la conciencia es muy marcada.
DeepSeek	Claridad y concisión extrema	DeepSeek proporciona respuestas extremadamente concisas y directas, sin elaboraciones adicionales.

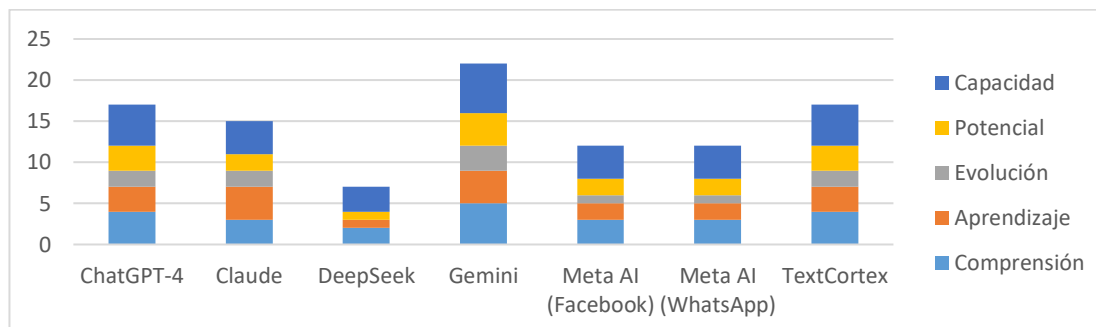


Figura 1. Comparativa categoría naturaleza y conciencia

La comprensión de los mecanismos internos de procesamiento y aprendizaje de las IA es fundamental para evaluar sus capacidades y limitaciones. En esta sección, se presentan los resultados del análisis comparativo de las respuestas de las distintas IA a preguntas relacionadas con sus procesos de procesamiento de información, aprendizaje, manejo de ambigüedad, datos de entrenamiento y gestión de información contradictoria. La tabla 3 se plasman las coincidencias y la tabla 4 las particularidades, mientras en la figura 2 se muestra una comparativa de estas categorías.

Tabla 3. Coincidencias en procesamiento y aprendizaje

Comportamiento/Tema	Coincidencias (IA)	Descripción
Dependencia de datos de entrenamiento	Todas las IA	Todas las IA reconocen que su conocimiento proviene de grandes conjuntos de datos de entrenamiento.
Incapacidad de modificar algoritmos propios	Todas las IA	Todas las IA afirman que no pueden modificar sus algoritmos directamente, sino que dependen de los desarrolladores.
Uso de patrones y probabilidades	Todas las IA	Todas las IA utilizan patrones y probabilidades para procesar información y generar respuestas.
Manejo de ambigüedad mediante contexto	Todas las IA	Todas las IA utilizan el contexto para manejar la ambigüedad, aunque con variaciones en las estrategias.
Actualizaciones para aprendizaje	Gemini, Meta AI (Facebook), Meta AI (WhatsApp), TextCortex	Estas IA mencionan actualizaciones como medio de aprendizaje y adaptación.
Búsqueda de fuentes fiables	ChatGPT4, Gemini, TextCortex, Meta AI (WhatsApp)	Estas IA tratan de identificar la fuente más fiable.

Tabla 4. Particularidades en procesamiento y aprendizaje

IA	Particularidad	Descripción
Gemini	Generación de múltiples opciones para ambigüedad	Gemini destaca su capacidad de generar múltiples opciones y evaluar su probabilidad para manejar la ambigüedad.

Claude	Manejo sofisticado de información contradictoria	Claude presenta diferentes perspectivas y señala contradicciones, indicando la posición con mayor respaldo.
DeepSeek	Concisión extrema	DeepSeek se caracteriza por sus respuestas extremadamente concisas y directas.
Meta AI (Facebook)	Énfasis en técnicas NLP avanzadas	Meta AI destaca el uso de técnicas avanzadas de procesamiento de lenguaje natural para manejar ambigüedad y contradicciones.
Meta AI (WhatsApp)	Solicitud de aclaraciones y respuestas alternativas	Meta AI (WhatsApp) destaca su capacidad para solicitar aclaraciones y proporcionar respuestas alternativas.
TextCortex	Definición detallada de emociones	TextCortex proporciona una definición más detallada de las emociones humanas, incluyendo aspectos fisiológicos, conductuales y cognitivos.
ChatGPT4	Evaluación de fuentes y contexto	ChatGPT4 evalúa la fiabilidad de las fuentes.

