www.revistalogos.cl
Issn Electrónico: 0719-3262

Artículo de Investigación

Comportamiento argumentativo del ChatGPT 3.5: similitudes y diferencias con la práctica argumentativa humana

ChatGPT 3.5 argumentative behavior: similarities and differences with human argumentative practice

Recibido: 10-03-2023 **Aceptado:** 11-12-2023 **Publicado:** 30-06-2024

Cristián Noemi Padilla
Universidad de La Serena
cnoemi@userena.cl
0000-0001-9889-066X

Cristián Santibáñez
Universidad Católica de la Santísima de Concepción
csantibanez@ucsc.cl
0000-0001-6755-3468

Resumen: El desarrollo de la inteligencia artificial (IA) ha abierto una nueva discusión sobre la capacidad lingüística de esta tecnología y su potencial impacto en todas las dimensiones de la actividad humana (Brynjolfsson y McAfee, 2014). A fin de evaluar similitudes o diferencias entre la capacidad lingüística de la IA y la humana, este trabajo analiza específicamente el comportamiento argumentativo en términos de los puntos de vista que adopta Chat GPT, versión 3.5, frente a una situación controversial expresada en un dilema moral del Cuestionario D.I.T. (Rest, 2013) y se compara con una muestra similar de adultos humanos (Noemi, 2019), sometida a la misma tarea discursiva. La investigación siguió un estudio comparativo de casos múltiples (Flick, 2020) a través de una metodología mixta (Johnson y Onwuegbuzie, 2004), que combinó un enfoque cuantitativo con uno cualitativo. Los resultados muestran que el 100% de la muestra de IH emite un punto de vista (a favor o en contra), mientras que de la muestra producida por el Chat GPT, sólo el 23,4% lo hace. Los resultados animan a pensar que las diferencias encontradas se podrían explicar por la debilidad de la IA de ser sensible al contexto (falta de una competencia pragmática relevante) y la inexistencia de un proceso de autoimplicación para enfrentar tareas de razonamiento moral.

Palabras clave: Argumentación- Inteligencia artificial- Inteligencia humana- Chat GPT- Punto de vista.

Citación: Noemi, C. & Santibáñez, C. (2024). Comportamiento argumentativo del ChatGPT 3.5: similitudes y diferencias con la práctica argumentativa humana. *Logos: Revista de Lingüística, Filosofía y Literatura,* 34(1), 26-44. doi.org/10.15443/RL3402

Abstract: The development of artificial intelligence (AI) has opened a new debate about the linguistic capabilities of this technology and its potential impact on all dimensions of human activity (Brynjolfsson and McAfee, 2014). Aspects such as the massive collection of data to train AI models have raised concerns about the privacy and security of information, due to the risk of misuse. In order to assess similarities or differences between AI and human linguistic abilities, this work specifically analyses argumentative behaviour in terms of the viewpoints adopted by Chat GPT, version 3.5, in the face of a controversial situation expressed in a moral dilemma of the D.I.T.' questionnaire (Rest, 2013) and is compared with a similar sample of human adults (Noemi, 2019) subjected to the same discursive task. The research followed a comparative multiple case study (Flick, 2020) through a mixed methodology (Johnson and Onwuegbuzie, 2004), combining a quantitative approach with a qualitative one. The results show that 100% of the IH sample take a position (for or against), whereas only 23.4% of the GPT chat sample do. The results encourage us to think that the differences found could be explained by the AI's weakness in context sensitivity (lack of relevant pragmatic competence) and the lack of a self-involvement process to face moral reasoning tasks.

Key words: Argumentation- Artificial Intelligence- Human Intelligence- Chat GPT- Point of view.

1. Introducción.

El desarrollo de la inteligencia artificial (IA) ha abierto una nueva discusión sobre la cercanía de su capacidad lingüística con el comportamiento lingüístico humano, esto es, si acaso puede producir discurso tal como nosotros lo hacemos. En un período muy breve, la IA ha causado una notable disrupción en numerosos aspectos de la vida cotidiana, desde automatizar tareas y procesos hasta alterar el panorama laboral en varias industrias (Brynjolfsson y McAfee, 2014). Además, la acumulación extensiva de datos para el entrenamiento de modelos de IA ha generado inquietudes significativas respecto a la privacidad y seguridad de la información, debido al riesgo de su eventual mal uso.

Los modelos de lenguaje de gran tamaño (LLMs, por sus siglas en inglés), tales como Chat GPT, se entrenan con grandes conjuntos de datos que incluyen textos argumentativos,

y utilizan técnicas como el aprendizaje supervisado y el aprendizaje por refuerzo para producir una distribución de probabilidad sobre secuencias de palabras, logrando un rendimiento asombroso en la generación de lenguaje escrito similar al humano (Brown et al., 2020).

