Criterios para la valoración de la creación literaria

de las inteligencias artificiales generativas

Tomás Eduardo Tena-Acosta*
Jorge Alan Flores-Flores**

(pensamiento)
(pensamiento), (palabra)... Y oBra
ISSN: 2011-804X · e-ISSN: 2462-8441

Fecha de recepción: 19 de febrero de 2025
Fecha de aprobación: 16 de junio de 2025
Fecha de publicación: 01 de julio de 2025

Para citar este artículo

Tena-Acosta, T. E. y Flores-Flores, J. A. (2025). Criterios para la valoración de la creación literaria de las inteligencias artificiales generativas, *(Pensamiento), (Palabra)… Y Obra,* (34), e22825. https://doi.org/10.17227/ppo.num34-22825

* Licenciado, Universidad Autónoma de Chihuahua, Chihuahua, México. p301357@uach.mx

** Doctor, Universidad Autónoma de Chihuahua, Chihuahua, México. jorgealanf@gmail.com

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

Resumen

Este artículo postula un aparato crítico que explora cuatro grupos de criterios que permiten valorar el fenómeno de las creaciones literarias de las inteligencias artificiales generativas. El primero de los criterios incursiona en la semántica y en la verificación que se ha hecho de la poca o nula competencia que profesan de esta las IAG, lo que guía a problemas de sesgos, por ejemplo. El segundo criterio es el de la semiótica textual, que se postula como un verificador de la creatividad u originalidad de la trama de una narrativa. El tercer criterio es el de la crítica cultural, con el concepto de razón instrumental, que abre la exploración de las IAG desde una perspectiva de su uso como finalidad y no como medio, y que abre a observar otras problemáticas más generales en la reproducción del arte. El cuarto criterio es ético, y postula la necesidad de control y responsabilidad humana sobre las creaciones de las IAG.

Palabras clave: inteligencias artificiales generativas; literatura; semántica; sesgos; razón instrumental; ética

Criteria for Evaluating the Literary Creation of Generative Artificial Intelligences

Abstract

This article proposes a critical framework that explores four groups of criteria for evaluating the phenomenon of literary creations by generative artificial intelligences. The first criterion delves into semantics and the verification of the limited or nonexistent competence that generative AIs demonstrate in this area, which leads to issues such as bias, for example. The second criterion is textual semiotics, which is proposed as a means to verify the creativity or originality of the narrative's plot. The third criterion is cultural criticism, with the concept of instrumental reason, which opens the exploration of generative AIs from the perspective of their use as an end rather than a means, and leads to the observation of more general issues in the reproduction of art. The fourth criterion is ethical, advocating for the necessity of human control and responsibility over the creations of generative AIs.

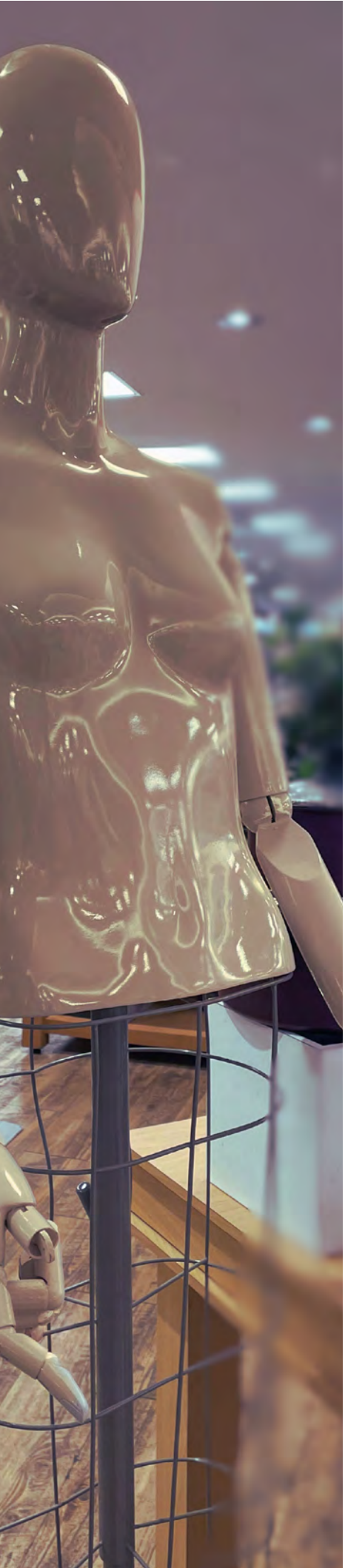
Keywords: generative artificial intelligences; literature; semantics; bias; instrumental reason; ethics

