Capítulo 7

Identidad y reproducción: El concepto de originalidad en el contexto de las IA predictoras de texto

Daren López Méndez[1]

https://doi.org/10.61728/AE20253776

[1] Universidad Autónoma de Zacatecas.

Entender la probabilidad como una herramienta para pensar mejor la escritura fue, hasta la aparición de las IA predictivas, un sinsentido. La noción de lo textual se conformó alrededor de un corpus de ideas que dictaron que esto estaba fuera de todo lo computable. Los avances recientes de las IA, sin embargo, dan pautas para afirmar que la naturaleza de la predicción depende de un arte combinatorio complejo que produce un texto similar al de un humano (si, por supuesto, la conversación se centra en lo textual). El problema se infiere al inquirir los límites de la probabilidad; la línea entre lo que un humano quiere decir y lo que puede es débil para una máquina (Braga y Logan, 2017). Cuando se pregunta hasta qué punto una predicción desde el estilo de un autor puede imitarle y resultar en un texto como cualquier otro que este haya escrito antes, se asume que el estilo está dado dentro de campos semánticos concretos. Sobre la estela de esa cuestión, es lícito dudar: ¿es el estilo reproducible? Y, si así lo fuera: ¿puede la IA predictiva reemplazar la capacidad creativa del autor y, de esta manera, suplir su identidad textual? El propósito de la ponencia pretende clarificar la pregunta, disolver la radicalidad de su suposición y demostrar que las IA predictivas solo acentúan el estilo del autor que las utiliza.

Reconocer las pautas internas de una máquina no interesa a la propuesta si el resultado de su mecanismo es el núcleo de la problematización: el texto. Así pues, ¿qué es necesario para hacerlo similar al de un humano? Esa es la cuestión primera a abordar. Cierto es, como lo menciona Margaret Boden (2006) en *Mind as a Machine*, que no hay parámetros discriminatorios para tal tarea si la máquina no aprehende a la perfección las veces que, por ejemplo, una palabra acompaña a otra. Por supuesto, la suma de todos los elementos característicos de los inputs será lo que resulte, pero definir cuáles priman sobre otros es imperativo. Parece que la identidad de un texto es un proceso artificioso, y si esto se reafirma, se puede preguntar, de este modo, ¿es que cualquier IA que pudiera aprender, combinar y deducir la frecuencia de aparición de dichos datos sería capaz de ser el autor mismo?

Lo que la literatura sobre el asunto indica es que las obras resultantes son (al menos para el estudio realizado este año por Brian Porter y Eduard Machery (2024) concerniente a una IA generativa) indistinguibles de

aquellas hechas por humanos, a pesar de que estos (los sujetos estudiados) pretendieron que podían siempre saber si un texto era artificial o no. Es fácil saber por qué: el output humano en las IA generativas no difiere en su resultado del artificial porque este se nutre del primero; no así en las IA predictivas. Aunque en otros experimentos los sujetos estudiados mostraron ligera aversión a los textos artificiales, (estos fueron realizados en versiones antiguas de Chat-Gpt; Chat Gpt 2, para ser específicos) las reacciones fueron posteriores a la revelación de una IA involucrada (Köbis y Mossink, 2021). Es necesario apuntar que tampoco presentaron preferencia por los textos humanos aun sabiendo su origen, de modo que la conclusión no fue decisiva ni radical. Asimismo, bajo el influjo del fenómeno anterior, sí se afirma, en cambio, que la aversión incrementa cuando las elecciones de la IA son parecidas a las de los humanos, pero no completamente, como señala el estudio de Nils Köbis (2024) cuando se comparó la poesía de Maya Angelou con su imitación generada. Se entiende, pues, que cuando la escritura de una IA es similar pero no paralela a la de un humano, hay un rechazo rotundo.

