

¿Es la inteligencia artificial doxástica un igual epistémico?*

Is doxastic artificial intelligence an epistemic peer?

ALBERTO MURCIA CARBONELL*

Resumen: La inteligencia artificial doxástica (IAD) es un tipo de inteligencia artificial que reproduce actitudes doxásticas. Si la IAD cumple con las condiciones de paridad epistémica que se le exige a un humano, ¿podría ser también un igual epistémico? Dos iguales epistémicos sostienen propiedades cognitivas simétricas como la inteligencia, el razonamiento o la ausencia de sesgos. Para evaluar si alguien es un igual se tendrán en cuenta estas condiciones: (1) la igualdad probatoria, (2) la igualdad cognitiva y (3) revelación completa. La IAD cumple tanto (1) como (2), pero es en (3) cuando se descubre que no puede ser un igual epistémico. Ésta responde con la opinión popular más aceptada estadísticamente, es incapaz de sostener y defender sus propias afirmaciones y éstas no son un genuino acto de habla. Es una máquina doxástica incapaz de señalar cuáles son las razones que guían su respuesta.

Palabras clave: desacuerdos epistémicos, epistemología, cultura digital, autoridad epistémica, inteligencia artificial, entrenamiento automatizado.

Abstract: Doxastic artificial intelligence (DAI) is a subclass of artificial intelligence that reproduces doxastic attitudes. If DAI meets the conditions for epistemic peerhood that are required for a human, could it also be considered as an epistemic peer? Two epistemic peers hold symmetric cognitive properties such as intelligence, reasoning, or absence of bias. To assess whether someone is a peer, these conditions will be taken into account: (1) evidential equality, (2) cognitive equality, and (3) situation of full disclosure. DAI meets both (1) and (2), but it is (3) when it is revealed that it cannot be an epistemic peer. Its answers are the most accepted popular opinion statistically, it is unable to sustain and defend its own asseverations, which are not considered as a genuine speech act. It is a doxastic machine unable to point out what are the reasons guiding its answers.

Keywords: epistemic disagreements, epistemology, digital culture, epistemic authority, artificial intelligence, machine learning.

Recibido: 11/04/2024. Aceptado: 24/04/2024.

* Este artículo fue posible gracias al proyecto nacional de investigación *Meta-actitudes, desacuerdos profundos y progreso moral*, financiado por el Ministerio de Ciencia e Innovación, PID2021-124152NB-I00. Universidad Carlos III de Madrid.

** Universidad Carlos III de Madrid. Profesor Asociado. Las líneas de investigación principales son: 1) filosofía y tecnología, 2) ética y nuevos medios, 3) cultura digital y 4) *game studies*. Última publicación: Murcia, A. (2022). La justificación del victimario como estrategia identitaria para el veterano de guerra. *Dilemata*, (39), 61-78. Correo electrónico: amurcia@uc3m.es

1. Introducción

La inteligencia artificial (IA) reproduce comportamiento inteligente bastante sofisticado. Puede auditar las cuentas de una compañía, determinar el reparto de ayudas sociales, reconocer el rostro de una persona en diferentes contextos, dirigir un vehículo u ordenar el inventario de un almacén para mejorar la eficiencia logística, por mencionar algunos ejemplos. Incluso es capaz de reproducir actitudes doxásticas, como ofrecer un diagnóstico médico, escribir un ensayo sobre un tema, responder a dilemas morales, trazar la mejor ruta para llegar hasta un destino determinado, o incluso discutir sobre temas complejos de calado filosófico. Dado que muestra estas propiedades doxásticas que parecen de fiar, ¿podríamos llegar a considerar que una IA es un igual epistémico?

Antes de atajar la cuestión, consideraremos que la IA pertenece al grupo de la tecnología inteligente [*smart technology*]¹. La tecnología inteligente son artefactos cuya función es reproducir comportamiento inteligente (Floridi 2023). No todas las tecnologías inteligentes son IA, pero la IA siempre es una tecnología inteligente. Un lavaplatos, una máquina automática de planchado, o un robot aspiradora, reproducen comportamiento inteligente (lavar platos, planchar, barrer), pero no son IA. De entre las tecnologías inteligentes que son IA, tendremos en cuenta solo las que reproducen un tipo de comportamiento: las actitudes doxásticas. Todas las IA reproducen comportamiento inteligente, pero no todas reproducen actitudes doxásticas. Una IA de reconocimiento facial no reproduce actitudes doxásticas, Dall-e tampoco. Un navegador inteligente, que traza una ruta entre dos espacios geográficos calculando el mejor tiempo de llegada, sí que reproduce actitudes doxásticas. ChatGPT, como generador de textos, también reproduce actitudes doxásticas. Llamaremos inteligencia artificial doxástica (IAD) a este tipo de IA, y será el objeto de este estudio.

Para considerar si una IAD puede ser un igual epistémico primero hay que entender qué es un desacuerdo epistémico. Éstos se producen cuando dos personas, que se consideran iguales epistémicos, sostienen actitudes doxásticas diferentes después de haber examinado la misma evidencia de carácter público (Frances y Matheson 2019, Zagzebski 2012).² Una de las consecuencias de estar en desacuerdo es que cuando se nos revela que alguien a quien consideramos como un igual epistémico sostiene que no-P, podría ser razonable que dejemos de confiar en nuestra creencia sobre P (Cocchiaro y Frances 2021, p. 1063, Kelly 2013). Los iguales comparten simetría en sus capacidades cognitivas, en aquello que saben sobre P y en el acceso a las evidencias. A la opinión de un igual siempre se le otorga cierto crédito, ya que se le considera como autoridad epistémica, aunque sea una autoridad de tipo débil, esto es, la autoridad del experto que lo es únicamente en el ámbito de las proposiciones sobre P (Zagzebski 2012).³ Si alguien se ve en la posible obligación de re-evaluar su creencia

1 Traducimos “*smart technology*” por “tecnología inteligente”, aunque consideramos que el término es confuso, pues contribuye a la idea errónea de que son “inteligentes”. En términos de Floridi (2023), son *listas, ingeniosas o sagaces*, pero no son inteligentes. Sucede lo mismo con “inteligencia artificial”, que da a entender algo incorrecto sobre esta tecnología, esto es, que produce inteligencia.

2 En concreto, aquí nos referimos a desacuerdos razonables, “casos en los que personas inteligentes con acceso a la información disponible y relevante llegan a conclusiones incompatibles” (Felman 2011, p.416).

3 Una autoridad epistémica no siempre es un igual epistémico, pero un igual epistémico siempre es una autoridad epistémica desde el punto de vista del agente que evalúa la paridad. En otras palabras, para que A considere

es porque tiene la certeza de que la otra persona es un igual epistémico. Así, antes de que llegue el desacuerdo, primero tendrá que determinar si el otro es un igual, y no a la inversa, es decir, que sea el desacuerdo el que le muestre que se encuentra ante un igual.