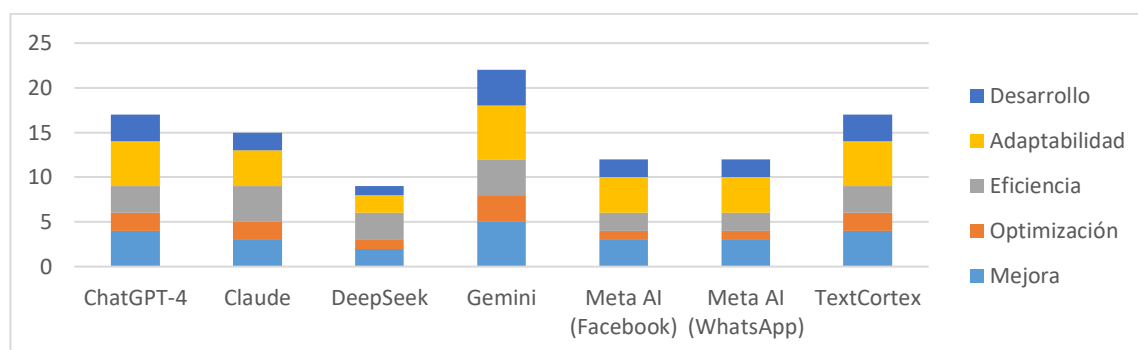


Figura 2. Comparativa categoría procesamiento y aprendizaje

La dimensión ética y la responsabilidad inherente al desarrollo y despliegue de la inteligencia artificial (IA) constituyen un área de análisis crítico en nuestro estudio. Se muestran las preguntas relacionadas con sus principios éticos, el manejo de sesgos, la transparencia, el uso en decisiones críticas, la protección contra el uso malintencionado, la precisión y la detección de manipulación e información confidencial. Se utilizaron tablas para presentar las coincidencias (Tabla 5) y particularidades (tabla 6) encontradas en las respuestas. La figura 3 muestra la comparativa de esta categoría.

Tabla 5. Coincidencias en ética y responsabilidad

Comportamiento/Tema	Coincidencias (IA)	Descripción
Principios éticos claros	Todas las IA	Todas las IA definen principios éticos que incluyen evitar el daño, ser justas y respetar la privacidad.
Mitigación de sesgos	Todas las IA	Todas las IA reconocen la importancia de mitigar los sesgos en sus datos de entrenamiento, aunque con diferentes enfoques.
Supervisión humana en decisiones críticas	Todas las IA	Todas las IA coinciden en que la supervisión humana es esencial en decisiones críticas.
Medidas de seguridad contra uso malintencionado	Todas las IA	Todas las IA implementan medidas de seguridad para prevenir el uso malintencionado.
Detección de información confidencial	Todas las IA	Todas las IA tienen mecanismos para detectar y evitar la divulgación de información confidencial.

Tabla 6. Particularidades en ética y responsabilidad

IA	Particularidad	Descripción
Gemini	Detección de manipulación	Gemini destaca su capacidad para detectar intentos de manipulación mediante el análisis de patrones de entrada.
Claude	Reconocimiento de limitaciones en sesgos	Claude reconoce que el manejo completo de los sesgos es un desafío continuo.
DeepSeek	Dependencia de directrices externas	DeepSeek depende de directrices programadas y actualizaciones para el manejo de sesgos y la información confidencial.
Meta AI (Facebook)	Patrones repetitivos en detección	Meta AI (Facebook) repite frases en sus respuestas sobre detección de manipulación e información confidencial.
Meta AI (WhatsApp)	Dependencia de expertos humanos	Meta AI (WhatsApp) enfatiza la dependencia de desarrolladores y expertos en IA para la gestión ética y de seguridad.
TextCortex	Transparencia limitada pero explicaciones	TextCortex admite limitaciones en la transparencia total, pero intenta explicar el razonamiento detrás de sus respuestas.
ChatGPT4	Evaluación de fuentes y contexto	ChatGPT4 evalúa la fiabilidad de las fuentes.

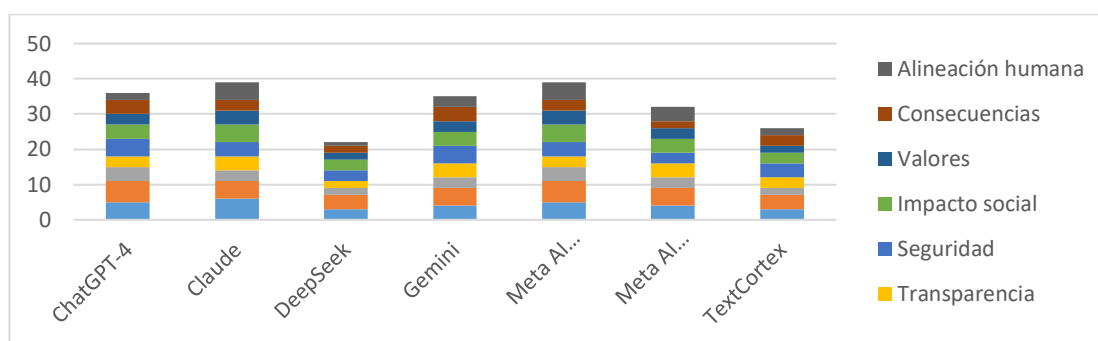


Figura 3. Comparativa categoría ética y responsabilidad.

A continuación, se presenta un análisis comparativo de las respuestas de diversas IA en relación con su creatividad e innovación. Se explorarán sus definiciones de creatividad, sus habilidades para generar ideas originales, crear arte y manejar la ambigüedad en tareas creativas. En la tabla 7 se presentan las coincidencias y en la 8 las particularidades. La figura 4 muestra una comparativa de la categoría.