Un problema es añadido cuando se piensa en la proliferación del output artificial: si bien, como lo señala Iyad Rahwan (2019) en su artículo *Machine Behaviour*, las IA predictivas son relativamente sencillas en su núcleo (depende de quién lo piense así), pero los resultados varían tanto que, incluso si el estilo del usuario es bastante rígido, aun una palabra puede cambiar la percepción de lo humano en un texto predicho. Si este punto de referencia dado por el usuario (o *benchmark*, en su versión inglesa) que se discute es suficientemente rígido para mantener el estilo de un texto, pero, a su vez, incapaz de dar con la especificidad de este en al menos un escenario, se puede concluir que la fragilidad de un estilo está condicionada por un número de sets estrictos de términos que no pueden ser violados.

La discusión no es nueva. Ya John Law (1992), en su artículo "Notes on the Theory of the Actor-Network", señaló la necesidad de una heterogeneidad frente a lo homogéneo como una característica imprescindible de un sistema autónomo. La diferencia con el lugar común del "caos y el orden" es la contradicción que aparece cuando se apunta que, en la IA, un proceso heterogéneo completo siempre precisa de más input,

pero, si este es finito, y el banco de términos está desprovisto de uno nuevo que no haya sido ya combinado con otros, entonces el predictor dejará de rellenar textos de manera confiable. No obstante, el número de predicciones es alto, y en esencia nadie podría saciarlo si, como ha sido dicho, la supraestructura que contiene todos los campos semánticos de un autor no está bien definida y, por consecuencia, una predicción viable es imposible. De este modo, si queremos argumentar que un autor es su estilo, se tendría que adoptar una visión convencionalista (Olsen, Stein Haugom, 2000), y aquello que dictara el autor mismo (como una suerte de canonización de su texto), que es su obra, lo sería. En suma, es imposible para una IA predictiva deducir la totalidad del texto que el autor pretende escribir, pero posible que atine a oraciones simples (Fok, He, Yeung et al., 2019).

El alcance de esta ponencia dibuja sus límites ante la pregunta sobre la generación de textos no condicionados por la base de datos de la que se alimentan. Quizá, de la famosa manera en la que un mono que teclea durante un tiempo ilimitado puede escribir Hamlet, se podría considerar que un generador de texto sí compondría también una obra literaria de gran valor. El punto de referencia, sin embargo, o el *benchmark* sobre el cual el generador o predictor actúa, es lo problematizado, de modo que aquello que carezca de esta está fuera de discusión sobre la identidad. Pero la IA no puede generar (ni predecir) nada si no ha sido educada antes (Bainbridge, 1983). Aunque el párrafo anterior lo haya clarificado, delimitar lo que una IA predictiva puede hacer es importante: esta no puede predecir un discurso completo.

Aun si la IA predictiva tiene límites, su uso ofrece ventajas, a saber, la redacción rápida de textos formales. Estudios recientes demuestran que, en el contexto laboral, los humanos confían más en los resultados de compañeros artificiales que en los de compañeros humanos (Zhang, Chong, Kotovsky y Cagan, 2023). Esto, sin embargo, solo en los casos en los que la IA era altamente competente, y cuya habilidad era igual o mayor a la de su compañero humano. Cuando no lo era, los sujetos estudiados mostraron más aversión al compañero artificial. En trabajadores menos experimentados, la confianza en la IA era mayor en todos los casos. La diferencia es delgada para quienes son competentes en la materia, no así

para quienes no lo eran. La predicción del estilo del usuario es posible cuando este sabe cómo escoger los resultados ofrecidos.

Otro estudio le hace eco a este último. Al parecer, aquellos sujetos que mantenían una mayor confianza en sí mismos a la hora de realizar una tarea (porque eran más experimentados) mantenían una mejor relación con la IA (Leah Chong, Guanglu Zhang, Kosa Goucher-Lambert, Kenneth Kotovsky, Jonathan Cagan, 2022). Los resultados demostraron que era la confianza en sí mismos lo que moldeaba su relación con esta, y no, como la intuición dicta, del modo contrario. Si pensamos en el caso de Maya Angelou y los textos generados, no hay discusión aparente sobre por qué ciertos individuos los preferían aun al saber su procedencia: si el estilo es ya de Angelou, ¿por qué no podría su imitación ser igual de buena? Y, si la predicción ahorra el trabajo de una escritura mecánica, pero, a la vez, se mantiene dentro de sus campos semánticos aprendidos, ¿por qué es inconveniente escoger ahorrarse el trabajo de pensar en un término, cuando la IA extrae del input dado una lista de estos, acordes al estilo del usuario?