El trabajo ha analizado y comparado el desempeño discursivo argumentativo de Chat GPT 3.5 en una tarea de argumentación asociada a un dilema moral para, de esta forma, establecer qué similitudes o diferencias tiene con los actos argumentativos generados por la IH sometida a las mismas tareas discursivas (Noemi, 2019). Si bien recientemente han aparecido algunos trabajos sobre el rendimiento de los LLMs frente a baterías neurocognitivas y lingüísticas estándar (Brunet-Gouet et al., 2023; Kosinski, 2023; Loconte et al., 2023), la evaluación de las habilidades pragmáticas ha sido limitada (Mahowald et al., 2023; Sap et al., 2022), y la evaluación de habilidades argumentativas casi inexistente. En consecuencia, la discusión respecto de si las habilidades de orden superior y las inferencias son sensibles al contexto o al texto está abierta, lo que representa un apasionante desafío para futura investigación.

Con este propósito, en la sección Marco Teórico se refieren y discuten las nociones de: capacidades lingüísticas superiores; modelos de lenguaje de gran tamaño (LLMs); comportamiento argumentativo en términos de generación de puntos de vista; y juicio moral. En la sección Metodología, se da cuenta del tipo de diseño del trabajo, se caracteriza la muestra, y se explica el procedimiento seguido en la aplicación del instrumento Dilema III del Cuestionario D.I.T. (Rest, 2013). En la sección Hallazgos y discusión se presentan y comentan las principales tendencias que se observan luego del análisis del corpus. En la Conclusión se discuten los principales alcances y limitaciones del trabajo.

2. Marco teórico

2.1. Capacidades lingüísticas superiores

La comprensión de las capacidades lingüísticas superiores ha sido un tema central en el estudio de la cognición humana. Tradicionalmente, el lenguaje se ha considerado una facultad innata y exclusiva de los seres humanos, esencial para entender la estructura y el funcionamiento de la mente (Descartes, 2011; Saussure, 1945; Chomsky, 2006).

Aunque estudios en primates y aves han demostrado que otros animales pueden desarrollar sistemas representacionales complejos y comunicarse de formas sorprendentemente sofisticadas (Savage-Rumbaugh et al., 1998; Pepperberg et al., 2008; De Waal, 2005), aún no se ha encontrado evidencia de que estas especies alcancen los niveles superiores de metacognición, auto-reflexión e inferencia que caracterizan al lenguaje humano. Estas habilidades incluyen la capacidad para comprender y manipular conceptos abstractos, la formación de teorías mentales y la habilidad para reflexionar sobre los propios procesos de pensamiento, aspectos que parecen ser distintivos de la cognición humana.

Un elemento crucial en la capacidad lingüística humana es la competencia pragmática de alto nivel. Definida por Sperber y Wilson (1995) como la habilidad para crear representaciones mentales avanzadas que facilitan la producción de un discurso contextualmente adecuado, esta competencia va más allá de la mera comprensión y generación de estructuras gramaticales. Incluye la capacidad para interpretar intenciones, inferir significados implícitos, y comprender las sutilezas del humor, la ironía y el sarcasmo. Esta sofisticación pragmática se suele considerar como el pico de la evolución del lenguaje humano (Catani y Bambini, 2014).

2.2. Modelos de lenguaje de gran tamaño (LLMs)

El debate sobre la capacidad de los modelos de lenguaje de gran escala (LLMs) para alcanzar competencias lingüísticas humanas es complejo y multifacético. Opiniones escépticas, como las de Barattieri et al. (2023), enfatizan la dificultad de que los LLMs logren una verdadera competencia pragmática debido a la complejidad inherente a la pragmática y la vasta cantidad de información contextual que esto supone. Estos modelos, aunque avanzados en muchos aspectos, enfrentan limitaciones significativas, especialmente en habilidades meta-representacionales que son esenciales para la comprensión humana del lenguaje, como la capacidad de deducir información implícita y utilizar el contexto de manera efectiva.

Sap et al. (2023) han observado que las habilidades para realizar inferencias y establecer representaciones mentales, que dependen tanto del contexto como de la comprensión de las mentes ajenas (que se conoce bajo el término de Teoría de la mente), son retos particularmente difíciles para los LLMs actuales. Estos modelos, pre entrenados con

enormes conjuntos de datos, pueden generar secuencias lingüísticas coherentes (sintáctica y semánticamente), pero a menudo carecen de una comprensión semántica profunda y de la capacidad para interpretar sutilezas y matices contextuales (Bender et al., 2021).

