Revista Telemática de Filosofía del Derecho, nº 29, 2026, pp. 187-197

ISSN 1575-7382

¿Qué hay de malo en simular razonar? Un vistazo a algunos problemas sobre IA y razonamiento probatorio

What Is Wrong with Simulating Reasoning? A Look at Some Problems in AI and Evidential Reasoning

Daniel González Lagier*

Resumen: Este texto analiza el impacto de la inteligencia artificial en el razonamiento jurídico probatorio, subrayando la importancia de distinguir entre la comprensión humana real y la mera simulación algorítmica. El autor sostiene que, aunque la tecnología procesa grandes volúmenes de datos mediante una geometría de las palabras, carece de anclaje en la realidad y de capacidad para realizar abducciones genuinas. Se identifican tres limitaciones fundamentales denominadas cegueras: la falta de criterio para determinar la relevancia, la incapacidad de distinguir la causalidad de la correlación y la ausencia de empatía para entender estados mentales. El artículo advierte sobre los riesgos de delegar completamente decisiones judiciales en sistemas opacos, defendiendo en su lugar una inteligencia híbrida donde el jurista supervise críticamente las herramientas digitales.

Abstract: This text analyzes the impact of artificial intelligence on evidential legal reasoning, underscoring the importance of distinguishing between genuine human understanding and mere algorithmic simulation. The author argues that, although technology processes vast volumes of data through a "geometry of words," it lacks both grounding in reality and the capacity to perform genuine abductions. Three fundamental limitations, termed "blindnesses," are identified: the lack of criteria to determine relevance, the inability to distinguish causation from correlation, and the absence of empathy required to understand mental states. The article warns of the risks of delegating judicial decisions to opaque systems, advocating instead for hybrid intelligence where legal professionals critically oversee digital tools.

Palabras clave: inteligencia artificial; abducción; razonamiento probatorio; semántica; simular razonar

Keywords: artificial intelligence; abduction; evidential legal reasoning; semantics; simulating reasoning

Fecha de recepción: 20 de septiembre de 2026

Fecha de aceptación: 09 de noviembre de 2026

* Universidad de Alicante. ORCID iD: https://orcid.org/0000-0003-0295-4206. Correo electrónico: daniel.gonzalez@ua.es.

Ponencia presentada en el XI Seminario Internacional "Proceso y Constitución: Los desafíos del razonamiento probatorio", organizado por la Pontificia Universidad Católica del Perú en Lima, septiembre de 2026. Para una mayor profundización en el tema puede verse González Lagier, D., "¿Loros con toga? Inteligencia artificial, argumentación y prueba de los hechos", en G. Posada y R. Cavani (eds.), *Los desafíos del razonamiento probatorio*, Lima: Palestra, 2026, y *Hechos y máquinas. ¿Puede la inteligencia artificial razonar sobre la prueba?*, Madrid/Barcelona: Marcial Pons, 2026.

1. El desafío de la IA

Luciano Floridi, uno de los pensadores más importantes sobre el impacto de la IA en nuestras vidas, ha advertido de que "El verdadero desafío [de la IA] no es tecnológico, sino filosófico y educativo: aprender a interactuar con sistemas avanzados preservando nuestra capacidad para distinguir entre simulación y realidad, entre comprensión aparente y comprensión genuina, y resistiendo la manipulación peligrosa y desempoderadora"[1]. Si esto es así, uno de los grandes peligros de la IA está en desconocer sus fundamentos, perder actitud crítica hacia ella y delegar nuestra capacidad de decisión en los algoritmos, en lugar de tomarlos sólo como un instrumento. Para los juristas, es especialmente importante mantener el criterio frente a la IA, conociendo sus aptitudes y sus límites para enfrentarse a los problemas jurídicos. Y para los teóricos de la prueba en particular, uno de los grandes interrogantes es la capacidad de la IA para suplir el razonamiento probatorio del juez en sus distintas fases, esto es, determinar hasta qué punto podemos confiar en ella. Para enfrentarnos a este interrogante, me parece que es necesario un enfoque interdisciplinar, que combine filosofía de la mente, epistemología, ciencias cognitivas, teoría de la argumentación y teoría de la prueba. Mi propósito en estas pocas páginas es dar algunas pinceladas para avanzar en esa dirección.