Critérios para a balanço da criação literária de inteligências artificiais generativas

Resumo

Este artigo propõe um aparato crítico que explora quatro grupos de critérios que permitem avalizar o fenômeno das criações literárias de Inteligências Artificiais Generativas (IAGs). O primeiro dos critérios aprofunda-se na semântica e na verificação que tem sido feita da pouca ou nenhuma competência que as IAGs professam neste sentido, o que leva a problemas de parcialidade, por exemplo. O segundo critério é o da semiótica textual, que é postulada como verificadora da criatividade ou originalidade do enredo de uma narrativa. O terceiro critério é o da crítica cultural, com o conceito de razão instrumental, que aborda a IAG a partir de uma perspectiva de sua utilização como fim e não como meio, e que abre à observação de outros problemas mais gerais na reprodução da arte. O quarto critério é ético e postula a necessidade de controle e responsabilidade humana sobre as criações dos IAGs.

Palavras-chave: inteligências artificiais generativas; literatura; semântica; vieses; razão instrumental; ética



Introducción

Este artículo se propone seleccionar y proponer cuatro criterios con los que valorar las creaciones literarias de las inteligencias artificiales generativas. Con esta propuesta se da por hecho algo que suele pasarse por alto: *estas tecnologías tienen la capacidad de crear*, incluso de ir más allá de lo que se espera de una herramienta para la creatividad. Componen poemas, crean cuentos, pueden componer música, generar imágenes digitales y, con el respaldo de la robótica, hasta pintar cuadros. Si bien valdría la pena intervenir en cuanto a qué significan estas creaciones, desde una perspectiva semiótica o incluso filosófica, aquí se da por hecho que existen, que son signos dentro de ese mundo de los signos y la comunicación que es la cultura y, por lo tanto, que se puede reconocer su pertinencia.

Para todas estas áreas del arte en las que pueden participar las inteligencias artificiales, existen distintos tipos de crítica especializada, encargados de resignificar algunas de las obras en las que se enfocan. Críticas que fijan la mirada, por ejemplo, en lo estético: las composiciones de las obras, su originalidad y su significado, e incluso críticas que añaden a la lista el valor cultural de una obra, considerada desde su espacio y su tiempo. Tan solo para la literatura pueden enumerarse varias de ellas: "Como sabemos, la crítica puede tener diversas modalidades: hay una crítica filológica, estética, sociológica, psicoanalítica; hay una crítica que expresa juicios de valor y otra que pone de manifiesto el trayecto de una escritura. Y otras más" (Eco, 1987, pp. 258-259).

Ahora bien, si las obras —del tipo que sean— producidas por o en coautoría con las inteligencias artificiales (IA) son signos de la cultura, estas no deberían estar exentas de ser examinadas de modo similar, con criterios semejantes a los que se emplean para obras de arte creadas —esta enunciación podría parecer chocante— por seres humanos. De modo que no parecería haber un mayor problema: se aplicarían los mismos criterios y conceptos de una de las varias críticas y se diría si tal pieza de X IA es interesante desde un punto de vista estético o cultural, y los porqués. Sin embargo, eso no atiende a un problema emergente: el del choque que socioculturalmente se ha suscitado con estas tecnologías. Y aquí podría introducirse una discusión sociológica o antropológica sobre cómo la tecnología ha levantado, desde siempre, estas alarmas en los seres humanos. En el caso particular de las inteligencias artificiales, sobre todo las del corte generativo (IAG), capaces de generar texto, audio, imágenes y video, estos encuentros con las áreas de lo humano se han manifestado a través de protestas, demandas y estudios que ponen frente a frente las capacidades de las IA y las de los seres humanos. Pese a estos hechos, que se suscitan por no haberse cruzado ciertos umbrales éticos o por la falta de comprensión de las tecnologías en cuestión, no dejan de producirse otros encuentros, al menos igual de problemáticos: concursos de renombre que otorgan premios a imágenes que no se sabía que fueron generadas mediante IA; estudios que revelan las preferencias de un grupo muestra por poesía escrita por una IAG frente a la de autores humanos, o que no pueden identificar quién ha escrito qué; la existencia

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

en la red de una cantidad ya desorbitada de creadores de contenido que proporcionan instrucciones para sacar provecho de estas tecnologías y crear, digamos, una novela por día, o que, como consecuencia de esto, la librería virtual de Amazon se abarrotara en 2023 de libros escritos mediante IA, etcétera. Estos fenómenos no hacen sino levantar preguntas cuyas respuestas apenas estamos aprendiendo a esbozar desde el campo de las humanidades. Y si lo que en verdad queremos es evitar futuros desencuentros, es necesario que partamos desde este campo para intentar contestar las preguntas en juego.

