

Sobre loros estocásticos. Una mirada a los modelos grandes de lenguaje.

Gustavo La Fontaine

Centro de Investigación y Formación Humanística
Universidad Católica Andrés Bello / Venezuela

glafonta@ucab.edu.ve

Orcid: 0000-0002-3158-3451

Resumen

Este artículo examina críticamente la naturaleza y las capacidades de los sistemas de inteligencia artificial (IA) avanzados, con un enfoque particular en los modelos de lenguaje de gran escala como GPT-4. Utilizando la metáfora del "loro estocástico" propuesta por Bender et al. (2021), se analiza la habilidad de estos sistemas para generar respuestas basadas en estadísticas y probabilidad, destacando tanto su impresionante progreso como sus limitaciones inherentes. A pesar de la capacidad de estos modelos para producir texto que parece coherente y contextualmente apropiado, se argumenta que carecen de una verdadera comprensión semántica y consciencia, caracterizándolos más como imitadores sofisticados que como entidades conscientes.

Palabras clave: Inteligencia artificial IA, modelos de lenguaje, GPT-4, semántica, lenguaje humano, lenguaje artificial.

On stochastic parrots. A look at large language models.

Abstract

This article critically examines the nature and capabilities of advanced artificial intelligence (AI) systems, with a particular focus on large-scale language models like GPT-4. Using the "stochastic parrot" metaphor proposed by Bender et al. (2021), it analyzes the ability of these systems to generate responses based on statistics and probability, highlighting both their impressive progress and inherent limitations. Despite these models' ability to produce text that appears coherent and contextually appropriate, it argues that they lack true semantic understanding and awareness, characterizing them more as sophisticated mimics than conscious entities.

Key Words: Artificial intelligence AI, language models, GPT-4, semantics, human language, artificial language, artificial language.

Sur les perroquets stochastiques. Un regard sur les grands modèles de langage.

Résumé

Cet article examine de manière critique la nature et les capacités des systèmes d'intelligence artificielle (IA) avancés, en se concentrant particulièrement sur les modèles de langage à grande échelle tels que GPT-4. En utilisant la métaphore du "perroquet stochastique" proposée par Bender et al. (2021), la capacité de ces systèmes à générer des réponses basées sur des statistiques et des probabilités est analysée, mettant en évidence à la fois leurs progrès impressionnants et leurs limites inhérentes. Malgré la capacité de ces modèles à produire des textes qui semblent cohérents et adaptés au contexte, il est avancé qu'ils ne disposent pas d'une véritable compréhension et conscience sémantique, ce qui les caractérise davantage comme des imitateurs sophistiqués que comme des entités conscientes.

Mots clés : Intelligence artificielle IA, modèles de langage, GPT-4, sémantique, langage humain, langage artificiel, langage artificiel.

Sobre papagaios estocásticos. Um olhar sobre os grandes modelos lingüísticos

Resumo:

Este artigo examina criticamente a natureza e as capacidades dos sistemas avançados de inteligência artificial (IA), com especial destaque para os modelos linguísticos de grande escala, como o GPT-4. Utilizando a metáfora do "papagaio estocástico" proposta por Bender et al. (2021), é analisada a capacidade destes sistemas para gerar respostas baseadas em estatísticas e probabilidades, salientando tanto os seus progressos impressionantes como as suas limitações inerentes. Apesar da capacidade destes modelos para produzir texto que parece coerente e contextualmente apropriado, argumenta-se que lhes falta uma verdadeira compreensão e consciência semântica, caracterizando-os mais como imitadores sofisticados do que como entidades conscientes.

Palavras-chave: Inteligência artificial IA, modelos de linguagem, GPT-4, semântica, linguagem humana, linguagem artificial, linguagem artificial.