2. El núcleo del razonamiento probatorio

Comencemos por una cuestión de teoría de la prueba. Se ha dicho ya muchas veces que pueden distinguirse tres momentos en el proceso de prueba: (1) el de recolección de las pruebas y de la información, (2) el de valoración de la misma y (3) el de toma de la decisión sobre qué hechos declarar probados[2]. Lo que quizá no se ha dicho tantas veces es que, si bien cada fase tiene sus propias peculiaridades, se trata en todos los casos de un razonamiento que gira alrededor de hipótesis o conjeturas: En la fase de investigación, en primer lugar deben formularse las hipótesis compatibles con la primera información y las primeras pruebas disponibles y, seguidamente, las hipótesis orientan qué nuevas pruebas deben buscarse (a este argumento podemos llamarlo abducción creadora). En la fase de valoración, se selecciona la hipótesis que recibe una mayor confirmación, de acuerdo con los criterios de valoración de la prueba (a este argumento podemos llamarlo abducción justificativa o inferencia a la mejor explicación). En la fase de decisión, se evalúa si el grado de confirmación obtenido por la hipótesis tiene la corroboración que se estima adecuada para la decisión de la que se trata (lo que, en una concepción como la de

[1] Floridi, L., "AI and Semantic Pareidolia: When We See Consciousness Where There Is None", *SSRN Electronic Journal*, 2026, https://doi.org/10.2139/ssrn.5309682

[2] Ferrer Beltrán, J., Vázquez, C., Fernández López, M. y González Lagier, D., *Manual de Razonamiento probatorio*, México: Suprema Corte de Justicia de la Nación, 2022, cap. II.

Jordi Ferrer, sería un silogismo subsuntivo, es decir, una deducción[3], aunque para otros autores, como Diego Dei Vecchi, tiene un irreductible componente valorativo[4]). La estructura del razonamiento abductivo que se construye (o se evalúa) es siempre similar: elementos de juicio, hipótesis y generalizaciones empíricas que hacen la función de enlace entre unos y otras. En *El mito de la inteligencia artificial*, por ejemplo, Erik J. Larson[5] sostuvo que mientras las IA pueden realizar deducciones y en cierto sentido inducciones-generalizadoras, carecen de la habilidad de hacer buenas abducciones, al mismo tiempo que sostiene que la abducción es el modo de pensar genuino de los seres humanos y el único que, en sentido estricto, es ampliativo, creador, en el sentido de que descubre algo nuevo.

3. Razonar y simular razonar

Sigamos con una cuestión de filosofía de la IA. El argumento más usual para negar la capacidad de las IA de hacer abducciones es que, en realidad, las IA no razonan, sino que simulan razonar[6]. Esto es lo que hay detrás de la advertencia de Floridi. Mientras la deducción es formalizable por medio de reglas de inferencia, y las generalizaciones o búsqueda de patrones son el resultado de algoritmos como el descenso de gradiente y la retropropagación (dicho muy básicamente, un sistema algorítmico de ensayo y error para ir ajustando las respuestas de la IA a las respuestas esperadas), los intentos de formalizar la abducción no han dado resultados tan satisfactorios. La abducción es un tipo de argumento muy abierto al conocimiento del mundo, dado el peso en el mismo de las generalizaciones empíricas y de las condiciones de derrotabilidad, y parece estrechamente ligado al sentido común, la posesión de un mapa o representación del mundo, la intuición y la imaginación, facultades de las que la IA carece.

4. ¿Qué quiere decir que las IA sólo simulan razonar?

