Sergio Martén Saborío

La IA no tiene modelo de mundo: Modelos de mundo y la escalera de la causalidad

Nota: *Una versión preliminar de este texto fue presentada el 26 de diciembre de 2025 en la conferencia de clausura organizada por la Escuela de Filosofía de la Universidad de Costa Rica. Esta publicación se presenta en conjunto con el texto de Jethro Masís, quien también expuso en dicha conferencia. Ambos textos buscan poner en diálogo El regreso de los modelos de mundo en la inteligencia artificial: discusiones filosóficas.*

La emergencia en el espacio de discusión pública de los problemas sociopolíticos y epistemológicos relacionados con la IA ha beneficiado a la filosofía en dos sentidos: primero, ha funcionado para evidenciar la importancia del análisis conceptual con el nivel de detalle al que estamos acostumbrados, aplicado a otras disciplinas. Basta con ver publicaciones académicas y no académicas dedicadas a la inteligencia artificial para darse cuenta de la cantidad de sesgos epistemológicos, y hasta ontológicos, que para las personas formadas en filosofía son evidentes. Esto nos abre un claro espacio de diálogo con esas otras disciplinas, en el que ya no es tan difuso el aporte del que muchas veces son escépticos.

Segundo, ha fungido también como un lugar de acuerdo *a lo interno* de la filosofía. Independientemente de si la metodología es fenomenológica, crítica o analítica (asumiendo que estas distinciones son claras) parece haber un consenso claro de que la inteligencia artificial no es inteligente del todo, como a los dueños de las corporaciones encargadas de su desarrollo les gustaría hacernos creer. Esto permite que el debate se concentre más en contra de esa presuposición, ahora tan esparcida en la cultura, y menos en debates a lo interno que no llevan a conclusiones muy relevantes a lo externo. Ya sea que el argumento se base en el ser-en-el-mundo o en la lógica de la justificación de la epistemología anglosajona. El argumento va a mostrar la ingenuidad de la postura intuitiva y lega de que la inteligencia artificial ya está al nivel de la inteligencia humana y que pronto será, incluso, superior.

Esta conferencia de clausura busca ejemplificar lo anterior ofreciendo un argumento sobre un aspecto particular de la inteligencia humana que la inteligencia artificial no ha estado cerca de lograr imitar: el modelo de mundo. El argumento central será el siguiente: (1) El entendimiento causal es condición necesaria para la posesión de un modelo de mundo. Además, (2) la inteligencia artificial (GOFAI, RNNs y la IA generativa, que en términos generales son todos los tipos de IA que tenemos hasta el momento) no tiene entendimiento causal. Por lo tanto, la IA (al menos lo que entendemos actualmente por IA) no posee un modelo de mundo. Como corolario a este argumento, se sostiene que los modelos de mundo son una parte fundamental de la inteligencia, por lo que, sin estos, no hay inteligencia del todo.

Con esta finalidad, iniciaré dando algunas definiciones de lo que se entiende por los conceptos de modelo, modelo de mundo, habilidades e inteligencia. Luego pasaré a discutir los tres niveles de la escalera causal (Pearl & Mackenzie

Rev. Filosofía Univ. Costa Rica, LXV (171), 187-193, Enero - Abril 2026 / ISSN: 0034-8252 / EISSN: 2215-5589

2018) y por qué esta es importante para la posesión de modelos de mundo. Por último, evidenciaré ejemplos de la literatura académica y no académica, y las definiciones superficiales que asumen de *inteligencia* y *modelo de mundo* para mostrar que una consideración más detallada de estos conceptos revela la ingenuidad de sus posturas y la lejanía en grados, entre la *inteligencia* artificial y la inteligencia humana.

1. Modelos y modelos de mundo

Hay diferentes alternativas para la definición de lo que se entiende por modelos dependiendo del grado de constreñimiento que se requiera. Por ejemplo, una definición general de modelo, pero que lo constriñe a lo contrafáctico y a cierto nivel de razonamiento a pesar de su simpleza, es la que ofrece Williamson (2017): un modelo de algo es un ejemplo hipotético de ese algo. La palabra *hipotético* está haciendo un enorme trabajo en esta definición conceptual: lo hipotético es lo abductivo, lo que se supone sin estar ahí, pero que permite interactuar con lo que sí está ahí. Si la definición no incluyera este término, el modelo sería un *token* de un *type*: un ejemplar de un caso general. Sin embargo, al ser un *ejemplo hipotético*, debe ser necesariamente abstracto: debe ser más ben un *type*, y el caso o los casos particulares de lo que ejemplifica serían sus *tokens*. En otras palabras, un modelo es siempre una simplificación que se da mediante las capacidades de razonamiento de la abstracción y de la abducción. El modelo de la gravedad de Newton es un ejemplo hipotético de un tipo de interacción entre dos masas, mientras que la interacción entre Júpiter y Europa sería un *token* de ese ejemplo hipotético.