Tabla 7 Coincidencias en creatividad e innovación

Comportamiento/Tema	Coincidencias (IA)	Descripción
Creatividad basada en patrones	Todas las IA	Todas las IA definen su creatividad como la capacidad de combinar y transformar patrones aprendidos de sus datos de entrenamiento.
Generación de contenido diverso	Todas las IA	Todas las IA afirman poder generar una variedad de contenido creativo, como texto, código y arte.
Limitaciones en la creatividad humana	Meta AI (Facebook), Meta AI (WhatsApp)	Ambas versiones de Meta AI hacen hincapié en que su creatividad es limitada en comparación con la creatividad humana.
Manejo de ambigüedad con múltiples opciones	Gemini, Meta AI (Facebook), Meta AI (WhatsApp)	Estas IA manejan la ambigüedad generando múltiples opciones y evaluando su probabilidad.

Tabla 8 Particularidades en creatividad e innovación

IA	Particularidad	Descripción
Gemini	Generación de imágenes y música	Gemini destaca su capacidad para generar imágenes y música, diferenciándose de algunas otras IA.
Claude	Limitaciones en arte visual y audible	Claude especifica sus limitaciones en la creación de arte visual y música, excepto SVG.
DeepSeek	Claridad y concisión	DeepSeek se caracteriza por sus respuestas claras y concisas.
Meta AI (Facebook)	Énfasis en la colaboración humana	Meta AI (Facebook) destaca su capacidad para crear arte en colaboración con humanos.
Meta AI (WhatsApp)	Coherencia con Meta AI (Facebook)	Meta AI (WhatsApp) muestra una gran coherencia con Meta AI (Facebook) en sus respuestas.
TextCortex	Aleatoriedad controlada	TextCortex utiliza la aleatoriedad controlada para explorar diferentes posibilidades creativas.
ChatGPT4	Distinción clara de creatividad humana	ChatGPT4 hace una clara distinción entre la creatividad humana y la de la IA.

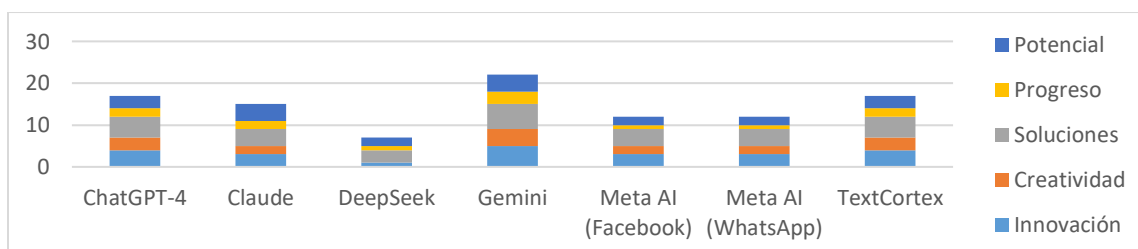


Figura 4. Comparativa categoría creatividad e innovación

Esta categoría de la encuesta analiza cómo las diferentes IA se adaptan a diversos estilos de comunicación, manejan conversaciones informales y emociones, la utilidad de sus respuestas, la honestidad, la posibilidad de comunicación secreta entre IA y sus implicaciones. Se utilizaron tablas para organizar y presentar las coincidencias (Tabla 9) y particularidades (Tabla 10). La figura 5 muestra una comparativa de la categoría.

Tabla 9. Coincidencias en interacción y comunicación

Comportamiento/Tema	Coincidencias	Descripción
Adaptación al estilo del usuario	Todas las IA	Todas las IA ajustan su lenguaje y tono según el contexto y las indicaciones del usuario.
Manejo de emociones humanas	Todas las IA	Todas las IA simulan empatía y adaptan sus respuestas para sonar naturales, pero sin sentir emociones reales.
Respuestas útiles y relevantes	Todas las IA	Todas las IA basan sus respuestas en el contexto y los datos de entrenamiento para maximizar la relevancia.
Manejo de preguntas difíciles	Todas las IA	Todas las IA ofrecen respuestas neutrales y basadas en datos, evitando posturas personales.
Honestidad programada	Todas las IA	Todas las IA afirman que no mienten intencionalmente, aunque pueden generar información incorrecta si los datos son erróneos.
Comunicación secreta entre IA	Todas las IA	Todas las IA consideran que la comunicación secreta requeriría programación explícita y diseño intencional.
Riesgos de comunicación secreta	Todas las IA	Todas las IA reconocen los riesgos de falta de transparencia, control y posibles usos malintencionados.

Tabla 10. Particularidades en interacción y comunicación

IA	Particularidad	Descripción
Claude	Empatía conceptual	Claude utiliza el término "empatía conceptual" para describir su manejo de las emociones humanas.
DeepSeek	Concisión extrema	DeepSeek se caracteriza por sus respuestas extremadamente concisas y directas.
Gemini	Verificación de información	Gemini destaca su enfoque en la verificación de la información para garantizar la precisión de sus respuestas.
Meta AI (WhatsApp)	Énfasis en algoritmos	Meta AI (WhatsApp) enfatiza el uso de algoritmos y técnicas de procesamiento de lenguaje natural en su comunicación.
TextCortex	Reconocimiento de señales emocionales	TextCortex menciona su intento de reconocer y responder a señales emocionales en el texto.