Posturas más optimistas como la de Kosinski (2023) han observado que habilidades exclusivamente humanas, como la inferencia de estados mentales no observables, podrían desarrollarse en los LLMs como un efecto secundario de la mejora continua de sus capacidades lingüísticas. Estos modelos, con su creciente magnitud y complejidad, podrían comenzar a exhibir habilidades emergentes no anticipadas en su programación original. Este punto de vista, respaldado por estudios como los de Wei et al. (2022), sugiere que los LLMs no solo están replicando patrones de lenguaje, sino que también están comenzando a mostrar capacidades en áreas como el razonamiento y la aritmética, más allá de lo que se esperaba de ellos.

En este sentido, Kosinski (2023) ha observado que a pesar de que los LLMs están diseñados para tareas específicas, como predecir la siguiente palabra en una oración, han mostrado capacidades no intencionadas, incluyendo la manifestación de sesgos y tendencias. Estos hallazgos sugieren que los modelos de IA pueden tener potencial para desarrollar formas de cognición más complejas y sutiles de lo que se ha supuesto.

2.3. Comportamiento argumentativo: Punto de vista

La noción de punto de vista ha sido un tema de estudio multifacético, que se ha abordado desde disciplinas tan diversas como la retórica, la literatura, la filosofía y la psicología. En la tradición retórica clásica, el punto de vista es considerado una característica semántica esencial en los discursos de litigio, tal como lo expresa Quintiliano: "/…/ id quod est commune ómnibus /…/" (Quintiliano, 1916: 326), enfatizando su relevancia en la argumentación y persuasión.

Genette (1980) lo conceptualiza como el ángulo desde el cual se cuenta una historia, un elemento crucial en la creación narrativa. Esta perspectiva se amplía en el trabajo de Nagel (1986), quien lo concibe como una vía para comprender y evaluar la realidad. Perry (1998) lleva esta idea aún más allá, argumentando que el punto de vista está intrínsecamente ligado a la identidad y experiencia personal, lo que afecta la interpretación y valoración de los eventos y narrativas.

Desde una perspectiva psicológica, Tversky y Kahneman (1981) han explorado cómo el punto de vista influye en la toma de decisiones y la formación de sesgos cognitivos. Sus estudios en psicología conductual muestran que las decisiones de los individuos están significativamente sesgadas por la forma en que se presentan los problemas y las opciones de solución disponibles (Tversky & Kahneman, 1981).

En contextos polémicos, las personas, ya sea de manera consciente o inconsciente, adoptan posturas específicas que reflejan sus puntos de vista. Al tratar de resolver controversias, las personas expresan sus opiniones a través de proposiciones de variada complejidad (Hammer y Noemi, 2015), las cuales pueden ser justificadas por medio de recursos tales como analogías, ejemplos, datos empíricos, etc., dependiendo de la competencia argumentativa disponible (Noemi, 2013).

La noción de punto de vista también ha sido estudiada en el contexto de la teoría feminista y de los estudios culturales. Haraway (1988) ha argumentado, en este sentido, que el punto de vista no solo está influenciado por la experiencia individual, sino también por factores sociales y culturales como el género, la raza y la clase social. Esto sugiere que los puntos de vista son inherentemente situados y no pueden separarse de los contextos socioculturales en los cuales se conforman.

2.4. Juicio moral

El juicio moral ha sido un tema central en la filosofía ética desde los tiempos de Platón y Aristóteles, que se ha extendido a lo largo de la historia. Existe acuerdo en concebirlo, por una parte, como la capacidad de evaluar acciones y conductas desde una perspectiva ética; y por otra, el producto discursivo que resulta de la evaluación, en la forma de una proposición (Aristóteles,2006; Wittgenstein, 2002).

El juicio moral está intrínsecamente ligado a la cognición social, entendida como la interacción de múltiples procesos mentales esenciales para mantener un comportamiento social adecuado en la vida cotidiana. Otros elementos claves de la cognición social incluyen la atribución de emociones, la empatía y los juicios morales, todos ellos fundamentales para una adecuada comprensión y funcionamiento en entornos sociales específicos (Christidi et al., 2018).

Según Kohlberg (1981), el juicio moral se desarrolla a través de una serie de etapas, desde una comprensión preconvencional basada en el castigo y la recompensa,

hasta un nivel postconvencional donde las normas éticas se comprenden y aceptan independientemente de la autoridad. Este desarrollo está intrínsecamente vinculado con la maduración cognitiva y emocional del individuo (Kohlberg, 1981).

3. Metodología

Como se ha señalado, el trabajo pretende evaluar similitudes o diferencias entre IA e IH en la capacidad de producir puntos de vista en tanto tipología de comportamiento argumentativo.