Este artículo propone que, como ya se anunció, no deberíamos utilizar las mismas herramientas críticas con las que comprendemos y valoramos el arte humano para las obras de las IAG. Antes que nada, por el motivo ya expuesto: estas tecnologías, como fenómenos culturales, nos han revelado desafíos por superar, empezando por la importancia de estos fenómenos creativos. Si no fuera este el caso, artistas y creadores no habrían alzado ya la voz, y organismos multinacionales, como la misma Unión Europea, no se tomarían la molestia de lanzar sendos manuales regulatorios aplicables a estas tecnologías —aunque prioritariamente abordan cuestiones en materia de ciberseguridad, esto nos dice que los retos en otras materias siguen abiertos y a la expectativa—. Por ello, esta propuesta se ocupará de la creación literaria —en momentos específicos de la narrativa— de las creaciones de las IAG, área de la literatura que, como las demás, enfrenta sus propios problemas.

Es probable que las inteligencias artificiales generativas terminen por convertirse en otra herramienta más que agregar al repertorio de las demandas actuales. Generalmente, se las promociona como herramientas para, por ejemplo, la productividad o el apoyo a la creatividad, sin que por ello se señalen ciertas deficiencias en cómo percibimos nuestra productividad y nuestra creatividad, pese a que críticos como Baudrillard (1985) no estarían del todo de acuerdo con esta visión: "Si los hombres crean o fantasean máquinas inteligentes es porque desesperan secretamente de su inteligencia, o porque sucumben bajo el peso de una inteligencia monstruosa e inútil" (p. 5). De cualquier modo, estas poderosas tecnologías están llamadas a prestarse como medio para quien tenga acceso a utilizarlas. El acceso a ellas y su utilización resultan especialmente sencillos. Sin embargo, al tratarse de un fenómeno naciente, cuya comprensión —sobre todo desde fuera de las áreas informáticas— apenas empieza a desarrollarse, es necesario repensar y revalorar su importancia. Proponemos, pues, establecer un aparato crítico que aspire a reducir cualquier sesgo de un área sobre otra, a cerrar las brechas abiertas por la falta de información que suele darse en campos ajenos al de las IA, y que estos criterios representen herramientas con las que la teoría y la crítica literarias logren abordar las creaciones de estas tecnologías con mayor profundidad y emitir juicios mejor sustentados. Es decir, un aparato crítico que no ignore las cualidades únicas del fenómeno y evite evaluarlas bajo criterios generales y genéricos, en favor de atender a las problemáticas comentadas y de lograr una mayor comprensión del actuar de las IA en la literatura. De hecho, esto representa una oportunidad para que la literatura se ponga a la vanguardia. Si se dejara pasar la oportunidad de que un área proponga sus propios criterios específicos frente a una tecnología emergente que ha suscitado ocasionales alarmas, estos problemas antes enumerados continuarán repitiéndose, como es lógico. Por el contrario, una teoría literaria de su tiempo debe observar los fenómenos culturales coetáneos y preguntarse cómo se reflejan en la literatura. Es similar a lo que ocurre cuando se efectúa un estudio del mercado editorial: deberían revisarse ciertas tendencias que prevengan a los editores sobre si es viable o no apostar por una obra. Lo mismo debería suceder con el fenómeno de la tecnología: lo ideal sería que se propongan estudios que permitan entender mejor la relación entre estas áreas en contacto.

En síntesis, este artículo abordará cuatro tipos de criterios con los que valorar las creaciones de las IAG, en favor de visibilizar los puntos problemáticos que estas tecnologías plantean ante la teoría literaria cuando se presentan como capaces de producir creatividad. Primero, se abordan dos criterios textuales: el criterio semántico y el criterio semiótico-textual, que fijan su atención en la composición del texto. Luego, se exponen dos criterios extratextuales, porque consideramos que sería poco fructífero permanecer dentro de los textos y no dar paso a la oportunidad de explorar los fenómenos tecnológicos desde una perspectiva cultural y filosófica, hecho que, creemos, no haría sino reforzar los tipos de crítica textual.