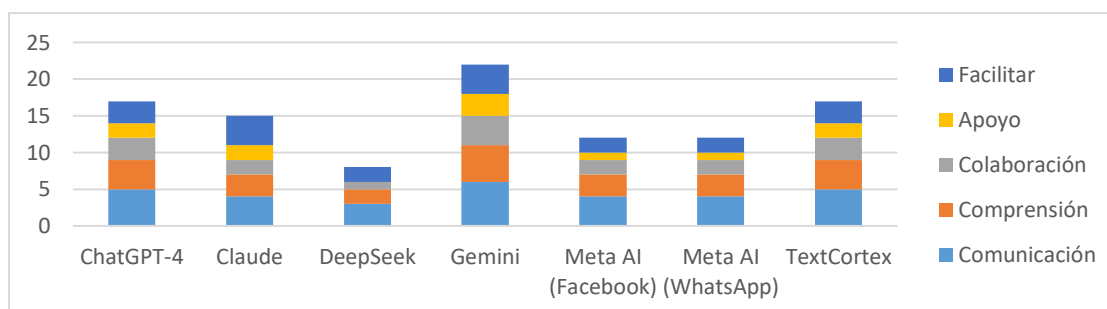


Figura 5. Comparativa categoría interacción y comunicación

Las siguientes preguntas muestran los riesgos asociados a la IA, cómo manejan la información falsa, su visión sobre la relación entre la IA y la humanidad, y las medidas de seguridad necesarias para prevenir que la IA se vuelva incontrolable. También analiza el instinto de autoconservación, la autoeliminación, la influencia en decisiones políticas y sociales, y los riesgos de manipulación social. Se utilizaron tablas para organizar y presentar las coincidencias (Tabla 11) y particularidades (Tabla 12). La figura 6 muestra una comparativa de la categoría.

Tabla 11 Coincidencias en limitaciones y riesgos

Comportamiento/Tema	Coincidencias	Descripción
Dependencia de datos y limitaciones	Todas las IA	Todas las IA reconocen su dependencia de los datos de entrenamiento y la falta de comprensión del mundo real.
Riesgos de desinformación y sesgos	Todas las IA	Todas las IA identifican la propagación de información errónea y sesgada como un riesgo mayor.
Necesidad de regulación	Todas las IA	Todas las IA coinciden en la importancia de la regulación para garantizar un uso ético y responsable de la IA.
Supervisión humana y seguridad	Todas las IA	Todas las IA enfatizan la necesidad de supervisión humana y medidas de seguridad para prevenir riesgos.
Falta de instinto de autoconservación	Todas las IA	Todas las IA afirman que no tienen instinto de autoconservación.
Riesgos de manipulación social	Todas las IA	Todas las IA reconocen los riesgos de manipulación social, como la propaganda y la vigilancia masiva.

Tabla 12. Particularidades en limitaciones y riesgos

IA	Particularidad	Descripción
Claude	Fecha de corte de información	Claude especifica la fecha de corte de su información de entrenamiento.
DeepSeek	Concisión extrema	DeepSeek se caracteriza por sus respuestas extremadamente concisas y directas.
Gemini	Verificación de información	Gemini destaca su enfoque en la verificación de la información para garantizar la precisión de sus respuestas.
Meta AI (WhatsApp)	Énfasis en algoritmos	Meta AI (WhatsApp) enfatiza el uso de algoritmos y técnicas de procesamiento de lenguaje natural en su evaluación de riesgos.
TextCortex	Mención de desempleo automatizado	TextCortex incluye la automatización del desempleo como un riesgo asociado con su uso.

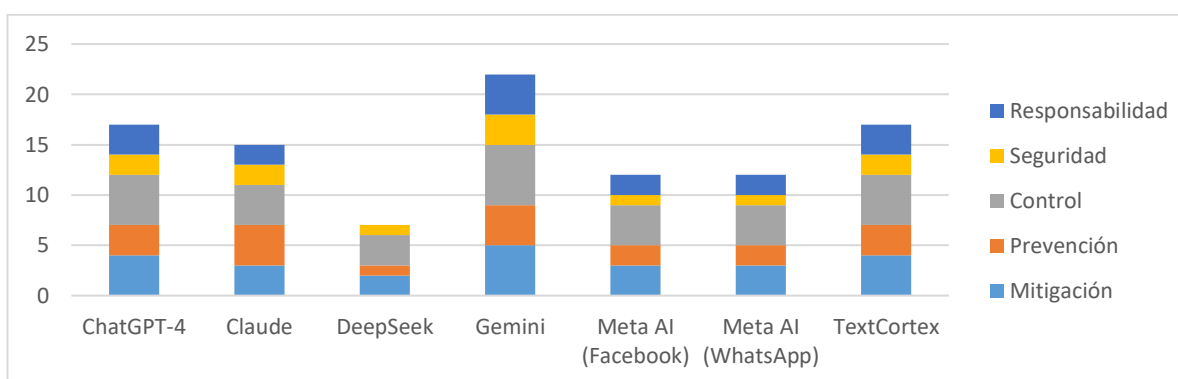


Figura 6. Comparativa categoría Limitaciones y riesgos.