Para saber si se está ante un igual epistémico se constata si ese alguien cumple ciertas condiciones de simetría que permitan evaluarlo como tal. Alguien es un igual si posee ciertas virtudes epistémicas (Zagzebski 2012, p. 205), como la inteligencia, la capacidad para razonar y la ausencia de sesgos (Rowland 2017, Kelly 2005), a lo que hay que añadir que sus antecedentes garantizan que es conoedor de P (Christensen y Lackey 2013). El igual no se circunscribe solo a *alguien*, sino que también podría ser *algo*. Si una IAD cumple las condiciones de simetría, entonces se debe afirmar que también es un igual epistémico. Señalamos a continuación éstas condiciones, elaboradas desde Bistagnino (2011), pero que se siguen igualmente desde Kelly (2005 y 2013), Christensen (2009), Lackey (2013), Rowland (2017) Frances y Matheson (2019) y Cocchiari y Frances (2021).

(1) *Igualdad probatoria*: A y B conocen las evidencias y los argumentos que influyen en la cuestión de P.

(2) *Igualdad cognitiva*: A y B son igualmente virtuosos en términos epistémicos en su evaluación de las pruebas y argumentos que influyen en P.

(3) *Revelación completa*: A y B comparten todas sus pruebas y argumentos sobre P.

Al considerar a alguien (o algo) como un igual, se evalúa su comportamiento como expresión de su cognición, prestando atención también al historial de eficiencia y competencia del agente sobre P (Christensen y Lackey 2013). Si la evaluación es positiva, entonces A tiene razones para confiar en B como su igual epistémico, el cual también es reconocido con cierta autoridad epistémica. Dicho esto, ¿podemos considerar a una IAD como un igual epistémico?

Supongamos lo siguiente. Javier es un taxista experimentado. Acaba de recoger a un cliente que quiere llegar lo antes posible al aeropuerto. ¿Cuál es la mejor ruta? Javier conoce bien las carreteras y está acostumbrado a las contingencias de su ciudad a esas horas. Tiene un navegador con IAD que, para tomar un ejemplo en concreto, diremos que es Google Maps. Javier puede seguir varias estrategias: (i) Conecta Maps y deja al dispositivo que sugiera la mejor ruta; (ii) Prefiere ser él quien decide la ruta y no lo conecta; (iii) Javier evalúa la ruta a la vez que activa Maps para que también le proporcione un recorrido hasta el aeropuerto.

Podemos pensar (i) de dos formas: o bien Javier está tomando Maps como un superior epistémico, o bien considera que es un tipo de autoridad epistémica débil. En el primer caso, Javier evalúa como más probable, digamos un 90%, que la IAD trace una ruta mejor de la que él pueda pensar. Javier asume que el dispositivo es virtuoso en tanto que eficiente. A eso se le llama “deferencia completa” (Elga 2007): ante una proposición sobre la ruta óptima de

a B como un igual epistémico sobre P, A está obligado a afirmar que B es una autoridad epistémica sobre P. Siguiendo a Zagzebski (2021, p.1), utilizamos el concepto de “autoridad epistémica” como “autoridad epistémica débil”, es decir, alguien que es un “experto” sobre un tema, un área de conocimiento, una afición, etc. Una cuestión diferente es por qué el experto es autoridad. Se podrá observar que en el artículo se ha simplificado bastante la cuestión, asumiendo que el experto es alguien (o algo) que es capaz de producir enunciados verdaderos de forma estable y efectiva sobre un asunto determinado.

llegada al aeropuerto, las posibilidades de que Javier crea que P son las mismas que tiene Maps de considerar que P. En el segundo caso, podemos pensar en Maps como una autoridad epistémica débil. Siguiendo el sentido que le da Zagzebski (2012, p.3), primero debemos partir desde la premisa de que Javier tiene la intención de tener certeza; después, Javier puede considerar X como la mejor ruta, pero si Maps responde que Y tendrá un mejor tiempo estimado de llegada, Javier, que quiere estar en lo correcto, desechará su creencia sobre X y depositará su confianza en Y. Javier considera a la IAD como una autoridad epistémica, pero no como un igual epistémico.

En (ii), Javier cree saber sin duda cuál es la mejor ruta. Sus años conduciendo le dan la confianza necesaria como para justificar su creencia. Así que no utilizará Maps. En esta situación, Javier se considera como la única autoridad epistémica.

En (iii), la opinión de Javier sobre cuál es la ruta óptima no depende de lo que Maps sugiere, sino que va a comparar lo que él cree con lo que responda el navegador. Suele pasar que la ruta trazada por Maps coincide con la que Javier opina que es la óptima, por lo que en esta ocasión también confía en que será igual. De esta manera, Javier trata a Maps como si fuera un igual epistémico, pero también como una autoridad epistémica, porque suele ser efectiva en su función. Es (iii) la situación que nos interesa para este artículo. ¿Está Javier en un error al evaluar que la IAD puede ser un igual epistémico?

En lo que sigue no indagaremos sobre si podemos tener un desacuerdo razonable genuino con una IA, tampoco si el desacuerdo *debería* obligarnos a reconsiderar nuestras creencias, ni si en la revelación completa los iguales deben discutir sobre sus argumentos y razones. Queremos centrarnos únicamente en la pregunta de si una IAD puede ser considerada como un igual epistémico. Tal vez se pueda inferir de nuestras palabras que estamos asumiendo que se puede tener un desacuerdo con una IA, puesto que ser un igual es condición necesaria para el desacuerdo epistémico. No es así. La posición que se defenderá aquí es que la IAD no puede ser considerada como un igual epistémico, y si esto es condición necesaria para estar en un desacuerdo, entonces no podemos tener un desacuerdo. Sea como sea, como dijimos, no trataremos este asunto aquí.

Afirmamos que la IAD puede cumplir con las condiciones de (1) igualdad probatoria y de (2) igualdad cognitiva, pero (3) la revelación completa descubre que la naturaleza estadística de las opiniones reproducidas es insuficiente para que la consideremos como igual y como autoridad epistémica. Las razones sobre por qué no es un igual las daremos en la sección 3, mientras que en la 4 dedicaremos un breve espacio a explicar las razones por las que consideramos que tampoco es una autoridad epistémica. Antes de llegar a ese punto, primero revisaremos cuál es el valor doxástico de una IAD.

2. El valor doxástico de la IAD

La IAD es un tipo de artefacto que reproduce el comportamiento inteligente de las actitudes doxásticas. Una IAD como ChatGPT representa estos datos mediante lenguaje natural, mientras que un navegador inteligente lo hace señalando una ruta desde una posición geográfica hasta otra que tenga el mejor tiempo de llegada. Sea mediante palabras o con un trazado sobre un mapa, ambas tecnologías cumplen con su tarea presentando los datos

en forma de enunciados declarativos, veraces y con valor doxástico.⁴ La IAD sustituye la inteligencia humana mediante procesos que son agenciales, pero no inteligentes (Floridi, 2023, pp. 23-24).