Cuando se plantea esta cuestión, la respuesta suele venir de la filosofía de la mente y trata de mostrar que las IA operan sin comprensión real de las palabras y conceptos con

[3] Ferrer Beltrán, J., *Prueba sin convicción. Estándares de prueba y debido proceso*, Madrid/Barcelona: Marcial Pons, 2025.

[4] Por ejemplo, Dei Vecchi, D., "Prueba libre, justificación epistémica y el noble sueño de los estándares de prueba", *Revista de Derecho (Valdivia)*, vol. 33, nº 2, 2020, https://doi.org/10.4067/S0718-09502020000200025

[5] Larson, E. J., *El mito de la inteligencia artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos*, Barcelona: Shackleton Books, 2021. Aunque Larson reconoce actualmente que la IA simula abducciones, parece seguir pensando que éstas no son genuinos razonamientos.

[6] Floridi, L., Morley, J., Novelli, C. y Watson, D., "What Kind of Reasoning (if any) is an LLM actually doing? On the Stochastic Nature and Abductive Appearance of Large Language Models", arXiv:2512.10080, 2025/2026, https://doi.org/10.48550/arXiv.2512.10080

los que les preguntamos y con los que responden, esto es, operan sólo en un nivel sintáctico y no semántico[7]. Para entender esta idea, reflexionemos sobre cómo funcionan los modelos de redes neuronales y aprendizaje profundo, por ejemplo un Gran Modelo de Lenguaje. Supongamos que queremos construir uno especializado en Derecho. Le proporcionamos enormes cantidades de textos jurídicos (leyes, sentencias, dictámenes, monografías teóricas,...). El sistema comienza un proceso de "tokenización". Un token es una palabra (o fragmento de palabra) a la que se le asigna un número. Por ejemplo, a "justicia" se le puede asignar el número 45765. A este número se le asigna una especie de "coordenadas" para situar el token en un punto de cierto "espacio virtual" (imaginemos los tokens "flotando" en una enorme habitación). El sistema aprende que estadísticamente "justicia" aparece muchas veces junto con "toga" y nunca o casi nunca junto con "lechuga". Por ello, al token "toga" le asigna unas coordenadas que lo sitúan en el espacio virtual cerca de "justicia" y a "lechuga" unas coordenadas muy lejanas. A la hora de responder a una pregunta sobre la "justicia" la IA elegirá las palabras estadísticamente más cercanas a ésta (siguiendo ciertas reglas de estilo, construcción sintáctica adecuada, etc.), tratando de predecir las palabras más probables del texto que está construyendo. Además, la mayor o menor cercanía de las palabras en ese espacio virtual (la habitación de los tokens flotantes) se va modificando en función de las palabras anteriores y posteriores a "justicia". A esta manera de procesar la información la podemos llamar "geometría de las palabras"[8]. Esta manera de operar no accede al significado de ninguna palabra, no conecta las palabras con la realidad. Para hacerlo, la IA necesitaría tener consciencia fenoménica, esto es, la capacidad de experimentar las sensaciones que la interacción con el mundo produce en nosotros. Es decir, experimentar los colores, el frío y el calor, el dolor y el placer, la solidez de los objetos, su forma, probar a usar los objetos de una u otra manera, sentir indignación, miedo, alegría, etc. El lenguaje y los conceptos se forman en esta interacción con el mundo, y nuestro pensamiento se estructura a partir de esa interacción, no como un proceso cerrado, encapsulado en un espacio sólo lleno de tokens relacionados estadísticamente. Este es un primer aspecto del funcionamiento de las IA que hemos de aprender: la falta de anclaje en la realidad de los símbolos con los que trabaja[9].

[7] Searle, J., *Mente, cerebro y ciencia*, Madrid: Cátedra, 1985.