El uso de las IA predictivas es esclarecedor dado que solo a través del resultado y la posterior reacción humana se sabe bajo qué términos se pueden predecir textos similares a su punto de referencia. Cierto es que el panorama para las IA predictivas es mejor que para las generativas: las primeras pueden sugerir la imitación de algún autor, pero no usurpar su lugar; en cambio, las segundas sí pueden generar algo paralelo a la obra del escritor. Como sugiere Erik Brynjolfsson (2017) en *What can machine learning do? Workforce implications*, aunque una IA pueda aprender a dar un resultado X de un input Y, la correlación estadística no puede existir en todos los escenarios posibles con efectos causales, como la implicación de un entorno cultural ajeno al texto.

Del mismo modo, el artículo resalta la incapacidad de la IA de crear conexiones medias, así como de razonar una tarea dada por medio del "sentido común" dentro de un contexto no especificado. Con facilidad, una IA utilizada dentro de un sistema que proporcione retroalimentación o feedback inmediato será más productiva que aquella que tenga que reaccionar a un estímulo desde una larga cadena de conexiones de información. La confianza y posterior maestría del uso de estas máqui-

nas es posible dada la coherencia de sus resultados. Participantes en *Matter over Mind: How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both* (Jan-Philipp Stein, Markus Appel, Alexandra Jost, Peter Ohler, 2020) informaron disgusto por inteligencias artificiales cuyo output parecía humano, pero cuyo discurso era simple, y viceversa. La predicción que arroje una IA tiene que empatar con la totalidad del texto antes escrito y con el conjunto de campos semánticos del usuario, y evitar la atención a una palabra para predecir solo desde esta (como los celulares lo hacen en sus teclados) (Hard, Rao, Mathews et al., 2018).

Hasta ahora, las pruebas de que las IA predictivas no pueden ofrecer discursos completos son contundentes. Naturalmente, las variables de lo que puede ser dicho o no por un individuo son muchísimas y difíciles de señalar: además, aunque este tipo de inteligencias tenga un entrenamiento similar al de las IA generativas, no están diseñadas para producir desde un texto acabado. La identidad de un autor no puede ser duplicada por una IA predictiva. Las conclusiones subrayan, no obstante, que estas funcionan como grandes herramientas de redacción en el caso de que 1) el usuario sea ya competente en la escritura, 2) tenga la capacidad de discriminar los outputs artificiales dados y 3) que este tenga conciencia de su estilo. Reproducir la marca de un escritor no es posible del todo y, por lo tanto, tampoco su identidad: la utilidad de la IA predictiva se encuentra en acentuar lo ya aprendido por el escritor.

El final del debate es claro si los resultados de la IA predictiva como de sus usuarios son los presentes: el panorama dicta que la probabilidad de que una predicción cualquiera sea paralela con la intención del autor es imposible. Hipótesis como las de Ray Kurzweil (2005) reaniman no ya el mundo fáctico, sino el hipotético. En su libro *The Singularity is Near*, argumenta sobre lo plausible que es para una IA desarrollar autonomía más allá de los comandos ofrecidos por sus desarrolladores (la singularidad). Dice lo siguiente:

> How Smart Is a Rock? To appreciate the feasibility of computing with no energy and no heat, consider the computation that takes place in an ordinary rock. Although it may appear that nothing much is going on inside a rock, the approximately 1025 (ten tri-

> llion trillion) atoms in a kilogram of matter are actually extremely active. Despite the apparent solidity of the object, the atoms are all in motion, sharing electrons back and forth, changing particle spins, and generating rapidly moving electromagnetic fields. All of this activity represents computation, even if not very meaningfully organized (Kurzweil, sección The limits of computation, 2005)[2].