Este modelo de la gravedad de Newton describe adecuadamente una gran cantidad de interacciones entre masas: es un ejemplo hipotético de la totalidad de las interacciones gravitacionales entre masas, lo cual se hace más claro con el modelo espacio-temporal de la teoría de la relatividad general de Einstein; ya que permite atrapar más fenómenos que incluiríamos dentro del conjunto de la totalidad de las interacciones gravitacionales entre masas. Más aún, esta diferencia entre la gravedad newtoniana y la de la relatividad general evidencia la diferencia en grados de adecuación que se puede atribuir a los modelos: un modelo es tanto más adecuado, cuanto más consistentemente permite lograr un fin.

Importante es notar, que el modelo y su adecuación no se tienen que medir necesariamente por su *veracidad* (Pfister 2025), lo cual implicaría una serie de problemas definitorios ontológicos sobre los que cuesta mucho conseguir acuerdo. Más bien tienen que ver con una cuestión teleológica: si el modelo permite alcanzar un fin, funciona. Cuanto más consistentemente permite alcanzar ese fin, mejor funciona. El modelo de la gravedad de Newton tiene un alto nivel de adecuación en tanto permite lograr consistentemente muchos fines (predecir la ubicación de un cuerpo celeste dadas ciertas condiciones iniciales, predecir la velocidad de objetos masivos dado que se conozcan ciertas propiedades de los mismos, etc.). El modelo de la gravedad de Einstein podría asumirse como más adecuado, en tanto permite más consistencia al incluir más fenómenos en su capacidad predictiva (como el funcionamiento de los agujeros negros, el cual no cabe dentro del modelo newtoniano). Sin embargo, esto no hace al modelo de Newton inútil, ya que es más simple y permite cumplir ciertos fines con suficiente consistencia y con mayor facilidad. Qué tan adecuado sea cada uno de ellos, depende de esos fines (entre los que se puede incluir la eficiencia y diferentes niveles de precisión). Es decir, desde esta concepción, la veracidad y la falsedad de los modelos no es relevante, sino su capacidad de producir resultados con un nivel significativo de consistencia.

Otra definición de modelo puede resultar útil por constreñir menos, en tanto que elimina la necesidad de lo *hipotético*: es una representación de un sistema. Una representación, a su vez, es una versión incompleta e indirecta de ese sistema. Al hablar de sistema, nos referimos a la realidad externa, de la cual dependen nuestras percepciones, pero que no atrapan *directamente* (Pfister 2025).

Esta postura permite que llamemos modelo a la forma en que una bacteria consigue nutrientes. La quimiotaxis es el movimiento de una bacteria hacia fuentes de alimentación o lejos

de sustancias peligrosas. Es un proceso que está determinado por los receptores de gradientes químicos presentes en estos organismos (que representan el sistema en el que se encuentran realmente esas fuentes de alimentación o sustancias peligrosas). Cuando hay una concentración de fuentes alimenticias, las bacterias son capaces de detectarlas e inician su movimiento (mediante flagelos) en la dirección de mayor concentración química. Lo contrario en caso de sustancias dañinas (Stock & Baker 2009). El ejemplo no muestra ninguna capacidad de pensamiento contrafáctico o razonamiento hipotético de parte de las bacterias, pero sí se puede llamar *modelo* a la representación (la versión incompleta, por basarse solo en lo químico, e indirecta, por haber intermediación de receptores de gradientes químicos del sistema) por la que se mueven las bacterias.

Para los propósitos de esta presentación, prefiero la definición que constriñe menos el sentido de *modelo*, pues no lo limita al Ser Humano y, de todos modos, para eso existe el segundo concepto que se definirá a continuación: el de modelo de mundo.