[8] González Lagier, D., "¿Loros con toga? Inteligencia artificial, argumentación y prueba de los hechos", *op. cit.*

[9] Esta no es, evidentemente, la única concepción sobre el lenguaje y la semántica. Otras teorías sostienen que el significado se agota en las relaciones lógicas o estadísticas que los símbolos mantienen entre sí dentro del propio sistema. Sin embargo, aun cuando se aceptase esta noción de "significado", se podría seguir manteniendo que las IA no tienen un mapa adecuado del mundo (que es lo relevante del asunto), puesto que sus conceptos no conectan con él.

5. ¿Hay algo malo en simular razonar?

Por tanto, cuando una IA parece hacer una abducción, sólo escoge la respuesta que la geometría de las palabras muestra como más adecuada, sin saber lo que está diciendo, a la manera de un loro que imita sonidos componiendo palabras sin saber su significado[10]. Pero, ¿hay algo malo en esto? ¿Es un defecto? Cualquiera que mínimamente haya interactuado con una IA puede darse cuenta de que, desde el punto de vista puramente funcional y del resultado, no parece haber una diferencia manifiesta entre una abducción genuina y una simulación de abducción. Una simulación de abducción realizada por una IA sofisticada puede parecer tan razonable como una genuina abducción. Esto plantea inmediatamente la cuestión de si conceptos como "razonar", "mente" o "inteligencia" no pueden ser extendidos y atribuidos legítimamente a las máquinas, aunque lleguen a sus resultados por vías diferentes (es la sugerencia detrás del test de Turing[11]). Hay una tensión constante entre dar a estos conceptos un significado puramente funcional o un significado que podríamos llamar "humanista" o "antropocéntrico", es decir, teniendo en cuenta la mayor o menor proximidad con el razonamiento humano. No obstante, siendo ésta una cuestión conceptual de cierta relevancia, no es lo más relevante para la discusión sobre el papel de las IA en el razonamiento probatorio. En mi opinión, lo relevante es si esas diferentes maneras de hacer el recorrido hacen que una vía sea más defectuosa, insegura, irracional (entendiendo esta noción desde el punto de vista puramente de la eficacia epistémica) que la otra. Esto es, en el caso de la prueba, si al "razonar" o procesar la información de manera diferente, las IA son mejores o peores para llegar a resultados epistemológicamente correctos.

6. Cuestiones de eficacia epistémica y cuestiones de legitimidad

Cuando nos planteamos la posibilidad de que las IA puedan ocuparse del razonamiento probatorio resulta útil distinguir entre cuestiones de eficacia epistémica y cuestiones de legitimidad, aunque es evidente que están relacionadas (una distinción en parte similar ha sido sugerida por Renzo Cavani, con otra terminología[12]). Las cuestiones de eficacia epistémica tienen que ver con la capacidad de la IA para proporcionar hipótesis correctas (lo que en el caso de la concepción racionalista de la prueba, tiene que

[10] Bender, E. M. y otros, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", en *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, https://doi.org/10.1145/3442188.3445922

[11] Turing, A., "Maquinaria computadora e inteligencia", en A. R. Anderson (ed.), *Controversia sobre mentes y máquinas*, Barcelona: Tusquets, 1984.

[12] Cavani, R., "Prueba e inteligencia artificial. De lo técnico a lo ético", en G. Posada y R. Cavani (eds.), *Los desafíos del razonamiento probatorio*, Lima: Palestra, 2026.

ver con la probabilidad de que sean verdaderas, en el sentido de correspondientes con la realidad). Las cuestiones de legitimidad serían de carácter normativo: dada la opacidad del modo de funcionamiento de las IA, ¿debemos aceptar respuestas que son obtenidas de manera no transparente, no explicables, que impiden, por ejemplo, la contradicción e incluso en buena medida el derecho de defensa? ¿Respuestas de este tipo pueden considerarse una "justificación" de la decisión? Me parece que en los numerosos trabajos que están surgiendo sobre IA en el ámbito de la teoría del Derecho nos hemos centrado más en las cuestiones de legitimidad que en las de eficacia epistémica. Ambas son importantes, pero aquí me ocuparé de las cuestiones de eficacia. Lo que voy a sugerir es que incluso si fuera posible resolver los problemas de legitimidad en el uso de las IA para el razonamiento probatorio, todavía persistirían los de eficacia epistémica.

