

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

Daren López Méndez

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.¹ 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? Este texto 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. Ciertamente es, como lo menciona Margaret Boden en *Mind as a Machine*,² 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 concerniente a una IA ge-

¹ A. Braga y R. K. Logan, «The emperor of strong AI has no clothes: limits to artificial intelligence» p. 156.

² Cfr. Margaret Boden, *Mind as Machine: A History of Cognitive Science*.

nerativa) indistinguibles de aquellas hechas por humanos,³ 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.⁴ 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 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.⁵

Un problema es añadido cuando se piensa en la proliferación del *output* artificial: si bien, como lo señala Iyad Rahwan en su artículo *Machine Behaviour*,⁶ 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* es-

trictos de términos que no pueden ser violados.

La discusión no es nueva. Ya en 1992 John Law, en su artículo *Notes on the Theory of the Actor-Network*,⁷ señaló la necesidad de una heterogeneidad frente a lo homogéneo como una característica imprescindible de un sistema autónomo. La diferencia para 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,⁸ 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.⁹

El alcance de este texto 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 la *benchmark* sobre la 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.¹⁰ Aunque

³ B. Porter, E. Machery, «AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably».

⁴ Nils Köbis, Luca D. Mossink, «Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry».

⁵ *Idem*.

⁶ I. Rahwan, M. Cebrian, *et al.* «Machine behaviour».

⁷ J. Law, «Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity».

⁸ Stein Haugom Olsen, «Conventions And Rules In Literature».

⁹ W. W. Fok, Y. S. He, *et al.*, «Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine».

¹⁰ L. Bainbridge, «Ironies of automation».

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.¹¹ 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 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.¹² 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 el 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 reac-

¹¹ Guanglu Zhang, Leah Chong, *et al.*, «Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation».

¹² Leah Chong, Guanglu Zhang, *et al.*, «Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice».

ción 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 en *What can machine learning do? Workforce implications*¹³ 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 máquinas 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*¹⁴ los 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).¹⁵

Hasta ahora, las pruebas de que las IA predictivas no pueden ofrecer discursos completos son contundentes. Naturalmente, las variables de

¹³ E. Brynjolfsson y T. Mitchell, «What can machine learning do? Workforce implications».

¹⁴ Jan-Philipp Stein, Markus Appel, *et al.*, «Matter over mind? How the acceptance of digital entities depends on their appearance, mental prowess, and the interaction between both».

¹⁵ A. Hard, K. Rao, *et al.*, «Federated learning for mobile keyboard prediction».

lo que puede ser dicho o no por un individuo son muchísimas y difíciles de señalar: además, aunque este tipo de inteligencias tengan un entrenamiento similar a 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.

Fuentes

Bainbridge, L. Ironies of automation. *Automatica* 19, 775–779 (1983). 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. In 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*, Volume 139, 2023, 107536, ISSN 0747-5632, <<https://doi.org/10.1016/j.chb.2022.107536>>. Law, J. Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice* 5, 379–393 (1992). <<https://doi.org/10.1007/BF01059830>>. Leah Chong, Guanglu Zhang, Kosa Goucher-Lambert, Kenneth Kotovsky,

Jonathan Cagan, Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice, *Computers in Human Behavior*, Volume 127, 2022, 107018, ISSN 0747-5632. Margaret Boden. 2008. *Mind as Machine: A History of Cognitive Science*. Oxford University Press, Inc., USA. Nils Köbis, Luca D. Mossink, Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry, *Computers in Human Behavior*, Volume 114, 2021, 106553, ISSN 0747-5632, <<https://doi.org/10.1016/j.chb.2020.106553>>. Olsen, Stein Haugom. "Conventions And Rules in Literature." *Metaphilosophy* 31, no. 1/2 (2000): 25–42. <<http://www.jstor.org/stable/24439296>>. Porter, B., Machery, E. AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Sci Rep* 14, 26133 (2024). <<https://doi.org/10.1038/s41598-024-76900-1>>. Rahwan, I., Cebrian, M., Obradovich, N. et al. Machine behaviour. *Nature* 568, 477–486 (2019). <<https://doi.org/10.1038/s41586-019-1138-y>>.