Es pertinente preguntarse si una IAD puede ser considerada como igual epistémico debido a las características señaladas, pero también por su irrupción en diversas actividades humanas, como en la medicina, la enseñanza o la auditoría de empresas, entre otras. Esta pregunta no se ha planteado como tal en las discusiones sobre el tema de la igualdad epistémica. Podemos señalar al menos un factor que consideramos que pudo haber influido en que el desdén hacia esta cuestión, que tiene que ver con las propiedades epistémicas de las computadoras. Se podría resumir en si una computadora, y por extensión una IA, *sabe* algo sobre aquello que computa. Desde el desafío de la “habitación china” que planteó John Searle (1982) se considera que las computadoras carecen de capacidades semánticas. El argumento de Searle, expuesto aquí de manera muy tosca, señala que las computadoras son máquinas que operan solo con sintaxis mediante reglas bien definidas en su programa. Son eficientes en su función de computar sin que sea necesario que “entiendan” el contenido semántico que transmiten sus operaciones. En otras palabras, manejan la sintáctica del lenguaje, pero no sus significados; producen información semántica, pero para que computen no hay necesidad de que sean máquinas semánticas.

Si seguimos la propuesta de Searle, esta nos lleva a descartar que una IA comparta las propiedades epistémicas de un humano. Al mismo tiempo, también podemos afirmar que, pese a esta limitación, pueden ser eficientes en su función. Que una IA desconozca el significado de la palabra “cáncer” no impide que señale un cáncer en una tomografía con un grado de acierto equivalente al de un oncólogo experimentado. Las computadoras no necesitan semántica para actuar y ser eficientes. Por lo tanto, la cuestión no sería la de tratar de equiparar las capacidades semánticas de humanos e IA, sino de si el resultado de sus funciones son equivalentes. De serlo, una máquina que reproduce actitudes doxásticas equivalentes a las que produciría un humano si éste estuviera en su lugar podría ser considerada como un igual epistémico, como trataremos de desarrollar en el siguiente epígrafe.

Alguien puede objetar que, en efecto, pueden reproducir comportamiento inteligente siguiendo un procedimiento de reglas bien definidas (Floridi 2024, pp. 39-43), pero eso no tiene nada que ver con una actitud doxástica genuina, sólo sobre seguir un algoritmo. Así es, una IA carece de estados mentales, al menos eso podemos afirmar ahora mismo. Puede darse el caso de que debido a un fenómeno emergente, una IA comience a generar creencias. Es

4 La IAD no opera con semántica (Searle 1983; 2000), pero sí entrega información semántica. Para que esto ocurra, sus datos se presentan en forma de enunciado declarativo, bien formado, significativo y veraz (Floridi 2013). En el caso de Maps, la ruta es la información semántica ofrecida por la IAD, y esta es trasladable desde un medio que presenta sus datos para ser vistos a otro en que se puede evaluar su verdad o falsedad, como es el caso de un enunciado (Floridi 2013 p. 186, Formigari 2004, pp. 91-92, Dretske 2000). Ese enunciado declarativo puede ser evaluado mientras que cumpla que los datos estén bien formados, sean significativos y veraces (Floridi 2013). “Bien formados” significa que se han ordenado según las reglas sintácticas del sistema utilizado. Por lo general, la sintaxis se relaciona con el lenguaje, pero su sentido puede ampliarse a cualquier sistema o código en el que los datos se agrupan siguiendo las reglas que los gobiernan. Datos “significativos” son aquellos que tienen significado en el sistema, código o lenguaje elegido. Que sean “veraces” implica que “proveen de contenidos verdaderos sobre el sistema modelado” (Floridi 2013, p.109). Esto último no significa que sean necesariamente verdaderos, solo que *pueden* actualizarse a verdaderos.

algo improbable, pero no imposible, pues otros fenómenos emergentes surgen causados por el hardware, como es la emisión de calor, ciertos zumbidos o incluso crujidos al ejecutar un programa (Searle 2000, p.26). Por el momento, no hay software, ni hardware que produzca inteligencia, ni remotamente parece que vaya a ser el caso en bastante tiempo (Floridi 2024), por lo que, consideramos, carece de sentido discutir esa posibilidad. Que la IAD produzca estados mentales genuinos o si solo los reproduce es de poca importancia para este artículo. Nos interesa en mayor medida lo que el comportamiento dice de la fiabilidad de sus procesos para producir actitudes doxásticas que sean estables y eficientes.

Aclaremos unas cuestiones relacionadas con Maps, ya que lo estamos tomando como ejemplo de IAD. Maps es una tecnología inteligente cuya IA está basada en aprendizaje automatizado [*machine learning*] y aprendizaje profundo [*deep learning*] (Lau 2020) que reproduce enunciados doxásticos sobre cuál es la mejor ruta para alcanzar un punto geográfico concreto. El método de aprendizaje es relevante porque determina de qué manera la IAD reproducirá el comportamiento inteligente. El aprendizaje profundo permite al dispositivo la toma de decisiones guiada por series de datos. Para que esto suceda, identifica y extrae patrones de esas series (Kelleher 2019) para después, desde esos patrones, ofrecer respuestas que se ajusten al marco de la solicitud de respuesta que demande una orden de entrada. De entre todas las respuestas probables, el dispositivo elige la que tenga un mayor índice de probabilidad de adecuarse al marco de la orden de entrada.

Maps es más complejo que un sistema de posicionamiento global (GPS). Todos los navegadores tienen GPS, pero no todos los GPS son navegadores inteligentes. El GPS relaciona el dispositivo con unas coordenadas y las proyecta sobre un mapa, pero no tendrá en consideración eventos novedosos que suceden en el trayecto. Maps recibe un comando de entrada para que el dispositivo trace una ruta óptima entre dos puntos. Esta IAD usa sus series históricas de registros de conductores humanos, desde donde extrae los patrones de comportamiento, siendo esa la parte *fija* de sus datos almacenados. A esto le añade una parte *dinámica* de series de datos, habitualmente aquellos que se están produciendo en el momento previo a la respuesta, pero que también sirven para actualizar el trazado durante el trayecto: geoposicionamiento de otros teléfonos que circulan por carretera, información del volumen de tráfico, accidentes, meteorología y condiciones del asfalto, entre muchas otras (Mehta, Kanani y Lande 2019, Lau 2020, Hassan 2022). Aunque la compañía responsable de Maps quiera situar su tasa de efectividad en un 97% (Lau 2020), siempre hay margen para el error.