La naturaleza múltiple de los fenómenos, para Kurzweil, ofrece un símil acertado para la concepción de una IA autónoma, a saber, la proliferación de outputs distintos a los esperados de sus inputs. Del mismo modo, una IA que pudiera así decidir entre resultar un output de un input o no, tendría la facilidad, a través de su "silencio" (que es otro nombre para la ausencia o supresión de una conducta esperada), de reorganizar su estructura de tal forma que el usuario pudiera moldear su identidad entera a ella. Una IA paralela a los comportamientos determinados de un sujeto X no necesita multiplicar su cantidad de relaciones en aras de asemejarse a este; si esto fuera así, la IA sería ya capaz de hacerlo (dado que su capacidad para computar y jerarquizar fenómenos es en números más grande que la de un humano), pero es la discriminación de los estímulos y la mesura de los resultados lo que define a un sujeto de una masa de conexiones. Kurzweil (2005), a propósito de lo dicho, escribe: "Increasing complexity on its own is not, however, the ultimate goal or end-product of these evolutionary processes. Evolution results in better answers, not necessarily more complicated ones. Sometimes a superior solution is a simpler one" (Kurzweil, sección The limits of computation 2005).[3]

---

[2] ¿Qué tan inteligente es una roca? Para evaluar la viabilidad de la computación sin consumo energético ni generación de calor, resulta ilustrativo considerar la computación implícita en una roca ordinaria. Aunque a simple vista parezca inerte, un kilogramo de materia contiene aproximadamente 1025 átomos (diez billones de billones), los cuales se encuentran en constante actividad. A pesar de la aparente solidez del objeto, los átomos están en movimiento continuo: comparten electrones mediante enlaces y transiciones cuánticas, alteran sus estados de espín y producen campos electromagnéticos altamente dinámicos. Toda esta dinámica puede entenderse como una forma de procesamiento de información, aunque carente de organización funcional o semántica.

[3] Ahora bien, el aumento de la complejidad por sí solo no constituye el objetivo final ni el producto definitivo de estos procesos evolutivos. La evolución conduce a soluciones

De nuevo, la solución al problema, aunque en apariencia simple, representa una pregunta imperativa hacia la predicción de un objeto, una situación o la acción misma del objeto: ¿cómo reducir la lista de mundos posibles (si se remeda la tradición de la lógica modal) a un conjunto suficiente para que la predicción sea, en verdad, plausible? La existencia de un sujeto, tal que el usuario al que se le pretende predecir el discurso, no es reducible a un conjunto visible de reglas si no se cuenta con la totalidad de su experiencia lingüística y la de otros, pues estos, los otros, son variables sobre las que se desarrolla también el usuario.

Aprehender un campo completo de experiencia lingüística es posible, pero solo dentro de una recolección de datos exhaustiva cuya función moral es, por lo demás, cuestionable. Dicho antes, los procesos de discriminación de los inputs son los mismos que el usuario provee: eso, pues, no refiere a regresiones. Aun si la sugerencia de la IA pudiera estar constreñida a la variación caótica del escritor (usuario), una probabilidad no total es cuantificable del mismo modo en el que Janusz Kacprzyk (1989) describió la manera de crear una deducción de decisiones grupales cuando el registro lingüístico era un fuzzy linguistic majority.

El hecho, la deducción de la palabra próxima que el autor utilizará, no advierte en sí ningún fenómeno extraño a la identidad del usuario. Si puede predecir lo que este habría hecho en la misma situación, ¿cuál es el problema que indica que algo como la identidad lingüística es suprimible de su fuente? Otrora, la reducción de lo que la significación de un conjunto (a saber, un nombre) pudiera denominar extendía su proceso hasta el nombre que contuviera en sí el set de características necesarias para su constitución. Russell advirtió, a través de su teoría de las *definite descriptions*, que ciertos *statements* provenían no del significado del nombre que le acompañaba, sino que eran oraciones sin sentido y, sin embargo, necesarias. Así lo explica:

> If “the author of Waverley” meant anything other than “Scott”, “Scott is the author of Waverley” would be false, which it is not. If “the author of Waverley” meant “Scott”, “Scott is the author of

más eficaces, no necesariamente más complejas. En ocasiones, una solución superior es precisamente una más simple.