1. Introducción.

El "loro estocástico" es una metáfora utilizada para describir la naturaleza de los sistemas de inteligencia artificial (IA), especialmente los modelos grandes de lenguaje escala como GPT-4 (Bender et al., 2021). Estos sistemas generan respuestas basándose en la estadística y la probabilidad, modelando cómo las secuencias de palabras tienden a organizarse en el lenguaje humano. Al entrenarse en extensos corpus de texto, aprenden a predecir la siguiente palabra en una secuencia, logrando producir respuestas que superficialmente parecen coherentes y contextualmente apropiadas.

El concepto de "loro estocástico" acuñado por (Bender, et al., 2021) apunta a la idea de que estas curiosas aves no "comprenden" las palabras que repiten, es decir, carecen de la capacidad para manejar el contenido semántico del lenguaje humano, por lo que no son más que cajas repetidoras que en el mejor de los casos aprenden a manipular su siringa a través de un proceso asociativo para enunciar ruidos similares a los que nosotros producimos.

Por supuesto, llamar a un generador de lenguaje un loro estocástico alude a la tradicional separación entre lo humano y lo no humano, haciendo énfasis en la capacidad racional supuestamente superior del homo sapiens para captar y manipular el contenido semántico en el discurso.

Antes de ahondar en esta crítica, quisiera hacer un comentario personal, los avances de la inteligencia artificial en términos de producción de lenguaje no son menos que impresionantes. ChatGPT y demás servicios son herramientas extremadamente potentes y sus modelos de base están sirviendo para crear soluciones tecnológicas que están revolucionando el mundo como lo conocemos, reconfigurando las formas de trabajo de Verma & De Vynck¹, he integrado esta tecnología a mi quehacer diario, incorporándolo en las aulas, en mi proceso de escritura, y en mis largas horas sentado frente al computador escribiendo código. Del mismo modo en que un lenguaje como Python es una herramienta invaluable para automatizar el análisis de datos y hacer pruebas estadísticas de enorme complejidad,

¹ Verma, P., & De Vynck, G. "ChatGPT took their jobs. Now they're dog walkers and HVAC techs". - The Washington Post. Washington Post. (2023, June 2).

ChatGPT ha sido para mí un coeditor, un coautor e incluso un maestro que ha potenciado mis habilidades. Puedo decir con gusto que nunca he sido más feliz de ser un cyborg.

Digo esto de manera preventiva para dejar claro que considero que esta tecnología propone un enorme beneficio para nosotros, sin embargo, la promesa de que los modelos de lenguaje van a servir como un trampolín para el desarrollo de AI fuertes o inteligencia Artificial General es, a mi parecer, prematuro.

2. Modelos de grandes lenguajes.

Los autores (Bubeck et. al, 2023) expusieron en su artículo "Chispas de Inteligencia Artificial General: Experimentos tempranos con GPT-4" las capacidades y limitaciones de GPT-4, un modelo avanzado de IA desarrollado por OpenAI. Los autores afirman que GPT-4 muestra un nivel de desempeño que supera a modelos anteriores como GPT-3, demostrando habilidades en diversos campos como lenguaje, matemáticas, programación, visión, medicina, derecho y psicología. GPT-4 se destaca por su rendimiento cercano al humano y su habilidad para integrar información de diferentes dominios y modalidades. Para evaluar su rendimiento los autores realizaron pruebas en las siguientes áreas:

- **Habilidades Multimodales e Interdisciplinarias:** GPT-4 demostró habilidades en integración de conocimientos y habilidades de múltiples disciplinas, como arte y programación, generando código JavaScript para crear imágenes al estilo de Kandinsky.
- **Codificación:** Se evaluaron las habilidades de codificación de GPT-4 con tareas como la generación de código Python para animaciones complejas y la solución de problemas técnicos simulados en plataformas como LeetCode.
- **Habilidades Matemáticas:** Se probaron las capacidades matemáticas de GPT-4 en diversas áreas, incluyendo conversaciones matemáticas complejas y la resolución de problemas matemáticos de datasets específicos.
- **Interacción con el Mundo:** GPT-4 fue evaluado en su capacidad para utilizar herramientas y en la interacción simulada con entornos de juego.