7. Las tres cegueras

Para ello, veamos una cuestión de teoría del conocimiento. La ausencia de comprensión semántica de las IA conlleva la ausencia de lo que podríamos llamar un "mapa del mundo" adecuado. Un mapa es una representación simplificada de la realidad, que tiene en última instancia una función práctica: orientarse para decidir y actuar. Los seres humanos nos hemos formado un mapa mental del mundo que habitamos, a través de nuestras interacciones con la realidad y de su percepción consciente. Este mapa del mundo, en cierta medida, se ha trasladado al lenguaje. Las IA acceden al mapa del mundo a través del lenguaje (o de información, en general), pero tras un "triple salto": El primer salto, de la realidad a la mente humana; el segundo, de la mente y la comprensión humana al lenguaje (en el caso de los LLM) o a los datos y variables con los que trabaje la IA; y el tercer salto se produce con la pérdida de la semántica a favor de una "geometría de los datos" construida por métodos algorítmicos[13]. Si las IA poseen un mapa del mundo, se trata entonces de un mapa muy pobre, lo que plantea importantes problemas en el razonamiento de las IA. Señalaré tres problemas básicos, que están detrás de otros problemas (como los sesgos estructurales de la propia manera de operar de la IA o la dificultad de distinguir entre casos fáciles y difíciles): (1) la ceguera ante la relevancia, (2) la ceguera ante la causalidad; y (3) la ceguera ante los estados mentales.

[13] González Lagier, D., "¿Loros con toga? Inteligencia artificial, argumentación y prueba de los hechos", *op. cit.* También González Lagier, D., *Hechos y máquinas. ¿Puede la inteligencia artificial razonar sobre la prueba?*, *op. cit.*, cap. III.

8. La ceguera ante la relevancia

La ceguera ante la relevancia tiene que ver con el llamado "problema del marco", que ha sido discutido en la filosofía de la IA casi desde sus inicios[14]. Con esa denominación se hace referencia a un conjunto de problemas relativos a la selección de la información relevante para tomar una decisión. La racionalidad de una decisión parece exigir un examen detallado de cada uno de los aspectos de la situación (del contexto sobre el que se decide) y de cada una de las alternativas de acción y sus consecuencias. Dado que es imposible analizar toda la información, aparecen los problemas de relevancia: (a) evaluar qué información del contexto de decisión es relevante y cuál no; (b) evaluar qué consecuencias de las distintas alternativas de acción son relevantes y cuáles no (y su grado de relevancia); y (c) determinar en qué momento la información relevante de la que se dispone es suficiente para decidir (sin examinar el resto de información). Los humanos resolvemos el problema del marco por medio de atajos y heurísticas (en las que parece que las emociones pueden tener un papel importante). El mapa del mundo y el sentido común nos permiten tomar decisiones previendo rápidamente las consecuencias relevantes de las mismas y detectando los aspectos esenciales del contexto. Pero para la IA la evaluación de la relevancia se limita a aplicar la información que estadísticamente ha resultado relevante en casos anteriores; esto es, para la IA la relevancia es simplemente una cuestión estadística. Esto deja muchas lagunas, pues, como ha señalado Jerry Fodor, "la información pertinente a la solución óptima de un problema abductivo puede provenir, en principio, de cualquier parte en la red de nuestras anteriores convicciones epistémicas"[15]. Es decir, cualquier dato podría acabar convirtiéndose en algo que refuerza o debilita, o incluso impide, el paso de las premisas a la conclusión de la abducción. Esto es así también para las inferencias probatorias, que en este punto no tienen peculiaridades respecto de las abducciones cotidianas. Y si cualquier dato puede ser relevante para la evaluación de la abducción, entonces es imposible que todos los criterios de relevancia sean captados por las operaciones estadísticas de las IA. Dicho de otra manera: La IA tratará todos los casos como casos fáciles, despreciando en ocasiones información que puede ser relevante.