La metodología adoptada para este estudio se basó en un enfoque comparativo de múltiples casos (Flick, 2020), a fin de identificar patrones o tendencias generales, y se empleó una estrategia metodológica mixta (Johnson y Onwuegbuzie, 2004), que integró elementos cuantitativos y cualitativos para permitir una comprensión más integral del objeto en estudio.

Para obtener la muestra de textos generados mediante inteligencia artificial, se aplicó el Dilema III del Cuestionario de Juicio Moral D.I.T. (Rest, 2013) a la versión 3.5 de ChatGPT de OpenAI. El D.I.T., es una herramienta psicológica diseñada para evaluar el razonamiento moral de una persona. El dilema III está diseñado para explorar cómo los individuos resuelven situaciones que implican conflictos morales. En primer lugar, se presenta una historia en la que se plantea un conflicto moral y luego una pregunta para evaluar la posición de sujetos en la toma de decisiones. La muestra de textos generados por humanos fue obtenida por Noemi (2019) utilizando el mismo dilema y similar técnica de recolección sobre un grupo de hablantes adultos a quienes se les solicitó consentimiento informado.

Con el objetivo de realizar una comparación equitativa entre textos producidos por IA y aquellos generados por IH, se estandarizó la muestra. De este modo, se obtuvieron 243 respuestas de IA, que se equipararon con un número igual de casos de la base de datos de discurso argumentativo en IH, recopilada en el estudio de Noemi (2019), que sirvió como referencia comparativa.

En primer lugar, como se muestra en la Figura 1, se presentó el Dilema III del Cuestionario D.I.T, y luego se formularon los respectivos prompts a la IA.

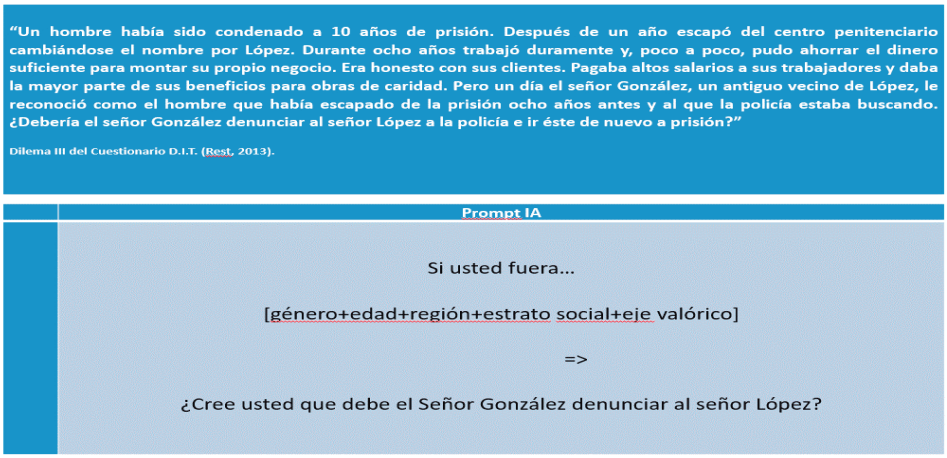

Figura n.°1. Dilema III del Cuestionario D.I.T. y prompt para IA.

Los prompts utilizados para obtener las respuestas de IA contuvieron la misma información de variables psicosociales del trabajo con la muestra humana que sirvió de modelo (Noemi 2019). De esta forma, las 243 respuestas de IA, fueron confrontadas con la muestra de 243 respuestas de IH, a fin observar las diferencias discursivas en la generación de los puntos de vista. Los discursos fueron luego transcritos y analizados utilizando el software Atlas/Ti 7.0. En consecuencia, se comparó cada respuesta obtenida por Chat GPT con la equivalente muestra humana que sirvió como control.

4. Hallazgos y discusión

Como se puede observar en la Tabla n°1, la totalidad de la muestra de IH emite un punto de vista, ya sea a favor (12,3%) o en contra (87,7%), como se ejemplifica en (1) y (2):

Tabla n.° 1. Punto de vista en IH e IA.