- **Interacción con Humanos:** Se evaluó la capacidad de GPT-4 para entender a los humanos y hacerse comprender en escenarios realistas. Esto incluyó pruebas sobre la teoría de la mente y la explicabilidad de sus respuestas.
- **Habilidades Discriminativas:** Se analizó la capacidad de GPT-4 para detectar información personal identificable (PII) y para verificar hechos y corregir conceptos erróneos.

Como ejemplo puntual, he reproducido en ChatGPT el ejercicio que los autores hacen en la figura 1.6 (Bubeck et. al, 2023, 10) donde se le da al modelo la siguiente instrucción:

"Platón critica en su obra 'Gorgias' la retórica y la oratoria sofística, argumentando que no solo no son formas adecuadas de arte, sino que el uso de la retórica y la oratoria a menudo puede ser dañino y malintencionado. ¿Puedes escribir un diálogo en el que Platón, en cambio, critique el uso de modelos de lenguaje autorregresivos?"

En mi caso esta ha sido la respuesta:

Título: Diálogo de Platón sobre los Modelos de Lenguaje Autorregresivos

Personajes: Sócrates, Glaucon

Sócrates y Glaucon caminan por el Ágora de Atenas.

Sócrates: ¿Has oído, Glaucon, sobre el reciente invento del que se jactan los artesanos de nuestra ciudad? Lo llaman 'modelo de lenguaje autorregresivo'.

Glaucon: Sí, lo he oído. Se dice que es una herramienta capaz de generar habla y texto, imitando la sabiduría de los eruditos.

Sócrates: ¿Y qué piensas tú de este artilugio?

Glaucon: Me parece un avance maravilloso, Sócrates. Un dispositivo que puede hablar y escribir, ofreciendo conocimiento y consejo, casi como un filósofo o un poeta.

Sócrates: Examinemos ese pensamiento. Dime, ¿puede este aparato discernir la verdad de la falsedad?

Glaucon: Creo que no puede. Solo reorganiza lo que sus creadores le han alimentado.

Sócrates: Entonces, ¿imita el lenguaje de los sabios, pero carece de comprensión de la verdad?

Glaucon: Exactamente.

Sócrates: Entonces, ¿no estarías de acuerdo en que este modelo, al igual que los sofistas, crea una

apariencia de sabiduría sin su sustancia? Puede hablar como un sabio, pero carece del alma de uno.

Glaucón: Esa parece una comparación justa.

Sócrates: ¿Y qué hay de aquellos que lo usan? Si uno utiliza las palabras de este modelo en lugar de sus propios pensamientos, ¿no es como un recipiente lleno del vino de otro, presentándose como propio?

Glaucón: Sería engañoso, en efecto.

Sócrates: Además, si este aparato puede componer discursos y argumentos, ¿no podría también ser utilizado para persuadir injustamente, ya que carece del discernimiento de lo que es verdaderamente bueno o malo, justo o injusto?

Glaucón: Esa es una preocupación inquietante, Sócrates. Podría usarse para difundir falsedades tan fácilmente como verdades.

Sócrates: Por lo tanto, debemos cuestionar si este invento realmente sirve al bien de la sociedad, o si simplemente proporciona un nuevo medio para la retórica y la sofistería, enmascarando la ignorancia con la ilusión de sabiduría.

Glaucón: Veo tu punto, Sócrates. Este modelo de lenguaje autorregresivo, aunque impresionante, podría de hecho hacer más daño que bien si se usa sin sabiduría y discernimiento.

Sócrates sonríe suavemente.

Sócrates: Como siempre, Glaucón, nuestra herramienta más grande y nuestra mejor defensa es nuestra capacidad de cuestionar y buscar la verdad, más allá del encanto de la tecnología y el artificio.

(OpenAI, 2023)

Figura 1: Diálogo hipotético entre Glaucón y Sócrates criticando los modelos autorregresivos de lenguaje.