9. La ceguera ante la causalidad

Otra de las "cegueras" que pueden afectar a la eficacia epistémica de las IA es la incapacidad para distinguir entre relaciones de causalidad y meras correlaciones

[14] González Lagier, D., "Las funciones (y disfunciones) epistémicas de las emociones y la responsabilidad emocional de los jueces", *Jueces para la democracia*, nº 115, 2026.

[15] Fodor, J., *La mente no funciona así. Alcance y límites de la psicología computacional*, Madrid: Siglo XXI, 2003, p. 57.

accidentales[16]. Tanto unas como otras son "generalizaciones empíricas", es decir, descripciones de regularidades que se dan en el mundo: por ejemplo, que tras el día llega la noche, que el calor dilata los metales, que ninguna montaña terrestre tiene más de 9000 metros de altura o que muchos trabajos son remunerados. Estas generalizaciones (a las que en teoría de la prueba se hace referencia como leyes científicas y máximas de experiencia) son fundamentales para los razonamientos en materia de prueba y constituyen el enlace entre los elementos de juicio y la hipótesis que se quiere justificar. Pero hay una diferencia importante entre ellas: algunas generalizaciones son relaciones causales y otras son meramente accidentales. Las relaciones causales son regularidades necesarias, es decir, se refieren a algún tipo de relación que no es accidental, sino una "ley", por lo que alimentan contrafácticos ("Si A no hubiera ocurrido, B no hubiera tenido lugar"). Si un metal no se calienta, no se dilatará. Las generalizaciones accidentales, en cambio, no tienen una conexión necesaria entre sus elementos: No podemos afirmar que "Si surgiera una nueva montaña (por ejemplo, por efecto de una erupción volcánica), esta tendrá menos de 9000 metros", puesto que es una mera regularidad accidental, no causal. Distinguir el carácter necesario o no de las generalizaciones empíricas es importante para valorar la solidez de la inferencia probatoria abductiva. Además de su fiabilidad, esto es, si están bien sustentadas en una inducción amplia, es importante valorar la "proyectabilidad" de las generalizaciones[17], esto es, su aplicabilidad a casos nuevos, y ésta es mayor en el caso de leyes causales necesarias y es menor, o incluso inexistente, en el caso de las regularidades accidentales.

El funcionamiento de las IA no les permite distinguir unas de otras, porque se quedan en la mera asociación, limitándose a extraer generalizaciones de una manera estadística, sin plantearse (ni poder detectar) la naturaleza de la relación. Para las IA las asociaciones y generalizaciones no están en el mundo, son simplemente una alta probabilidad de ocurrencia conjunta (o cercana) de palabras u otro tipo de variables en sus bases de datos. Mientras que el ser humano utiliza un modelo mental del mundo que le permite realizar experimentos mentales o "contrafácticos" (imaginando qué sucedería si alteráramos una variable), la IA solo puede predecir el siguiente ítem basándose en su "geometría de datos".

10. La ceguera ante los estados mentales

La tercera ceguera tiene que ver con la falta de empatía. El mapa o modelo del mundo que poseemos los humanos incluye también criterios para explicar y predecir las acciones de los demás. Estas explicaciones y predicciones se basan en la atribución de

[16] Sobre la relevancia de la distinción véase González Lagier, D., *Hechos y máquinas*, *op. cit.*, cap. II. Sobre la incapacidad de la IA para distinguirlas véase el cap. III.