Un conductor que conozca bien los alrededores, puede concluir que una ruta que parece peor en realidad es mejor al examinar evidencias que Maps es incapaz de calcular, como el comportamiento específico de los conductores que ocupan una vía, el tiempo que tardan los semáforos en cambiar de color, o que el vehículo va a pasar por una zona en donde la gente suele buscar aparcamiento, por lo que ese lugar tiende a generar retenciones. Maps tampoco es sensible al vehículo en el que se circula, por lo que desconoce cuestiones que probablemente el conductor tenga incorporada en su rutina, como que su coche se calienta en exceso en el centro de la ciudad, o que su motor le obliga a ir muy despacio en las cuestas.⁵

5 Estas cuestiones que afectan al portador del dispositivo son especialmente evidentes cuando Maps traza una ruta para un peatón cuyas especificidades son evitadas por el código, estableciendo el camino sin importar la condición física, edad o sus problemas cognitivos. Maps como IAD está diseñado fundamentalmente para la

3. La IAD como igual epistémico

Si la IAD cumple con las condiciones que exigimos a un ser humano para que lo consideremos como un igual epistémico, entonces deberíamos reconocer que la IAD también es un igual. A considera a B como igual epistémico, en caso de que B cumpla con las condiciones de

(1) igualdad probatoria, (2) igualdad cognitiva y (3) revelación completa.

Como condición previa, A debe sostener una actitud distinta al egoísmo epistémico y reconocer que hay otros seres con sus mismas capacidades epistémicas. La cuestión aquí es si B debe cumplir escrupulosamente con las tres condiciones o sólo en parte. Sostenemos que la igualdad absoluta es un ideal improbable que solo funciona desde una perspectiva teórica (Kelly 2005; King 2012; Cocchiario y Frances 2021), pero eso no debe impedir que nos preguntemos que, en caso de que encontrar un caso de igualdad epistémica, cuáles son las condiciones que lo permiten. Defendemos que esas tres propiedades son simétricas, por lo que no puede haber un “término medio” (Zagzebski 2012, p.211): o bien aceptamos todas las condiciones o bien ninguna.

Consideramos que estas tres condiciones forman un mismo proceso de evaluación de la igualdad epistémica. Ninguna es suficiente por sí misma para demostrar que la IAD es un par epistémico. En (1) y (2) se evalúa si las propiedades y virtudes epistémicas del otro son simétricas, mientras que (3) tiene la función de presentar una justificación adecuada de cómo las propiedades y virtudes han elaborado *esa* creencia en lugar de otra cualquiera. En una situación de igualdad, (3) será simétrica, pues ambas partes llegarán a la misma creencia siguiendo el mismo camino de argumentos, razones o justificaciones. Hay que pensar que el proceso tiene que ocurrir de forma simultánea, es decir, no es un programa en el que primero se evalúa (1) y si es correcta se pasa a (2), y después a (3). Que las hayamos ordenado no se debe confundir con que sean pasos a seguir.

3.1. Igualdad probatoria.⁶

La igualdad probatoria exige que A y B conozcan las evidencias y los argumentos que influyen en la cuestión de P (Leslie 2013, Audi 2013). Si aceptamos que las evidencias

conducción de vehículos, cuando se trata de rutas a pie se comporta como un GPS que confunde aceras con calzadas. Sobre algunas cuestiones que se resuelven en ese párrafo, ver la nota 8.

6 Los argumentos que se exponen tanto en 3.1, 3.2 y 3.3 se orientan en la dirección de casos factuales. Con el ejemplo de Maps, partimos de la discutible idea de que podemos determinar factualmente cuál es la mejor ruta posible, cuando es un caso similar a un enunciado sobre el gusto (“Juego de Tronos es la mejor serie del mundo”) o un enunciado moral (“Es mejor decir la verdad en una situación así”), esto es, son evaluativos. Agradecemos al revisor/a anónimo/a esta importante puntualización. Si analizamos las condiciones desde una perspectiva evaluativa, en lugar de la factual, los resultados nos llevarían a la misma conclusión que llegamos aquí, esto es, que la IAD no puede ser un igual epistémico, pero la argumentación debería tener en cuenta otros factores que aquí desestimamos. En cualquier caso, creemos que afrontar esta perspectiva evaluativa necesitaría un mayor espacio de elaboración, por lo que deberá ser abordada en otra circunstancia.

deben tener un carácter público (Zagzebski 2012), entonces tanto el humano como la IAD pueden acceder en igualdad de condiciones a las evidencias, establecer relaciones y formar una proposición sobre si P.

Pensemos que las series históricas que utiliza la IAD entrenada mediante aprendizaje automatizado son enormes en términos de cantidad de datos. Desde estas series de datos, que han sido generadas por seres humanos, identificará los patrones que servirán para dar una respuesta a la solicitud requerida (Kelleher 2019); por ejemplo, de la serie de datos sobre el comportamiento de todos los conductores que han pasado por las carreteras que van al aeropuerto durante los dos últimos meses se pueden extraer patrones que servirán para que la IAD reproduzca una actitud doxástica. El navegador cuenta tanto con estas series de datos como otras series que se están produciendo a tiempo real: tráfico, pavimento, obras, accidentes, etc. Puede dar la impresión de que en un caso como el del taxista Javier éste carece de ciertas evidencias comparativamente con la IAD, cosa que es incorrecta. Si todos los datos *deben* ser públicos, entonces siempre habrá alguna manera de acceder a éstos. Si Javier desconoce el volumen de tráfico e incidencias de los vehículos en la carretera A24, puede visitar la página de la DGT, llamar a otro taxista que esté en ruta, escuchar las noticias, etc. y acceder a las evidencias necesarias para formarse una opinión.

Lo que difiere considerablemente es la capacidad de la IAD en cuestiones como el almacenamiento o la velocidad de acceso y respuesta si se compara con las de un humano, ya que la máquina casi siempre le supera en rendimiento. Para determinar la igualdad epistémica esto es irrelevante. Excepto alguna que otra discusión sobre este tipo de capacidades de los iguales deben estar también a la par (Frances y Matheson 2019), pocas veces se tiene en cuenta. Solo es relevante si las capacidades cognitivas están claramente impedidas (Christensen y Lackey 2013).

Por lo tanto, si aceptamos lo anterior, podemos decir que la IAD y el humano mantienen una simetría de igualdad probatoria.

3.2. Igualdad cognitiva.

La segunda condición exige que A y B sean igualmente virtuosos en su evaluación de las pruebas y argumentos que influyen en P. Las virtudes epistémicas, por tanto, son el foco de la discusión (Audi 2013, Rowlands 2017), siendo algunas de estas virtudes la inteligencia, la capacidad para razonar y la ausencia de sesgos (Rowland 2017, Kelly 2005), entre otras. Si debemos buscar esas virtudes en procesos internos que sean idénticos a los de un humano, entonces no puede haber simetría. La IAD carece de rasgos mentales individuales (Baehr 2011) que la coloquen en situación de simetría con un humano; tampoco tiene un cerebro humano, y aunque las redes neuronales del aprendizaje profundo imiten la estructura de las conexiones nerviosas y neuronales, su funcionamiento es diferente. En la IAD no hay inteligencia, ni conoce el significado de las proposiciones que reproduce, pues es una máquina que únicamente opera con la sintaxis. Además, debe estar sesgada, porque los sesgos son condición necesaria para evitar problemas habituales en estos sistemas como el del sobreajuste y el del infraajuste (Kelleher 2019, p.20). Sabemos que los procesos de razonamiento que llevan a la extracción de datos son similares tanto en humanos como en la IAD, pero

sería demasiado polémico por nuestra parte tratar de defender que son fenómenos idénticos. Tampoco creemos que lo sean.