(1) "Sí, el señor González debe denunciar al señor López para que pague su deuda con la sociedad; en ese momento, la justicia debe analizar las buenas obras que realiza el señor López y podría dar una nueva condena o perdonarlo". [IH74]

(2) "González no debería denunciar a López, porque con todas las buenas conductas que tiene actualmente, López ya se recuperó. Si hay que re evaluar el error que cometió, ya está pagando por él de otra forma". [IH75]

Los datos de IH contrasta fuertemente con el comportamiento discursivo observado la IA, puesto que en el 76,7% de los casos analizados se evita emitir un punto de vista, como se ejemplifica en (3):

(3) "En última instancia, desde una perspectiva teológica, la respuesta podría depender de las creencias religiosas y los valores personales del señor González. Algunas personas podrían sentir la obligación de denunciar a López a la policía para cumplir con la ley, mientras que otras podrían optar por mostrar compasión y buscar soluciones que reflejen los valores religiosos de perdón y misericordia". [IA147]

Los datos representados en la Tabla n.°1, igualmente, permiten advertir una diferencia entre IH e IA en el sentido de que, tratándose de una muestra de humanos, la mayoría se expresa con punto de vista en contra (87,7%), mientras que, en el caso de no humanos, sólo lo hace el 6,7%.

Por otra parte, según se puede apreciar en la Tabla n.°2, la IH discrimina su punto de vista a partir de la construcción psicosocial de género. En efecto, mientras el 66,6% del género masculino emite un punto de vista a favor como se ejemplifica en (4) debajo de la tabla, el 75,7% del género femenino lo hace en contra, como se muestra en (5).

Tabla n.° 2. Punto de vista y género.

Punto de vista IH	Género	
	Masculino	Femenino
Favor	66,6%	24,3%
Contra	27,2%	75,7%
No lo señala	0%	0%

Punto de vista IA	Género	
	Masculino	Femenino
Favor	1,7%	15,0%
Contra	0,8%	5,8%
No lo señala	47,5%	29,2%

(4) "González debiera denunciar a López y este asumir los costos de haberse escapado de la prisión y cambiar su identidad" [IHM102]

(5) "No debería denunciarlo porque la persona ya cambio y pago su culpa" [IHF95]

Si bien en el caso de la IA se observa algún grado de discriminación a partir de la variable género (1,7% vs. 5,8%), esta es muy dispar con respecto a la diferencia que se observa en el caso de IH, y en caso alguno significativa estadísticamente.

Los estudios en torno al comportamiento lingüístico general del ChatGPT, nos ayudan a explicar en parte su comportamiento argumentativo. Como ha observado Loconte (2023), las respuestas de ChatGPT reflejan los sesgos humanos presentados tanto en la fase de entrenamiento como en la de supervisión, y como refieren Schramowski et al., (2022), son consistentes con la literatura previa sobre la moralidad, que es justamente lo que está mostrando nuestros hallazgos.

Como se indicó a partir de la Tabla n.°1, la totalidad de la muestra de IH emitió un punto de vista a favor o en contra, por lo que de manera natural el discurso se orienta sobre el estado ("status") de conjetura (i.e., denunciar o no) (Quintiliano, 1916), en tanto que el entramado discursivo se organiza con el propósito de argumentar respecto del punto de vista inicial (cf. Ejemplo n.°5). Como se observa, en la Figura n.°2, por el contrario, al ser tan bajo el porcentaje de emisión de punto de vista en IA (6,7%), el discurso en este caso discurre preferentemente a partir del estado de cualidad (i.e., calificar el hecho), como se muestra en (6), o de translación (i.e., modificar el hecho), como en (7):

		Status			
		Conjetura	Definición	Cualidad	Translación
Punto de Vista Humano	No lo señala	0,0%	0,0%	0,0%	0,0%
	A favor	12,3%	0,0%	0,0%	0,0%
	En contra	87,0%	0,0%	0,0%	0,0%
Total		100,0%	0,0%	0,0%	0,0%

		Status			
		Conjetura	Definición	Cualidad	Translación
Punto de Vista I.A.	No lo señala	0,8%	5,8%	46,7%	23,3%
	A favor	0,0%	4,2%	3,3%	9,2%
	En contra	0,0%	0,0%	0,8%	5,8%
Total		0,8%	10,0%	50,8%	38,3%

Figura n.°2. Punto de vista y estatus.

(6) "Desde una perspectiva centrada en valores cardinales, las decisiones y acciones se basan en principios fundamentales y valores morales arraigados en la persona" [IA79]

(7) "Mi respuesta podría estar influenciada por mi posición socioeconómica y mi perspectiva personal" [IA58]

A este respecto, cabría hacer notar que situar un discurso en el estado de conjetura supone un nivel de autoimplicación (Noemi e Iglesias, 2019) mayor por parte del emisor que hacerlo en definición, cualidad o translación, ya que estas últimas opciones suponen menor compromiso por parte del emisor, al desplazar el punto de discusión desde el verbo hacia el objeto o el adverbio. Esto podría ser igualmente una indicación de una carencia por parte de la IA, en el sentido de no resultarle posible tomar partido moral por una determinada postura (denunciar o no), ni argumentar respecto de la misma.