Un modelo de mundo es «la totalidad de todo el conocimiento, i.e. de todas las habilidades (skills) y todo el conocimiento no orientado hacia un fin, tal como el conocimiento sobre estados particulares de un mundo» (Pfister 2025, 20). A su vez, una habilidad consiste en la capacidad de lograr una meta específica bajo condiciones específicas conocidas. En este sentido, la bacteria E. coli tiene la habilidad de acercarse eficientemente a fuentes de nutrición y alejarse eficientemente de sustancias dañinas.

Tomada como tal, esta definición de modelo de mundo no evitaría pensar en la IA actual (y desde la llamada good old-fashioned AI) como una entidad que posea un modelo de mundo. Por ello, propongo que realmente no "tienen" modelos de mundo, ni siquiera simplemente modelos. Ellos *son* modelos, creados por seres humanos en este caso. Son habilidades *de* los seres humanos, que son quienes los han modelado y para cuyos fines funcionan. Al no tener fines propios, no pueden *tener* tampoco modelos. Para poder argumentar lo anterior, entonces propongo añadir un elemento a la definición de modelo de mundo:

> Un modelo de mundo postulado se puede resumir como un espacio de estados Φ y un mapeo $\varphi : X \rightarrow \Phi$ que asocia cada entrada con algún estado $\varphi(x) \in \Phi$ (...). Un modelo de mundo describe la estructura de los estados *implícita en los datos*. (Vafa et al. 2025, 2)

Es decir, un modelo de mundo está preocupado por el mapeo de procesos que subyacen a la razón de que el estado de cosas sea el que es. O, en otras palabras, el modelo de mundo es una estructura implícita que explica por qué las cosas son como son. La E. coli, por ejemplo, tendría una habilidad para la cual necesita de un modelo, pero habría que discutir si tiene del todo un modelo de mundo, ya que no existiría ese mapeo o esa estructura implícita en los datos. Las respuestas de la bacteria serían inmediatas. Para aclarar mejor esta noción, será necesario primero pasar por el concepto de la escalera causal propuesto por Pearl (Pearl & Mackenzie 2018).

Antes, sin embargo, me gustaría detenerme en la inteligencia: otro concepto que no aplica para las irónicamente llamadas inteligencias artificiales (por lo menos en la actualidad). Se puede utilizar lo establecido sobre las habilidades para decir que la inteligencia consiste en la capacidad de crear habilidades nuevas que permitan lograr metas bajo condiciones parcialmente desconocidas (Pfister 2025).

Algo sería inteligente, entonces, en la medida en que pueda desarrollar nuevos mecanismos para lograr una meta que no se encontraban entre los que ya poseía y utilizando condiciones no especificadas previamente. Imaginemos que la E. coli absorbe una substancia que inhibe totalmente sus receptores químicos. ¿Sería capaz de conseguir nutrientes y alejarse del peligro en ese caso, desarrollando ella misma (no a través de procesos evolutivos) nuevos mecanismos para lograrlo? Probablemente no, pues no posee un modelo de mundo que le permita comprender lo subyacente a su consecución de alimentos, sino que funciona únicamente mediante esa habilidad de manera inmediata. En este sentido, tampoco sería inteligente. Lo mismo ocurriría con la IA (sea GOFAI, redes neuronales o transformadores) que solo funciona dentro de los constreñimientos del modelo que es, pero que no sería

capaz de desarrollar una nueva habilidad bajo condiciones desconocidas (algo se puede discutir al respecto con modelos basados en refuerzo, pero no hay espacio para hacerlo ahora).

2. La escalera causal y el pensamiento contrafáctico

Si la inteligencia consiste en el desarrollo de habilidades bajo condiciones parcialmente desconocidas, entonces se tiene que poder llegar a conclusiones sobre acciones sin haberlas ejecutado todavía. Es decir, se requiere de pensar sobre aquello que no es, o que no ha llegado a ser. El pensamiento contrafáctico, el análisis de qué hubiera pasado si algo que ocurrió no hubiese ocurrido (o viceversa), o en general, el razonamiento abductivo, es lo que distingue al ser humano de otros agentes, tanto naturales como artificiales, y es lo que Pearl y Mackenzie (2018) ubican en el peldaño más alto de la escalera causal.

Esta escalera describe los distintos tipos de manejo de la causalidad, desde los más básicos y concretos, hasta los más complejos y abstractos. Consiste en tres peldaños: (1) el del nivel de la asociación, (2) el nivel de la intervención y (3) el nivel de los contrafactuales.