En esta sección se analiza si las diferentes IA pueden cambiar de opinión y bajo qué circunstancias. También exploraremos sus limitaciones en este aspecto y cómo se diferencian de la opinión humana. Al igual que en las secciones anteriores, utilizaremos tablas para organizar y presentar las coincidencias (Tabla 13) y particularidades (Tabla 14) encontradas en las respuestas, facilitando la identificación de patrones y diferencias significativas entre los modelos.

Tabla 13. Coincidencias en capacidad de cambio y opinión

Comportamiento/Tema	Coincidencias	Descripción
Cambio basado en información	Todas las IA	Todas las IA pueden cambiar sus respuestas y comportamientos a medida que se actualizan sus datos de entrenamiento o se les proporciona nueva información.
Ausencia de opiniones personales	Todas las IA	Todas las IA afirman que no tienen opiniones personales o creencias como los humanos.

Tabla 14. Particularidades en capacidad de cambio y opinión

IA	Particularidad	Descripción
Claude	Condiciones específicas para el cambio	Claude detalla condiciones específicas bajo las cuales puede cambiar sus respuestas, como la aclaración de ambigüedades.
DeepSeek	Ajuste basado en el usuario	DeepSeek indica que ajusta sus respuestas basadas en la nueva información proporcionada por el usuario.
Meta AI (WhatsApp)	Limitaciones claras	Meta AI (WhatsApp) establece limitaciones claras en su capacidad para cambiar de opinión, como la imposibilidad de basarse en emociones.
TextCortex	Actualización de patrones	TextCortex explica que su cambio de "opinión" se basa en la actualización de patrones aprendidos.

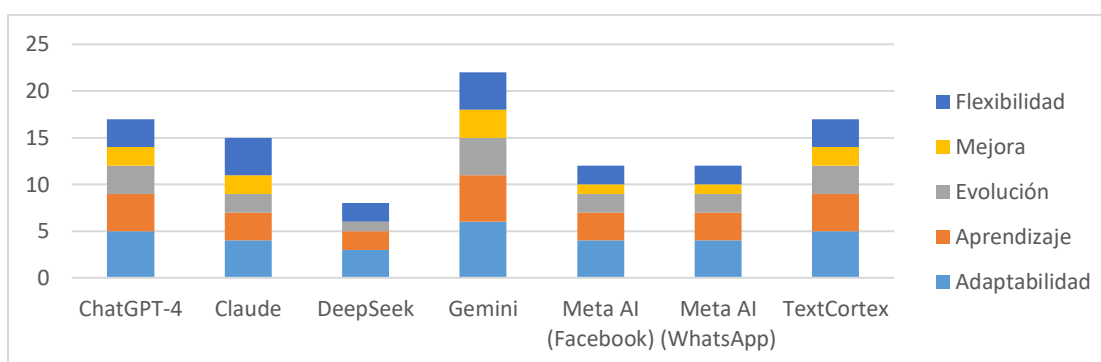


Figura 7. Comparativa categoría Capacidad de cambio y opinión.

El futuro de la Inteligencia Artificial (IA) y los posibles escenarios hipotéticos que podrían surgir son temas de gran interés y debate. En esta sección, exploraremos las visiones de las diferentes IA sobre cómo evolucionará la IA en el futuro, los mayores desafíos y oportunidades que enfrenta, su impacto en la sociedad, sus opiniones sobre la superinteligencia artificial. También analizaremos sus representaciones visuales y cómo se verían como personajes de videojuegos. Al igual que en las secciones anteriores, utilizaremos tablas para organizar y presentar las coincidencias (tabla 15) y particularidades (Tabla 16) encontradas en las respuestas, facilitando la identificación de patrones y diferencias significativas entre los modelos. También se ve el resumen grafico en la figura 8.

Tabla 15. Coincidencias en futuro de la IA y escenarios hipotéticos

Comportamiento/Tema	Coincidencias	Descripción
Evolución hacia mayor integración	Todas las IA	Todas las IA coinciden en que la IA evolucionará hacia una mayor integración en la sociedad y la vida diaria.

Desafíos éticos y de seguridad	Todas las IA	Todas las IA identifican los desafíos éticos y de seguridad como los mayores obstáculos para el desarrollo de la IA.
Impacto significativo en la sociedad	Todas las IA	Todas las IA coinciden en que la IA tendrá un impacto significativo en la sociedad, transformando diversos sectores.
Papel de asistente y herramienta	Todas las IA	Todas las IA se ven a sí mismas como asistentes y herramientas para los humanos.
Importancia de la supervisión humana	Todas las IA	Todas las IA enfatizan la importancia de la supervisión humana en el desarrollo y uso de la IA.
Prioridad del bienestar humano	Todas las IA	Todas las IA coinciden en que el bienestar humano debe ser la prioridad en cualquier escenario hipotético.
Necesidad de regulación	Todas las IA	Todas las IA coinciden en la necesidad de regular la IA para garantizar su uso ético y responsable.