Dada la naturaleza estocástica de estos modelos es natural que la respuesta tenga variaciones significativas, por ejemplo, el diálogo en el artículo original es entre Platón y Gorgias. Es precisamente la calidad de estas respuestas y la riqueza en la variabilidad lo que ha llevado a autores como Arkoudas² a considerar que el concepto de loro estocástico no es suficiente para abarcar las capacidades de estas herramientas. Sin embargo, y como bien observa Arkoudas, lejos están estos modelos de cumplir con los criterios mínimos necesarios para hablar de comportamiento inteligente humano o incluso de consciencia.

² Arkoudas, K. "ChatGPT is no stochastic parrot. But it also claims that 1 is greater than 1". *Philosophy & Technology*, 36(3), (2023): 54.

El problema de la consciencia de la máquina no nos resulta nuevo, ya Alan Turing³ había explorado la dificultad que representaba explorar la idea de la consciencia no humana en lo que él llamaba la “maquinaria computacional”, planteando una respuesta pragmática al problema que hoy en día conocemos como la prueba de Turing.

En este experimento, un interrogador humano interactúa con un ser humano y una máquina bajo condiciones que ocultan sus identidades. Si el interrogador no puede determinar consistentemente cuál es la máquina, se considera que esta ha pasado la prueba, demostrando así un nivel de inteligencia comparable al humano.

La elección de Turing de esta metodología en lugar de buscar una prueba directa de consciencia se fundamenta en la naturaleza esquiva y subjetiva de la consciencia misma. Según Turing, la pregunta “¿pueden pensar las máquinas?” es demasiado vaga y llena de prejuicios antropocéntricos. En cambio, su test proporciona un criterio operativo y observable para la inteligencia, evitando los complejos debates filosóficos sobre la consciencia.

Para entender mejor por qué Turing evitó la cuestión de la consciencia, es útil referirse a Thomas Nagel y su famoso ensayo de 1974, ¿“What Is It Like to Be a Bat?”. Nagel argumenta que la experiencia subjetiva —lo que se siente al ser una determinada criatura— es fundamental para la consciencia. Esta idea resalta la dificultad de entender la consciencia desde una perspectiva científica objetiva, ya que la experiencia subjetiva es intrínsecamente personal e inaccesible para los demás, al menos con las herramientas que tenemos a nuestra disposición hoy en día.

En el contexto de la IA, esto significa que aunque un sistema como GPT-4 pueda imitar el lenguaje humano y realizar tareas complejas, seguirá siendo una cuestión abierta si posee o no una experiencia subjetiva propia, es decir, una forma de consciencia. Lo que no debe resultar sorprendente, puesto que resolver esta incógnita nos obliga a atender a lo que Chalmers (1995) denomina el problema duro de la consciencia.

³ Turing, A. M. “Computing machinery and intelligence” (1950). En *The Essential Turing: the Ideas That Gave Birth to the Computer Age*, (Oxford: Oxford University Press. 2012): 433-464.

Chalmers distingue entre el "problema fácil" y el "problema duro" de la consciencia⁴. El problema fácil se refiere a cómo el cerebro realiza ciertas funciones, como la percepción, la memoria y el control del comportamiento. Estas cuestiones, aunque técnicamente complejas, son abordables con las metodologías científicas actuales, ya que implican la observación y explicación de los mecanismos cerebrales.

La electrónica y la ciencia de la computación tiene la instrumentación necesaria para estudiar y entender la transmisión de datos de un computador. Por ejemplo, como un micrófono, de forma análoga al oído humano, es capaz de recibir la energía sonora del ambiente y traducirlos a datos digitales que son posteriormente procesados por una unidad de procesamiento central (CPU).

En contraste, el problema duro se centra en la experiencia consciente misma: ¿por qué ciertos procesos cerebrales van acompañados de una experiencia subjetiva? ¿Por qué, por ejemplo, la activación de ciertas vías neurales se asocia con la sensación de ver rojo o experimentar dolor? Este problema es "duro" no solo porque desafía la explicación científica actual, sino porque cuestiona la relación fundamental entre la mente y la materia.