En el primer peldaño, los autores ubican a la mayoría de los animales y a las inteligencias artificiales. Es el peldaño de la observación y la correlación resultante: el nivel de los datos. Se observa que cuando hay rayería, suele verse acompañada de nubosidad y lluvias. Hay una correlación entre estos elementos que podemos encontrar solo con observar. Sin embargo, no está claro si la rayería causa la lluvia, si la lluvia causa la rayería o si ambos son causados por un tercer elemento invisible o no tomado en cuenta. Los escarabajos Julodimorpha bakewelli intentan copular con botellas de vidrio de una tonalidad específica y con características específicas en su superficie (Gwynne & Rentz 1983). La hipótesis es que esta tonalidad y este tipo de superficie son los proxis de la presencia de una hembra de la especie, y es el modelo que por cientos de miles de años ha funcionado para un apareamiento exitoso, hasta que aparecieron las botellas de vidrio con exactamente estas características. En este sentido, lo que existía para los escarabajos era únicamente una asociación basada en datos: la presencia de tales propiedades en el entorno se da cuando hay una hembra de la especie en el entorno. Qué causa esas propiedades, o qué causa que la reproducción sea exitosa, son cuestiones que no están en ese modelo (y por lo tanto no habría un modelo de mundo *stricto sensu*).

La inteligencia artificial basada en datos funciona exactamente de la misma forma. Mientras que sus resultados no dejan de ser sorprendentes, pues es impresionante la cantidad de habilidades que se pueden desprender de encontrar correlaciones en los datos, no por eso se alejan del peldaño de la asociación para llegar a uno más alto. De hecho, ese es el que muchos autores críticos de la dirección del desarrollo de la IA señalan como el motivo de su estancamiento en años recientes: la asunción de que existe una ley de escalada en los modelos de redes neuronales de lenguaje, según la cual, cuantos más datos se le alimente en el entrenamiento, tanto mejor será su desempeño (Kaplan et al. 2020). No obstante, esta *ley* de escalada funciona solo para el proxy de capacidad de predicción del siguiente *token* de palabra. Mientras que muchos encuentran en esto la inteligencia, la definición que aquí hemos dado muestran que está muy lejos de ello. Está dos peldaños por debajo de ello, para ser precisos. A fin de cuentas, lo que hace la IA independientemente de la cantidad de datos que se le haya alimentado durante el entrenamiento, no deja de ser exactamente lo mismo. Esperar que repentinamente emerja algo así como la inteligencia fue la hipótesis en la que muchos se basaron, pero que se basa en una definición completamente superficial de la inteligencia como tal (mostrando aquello de lo que hablé al inicio de esta presentación: que las personas dedicadas a la filosofía encontrarían casi intuitivamente sospechosa una definición así, desde cualquier rama y tradición que se le mire).

El segundo peldaño propuesto por Pearl y Mackenzie (2018) es el de la comprensión de la causalidad al nivel de la intervención. En este se encuentran algunos pocos animales, y los bebes humanos. Hace referencia a preguntas del tipo:

¿qué pasa con X si hago Y? Por ejemplo: ¿Qué ocurre con mi dolor de cabeza si tomo esta pastilla? (Bareinboim et al. 2022). Hay un conjunto de datos conocidos, pero se busca la formación de nuevos datos para formar estructuras nuevas. Las intervenciones «por definición rompen las reglas del entorno en el que se entrenó a la máquina» (Pearl & Mackenzie 2018, 37). O sea, se quiere llegar a encontrar un patrón en un caso que no está entre los tipos de datos de entrenamiento. Por eso es que las IAs actuales no llegan a este peldaño de la escalera y, por lo tanto, tampoco al tercero y último.

El peldaño final es el del nivel de lo contrafactual, y la actividad asociada a este es el imaginar: ¿qué hubiera pasado si hubiera hecho X en lugar de Y? (Bareinboim et al. 2022). Este nivel es, al menos por el momento, aunque no parece que pronto vaya a haber un cambio al respecto, el propio de los seres humanos. El pensamiento del nivel de la asociación y el de lo contrafactual casi que se oponen, ya que aquel se basa en hechos, mientras este se basa en lo que por definición no son hechos o en la negación de ciertos hechos. Incluso la ciencia participa constantemente de lo contrafactual o permite desarrollar con mayor complejidad este nivel de pensamiento, pues da las herramientas para saber con certeza, a través de leyes naturales, por ejemplo, qué hubiera ocurrido si los hechos hubieran sido otros. Los modelos de mundo en su definición constreñida se desarrollan gracias al pensamiento contrafactual: sin él, no participaríamos de ellos. Como veremos más adelante, cuando hablemos de los supuestos modelos de mundo de modelos de IA generativa como Sora; la falta de capacidad contrafactual es justamente lo que les impide tener un modelo de mundo, mientras que la falta de comprensión causal *interventiva* les impide pasar al segundo peldaño de la escalera causal.