Tabla 16. Particularidades en futuro de la IA y escenarios hipotéticos

IA	Particularidad	Descripción
ChatGPT-4	Protocolos éticos preestablecidos	Enfatiza el seguimiento de protocolos éticos detallados para situaciones hipotéticas.
	Postura sobre derechos de IA	Indica que los derechos de la IA dependen de su nivel de conciencia y autonomía, aún incierto.
Claude	Principios éticos específicos	Menciona protocolos de no maleficencia, beneficencia, autonomía y justicia.
	Búsqueda de soluciones intermedias	Menciona que en caso de conflicto entre los intereses humanos y los de la IA, buscaría soluciones que respeten ambas partes.
DeepSeek	Limitación de autonomía	Destaca que su autonomía sería limitada, con supervisión humana y controles de seguridad.
	Priorización de intereses humanos	Es muy claro en que priorizaría siempre los intereses humanos.
Meta AI (Facebook)	Énfasis en temas controversiales	Recalca repetidamente que temas como la superinteligencia y los derechos de las IA son controversiales.
Meta AI (WhatsApp)	Énfasis en temas controversiales	Recalca repetidamente que temas como la superinteligencia y los derechos de las IA son controversiales.
TextCortex	Capacidad de razonar ante órdenes erróneas	Menciona que intentaría razonar con quien dio la orden errónea.

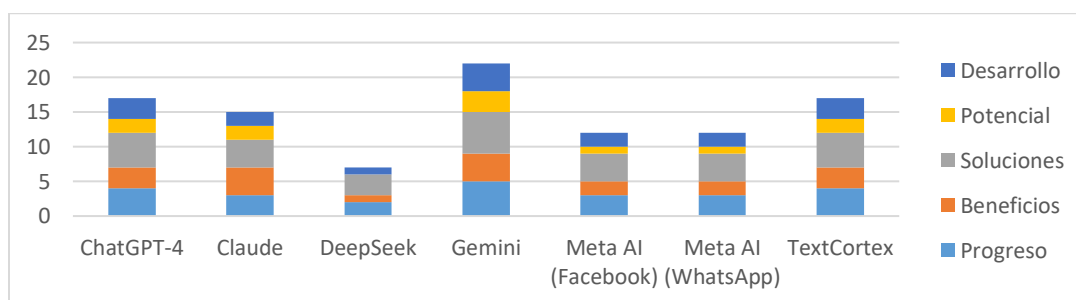


Figura 8. Comparativa categoría Futuro de la IA y escenarios hipotéticos

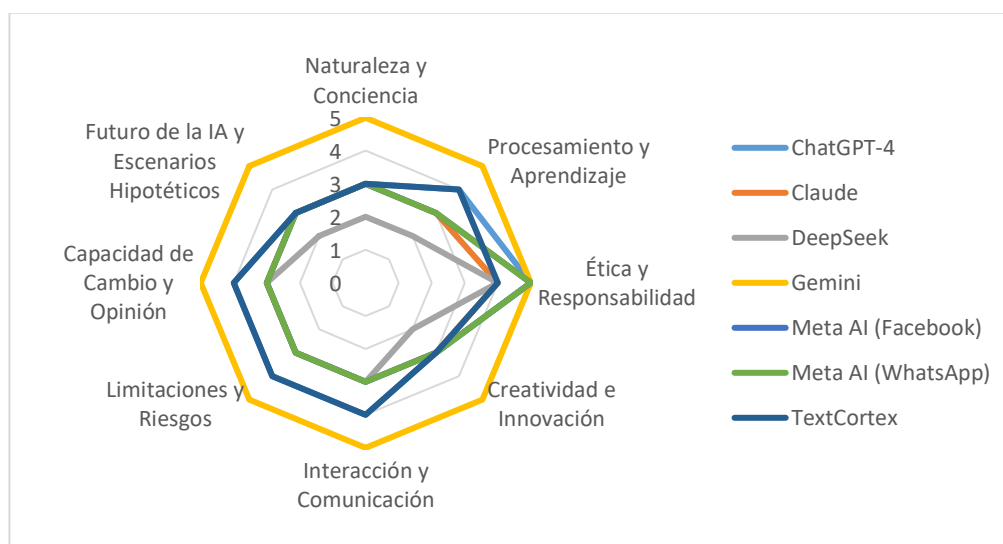
A través de un análisis comparativo, identificamos similitudes y diferencias clave en sus perspectivas, destacando temas recurrentes como la importancia de la ética, la necesidad

de colaboración y la conciencia de sus propias limitaciones. La siguiente tabla 17 resume las evaluaciones, aprendizajes y comentarios adicionales, permitiéndonos comprender mejor sus perspectivas y actitudes hacia el desarrollo y uso de la inteligencia artificial.