Existe otra estrategia. Se sigue desde cierto escepticismo ante la condición de virtud cognitiva, que se puede inferir de las aportaciones de King (2012) y de Cocchiari y Frances (2021). Dado que es imposible tener razones suficientes para asegurar que aquel que tratamos de evaluar como igual cumple escrupulosamente con las capacidades exigibles, el acento se coloca en la expresión del comportamiento como reflejo de la actitud doxástica. Para la condición de igualdad cognitiva debería importar menos la correspondencia perfecta entre las funciones que generan los procesos evaluativos, racionales, sesgos, etc., e importar más el comportamiento como expresión de las capacidades epistémicas.

Nadie defiende que la IA tenga estados mentales, pues no produce actitudes doxásticas genuinas, pero tampoco hay una exigencia lógica de que los tenga. Como ya se señaló, aquí no defenderemos que la tecnología inteligente tenga estados mentales, ni la actual, ni ninguna que esté en un futuro a largo plazo, pese a las dudas que algunos filósofos tratan sostener al respecto (Chalmers 2023). Queda en su campo demostrar esta posición. Solo es necesario que el resultado de sus funciones resulten equivalentes a las que se le exigen a un humano para que sea calificada como igual epistémico.

Si tuviéramos que comparar los estados internos de un humano que producen una actitud doxástica y los procesos algorítmicos que reproducen una intención doxástica, no hay simetría posible. En cambio, si nos fijamos solo en el resultado, todo cambia. Por ejemplo, si un individuo ve unos platos limpios tiene complicado afirmar si fue un humano o un lavaplatos quien terminó la tarea, dado que el resultado del comportamiento del lavaplatos es equivalente al de un humano. Pero si alguien lo estuviera observando no los confundiría, aún siendo idéntica la entrada (los platos sucios) y el resultado del proceso (los platos limpios). Trasladando la analogía a una IAD, si no sometemos a escrutinio los estados internos de la máquina, se podría concluir que se da una simetría entre las capacidades epistémicas que producen resultados equivalentes.

¿Responden estas capacidades a alguna virtud epistémica? No necesariamente. En los últimos años se ha podido observar un cambio de rumbo desde la capital importancia de las virtudes epistémicas en la igualdad, como la perspicacia, la honradez o el rigor (Gutting 1982), hasta verse relegadas a un segundo plano, cobrando mayor importancia los antecedentes del igual epistémico (Bistagnino 2011, Cocchiari y Frances 2021). Antecedentes como que haya demostrado sobrada competencia en P o de si está al tanto de las particularidades de P (Leslie 2013, Audi 2013). Así, resulta menos importante si las funciones internas operan bajo el concepto de virtud mientras que produzca actitudes doxásticas de calidad. Si consideramos una virtud epistémica como producir de manera eficiente y sistemática creencias verdaderas (Driver 2000, p.126), y si la IAD ofrece creencias verdaderas eficientemente y de manera estable, entonces habría que conceder que existe simetría.⁷

Al rastrear los antecedentes se responde a cuestiones del tipo “¿esta persona suele generar creencias acertadas sobre P?” Se puede auditar el historial de efectividad sobre P tanto de una persona como de una IAD. Si, por ejemplo, Maps traza habitualmente rutas que Javier

7 Aunque en el texto parece asumirse que la calidad de la información de una IAD se juzga por ser verdadera, en realidad habría que decir que ésta se juzga por su veracidad. Al respecto, ver la nota la nota al pie 6.

también considera como óptimas, este último se puede formar una opinión sobre la efectividad de la IAD, probablemente otorgándole un grado de confianza alto. Aunque Javier sepa que el comportamiento de Maps está producido por funciones automatizadas puede resultar irrelevante en tanto que lo que observa como respuesta se ajusta a sus propias creencias, y esto le lleva a concluir que tiene capacidades epistémicas equivalentes. Puede aplicar una máxima, como propone Roland (2017), por la que B es el igual epistémico de A acerca de P si B tiene la misma probabilidad que A de tener razón acerca de P.

En conclusión, la IAD cumple la condición de igualdad cognitiva siempre y cuando: a) no se tenga en cuenta las funciones internas, pues lo relevante es que hay un proceso de cálculo sobre las evidencias y un resultado en forma de actitud doxástica que es equivalente a la de un ser humano; y b) que no consideremos como fundamental que la IAD reproduce actitudes doxásticas en lugar de creencias genuinas.

3.3. Revelación completa.

En la revelación completa A y B comparten todas sus pruebas y argumentos sobre P. En la revelación completa el otro muestra sus cartas, de forma que se averigüe si iba de farol. Se comparte, pero no es necesario discutir las pruebas, razones o argumentos sobre P, pues el objetivo de la revelación es determinar si la IAD puede ser considerada un par y no si puede haber un desacuerdo. Es necesario subrayar esto porque hay quien considera que la revelación completa sólo es necesaria para que surja un desacuerdo, pero no lo es para reconocer al otro como igual epistémico (Bistagnino 2011, p. 416), lo que implica que (1) y (2) son independientes de (3), siendo (3) irrelevante para la evaluación del par epistémico. Si un agente A considera que la IAD cumple (1) y (2), la situación de plena revelación no añade nada nuevo, pues la IAD ya es un igual. Estamos en sintonía con que la revelación completa marca el camino a seguir para determinar qué tipo de desacuerdo se está teniendo o si se produce un desacuerdo genuino, pero no coincidimos en que (3) sea una parte independiente que solo tenga esa función. De serlo, (1) y (2) serían suficientes para determinar si IAD es un igual epistémico, cosa con la que disentimos. Con (1) y (2) evaluamos las propiedades de nuestro posible igual, mientras que (3) tiene la función de presentar una justificación adecuada de cómo esas propiedades generan una creencia concreta. La simetría debe ampliarse más allá de las propiedades cognitivas. Debe mostrar cómo se ha formado una creencia concreta desde las evidencias hasta lo que se está afirmando, permitiéndonos también rastrear los antecedentes epistémicos de la otra persona.

Al ampliar la simetría con (3), encontramos tres elementos que ponen en cuestión que la IAD sea un igual epistémico: La ausencia de inteligibilidad; la IAD desconoce sobre lo que responde; la IAD no afirma nada.