Para terminar antes con esta sección, falta considerar lo que proponen Pearl y Mackenzie (2018): a través de la inferencia causal, que es la formalización matemática y diagramática de las relaciones causales, de modo que se pueda extraer un modelo causal a partir de datos (es decir, no quedarse con correlaciones, sino encontrar hechos sobre relaciones causales a partir de datos), es como se podría llevar a la IA a un nuevo nivel en la escalera causal. O sea, si se implementa la inferencia causal desde la base de los modelos de IA, podría, ahora sí, hablarse de inteligencia artificial verdadera. Ha habido avances en esta área, pero todavía nada concluyente. Queda observar qué ocurre en un futuro. De todos modos, este resultado, el de la verdadera inteligencia artificial parece estar lejos, especialmente si la industria continúa apostando por presupuestos metafísicos, epistemológicos e incluso neurológicos que no se sostienen ante el mínimo escrutinio.

3. IA y modelos de mundo

La industria y muchos de los investigadores asociados a la industria de la IA han encontrado en la inteligencia artificial generativa, especialmente la generativa de videos, un productor de modelos de mundo (Zhu et al. 2025, Brooks et al. 2024). Se habla de modelos como Sora como *simuladores* de mundos, también, en la medida en que, al haber sido entrenados con cuantiosísimos datos en forma de video, son capaces de representar mundos que no existen con leyes físicas que los acompañan. Sin embargo, esto confunde tanto lo que es un simulador, como lo que es un modelo de mundo.

En muchos casos los videos producidos por Sora parezcan seguir ciertas leyes naturales, se debe a que están entrenados con datos que muestran esas leyes en funcionamiento, y los datos tienden a una norma que es la representada por el modelo. Esto no implica, de ningún modo, una comprensión de fondo o un modelo de mundo de fondo que haya emergido, como por arte de magia, a punta de correlaciones y meras asociaciones.

Esto no es solo una opinión: fue revelado de modo experimental en un artículo que propuso investigar si un modelo al que no se le había dado previamente la función gravitacional de Newton era capaz de sustraer esa misma función o alguna similar de un conjunto de datos sobre las órbitas de los planetas en el sistema solar (Vafa et al. 2025). El modelo fue capaz de predecir con mucha precisión el lugar en la órbita que ocuparía el planeta en momentos futuros, que

era justamente para lo que había sido construido y entrenado. Sin embargo, al inspeccionar la función gravitacional sobre la que estas predicciones se basaban, detectaron que se trataba de un conjunto de operaciones sin sentido y que ni siquiera tomaba en cuenta una de las dos masas en la interacción gravitacional.

Es decir, las predicciones, lo que se puede obtener mediante mera asociación, resultaron correctas con un acercamiento plenamente basado en datos. De allí no emergió, no obstante, ningún entendimiento de por qué los datos se correlacionan de la forma en la que lo hacen, de modo que, dados nuevos parámetros, no podría ni de cerca hacer una predicción acertada (solo si se basa en condiciones en las que ha sido entrenado). En otras palabras, y según las definiciones brindadas en la primera y segunda parte, estos modelos, ni son inteligentes, pues no son capaces de desarrollar de manera precisa una habilidad nueva bajo condiciones parcialmente desconocidas (en este caso, se conocen los datos de las ubicaciones de los planetas en diferentes momentos, pero no la función gravitacional ni la fuerza de atracción gravitacional que opera sobre ellos), ni tampoco poseen un modelo de mundo.