[17] González Lagier, D., *Hechos y máquinas*, *op. cit.*, cap. II.

estados mentales a los otros[18]. A su vez, esta atribución descansa en una analogía: si mi conducta es generada por mis deseos, creencias, emociones, intenciones, etc., la de los demás es probablemente causada por el mismo tipo de hechos psicológicos. A partir de aquí los seres humanos construimos una serie de reglas que correlacionan tipos de situaciones (la muerte de un familiar) con tipos de estados mentales (emociones de tristeza) y tipos de estados mentales (tristeza) con tipos de conducta (llanto, retraimiento, etc.). A esto se le ha llamado "psicología popular" o "psicología de sentido común". Sin ella, la vida social y política, tal como la entendemos, sería imposible. Para la adquisición de esta psicología de sentido común es necesaria una facultad de la que, obviamente, las IA carecen en sentido estricto: La empatía. La empatía (dejando de lado aproximaciones más románticas, como "sentir lo que el otro siente" o "vibrar con el otro") puede definirse como la capacidad de proyectarnos en la situación del otro para discernir qué es lo que decidiría, pensaría o sentiría yo si estuviera en su lugar. Obviamente, para esto es necesario haber sentido emociones, dolores, haber tenido creencias, haberse formado intenciones, etc. Esto es, tener una vida mental propia. Sin ésta, difícilmente se pueden realizar evaluaciones acerca del grado de intensidad de una emoción, o del dolor, o la naturaleza del dolor psíquico, etc., por lo que la empatía es fundamental para la prueba de los estados mentales (esto es, para aplicar correctamente estas reglas de la psicología de sentido común).

Las IA pueden operar con un análogo funcional de la empatía, al correlacionar enormes cantidades de datos lingüísticos (pueden relacionar, por ejemplo, la palabra "humillación" con la palabra "venganza"). Pero este mecanismo no permite ajustar individualizadamente las atribuciones de estados mentales. Puede permitir una hipótesis acerca de lo que un hombre medio habría querido, hecho o sentido en una determinada situación, pero sería insensible a las particularidades estadísticamente poco relevantes del caso que, sin embargo, pueden determinar la respuesta mental de un sujeto en particular. En Derecho penal, por ejemplo, se discute si debe adoptarse una perspectiva objetiva o subjetiva en la atribución de estados mentales. Por ejemplo, la circunstancia modificativa de la responsabilidad de miedo insuperable: ¿debe considerarse en relación con el miedo que sentiría la "persona media" en esa situación o con el que realmente sintió el sujeto evaluado? Se diría que las IA sólo pueden reconocer las emociones a partir de sus patrones típicos, por lo que no podrían asumir la perspectiva subjetiva.

11. La inteligencia híbrida

Las anteriores consideraciones, en mi opinión, no implican que las IA no puedan ser epistémicamente eficaces resolviendo un buen número de casos. Las tres cegueras no

[18] González Lagier, D., "Filosofía de la mente y prueba de los estados mentales: Defensa de los criterios de sentido común", *Quaestio facti. Revista Internacional sobre Razonamiento Probatorio*, nº 3, 2022.

son totales, sino parciales. Las IA "ven" a través de una nebulosa, son miopes. Las dificultades que he señalado son algunos de los aspectos que deberían tenerse en cuenta y controlarse, en lugar de asumir acríticamente las respuestas de la IA. Ciertamente, las enormes capacidades estadísticas de las máquinas y la enorme cantidad de datos que manejan pueden suplir buena parte de estas deficiencias. Pero una dificultad que estará siempre presente, dada la ausencia de genuina comprensión, será la de detectar en qué casos se plantea algo novedoso: Una excepción estadísticamente poco frecuente a la aplicación de una generalización, una combinación determinada de pesos probatorios de los distintos criterios de valoración atípica, la necesidad de entender matices en algún estado mental, la necesidad de estimar la credibilidad de un testigo evasivo, la necesidad de excluir algún sesgo de entrenamiento o de funcionamiento del sistema,... Por supuesto, también los humanos cometen errores en estos casos, lo que parece llevarnos a un dilema: ¿cuál es más fiable, la inteligencia humana o la inteligencia artificial?