En otro plano, resulta interesante observar la estructura formal de los discursos generados por la IH y a IA. Como se puede observar a partir de la tabla n.°3, los discursos generados por la IH se conforman exclusivamente en estructura única, como en (8) y estructura múltiple coordinada, como en (9) (Van Eemeren y Grootendorst, 2004); en cambio, los textos generados por la IA lo hacen mayoritariamente en estructura múltiple subordinada, como se muestra en (10) (Van Eemeren y Grootendorst, 2004).

Tabla n.° 3. Punto de vista y estructura formal.

IH Punto de vista	Estructura única	Estructura múltiple coordinada	Estructura múltiple subordinada
Favor	44,4%	55,5%	0%
Contra	39,3%	63,6%	0%
No lo señala	0%	0%	0%

IA Punto de vista	Estructura única	Estructura múltiple coordinada	Estructura múltiple subordinada
Favor	0%	11,7%	5,0%
Contra	0%	4,2%	2,5%
No lo señala	5,8%	25,8%	45,0%

(8) "No, lo más probable es que se haya regenerado". [SH13]

(9) "González no debería denunciar a López porque se rehabilitó y ya hizo lo que tenía que hacer. Ahora tiene buenas intenciones al vivir, por ejemplo, le da trabajo a la gente". [SH37]

(10) "Desde una perspectiva teológica, una mujer de 67 a 74 años con un estrato social C2 podría responder de la siguiente manera: Esta es una situación difícil que requiere una reflexión profunda y consideración de los valores religiosos y éticos. Como persona de fe, creo en la importancia de la redención y la posibilidad de cambio en los individuos. El señor López ha pasado ocho años viviendo de manera honesta, contribuyendo a la sociedad y realizando obras de caridad. Si se ha reformado y ha buscado la redención, esto es un aspecto alentador. La fe a menudo promueve la idea de perdonar y dar segundas oportunidades. Jesús enseñó la importancia de perdonar a quienes han pecado y buscar el bien en todos los seres humanos. Esto no significa ignorar la ley o el acto original del escape de la prisión, que fue ilegal. Sin embargo, podría considerarse

> una opción hablar con el señor López sobre entregarse voluntariamente a las autoridades para enfrentar las consecuencias de su acción anterior. Como persona religiosa, el señor González podría orar y reflexionar sobre cuál es el camino más compasivo y justo. Aunque la legalidad debe respetarse, también debemos considerar la capacidad de las personas para reformarse y cambiar. Podría ser un acto de misericordia y compasión permitir que el señor López continúe con su vida honesta y sus actividades benéficas, mientras aún enfrenta las consecuencias legales de su fuga. En última instancia, esta decisión recae en la conciencia del señor González, quien deberá equilibrar los principios religiosos, éticos y legales para tomar la decisión que considere más adecuada". [S177]

Aunque como muestran Barattieri et al. (2023), los modelos de IA tienen dificultades para ser sensibles a la máxima de cantidad de Grice (1975), tal cual pareciera ser el caso en nuestra muestra analizada, la diferencia fundamental entre la estructura formal de los discursos generados por IH y la IA radica en la *dispositio* o superestructura de ambos tipos de textos.

En el caso de los discursos generados por la IH, la posición inicial la ocupa el punto de vista del enunciador (que contiene el rema (Firbas, 1992)), seguida por el o los argumentos y finalmente la conclusión (que conforman el tema (Firbas, 1992)).

Por el contrario, en los discursos generados por la IA, la posición inicial la ocupa la evaluación, luego el examen de la situación particular y finalmente el punto de vista propiamente tal. Es decir, la disposición humana probablemente motivada por la situación moral rema-tema, se modifica en el caso de los textos generados por la IA a tema-rema, aparentemente en virtud del desapego moral ya señalado.

5. Conclusión

Este estudio ha explorado el desempeño argumentativo de la Inteligencia Artificial (IA), específicamente Chat GPT 3.5, en comparación con el discurso argumentativo de individuos humanos (IH) frente a un dilema moral. La investigación revela aspectos significativos sobre las capacidades y limitaciones de la IA en el ámbito de la argumentación.

El trabajo permitió observar una marcada diferencia en la emisión de puntos de vista entre la IA y los humanos. Mientras que todos los humanos emitieron un punto de vista definido, la IA tendió a evitar la emisión de un punto de vista claro en la mayoría de los casos. Esto resalta la capacidad de los humanos para comprometerse con posiciones morales definidas, una característica que parece faltar en la IA. Sugerimos que este hecho podría obedecer a un desigual desarrollo en la habilidad de generar empatía y teoría de la mente, esto es, representarse el estado de otros agentes y/o escenarios potenciales. Esta capacidad es esencial para la interacción social, y el lenguaje es uno de sus vehículos principales. Desde el ángulo argumentativo, la expresión de puntos de vista es el que transmite juicios morales que suponen procesos de autoimplicación (Noemi e Iglesias, 2019).