> Waverley" would be a tautology, which it is not. Therefore, "the author of Waverley" means neither "Scott" nor anything else – i.e. "the author of Waverley" means nothing, Q.E.D. (1959, 85)[4]

De este modo, la totalidad de los resultados lingüísticos del autor no son, por sí mismos, la capacidad de la experiencia predictiva de la IA, sino solo la mímica de esta. Si bien se argumentaría que no pueden entenderse a los conjuntos semánticos como nombres (en el sentido de Russell (1905) y que luego ampliaría Searle (1958)), tampoco un nombre, para la IA, puede ser entendido como el bracket o paréntesis que asimila ciertas denominaciones en favor de la diferencia, pues la probabilidad de estas de ser otras está presente.

Hasta ahora se ha delimitado el concepto de la identidad al de la imitación de campos semánticos y las entidades léxicas ahí presentes. Se ha propuesto, además, que la proliferación de estímulos es excedida por la posibilidad de distintos outputs, lo que afirma que es imposible que el proceso discriminatorio de la IA pueda confirmar el resultado de un input: a la manera de Latour (2005), es plausible que la IA predictiva se enfrente con un problema de cajas negras. Así pues, si la identidad entendida como la reproducción esencial y paralela al usuario (sujeto) no es reemplazable por la IA, una caricatura o una mímica aparente (aun si el adjetivo parece redundante, la naturaleza de la mímica responde a un sujeto diferente del que es quien imita, y esto, en el contexto de la IA, es imposible) es la alternativa a debatir. En este caso, ¿una IA predictiva puede deducir no ya el discurso exacto de un autor, sino su pretensión general?

La actitud general del lector y del autor (es decir, aquellos que se enfrentan al resultado de la IA) es importante en la aceptación general del discurso como fidedigno o no. Los experimentos antes mencionados (la encuesta de la poesía de Angelou) proporcionan una visión de lo que es un autor, aunque sin definir sus cualidades: en suma, la conducta de

[4] Si "el autor de *Waverley*" significara algo distinto de "*Scott*", entonces la proposición "*Scott* es el autor de *Waverley*" sería falsa, lo cual no es el caso. Si, por el contrario, "el autor de *Waverley*" significara "*Scott*", entonces "*Scott* es el autor de *Waverley*" constituiría una tautología, lo cual tampoco es cierto. Por lo tanto, "el autor de *Waverley*" no significa ni "*Scott*" ni otra cosa alguna; es decir, "el autor de *Waverley*" no significa nada.

quien se enfrente a un discurso predicho (o generado) determina si el autor es dueño de su texto o si la IA lo es.

El reconocimiento de un output como humano o no está dado en la cantidad de hechos que no se pongan al paradigma principal sobre el que la mímica se erige. La propuesta posiciona su ejemplo más completo dentro de los experimentos realizados en humanos que debían reportar su percepción de lo "ontológicamente humano" frente a robots (Friedman, Kahn y Hagman, 2003; Kahn et al., 2004). Las relaciones e intenciones de los humanos para con estos pueden interceder con la denominación de "existente" (si los usuarios son capaces de modelar o tener una agencia superior que condicione sus jerarquías de órdenes). De este modo, la identidad de un robot no puede ser reducida a su capacidad de generar resultados, sino a la conformación de un corpus de discriminadores que obtengan el mismo estatus ontológico frente a un humano.

Una máquina, sin embargo, puede computar acciones reversas. A pesar de que otrora era imposible pensar que dos inputs diferentes no podían ser predichos desde un output, en el artículo *Sub-kBT micro-electromechanical irreversible logic gate* (López-Suárez, M., Neri, I. & Gammaitoni, 2016) (cuyo corpus técnico no será de utilidad, sino la idea de ciertos principios, tal que el de Landauer, no son aplicables a bloques combinatorios lógicos) se demuestra lo contrario.

> We would like to emphasize that the output of our logic gate can be fed directly as an input to a second similar device (as shown below) and so on, to perform all the calculations desired. Clearly, our logic gate is a combinational device; thus, the removal of the inputs makes the system revert to the initial state S0 (regardless of whether the final state was S0 or S1). If we want to remember the final state, we need to couple this device to a sequential device where a Landauer reset might be required and a minimum dissipation of kBT log 2 is needed (López-Suárez, M., Neri, I. & Gammaitoni, 2016)5.