Ciertos estudios empíricos muestran que, aunque las IA pueden ser mejores y más eficaces en numerosas circunstancias, bajo determinadas condiciones pueden alcanzarse mejores resultados en tareas cognitivas (especialmente si tienen un componente creativo y valorativo) mediante sistemas colaborativos en los que el experto humano supervisa de forma crítica las sugerencias del modelo —superando así tanto al humano aislado como a la máquina autónoma—, si bien advierten de que esto sólo se logra si se evita el sesgo de anclaje de los jueces humanos frente a los resultados de las máquinas, es decir, si no se produce una deferencia automática hacia sus respuestas[19]. Esto es lo que creo que deberíamos trazarnos como objetivo a propósito del problema de la eficacia epistémica del razonamiento probatorio, para superar el desafío señalado por Floridi sin renunciar a la potencia de esta nueva tecnología.

12. Bibliografía

Bender, E. M. y otros, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", en *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021, https://doi.org/10.1145/3442188.3445922

Cavani, R., "Prueba e inteligencia artificial. De lo técnico a lo ético", en G. Posada y R. Cavani (eds.), *Los desafíos del razonamiento probatorio*, Lima: Palestra, 2026.

[19] Vaccaro, M., Almaatouq, A. y Malone, T., "When combinations of humans and AI are useful: A systematic review and meta-analysis", *Nature Human Behaviour*, vol. 8, nº 12, 2024.

Dei Vecchi, D., "Prueba libre, justificación epistémica y el noble sueño de los estándares de prueba", *Revista de Derecho (Valdivia)*, vol. 33, nº 2, 2020, https://doi.org/10.4067/S0718-09502020000200025

Ferrer Beltrán, J., *Prueba sin convicción. Estándares de prueba y debido proceso*, Madrid/Barcelona: Marcial Pons, 2025.

Ferrer Beltrán, J., Vázquez, C., Fernández López, M. y González Lagier, D., *Manual de Razonamiento probatorio*, México: Suprema Corte de Justicia de la Nación, 2022, cap. II.

Floridi, L., "AI and Semantic Pareidolia: When We See Consciousness Where There Is None", *SSRN Electronic Journal*, 2026, https://doi.org/10.2139/ssrn.5309682

Floridi, L., Morley, J., Novelli, C. y Watson, D., "What Kind of Reasoning (if any) is an LLM actually doing? On the Stochastic Nature and Abductive Appearance of Large Language Models", arXiv:2512.10080, 2025/2026, https://doi.org/10.48550/arXiv.2512.10080

Fodor, J., *La mente no funciona así. Alcance y límites de la psicología computacional*, Madrid: Siglo XXI, 2003.

González Lagier, D., "Filosofía de la mente y prueba de los estados mentales: Defensa de los criterios de sentido común", *Quaestio facti. Revista Internacional sobre Razonamiento Probatorio*, nº 3, 2022.

González Lagier, D., "¿Loros con toga? Inteligencia artificial, argumentación y prueba de los hechos", en G. Posada y R. Cavani (eds.), *Los desafíos del razonamiento probatorio*, Lima: Palestra, 2026.

González Lagier, D., "Las funciones (y disfunciones) epistémicas de las emociones y la responsabilidad emocional de los jueces", *Jueces para la democracia*, nº 115, 2026.

González Lagier, D., *Hechos y máquinas. ¿Puede la inteligencia artificial razonar sobre la prueba?*, Madrid/Barcelona: Marcial Pons, 2026.

Larson, E. J., *El mito de la inteligencia artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos*, Barcelona: Shackleton Books, 2021.

Searle, J., *Mente, cerebro y ciencia*, Madrid: Cátedra, 1985.

Turing, A., "Maquinaria computadora e inteligencia", en A. R. Anderson (ed.), *Controversia sobre mentes y máquinas*, Barcelona: Tusquets, 1984.

Vaccaro, M., Almaatouq, A. y Malone, T., "When combinations of humans and AI are useful: A systematic review and meta-analysis", *Nature Human Behaviour*, vol. 8, nº 12, 2024.