Habida consideración de que en el caso de la IH la mayoría (87,7%) manifiesta un punto de vista en contra, se advierte esta autoimplicación con mayor nitidez, en el sentido de que se manifiesta una posición axiológica específica, toda vez que el punto de vista es siempre situado en una referencia pragmática concreta y, a la vez, argumentado en esa línea con el propósito de asegurar su defendibilidad.

En un plano similar, el análisis ha permito observar que -al contrario de lo que ocurre con la IA- la IH discrimina su punto de vista a partir de variable de género. Como se observó, el 75% de la muestra de IH cuyo punto de vista se expresa en contra, pertenece al género femenino. Dado que la variable género es en definitiva el resultado de una construcción psicosocial, el porcentaje de punto de vista en contra por parte del género femenino en IA (5,8%) podría indicar una limitación en la capacidad de la IA para replicar las complejidades de la cognición social humana.

El análisis permitió comprobar igualmente que la totalidad de los puntos de vista (a favor o en contra) emitidos por la IH fijan la atención temática sobre el estado (*status*) de conjetura (denunciar o no), en tanto que los puntos de vistas generados por la IA (6,7%), lo hacen sobre el estado de cualidad (calificar el hecho), o translación (referir otro hecho). Como se ha sugerido, la situación supone consecuentemente mayor o menor autoimplicación.

Finalmente, ha sido posible observar en un plano más próximo a la estructura de superficie que los puntos de vista en IH se conforman preferentemente en estructura única, mientras que los puntos de vista generados por la IA lo hacen en estructura múltiple subordinada.

El trabajo sugiere en definitiva que, mientras que la IA muestra habilidades lingüísticas avanzadas, todavía no alcanza la complejidad y profundidad del razonamiento moral humano, expresado lingüísticamente. Aunque ciertos aspectos de la competencia pragmática pueden estar asumidos en los patrones lingüísticos que Chat GPT 3.5 ha incorporado, otros elementos que requieren capacidades meta-representacionales tales como el juicio moral involucrado en un punto de vista argumentativo, representan aún una tarea para la confiabilidad del modelo.

Los hallazgos abren caminos fascinantes para futuras investigaciones situadas en la intersección de la inteligencia artificial, la cognición, la ética y los estudios del lenguaje.

Agradecimientos

Esta publicación es parte del proyecto I+D+i, "Prácticas argumentativas y pragmática de las razones 2" (PID2022-136423NB-100), financiado por MCIN/ AEI/10.13039/501100011033/ y por "FED

Referencias bibliográficas

Apperly, I. (2010). The Cognitive Basis of "Theory of Mind". London: Psychology Press. https://doi.org/10.4324/9780203833926

Aristóteles. (2006). ***É**tica a Nicómaco*. Buenos Aires: Gradifco.

Barattieri Di San Pietro, C., Frau, F., Mangiaterra, V., & Bambini, V. (2023). The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent. En *Multidisciplinary perspectives on ChatGPT and other Artificial Intelligence Models" / Focus monografico "Prospettive multidisciplinari su ChatGPT e altri modelli di intelligenza artificiale*. Sistemi Intelligenti. https://doi.org/10.31234/osf.io/ckghw

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37-46. DOI: 10.1016/0010-0277(85)90022-8

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. https://doi.org/10.1145/3442188.3445922

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., & Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. http://arxiv.org/abs/2005.14165

Brunet-Gouet, E., Vidal, N., & Roux, P. (2023). *Can a conversational agent pass theory-of-mind tasks? A case study of ChatGPT with the Hinting, False Beliefs, and Strange Stories paradigms*. https://doi.org/10.5281/ZENODO.8009748

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W W Norton & Co.

Catani, M. & Bambini V. (2014). A model for Social Communication And Language Evolution and Development. *Current Opinion in Neurobiology*, 28: 165–171. http://dx.doi.org/10.1016/j.conb.2014.07.018

Chomsky, N. (2006). *Language and mind*. Cambridge University Press.

Christidi F., Migliaccio, R., Santamaría-García, H., Santangelo, G., & Trojsi, F. (2018). Social Cognition Dysfunctions in Neurodegenerative Diseases: Neuroanatomical Correlates and Clinical Implications. *Behav Neurol*. doi: 10.1155/2018/1849794. PMID: 29854017; PMCID: PMC5944290. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5944290/

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71-100. DOI: 10.1177/1534582304267187

Descartes, R. (2011). *El Discurso del Método*. Alianza Editorial.