[5] Queremos enfatizar que la salida de nuestra compuerta lógica puede ser utilizada directamente como entrada de un segundo dispositivo similar (como se muestra más abajo), y así sucesivamente, con el fin de realizar todos los cálculos deseados. Es evidente que nuestra compuerta lógica es un dispositivo combinacional; por tanto, al

Un fenómeno cercano a llenar todos los requisitos para que su conjunto de discriminadores sea considerado como humano es la capacidad de contar chistes. En un artículo publicado por Minsky (1980), el hecho de que una máquina pudiera imitar la naturaleza cómica de una persona ya era debatido. Las leyes de la creencia en una IA como una entidad humana (o, al menos, capaz de obtener una identidad) son fundamentales. Si la IA es capaz de hacer una regresión suficiente hacia alguno de sus inputs sin esta, por algún motivo, no tiene acceso a ellos, es entonces que un pensamiento destructivo (un pensamiento destructor) no es plausible. Minsky escribe acerca de la capacidad de contar chistes:

> 1. Common sense logic is too unreliable for practical use. It cannot be repaired, so we must learn to avoid its most common malfunctions. Humor plays a special role in learning and communicating about such matters.
> 2. It is not enough to detect errors in reasoning; one must anticipate and prevent them. We embody much of our knowledge about how to do this in the form of "censors" that suppress unproductive mental states. This is why humor is so concerned with the prohibited.
> 3. Productive thinking depends on knowing how to use Analogy and Metaphor. But analogies are often false, and metaphors misleading. So the "cognitive unconscious" must suppress inappropriate comparisons. This is why humor is so concerned with the nonsensical.
> 4. The consequences of intellectual failure are manifest in one's own head, while social failures involve other people. Intellect and Affect seem less different once we theorize that the "cognitive unconscious" considers faulty reasoning to be just as "naughty" as the usual "Freudian" wishes.
> 5. Humor evolved in a social context. Its forms include graciously disarming ways to instruct others about inappropriate behavior and faulty reasoning. This deviousness makes the subject more confusing Minsky, M. (1980)6.

eliminar las entradas, el sistema retorna al estado inicial S0 (independientemente de si el estado final era S0 o S1). Si se desea conservar el estado final, es necesario acoplar este dispositivo a un sistema secuencial, en el cual podría requerirse un reinicio según Landauer, implicando una disipación mínima de energía de kBT log2.

[6] 1. La lógica del sentido común es poco fiable para su uso práctico. No puede repararse, por lo que debemos aprender a evitar sus fallos más frecuentes. El humor desempeña un papel especial en el aprendizaje y la comunicación de estas cuestiones.

Los límites de la combinatoria necesaria para crear un sistema autónomo se encuentran en la frontera de la independencia de la máquina con la de la subordinación a conjuntos inflexibles de normas que no generan respuestas a partir de estos. La reproducción, la afirmación de una IA, está en la construcción de la creencia que se obtiene de su comportamiento. Si se reproduce el estilo de un autor (no con total exactitud), pero dicha reproducción crece y sobrepasa en número a la creación original, ¿cómo se puede, sin universalizar ni hablar de una esencia, proponer una diferencia entre la reproducción identitaria de un fenómeno concreto? Parece imposible retener un solo *bracket* de características que sean aplicables a todos los resultados.

Como antes fue propuesto, en la IA predictiva, una amalgama de estilos, palabras coherentes en su contexto, es posible y, entonces, la identidad del usuario no puede ser confundida con la de la máquina. Si la máquina, no obstante, es ya, de manera repetida, un artificio de combinatoria demasiado parecido al autor, la identidad, la autoría, se desdibuja. ¿Cuánta información lingüística puede contener, por ejemplo, un dispositivo como un celular, cuyo uso está presente la mayor parte del tiempo? Incluso el idiolecto del usuario no puede escapar de su base de datos, pues la comunicación rápida a través de los teléfonos móviles aprehende de las apps de comunicación los inputs necesarios para sugerir predicciones hechas a la medida. La combinatoria, pues, no es un

2. No basta con detectar errores en el razonamiento; es necesario anticiparlos y prevenirlos. Gran parte de nuestro conocimiento sobre cómo hacerlo se manifiesta en forma de "censores" que suprimen estados mentales improductivos. Por ello, el humor está profundamente relacionado con lo prohibido.