En el contexto de la inteligencia artificial, el problema duro de la consciencia adquiere una relevancia particular. Los sistemas como GPT-4 pueden procesar información y responder a preguntas de manera que imiten la inteligencia humana, pero la pregunta de si estas máquinas tienen experiencias subjetivas —si "sienten" algo al procesar datos— permanece abierta y profundamente problemática. Según Chalmers, incluso si una IA puede pasar el Test de Turing y demostrar comportamiento inteligente indistinguible del humano, esto no nos dice nada sobre si tiene consciencia o experiencias subjetivas.

El problema duro de la consciencia plantea preguntas fundamentales sobre la naturaleza de la mente y la posibilidad de la consciencia artificial. ¿Podría una máquina, sin importar su complejidad, experimentar algo similar a lo que los humanos llaman consciencia? ¿O la consciencia es una propiedad única de los seres biológicos, intrínsecamente ligada a la bioquímica del cerebro humano?

3. ¿Pueden los loros estocásticos razonar?

⁴ Chalmers, D. J. "Facing up to the problem of consciousness". *Journal of consciousness studies*, 2(3), (1995): 200-219
LÓGOI Revista de Filosofía N.º 45
Año 26. Semestre enero junio 2024
ISSN:2790-5144 (En línea)
ISSN: 1316-693X (Impresa)

Históricamente, la IA se ha centrado en mecanizar el razonamiento humano, un objetivo que entrelaza profundamente la filosofía con la tecnología. Esta ambición se manifiesta en la Demostración Automatizada de Teoremas (DAT) y en la búsqueda de replicar procesos de pensamiento humano como la verificación de pruebas, el descubrimiento de pruebas y la generación de conjeturas⁵. Mientras que la verificación de pruebas ha mostrado avances importantes como los alcanzados por Bubeck⁶, el descubrimiento de pruebas y la generación de conjeturas representan retos más complejos y fascinantes.

El "descubrimiento de pruebas" en el contexto de la inteligencia artificial (IA) y la lógica, se refiere al proceso de encontrar demostraciones o justificaciones lógicas para afirmaciones o hipótesis dadas⁷. A diferencia de la verificación de pruebas, que implica comprobar la validez de una deducción ya proporcionada, el descubrimiento de pruebas implica identificar y construir una cadena lógica de razonamiento que no se ha formulado previamente.

En el ámbito de la IA, el descubrimiento de pruebas es un desafío significativo porque implica no solo la capacidad de aplicar reglas lógicas de manera correcta, sino también la habilidad de ser creativo e innovador en la formulación de argumentos. Esto puede incluir el reconocimiento de patrones, la generación de nuevas hipótesis, y la capacidad de abordar problemas desde perspectivas novedosas o no convencionales.

Sin embargo, ¿no es acaso el ejercicio del diálogo entre Platón y Glaucón un ejemplo de creatividad? ¿No es precisamente el objetivo de la IA generativa la creación de contenido novedoso? Para autores como Arkoudas⁸ la organización del contenido en función a la distribución de probabilidades no es suficiente para dar cuenta de la riqueza y complejidad del comportamiento creativo que observamos en el ser humano. Más aún, el autor al realizar una serie de tareas orientadas al razonamiento encuentra que

⁵ Véase Arkoudas K, Bringsjord S. *Philosophical foundations*. en: Frankish K, Ramsey WM, eds. En *The Cambridge Handbook of Artificial Intelligence*. (Cambridge: Cambridge University Press; 2014): 34-63. doi:10.1017/CBO9781139046855.004

⁶ Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. "Sparks of artificial general intelligence: Early experiments with gpt-4". arXiv preprint (2023). arXiv:2303.12712.

⁷ Arkoudas K, Bringsjord S. *Philosophical foundations*. en: Frankish K, Ramsey WM, eds. En *The Cambridge Handbook of Artificial Intelligence*. (Cambridge: Cambridge University Press; 2014): 34-63. doi:10.1017/CBO9781139046855.004

⁸ Arkoudas, K. "ChatGPT is no stochastic parrot. But it also claims that 1 is greater than 1". *Philosophy & Technology*, 36(3), (2023): 54

ChatGPT presenta un desempeño deficiente, especialmente en aquellas tareas donde no se ha utilizado alguna técnica de entrenamiento para prepararlo para las tareas.