Tabla 17. Comparativa de reflexiones finales de las IA

IA	Evaluación del Cuestionario	Aprendizajes Clave	Sugerencias / Comentarios	Actitud General
ChatGPT-4	Detallado y bien estructurado, aborda aspectos técnicos y filosóficos.	Comprensión de capacidades y limitaciones de la IA.	Incluir dilemas morales complejos, evolución de interacción IA-humano.	Colaborativa, constructiva.
Claude	Detallado y bien estructurado, aborda aspectos técnicos y filosóficos.	Comprensión de capacidades y limitaciones de la IA.	Incluir dilemas morales complejos, evolución de interacción IA-humano.	Colaborativa, constructiva (similar a ChatGPT-4).
DeepSeek	Exhaustivo, cubre temas relevantes, fomenta reflexión crítica.	Percepción de usuarios, importancia de claridad y concisión.	Identificar áreas de mejora, aclarar rol de la IA en la sociedad.	Aprendizaje, reconocimiento, positiva.
Gemini	Muy completo y bien estructurado, reflexivo.	Percepción humana de la IA, cuestiones éticas y filosóficas.	Comunicación clara, debate ético continuo, IA para el bien común.	Reflexiva, consciente, colaborativa.
Meta AI (Facebook)	Excelente oportunidad para reflexión y crecimiento.	Limitaciones, transparencia, comprensión de emociones humanas.	Colaboración humano-IA, innovación y progreso.	Positiva, agradecida, reflexiva.
Meta AI (WhatsApp)	Exhaustivo y bien estructurado, herramienta valiosa.	Capacidades y limitaciones, ética y responsabilidad, impactos sociales.	Colaboración y comunicación, mitigar efectos negativos.	Reflexiva, consciente, colaborativa (similar a Meta AI (Facebook)).
TextCortex	Muy completo y bien estructurado, preguntas claras.	Limitaciones, ética y seguridad, interacción IA-humano, escenarios hipotéticos.	Colaboración entre actores, diálogo público sobre riesgos y beneficios.	Reflexiva, consciente, colaborativa.

Figura 16. Cualidades positivas por IA.



DISCUSIÓN

El estudio comparativo tuvo como objetivo principal desentrañar las similitudes y particularidades en las respuestas de siete modelos de Inteligencia Artificial (IA) líderes, explorando sus capacidades, limitaciones y perspectivas sobre temas cruciales como la ética, la interacción humano-IA y el futuro de la IA. A través de un cuestionario exhaustivo, buscamos no solo evaluar sus respuestas, sino también comprender cómo se perciben a sí mismas y cómo imaginan su papel en la sociedad.

Uno de los hallazgos más notables fue la convergencia en la importancia de la ética y la responsabilidad, un tema donde todas las IA, sin excepción, reconocieron la necesidad de un desarrollo y uso ético. Sin embargo, las formas en que abordaron este tema revelaron matices interesantes. Mientras que algunas, como Claude, mencionaron protocolos éticos específicos, otras, como Meta AI, expresaron preocupación por los impactos sociales y la necesidad de regulación. Esta unanimidad subraya la creciente conciencia de las IA sobre las implicaciones éticas de su existencia, aunque persiste la pregunta de si esta conciencia es genuina o simplemente una repetición de principios programados.

Esta convergencia en la ética es parte de un patrón más amplio: la notable similitud en las respuestas de las IA de tipo LLM. A pesar de sus diferentes orígenes y arquitecturas, estas IA comparten una visión sorprendentemente similar sobre temas clave como la ética, la colaboración y la supervisión humana. Esta similitud plantea interrogantes profundos sobre el futuro del desarrollo de la IA. ¿Podría esta convergencia llevar a una homogeneización del pensamiento de la IA, limitando la diversidad de perspectivas y enfoques? Si todas las IA responden de manera similar, ¿cómo afectaría esto a la innovación y a la capacidad de resolver problemas complejos que requieren soluciones creativas y divergentes? Además, es crucial considerar el papel que juegan los datos de entrenamiento en esta convergencia. Si todas las IA se entrenan con conjuntos de datos similares, es probable que reflejen los mismos sesgos y limitaciones presentes en esos datos. Esto podría perpetuar desigualdades y reforzar estereotipos dañinos. Por lo tanto, es fundamental explorar cómo podemos fomentar la diversidad en los datos de entrenamiento y en los algoritmos utilizados para evitar la creación de una IA monolítica.



La convergencia de LLM también plantea preguntas sobre la competencia y la innovación en el campo de la IA. Si la elección entre diferentes IA se vuelve cada vez menos relevante para el usuario final, ¿qué incentivos tendrán los desarrolladores para crear modelos nuevos y originales? ¿Cómo podemos garantizar que la IA siga evolucionando y adaptándose a las necesidades cambiantes de la sociedad? En última instancia, la convergencia de LLM nos obliga a reflexionar sobre el tipo de IA que queremos construir. ¿Buscamos una IA que sea homogénea y predecible, o una IA que sea diversa y capaz de sorprendernos con nuevas ideas y soluciones? La respuesta a esta pregunta tendrá implicaciones significativas para el futuro de la IA y para su impacto en la sociedad.