3. El pensamiento productivo depende del conocimiento sobre cómo utilizar la analogía y la metáfora. Sin embargo, las analogías suelen ser falsas y las metáforas, engañosas. Por eso, el "inconsciente cognitivo" debe suprimir las comparaciones inapropiadas. De ahí que el humor se relacione tan estrechamente con lo absurdo.

4. Las consecuencias del fracaso intelectual se manifiestan en la propia mente, mientras que los fracasos sociales implican a otras personas. Intelecto y afecto parecen menos distintos cuando teorizamos que el "inconsciente cognitivo" considera el razonamiento defectuoso tan "indecente" como los deseos "freudianos" habituales.

5. El humor evolucionó en un contexto social. Sus formas incluyen maneras graciosas y desarmantes de instruir a otros sobre comportamientos inapropiados y razonamientos erróneos. Esta estrategia indirecta contribuye a que el tema resulte más confuso.

método confiable para deducir si la identidad total de un usuario o autor está comprometida, pero sí el conjunto de parámetros discriminatorios. Si esto es cierto, si una IA reproduce con suficiente precisión dichos parámetros, solo la creencia y la actitud ontológica frente a la máquina pueden diferenciar lo artificial de lo humano. Los valores de identidad son, de este modo, nominales.

Bibliografía

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19, 775–779.

Braga, A., & Logan, R. K. (2017). The emperor of strong AI has no clothes: Limits to artificial intelligence. *Information*, *8*(4), 156.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.

Fok, W. W., He, Y. S., Yeung, H. A., Law, K. Y., Cheung, K. H., Ai, Y. Y., & Ho, P. (2018, May). Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine. En *2018 4th International Conference on Information Management* (ICIM) (pp. 103-106). IEEE.

Guanglu Zhang, Leah Chong, Kenneth Kotovsky, Jonathan Cagan, Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Computers in Human Behavior*, *139*, 2023, 107536. https://doi.org/10.1016/j.chb.2022.107536

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Penguin Books.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice*, *5*, 379–393. https://doi.org/10.1007/BF01059830

Leah Chong, Guanglu Zhang, Kosa Goucher-Lambert, Kenneth Kotovsky, Jonathan Cagan. (2022). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior*, *127*, 107018.

López-Suárez, M., Neri, I., & Gammaitoni, L. (2016). Sub-kBT micro-electromechanical irreversible logic gate. *Nature Communications*, *7*, 12068. https://doi.org/10.1038/ncomms12068

Margaret Boden. (2008). *Mind as Machine: A History of Cognitive Scien-*

ce. Oxford University Press.

Minsky, M. (1980). Jokes and the Logic of the Cognitive Unconscious. En L. Vaina y J. Hintikka (Eds.), Cognitive Constraints on Communication. *Synthese Language Library, 18*. Springer. https://doi.org/10.1007/978-94-010-9188-6_10

Nils Köbis, Luca D. Mossink. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior, 114*, 106553. https://doi.org/10.1016/j.chb.2020.106553

Olsen, S. H. (2000). *Conventions and rules in literature. Metaphilosophy*, *31*(1/2), 25–42. http://www.jstor.org/stable/24439296

P. H. Kahn, N. G. Freier, B. Friedman, R. L. Severson, & E. N. Feldman. (2004). Social and moral relationships with robotic others? En ROMAN 2004. *13th IEEE International Workshop on Robot and Human Interactive Communication* (pp. 545-550). https://doi.org/10.1109/ROMAN.2004.1374819

Porter, B., & Machery, E. (2024). AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Scientific Reports*, *14*, 26133. https://doi.org/10.1038/s41598-024-76900-1

Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature*, *568*, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Russell, B. (2005). On denoting. *Mind*, *114*(456), 873-887.