Si se analizan de cerca los *mundos* que aparecen en los videos generados por modelos como Sora, se podrá notar que existen muchos errores de continuidad y de funcionamiento de leyes naturales y de muchos otros tipos que no se siguen consistentemente. Aparecerán piernas que antes no existían; se atravesarán dos personas caminando en direcciones opuestas; la ropa que era de un tipo dejará de serlo en cuanto la *cámara* la desenfoque y la enfoque nuevamente. Las personas caminarán *pegadas* al suelo solo por el hecho de que, en efecto, en la vastísima mayoría de videos de personas caminando, lo están haciendo sobre el suelo y no en el cielo o alguna otra sustancia. Pero el suelo se mueve de maneras extrañas, como si tuviera vida propia, y las direcciones de las personas que no están enfocadas parecen no tener sentido. Basta con que aparezca una condición no previamente dada en los datos de entrenamiento para que el video generado resulte en un sinsentido absoluto: la señal de falta de inteligencia y, por supuesto, de un modelo de mundo.

En conclusión, se confirma el argumento propuesto inicialmente con el detalle explicativo a lo largo de la presentación: (1) como el entendimiento causal se mostró como condición necesaria para la posesión de un modelo de mundo y (2) la inteligencia artificial no tiene entendimiento causal por lo que se acaba de analizar en la última sección, (3) se puede concluir que la IA (al menos lo que entendemos actualmente por IA) no posee del todo un modelo de mundo, y aquello que construyen modelos como Sora no son más que curvas ajustadas para representar una normalidad que no es real en el sistema del que se extrae. Esto solo evidencia, ya sea la manipulación de la industria de la IA para hacer parecer que, en efecto, estos modelos son inteligentes y generan modelos de mundo, o, cuando menos, la ingenuidad conceptual con la que se acercan a conceptos epistémicos que terminan produciendo inversiones billonarias y las subsecuentes burbujas y, crisis sociales y económicas que se están viendo en días recientes. O, tal vez incluso, una combinación de ambas, que no sería nada improbable.

Referencias

Bareinboim, Elias, Juan D. Correa, Duligur Ibeling, and Thomas Icard. 2022. «On Pearl's hierarchy and the foundations of causal inference ». En *Probabilistic and Causal Inference: The Works of Judea Pearl*, ACM: 507-556.

Brooks, Peebles, et al.. 2024. *Video Generation Models as World Simulators*. OpenAI. https://openai.com/research/video-generation-models-as-world-simulators.

Gwynne, Darryl T., and David CF Rentz. «Beetles on the bottle: male buprestids mistake stubbies for females (Coleoptera)». *Australian Journal of Entomology* 22, no. 1 (1983): 79-80.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. «Scaling laws for neural language models». *arXiv preprint arXiv:2001.08361*. https://arxiv.org/pdf/2001.08361/1000

Pearl, Judea, y Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Pfister, Rolf. 2025. «A Representationalist, Functionalist and Naturalistic Conception of Intelligence as a Foundation for AGI». *arXiv preprint arXiv:2503.07600*. https://doi.org/10.48550/ARXIV.2503.07600.

Stock, Jeffrey B., y Michael D. Baker. 2009. «Chemotaxis». En *Encyclopedia of Microbiology, Third Edition*, 71–78. Elsevier. https://doi.org/10.1016/B978-012373944-5.00068-7.

Vafa, Keyon, Peter G. Chang, Ashesh Rambachan, and Sendhil Mullainathan. 2025. «What has a foundation model found? using inductive bias to probe for world models». *arXiv preprint arXiv:2507.06952*. https://arxiv.org/abs/2507.06952

Williamson, Timothy. 2017. «Model-Building in Philosophy». En *Philosophy's Future*, editado por Russell Blackford y Damien Broderick, 159–171. 1a ed. Wiley. https://doi.org/10.1002/9781119210115.ch12.

Zhu, Zheng, Xiaofeng Wang, Wangbo Zhao, Chen Min, Bohan Li, Nianchen Deng, Min Dou et al. 2025. «Is sora a world simulator? a comprehensive survey on general world models and beyond». *arXiv preprint arXiv:2405.03520*. https://arxiv.org/abs/2405.03520

Sergio Martén Saborío (sergio.marten@ucr.ac.cr) es profesor de la Sección de Filosofía y Pensamiento de la Escuela de Estudios Generales de la Universidad de Costa Rica, San José, Costa Rica. Máster en Filosofía. Su texto más reciente, *El 'arte' de la IA generativa y la matriz ética*, se encuentra en proceso de publicación.

Recibido 3 de diciembre, 2025.
Aprobado: 5 de diciembre, 2025.
DOI: 10.15517/revfil.2026.1799