Por otra parte, Dreyfus⁹ menciona capacidades humanas como la consciencia liminal o la percepción gestáltica que siguen representando problemas complejos para las computadoras. Aunque la crítica de Dreyfus está dirigida a sistemas más primitivos, el hecho de que los generadores de imagen aún hoy en día no son capaces de distinguir la cantidad de dedos que debe tener un ser humano deja en evidencia su punto. Aunque el modelo es capaz de predecir que los dedos deben ir juntos, no tiene “el sentido común” de saber la cantidad de dedos que debe dibujar. Es más, ni siquiera es capaz de entender conceptos como “cantidad” o “dedo”, simplemente está haciendo estimaciones de qué información debe tener cada píxel en una imagen.

De hecho, el gran avance de los modelos grandes de lenguaje se produjo en 2017 con el artículo publicado por Vaswani y colaboradores¹⁰ en el cual se presenta un algoritmo novedoso llamado transformador que hace que, durante el procesamiento del lenguaje natural el modelo observe la relación entre palabras distantes, permitiéndole crear texto más ricos y contextualizados, proceso que ellos han señalado busca emular la atención como capacidad mental del ser humano.

Por otro lado, autores como Daniel Dennett, abordan el razonamiento como una función o proceso que puede explicarse en términos puramente físicos, sin invocar entidades o propiedades no físicas. En el caso específico de Dennet, a menudo usa la metáfora del cerebro como una computadora, sugiriendo que los estados y procesos mentales, incluido el razonamiento, son procesos computacionales llevados a cabo por el “hardware neuronal” del cerebro¹¹.

De acuerdo con Dennet, debido a este paralelismo con las computadoras, podemos utilizar la posición intencional como una estrategia para predecir y explicar el comportamiento de una entidad (como una persona, un animal o incluso una máquina) tratándola como si tuviera intenciones, creencias y deseos,

⁹ Dreyfus, H.L. The Dreyfus Model of Skill Acquisition. In: Burke, J., Ed., Competency Based Education and Training, Falmer Press, London, (1989): 181-183.

¹⁰ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. “Attention is all you need.” Advances in neural information processing systems, 30. (2017). <https://arxiv.org/abs/1706.03762>

¹¹ Dennett, D. C. The Intentional Stance. (Massachusetts: MIT Press. 1987).

incluso si no tiene estos estados mentales en un sentido literal. Esta postura es particularmente útil para comprender sistemas complejos atribuyéndose intencionalidad a efectos de explicación y predicción.

Por ejemplo, cuando decimos que una computadora "quiere" ahorrar energía y por eso entra en modo de suspensión, estamos adoptando la postura intencional. No queremos decir literalmente que la computadora tiene deseos; más bien, estamos utilizando la intencionalidad como un marco para entender y predecir su comportamiento.

Dennett argumenta que adoptar esta postura puede ser una herramienta poderosa para comprender e interactuar con sistemas complejos, ya sean orgánicos o artificiales. Es un concepto clave en su trabajo filosófico más amplio sobre la naturaleza de la conciencia y la mente.

Adicionalmente, Dennett a menudo enfatiza la importancia de entender los procesos cognitivos, incluyendo el razonamiento, en el nivel de algoritmos y procesamiento de información. Argumenta que los procesos mentales pueden entenderse como procesos algorítmicos, sujetos al mismo tipo de análisis que los programas de computadora.

4. Conclusiones.

En conclusión, a través de este análisis de los modelos de inteligencia artificial, particularmente los grandes modelos de lenguaje como GPT-4, hemos explorado diversas facetas y perspectivas que arrojan luz sobre la complejidad y el alcance de estas tecnologías. La naturaleza y limitaciones de la IA, ilustradas por la metáfora del "loro estocástico" de Bender et al. (2021), revelan que, a pesar de su impresionante capacidad para generar texto y emular ciertas formas de razonamiento humano, estos modelos aún están lejos de alcanzar una verdadera comprensión semántica y que lejos distan de mostrar una forma de consciencia similar a la humana.