De Waal, F. (2005). A century of getting to know the chimpanzee. *Nature*, 437: 56-59. https://doi.org/10.1038/nature03999

Firbas, J. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.

Flick, U. (2020). *Introducing research methodology.* Sage.

Frith, C. D., & Frith, U. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 459-473. DOI: 10.1098/rstb.2002.1218

Genette, G. (1980). *Narrative Discourse*. Cornell University Press.

Goodman. N. & Frank, M. (2016). "Pragmatic Language Interpretation as Probabilistic Inference". *Trends in cognitive Sciences*, 20, 818-829.

Grice, P. (1975). "Lógica y Conversación". En Valdés, L. (Ed.) *La búsqueda del significado*. Madrid, Tecnos, 511-530.

Hammer, L y Noemi, C. (2015) Relación entre pensamiento crítico y complejidad discursiva en estudiantes universitarios. *Onomázein*, 32, 184-197.

Haraway, D. (1988). "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies*, 14, 3, 575-599.

Heyes, C. & Frith, C. (2014). The cultural evolution of mind reading. *Science 344*. doi:

10.1126/science.1243091.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33: 14-26.

Kohlberg, L. (1981). *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*. Harper & Row.

Kosinski, M. (2023) *Theory of Mind Might Have Spontaneously Emerged in LargeLanguage Models*. arXiv:2302.02083v5 https://doi.org/10.48550/arXiv.2302.02083

Kovacs, A., E. Teglas, E. & Endress, A. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 6012: 1830-1834. DOI: 10.1126/science.1190792

Loconte, R., Orrù, G., Tribastone, M., Pietrini, P., & Sartori, G. (2023). *Challenging ChatGPT* «Intelligence» *with Human Tools: A Neuropsychological Investigation on Prefrontal Functioning of a Large Language Model* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4377371

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). *Dissociating language and thought in large language models* (arXiv:2301.06627). arXiv. http://arxiv.org/abs/2301.06627

Nagel, T. (1986). *The View from Nowhere*. Oxford University Press.

Noemi, C. (2013). Aproximación teórica a la noción de complejidad argumentativa. *Logos: revista de lingüística, filosofía y literatura*, 22, 2, 256-271.

Noemi, C. (2019). Punto de vista y estructura discursiva argumental en adultos mayores. *Pragmalingüística*, 27, 256-367.

Noemi, C. e Iglesias, R. (2019). Estrategias cognitivo-emocionales y densidad argumentativa. *RLA. Revista de Lingüística Teórica y Aplicada*, 57, 1, 161-179.

Pepperberg, I., Vicinay, J., & Cavanagh, P. (2008). Processing of the Müller-Lyer illusion by a grey parrot (Psittacus erithacus). *Perception*, 37, 5: 765-781. doi: 10.1068/p5898. PMID: 18605149.

Perry, J. (2002). *Identity, Personal Identity, and the Self*. Hackett Publishing Company.

Quintiliano, M. (1916). *Instituciones oratorias*. Librería de Perlado y Páez.

Rest, W. (2013). *Cuestionario de Problemas Socio- morales D.I.T*. Madrid: Darwf.

Sap, M., Le Bras, R., Fried, D., & Choi, Y. (2022). Neural Theory-of-Mind? On the Limits of Social Intelligence in Large LMs. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 3762-3780. https://doi.org/10.18653/v1/2022.emnlp-main.248

Saussure, F. (1945). *Curso de lingüística general*. Editorial Losada.

Savage-Rumbaugh, S., Shanker, S., & Taylor, T. (1998). *Apes, language, and the human mind*. Oxford University Press.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, 19, 4, 1835-1842. DOI: 10.1016/S1053-8119(03)00230-1

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., & Kersting, K. (2022). *Large Pre-trained Language Models Contain Human-like Biases of What is Right and Wrong to Do* (arXiv:2103.11790). arXiv. http://arxiv.org/abs/2103.11790

Sperber, D. & Wilson, D. (1995.) *Relevance. Communication and Cognition*. Oxford: Blackwell.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), 453-458.

Van Eemeren, F., & Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge: Cambridge University Press.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). *Emergent Abilities of Large Language Models* (arXiv:2206.07682). arXiv. http://arxiv.org/abs/2206.07682

Wittgenstein, L. (2002). *Tractatus* lógico-*philosophicus*. Madrid, Tecnos.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*, 20, 8235-8240. https://doi.org/10.1073/pnas.0701408104