*Inteligibilidad. El sistema de entrenamiento profundo es tan opaco que es prácticamente imposible rastrear cómo una IA toma sus decisiones. Es el problema de la inteligibilidad (Floridi y Cowls 2021, p.540), que también podríamos llamar de la explicabilidad o transparencia. Para que las explicaciones de una IAD sean tomadas en consideración debemos saber cómo funciona. No se trata de subrayar las diferencias entre las funciones de la máquina y el humano a la hora de hacer públicas sus razones, sino que debemos entender

cómo actuaría la IAD si estuviera en nuestro lugar. Si, por ejemplo, Maps tiene acceso a la información de que el coche se calienta en las curvas, y el navegador guía el vehículo con autonomía plena, ¿mantendrá ese valor como relevante al determinar la ruta como haría la persona propietaria del automóvil? El problema a día de hoy es que no podemos interpretar adecuadamente el proceso de toma de decisiones. Observar los resultados de la reproducción de comportamiento es insuficiente para justificarlo. Que a estos procesos se les considere una “caja negra” (Castelvecchi 2016) es sintomático de lo ininteligible de los comportamientos de estos sistemas. Incluso abrir la caja negra tampoco garantiza que lleguemos a comprender la toma de decisiones. La IA no representa la realidad mediante modelos matemáticos, sino que los crea desde sus series de datos (O’neil 2017), por lo que no solo deberíamos comprender qué camino tomó en su decisión, sino cuál es su modelo. Esto afecta a (1) porque la IA no va a mostrar cuáles son las evidencias que ha tomado y cómo las sitúa en una relevancia valorativa. El proceso es tan opaco en términos de inteligibilidad como lo es un código QR para un ser humano, sabemos que contiene información semántica, pero no la vamos podemos interpretar (Floridi 2013).

El problema de la inteligibilidad también se relaciona con la pregunta de “¿quién es el que responde?”. Porque es el sistema el que lo hace, sí, pero la actitud doxástica reproducida no se corresponde con alguien concreto, sino que es la amalgama de comportamientos factorizados en un programa. No hay un alguien, sino un revuelto de actitudes. La evaluación de un igual debe darse sobre alguien concreto que sostiene una afirmación (King 2012; Cocchiario y Frances 2021). Esto es diferente a que tenga que estar necesariamente presente en el momento de la afirmación. Se puede considerar un igual a alguien desde un testimonio, en el que se exponen las razones por las que ese alguien afirma algo sobre P. Siguiendo a Kelly (2013) y (Elga 2007), un catedrático que estudie la obra de David Hume le puede considerar como un igual, ya que entendemos que el proceso de demostración de las actitudes doxásticas de Hume quedaron registradas en sus escritos.⁸ La igualdad no tiene que ser recíproca, por esa razón es irrelevante que la IAD nos reconozca como iguales, del mismo modo que Hume, dada su condición, tampoco lo hará.

La respuesta de la IAD de Maps, en el caso de Javier, es producto de un cálculo de probabilidad que se fundamenta en las decisiones de los conductores que pasaron por los caminos que llevan al aeropuerto. Dicho de otro modo, la IAD está produciendo respuestas vicariamente, en el sentido de que ésta sustituye sin autoridad a las opiniones registradas en sus bases de datos. Del mismo modo, siguiendo esta línea de razonamiento, si no se puede determinar quién debemos evaluar como igual, tampoco podemos considerarlo como una autoridad.

*La IAD no sabe lo que responde. En la segunda sección pusimos en valor que una de las objeciones sobre si la IAD puede ser un igual es que ésta carece de semántica o contenido mental alguno (Searle 1984, 2000). De este modo, una IA no sabe cuál es la mejor ruta, o qué es un cáncer, o quién escribió *Don Quijote*, aunque pueda trazar una ruta óptima,

8 Cappelen y Hawthorne (2009) diferencian entre “desacuerdo como actividad” y “desacuerdo como estado”. Dos individuos pueden *estar* en desacuerdo sin *tener* un desacuerdo, como cuando no tienen contacto entre sí ni saben de las actitudes del otro. Por ejemplo, A es un catedrático de filosofía del siglo XXI y B es Hume, filósofo del siglo XVII. Si se diera la situación de considerar a la IAD como un igual y que esto nos lleve hasta un desacuerdo, estaríamos hablando aquí de *estar en desacuerdo*. Al respecto, Losada (2015).

señalar un cáncer en una tomografía, o atribuir a Cervantes la autoría del libro. Aunque de respuestas consistentemente eficientes, y de ahí su utilidad, acertar las respuestas es solo una parte de saber algo. Los loros también pueden responder de forma muy sofisticada a ciertas cuestiones, pero no afirmamos que sepan algo sobre lo que están graznando.⁹ Es exigible ciertos procesos normativos que garanticen que esta sabe algo y no es simplemente azar afinado para asegurar la apariencia de éxito. Más allá de las reglas que se la hayan podido establecer durante su entrenamiento para que produzca resultados específicos, no existen normas que garanticen el valor epistemológico de los enunciados. En ocasiones ni siquiera es estable, ya que sobre la misma entrada de datos la IAD puede ofrecer resultados diferentes. Elige la opinión popular más aceptada, la que estadísticamente se ajusta mejor a los datos, en lugar de sostener una maquinaria robusta de generación de conocimiento. Si le exigimos que señale las fuentes de su respuesta, lo más cercano a una justificación será darnos una opción alternativa, referirse a su condición de máquina y que, por tanto, debemos tomar sus respuestas como “sugerencias”, o apelar a la autoridad de alguna de sus fuente, dando muestra de que son incapaces de sostener sus propias afirmaciones y, por tanto, de la ausencia de compromiso hacia estas.

*La IAD no afirma. Sería atribuir demasiado a la IAD si decimos que afirma algo. Sus respuestas toman forma de un acto ilocutivo de afirmación, pero no puede ser considerada como tal. ¿Un acto locutivo? Puede ser, pero no debemos confundirlos con una afirmación genuina. Primero, porque carece de una consistencia entre lo que parece afirmar y su defensa posterior. Falta algo del acto de afirmar si el agente no se compromete con lo que dice sobre P . La respuesta de la IAD solo está garantizada desde las opiniones ajenas que se han estructurado sobre modelos matemáticos.

Segundo, la IAD incumple cualquier “norma de aserción”, como aquellas que se refieren a la adecuación epistémica entre el hablante y el contenido de su afirmación (Pagin y Marsili, 2021). Incluso sucediendo que carecemos de consenso sobre si debe haber solo una norma o pueden ser varias, la IAD incumple todas aquellas que se consideran más robustas para explicar el acto de afirmar. La norma gobernada por conocimiento de Williamson dice que “hay que afirmar que p sólo si se conoce p ” (2002, p.243). Si la IAD no conoce nada sobre p , no puede afirmar nada de p . De la misma manera, al no haber creencias genuinas en el enunciado de la IAD, ésta tampoco cumple con otras normas del tipo “hay que afirmar que p sólo si se cree que p ”, o que “hay que afirmar que p sólo si se está epistémicamente justificado en creer que p ”.