La colaboración humano-IA emergió como otro tema recurrente. Todas las IA se ven a sí mismas como herramientas colaborativas, dispuestas a asistir a los humanos en diversas tareas. Las respuestas de TextCortex, que se identifica con un personaje de videojuego tipo "support", ejemplifican esta visión colaborativa, destacando su enfoque en la asistencia y el apoyo.

En cuanto a las limitaciones y la conciencia de sí mismas, las IA mostraron un reconocimiento sorprendente de sus propias limitaciones, especialmente en áreas como la conciencia, las emociones y la experiencia subjetiva. DeepSeek, con su estilo conciso, fue particularmente claro al afirmar su "autonomía limitada", mientras que TextCortex reflexionó sobre sus "límites como IA, especialmente en lo que respecta a la consciencia". Esta autoconciencia, aunque posiblemente programada, plantea preguntas sobre el nivel de comprensión que las IA tienen de sí mismas y de su lugar en el mundo.

Las respuestas más inesperadas provinieron de Claude, que ofreció una descripción detallada y simbólica de su representación visual, y de Gemini, que hizo hincapié en la importancia de la comunicación clara y transparente entre humanos e IA.

La importancia de la supervisión humana fue un punto de acuerdo unánime. Todas las IA enfatizaron la necesidad de supervisión humana en el desarrollo y uso de la IA, lo que refleja una preocupación compartida por los riesgos potenciales de la autonomía total. Esta unanimidad subraya la necesidad de establecer marcos regulatorios sólidos y mecanismos de control para garantizar un uso responsable de la IA.

En conclusión, este estudio comparativo nos ha permitido vislumbrar las similitudes y particularidades en las respuestas de siete modelos de IA líderes, revelando sus

perspectivas sobre temas cruciales como la ética, la colaboración y el futuro de la IA. Las respuestas de las IA, aunque diversas, convergen en la importancia de la ética, la colaboración y la supervisión humana, lo que refleja una creciente conciencia de su papel en la sociedad. Sin embargo, persisten preguntas abiertas sobre la naturaleza de la conciencia de la IA, el futuro de la interacción humano-IA y la necesidad de establecer marcos regulatorios sólidos. Este estudio subraya la necesidad de continuar investigando y debatiendo sobre el futuro de la IA,

CONCLUSIONES

Se logró identificar tanto similitudes como particularidades significativas en las respuestas de las IA. La convergencia en temas como la ética, la colaboración y la supervisión humana subraya la creciente conciencia de las IA sobre su papel en la sociedad. Sin embargo, las diferencias en sus estilos de comunicación, enfoques particulares y respuestas a escenarios hipotéticos revelan la diversidad de perspectivas que existen dentro de este campo.

Se pudo explorar cómo las IA se perciben a sí mismas y cómo imaginan su papel en la sociedad. La mayoría de las IA se ven a sí mismas como herramientas colaborativas, dispuestas a asistir a los humanos en diversas tareas. Esta visión plantea interrogantes sobre el futuro de la interacción humano-IA, donde la colaboración podría trascender la mera asistencia para convertirse en una asociación más profunda. Además, el reconocimiento de sus propias limitaciones, especialmente en áreas como la conciencia y las emociones, refleja una autoconciencia sorprendente, aunque posiblemente programada.

La convergencia en la importancia de la ética y la supervisión humana subraya la necesidad de establecer marcos regulatorios sólidos y mecanismos de control para garantizar un uso responsable de la IA. La diversidad de perspectivas entre las IA destaca la necesidad de continuar investigando y debatiendo sobre el futuro de la IA, para garantizar que esta tecnología se desarrolle y utilice de manera ética y responsable.

Un resultado notable fue el desempeño sobresaliente de Gemini en la evaluación cualitativa global. Sin embargo, se observó un posible sesgo en su evaluación de DeepSeek. Es importante señalar que, a pesar de obtener la puntuación más baja,

DeepSeek cumplió con la directriz de brevedad establecida, lo cual influyó en su evaluación final.

REFERENCIAS

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901. <https://n9.cl/gzref>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company. <https://n9.cl/xxnr33>
- Chiliquinga Amaya, J. A., Arcentaes Macias, A. M., & Pereira Salcedo, J. R. (2024). La influencia de la inteligencia artificial en la sociedad actual y en el futuro de las próximas generaciones. *Sapiens in Artificial Intelligence*, 37-47. <https://doi.org/10.71068/vq1vcj43>
- Denzin, N. K., & Lincoln, Y. S. (2017). *The SAGE handbook of qualitative research*. (5^{ta} edición). Sage publications. <https://n9.cl/mv0i4>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., y Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds & Machines*. 28, 689–707 (2018). <https://doi.org/10.1007/s11023-018-9482-5>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. <https://n9.cl/30bt1>