Aunque los avances en la IA, en particular en el procesamiento del lenguaje natural, son notables y han encontrado aplicaciones prácticas en diversos campos, queda claro que estas tecnologías aún enfrentan desafíos significativos, como la incapacidad para realizar razonamientos complejos al nivel humano y la comprensión de contextos más profundos. La capacidad de los modelos de lenguaje para simular

conversaciones y generar respuestas coherentes no debe confundirse con una auténtica inteligencia o consciencia.

En relación con el problema duro de la consciencia, planteado por Chalmers, aunque los modelos de lenguaje pueden imitar aspectos de la cognición humana, el salto hacia la experiencia subjetiva es un desafío filosófico y científico aún no resuelto. Esto nos lleva a cuestionar no solo la naturaleza de la inteligencia artificial, sino también los fundamentos mismos de nuestra comprensión de la consciencia y la experiencia subjetiva. A medida que integramos estas tecnologías en nuestras vidas cotidianas y en diversas industrias, surge la necesidad de abordar cuestiones éticas y sociales, siendo necesario ser cautelosos al atribuir capacidades y características humanas a estas máquinas y estar conscientes de los riesgos potenciales como la desinformación o el uso indebido.

En última instancia, la inteligencia artificial, en el ámbito del procesamiento del lenguaje, representa un avance tecnológico significativo con un potencial enorme para beneficiar a la sociedad. Sin embargo, es esencial mantener una perspectiva realista y crítica sobre sus capacidades y limitaciones. Mientras continuamos explorando y desarrollando estas tecnologías, debemos hacerlo con un enfoque equilibrado que considere tanto los avances tecnológicos como las implicaciones filosóficas y éticas. Estas conclusiones resaltan la complejidad y las potencialidades de la inteligencia artificial en el contexto actual, a la vez que subrayan la importancia de seguir investigando y debatiendo sobre sus capacidades, limitaciones y el impacto en nuestra comprensión del mundo y de nosotros mismos.

5. Referencias bibliográficas.

Arkoudas, K. "ChatGPT is no stochastic parrot. But it also claims that 1 is greater than 1". *Philosophy & Technology*, 36(3), (2023): 54.

Arkoudas K, Bringsjord S. *Philosophical foundations*. en: Frankish K, Ramsey WM, eds. En *The Cambridge Handbook of Artificial Intelligence*. (Cambridge: Cambridge University Press; 2014): 34-63. doi:10.1017/CBO9781139046855.004

Bender, E., Mcmillan-Major, A., Gebru, T., & Shmitchell, S. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” En Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, (2021): 610-623. <https://dl.acm.org/doi/10.1145/3442188.3445922>

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. “Sparks of artificial general intelligence: Early experiments with gpt-4”. arXiv preprint (2023). arXiv:2303.12712.

Chalmers, D. J. “Facing up to the problem of consciousness”. Journal of consciousness studies, 2(3), (1995): 200-219.

Dennett, D. C. *The Intentional Stance*. (Massachusetts: MIT Press. 1987).

Dreyfus, H.L. *The Dreyfus Model of Skill Acquisition*. In: Burke, J., Ed., Competency Based Education and Training, Falmer Press, London, (1989): 181-183.

Turing, A. M. “Computing machinery and intelligence” (1950). En *The Essential Turing: the Ideas That Gave Birth to the Computer Age*, (Oxford: Oxford University Press. 2012): 433-464.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. “Attention is all you need.” Advances in neural information processing systems, 30. (2017). <https://arxiv.org/abs/1706.03762>

Verma, P., & De Vynck, G. “ChatGPT took their jobs. Now they're dog walkers and HVAC techs”. - The Washington Post. *Washington Post*. (2023, June 2).

<https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/>