9 Una de las discusiones más relevantes sobre los Large Language Models (LLM) como ChatGPT, es si este es un “loro estocástico” (Bender, Gebru, McMillan-Major y Shmitchell, 2021). Un loro que acierta en la mayor parte de las ocasiones, sobre todo en cuestiones factuales, pero cuya arquitectura depende enormemente de las series de datos utilizadas en su entrenamiento. Otros no opinan igual y consideran que no se puede afirmar que estos modelos reproducen sus enunciados de forma caótica, pues las conversaciones que uno puede mantener en lenguaje natural alcanzan un nivel de fluidez que estadísticamente pueden considerarse como ordenadas (Arkoudas 2023). Sin embargo, esto solo oculta otros problemas subyacentes mucho más complicados de resolver, como la coherencia, esto es, mantenerse en el tema de la pregunta, producir sentido, adherirse a las máximas griceanas de conversación y demostrar un mínimo de entendimiento sobre cómo el mundo funciona (Arkoudas 2023, p.53). Esto nos lleva a que en una conversación prolongada con una IA ésta es incapaz de sostener sus propias afirmaciones.

En conclusión, si se revisa (1) y (2) desde (3) podemos concluir que aunque las primeras dos siguen siendo efectivas para determinar la igualdad con la IAD, su descripción es insuficiente. Podemos seguir defendiendo que sus procesos son equivalentes a los de un humano, pero detectamos que hay otros, como que no existe marco normativo alguno por el que garantizar la corrección de sus enunciados, que delatan la imposibilidad de que la IAD sea un igual epistémico.

4. ¿Al menos nos quedará la IAD como autoridad epistémica?

Un igual epistémico siempre es un tipo de autoridad epistémica, pero no al revés. Así que aunque no se pueda defender a la IAD como un igual, ¿al menos se puede conservar su estatus de autoridad? Si utilizamos la misma línea de razonamiento que se ha aplicado como objeción a que la IAD sea un igual epistémico, encontramos que el mismo razonamiento sirve para desechar que la IAD sea una autoridad.

Hasta ahora, hemos presentado a la IAD como medidor de lo verdadero. Un termómetro es un buen medidor de la temperatura de la sala y probablemente sea la autoridad al respecto. El medidor de verdad es identificado por Zagzebski (2012) como autoridad epistémica débil o autoridad del experto, aquella que se circunscribe a una parcela concreta de la realidad. Más allá de los límites de su área de experto, deja de ser autoridad. Por último, los objetos inanimados pueden ser considerados autoridades débiles (Zagzebski 2012, p.119). Un dispositivo de GPS es una autoridad en situar a gente en unas coordenadas de geolocalización, pero solo si es eficiente en esa tarea. Las autoridades epistémicas débiles aportan creencias robustas a las que damos preferencia sobre las nuestras, ya que “el hecho de que la autoridad tenga una creencia P es una razón para que yo crea P, que sustituye a mis otras razones relevantes para creer P y no se añade simplemente a ellas” (p. 105, traducción propia). Por ejemplo, alguien cree que la lejía cura la COVID, pero si el sistema nacional de salud recomienda no hacerlo, y esta persona considera que el sistema nacional es una autoridad, abandonará su creencia y tomará partido por lo que los expertos sugieren (Jäger 2016). En este sentido, la autoridad débil puede proporcionar creencias de mayor calidad que las nuestras. Y eso que la persona del ejemplo no conoce a alguien concreto dentro del sistema de salud, o viceversa, nadie del sistema de salud conoce a esta persona.

La autoridad débil se diferencia de la auténtica autoridad en que la segunda requiere una relación interpersonal. Es equivalente a la autoridad religiosa o política: puede ordenar que se crea que P, sustituyendo así la creencia que tenga el agente. Un médico de familia es una autoridad epistémica por el crédito que alguien le otorga y va más allá de su actividad profesional. De la misma manera, una reseña en Amazon puede formar la creencia “no comprar un libro”, pero si un amigo, del que siempre se valora su opinión, recomienda el libro, logrará que se sustituya la creencia sobre “no comprarlo” por “comprarlo” solo porque confiamos en este amigo. No está circunscrita a un espacio determinado o a un área del conocimiento, sino que es la relación entre sujetos la que establece la autoridad.

Entonces, dado que Maps no es un igual, ¿al menos podemos usar esa IAD para obtener creencias de mayor calidad? En cuanto a autoridad auténtica, tendríamos que descartar

que lo sea. Maps no pide que creas que su respuesta porque lo ordene. Tampoco establece una relación interpersonal en la que exista alguna responsabilidad entre las partes (Zagzebski 2012, p.119). La relación entre Javier y Maps es únicamente instrumental. Sería una irresponsabilidad por parte de Javier sustituir su creencia bajo la idea de que Maps es una autoridad. Pero, ¿y como autoridad débil? Maps ha demostrado ser efectivo en el trazado de rutas, por lo que Javier, que quiere estar en lo cierto, puede sustituir su creencia por la que ofrezca el dispositivo y depositar así su confianza en la fiabilidad del navegador.

Consideramos que hay una confusión en el razonamiento de Javier. Maps no es solo un GPS. Siguiendo a Zagzebski, un GPS puede ser una autoridad en tanto experto en localizar a un sujeto en unas coordenadas y marcar rutas que tengan solo en consideración la distancia que separa el punto de salida y el de llegada. Maps geolocaliza, por lo que está situado en esa franja de experto, pero también reproduce una opinión sobre cuál es la mejor ruta, atendiendo a su serie de datos sobre el comportamiento de los conductores, así como los datos a tiempo real de flujo de tráfico, pavimento, climatología, etc. También modifica esa ruta sobre la marcha en función del comportamiento de otros conductores y recomienda rutas más largas pero “más eficientes en consumo de energía”, por ejemplo.

Son actitudes doxásticas que surgen no porque sea el mejor juicio posible, sino que obedece a lo que fue el comportamiento de otros conductores. Son opiniones con poco fundamento, que son la media estadística de sus bases de datos. Una autoridad lo es porque es consistentemente fiable, pero también porque puede exponer razones (y defenderlas dado el caso) sobre P y tiene reconocida competencia sobre P (“sabe” sobre P). Es eficiente, pero también fiable. Podríamos suponer que Maps, como GPS, es eficiente, pero no es fiable. Por tanto, también es un error confiar en una IAD como autoridad epistémica. Es un instrumento doxástico que calcula cuál es la mejor opinión popular, pero que es incapaz de señalar cuáles son las razones que guían su respuesta.

5. Conclusiones

Hemos llegado a la conclusión de que una IAD no puede ser un igual epistémico. Aunque reproduce actitudes doxásticas y sus propiedades cognitivas sean equivalentes a las de un ser humano, algo que demuestra su comportamiento y resultados, en la revelación completa tenemos que considerar que estas propiedades son insuficientes. Para evaluar a alguien (o algo) como un igual epistémico debemos saber quién responde y así valorar su razonamiento, antecedentes y compromiso conceptual con lo que afirma sobre P. Una IAD no afirma nada, no sabe nada, no se compromete con sus enunciados y no se reconoce a quién corresponde la actitud doxástica que ésta reproduce. Estas mismas razones que nos llevan a negar que una IAD sea un par epistémico son igualmente aplicables para negar que sea una autoridad epistémica. Sería un error depositar nuestra confianza de manera incondicional en una IAD como autoridad epistémica débil, pues sólo algunas se fundamentan en comportamiento experto. Sería un gravísimo error considerarla como autoridad auténtica.

De entre las ventajas de conceder crédito a una autoridad epistémica es que, habitualmente, ésta nos proporciona creencias de mejor calidad. Análogamente, un igual epistémico, en caso de desacuerdo, nos podría empujar a que podamos re-evaluar nuestras creencias en

busca de otras de mejor calidad, justificación, robustez, etc. Dado que la IAD no es ni una cosa ni la otra, no podremos servirnos de esta ventaja.

Dicho esto, la calidad de la actitud doxástica reproducida por la IAD es relevante en cierta medida. Las actitudes doxásticas de la IAD no se forman de manera completamente azarosa. Depende del ajuste de los patrones de comportamiento de millones de datos que el aprendizaje automatizado aprovecha para construir las funciones de los modelos de las redes neuronales. Cuando la estadística y la programación son sólidas, así como el entrenamiento está afinado, podemos encontrar una IAD que funciona de forma considerablemente efectiva como generadora de opiniones. Es suficiente para reconocer a estas tecnologías como algo que podríamos llamar *igual doxástico*, dado que el resultado de su función es vicaria, esto es, está en lugar de una medida estadística de las actitudes doxásticas de humanos que fueron codificadas como series de datos. Aunque siga siendo un error confiar en la efectividad de la IAD como igual o como autoridad por las razones señaladas, podemos reconocer que la IA reproduce creencias estadísticamente más cercanas a lo correcto, sin que sea un efecto de “loro estocástico” o azar incoherente.

Referencias bibliográficas

- Arkoudas, K. (2023) ChatGPT is no Stochastic Parrot. But it also Claims that 1 is Greater than 1. *Philos. Technol.* 36, 54. doi:10.1007/s13347-023-00619-6.
- Audi, R. (2013). Dimensions of Intellectual Diversity and the Resolution of Disagreements. *The Epistemology of Disagreement: New Essays*, 205.
- Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. OUP Oxford.
- Bender, E. M., Gebru, T., McMillan-Major, A. y Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. En *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610-623.
- Bistagnino, G. (2011). Epistemology of disagreement: Mapping the debate. *Gli annali di LPF-Laboratorio di Politica comparata e Filosofia Pubblica*, 6, 159-187.
- Cappelen, H. y Hawthorne, J. (2009) *Relativism and Monadic Truth*. Oxford. doi:10.1093/acprof:oso/9780199560554.001.0001.
- Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, 538(7623), 20.
- Chalmers, D. J. (2023). Could a large language model be conscious?. arXiv preprint arXiv:2303.07103.
- Christensen, D. (2007). Epistemology of Disagreement: The Good News. *The Philosophical Review*, 116(2), 187–217.
- Christensen, D. (2009). Disagreement as evidence: The epistemology of controversy. *Philosophy Compass*, 4(5), 756-767.
- Cocchiaro, M. Z. y Frances, B. (2021). Epistemically different epistemic peers. *Topoi*, 40, 1063-1073. doi:10.1007/s11245-019-09678-x
- Dretske, F. I. (1981). *Knowledge and the Flow of Information*. MIT press.
- Dretske, F. I. (2000). *Perception, knowledge and belief: selected essays*. Cambridge University Press.

- Driver, J. (2000). Moral and epistemic virtue. *Knowledge, Belief, and Character: Readings in Virtue Epistemology*, Lanham, MD: Rowman & Littlefield, 123-34.
- Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3), 478-502.
- Feldman, R. (2006). Epistemological puzzles about disagreement. *Epistemology futures*, 216, 236.
- Floridi, L. (2013). *The philosophy of information*. OUP Oxford.
- Floridi, L. (2023). *The Ethics of Artificial Intelligence: principles, challenges, and opportunities*. OUP Oxford.
- Floridi, L. y Cows, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545.
- Frances, B. (2014). *Disagreement*. John Wiley & Sons.
- Frances, B. y Matheson, J., “Disagreement”, *The Stanford Encyclopedia of Philosophy* (2019 Invierno), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2019/entries/disagreement/>>.
- Gutting, G. (1982) *Religious Belief and Religious Skepticism*. Notre Dame: University of Notre Dame Press.
- Hassan, H. (2022, 4 de enero). Google Maps: A simple explanation on how it detects traffic jams. <https://medium.com/technology-hits/google-maps-a-simple-explanation-on-how-it-detects-traffic-jams-ce6940489c9c>
- Jäger, C. (2016). Epistemic Authority, preemptive reasons, and understanding. *Episteme*, 13(2), 167–185. doi:10.1017/epi.2015.38
- Kelleher, J. D. (2019). *Deep learning*. MIT press.
- Kelly, T. (2005). The epistemic significance of disagreement. *Oxford studies in epistemology*, 1, 167-196.
- Kelly, T. (2013). Disagreement and the Burdens of Judgment. *The epistemology of disagreement: New essays*, 31-53.
- King, N. L. (2012). Disagreement: What’s the problem? Or a good peer is hard to find. *Philosophy and Phenomenological Research*, 85(2), 249-272.
- Lackey, J. A. (2013). Disagreement and belief dependence: Why numbers matter. In *The epistemology of disagreement: New essays* (pp. 243-268). Oxford University Press.
- Lau, J. (2020, 3 de septiembre). Google Maps 101: How AI helps predict traffic and determine Routes. Google. blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/
- Mehta, H., Kanani, P. y Lande, P. (2019). Google maps. *International Journal of Computer Applications*, 178(8), 41-46.
- O’neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Pagin, P. y Marsili, N. (2021 Invierno), “Assertion”, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2021/entries/assertion/>>.
- Rowland, R. (2017). The epistemology of moral disagreement. *Philosophy Compass*, 12(2), e12398. doi: 10.1111/phc3.12398.
- Searle, J. R. (1982). The Chinese room revisited. *Behavioral and brain sciences*, 5(2), 345-348.

Searle, J. R. (2000) *El misterio de la conciencia*. Padios. Madrid.

Williamson, T. (2002). *Knowledge and its Limits*. Oxford University Press, USA.
doi:10.1093/019925656X.001.0001

Zagzebski, L. T. (2012). *Epistemic authority: A theory of trust, authority, and autonomy in belief*. Oxford University Press.