Dicho esto, retomamos la idea ya establecida: una teoría y una crítica literarias que realmente valoren el fenómeno literario deberían proponer una serie de herramientas críticas dirigidas a este fenómeno emergente, para lograr así un mayor alcance comprensivo desde sí mismas. No es una cuestión muy distinta de lo que, en primera

instancia, movilizó a organismos como la UE en lo que luego desembocó en un extenso documento de regulaciones: se trata de un punto de partida, si bien no necesariamente debe culminar en lo mismo. Aquí se proponen únicamente cuatro criterios para valorar y criticar formalmente la incursión de las IA en el campo de la creación literaria. Este aporte no es sino un *continuum* de posibles herramientas. Con ellas se establece la posibilidad de hacer frente, desde la teoría, a los nuevos posibles encuentros entre un área y otra, con la finalidad de comprender mejor las problemáticas suscitadas, y también en la expectativa de que, tras las respectivas investigaciones y tomas de decisiones, las inteligencias artificiales se conviertan en herramientas para la literatura y no en algo perjudicial, no en un fin en sí mismo.

Las críticas textuales

La selección de criterios no se limita, como ya se dijo, a los campos teóricos aquí evocados, pero se los elige por su relevancia y permisividad para la teorización y el análisis. Para comprender la relevancia de las implicaciones lingüísticas de un texto[1] creado por una IAG, acudimos a la semántica, que permite explorar las implicaciones de que no exista una verdadera comprensión del significado de lo que escriben las máquinas. Ante esto, se observa que, por supuesto, el autor humano —que es en todo momento, al menos hasta ahora, el operador de la IA— puede pasar por alto esta ausencia de competencia semántica, lo que le impediría establecer una comprensión más profunda de los resultados pragmáticos de los textos y, por ende, le dificultaría corregir o eliminar la mayor cantidad posible de sesgos o el peso mismo de estos.

Con el primero de los criterios, el semántico, se buscará exponer el argumento de que existe una ausencia de comprensión semántica por parte de las IA generativas, hecho que resulta particularmente relevante si se lo considera en el contexto de la creación literaria —y, por supuesto, fuera de esta también—. Es bien sabido ya en los estudios sobre IA que existen sesgos (*bias*) implícitos en su funcionamiento. La pregunta emergente es, por supuesto, ¿qué determina estos sesgos en una máquina? Sin embargo, la pregunta que exploraremos tiene más bien una inclinación hacia la pragmática: ¿qué implicaría para la creación literaria desde las IAG la presencia de estos sesgos o la ausencia de una inteligencia semántica real?

El segundo criterio textual se fundamenta en la semiótica textual, tal como la delinea Umberto Eco en su obra. De los cuatro criterios, este es quizá el único que puede considerarse verdaderamente estético. A través de conceptos clave de este campo de estudio, se propone explorar la creatividad de las IAG en función de las tramas que urden y las disyunciones que producen en sus creaciones. Así, en lugar de entender la creatividad como una cualidad general o superficial del texto, este enfoque permite conceptualizarla y definirla con mayor precisión, a partir del análisis y verificación de las decisiones tomadas antelas disyuntivas que orientan la trama de un texto.

El criterio semántico

Al ser la semántica una rama de la lingüística con la que, de hecho, las ciencias de la IA deben trabajar estrechamente, lo justo sería esperar un manejo pertinente y competente de la semántica del lenguaje verbal y escrito. Aunque hay otros fundamentos en el funcionamiento de la IA que se encargan de los tratamientos pertinentes de la información en general, las competencias lingüísticas son fundamentales, sobre todo si se analiza el fenómeno desde la disciplina de la programación del lenguaje natural —una de las grandes ramas de la IA—, como sugiere el siguiente artículo:

> La lingüística computacional es un área interdisciplinaria cuyo principal objetivo es especificar una teoría sobre la producción y comprensión del lenguaje natural, tan definida y exhaustiva que sirva de base para diseñar modelos computables, y a partir de éstos sea posible escribir programas mediante los cuales un computador pueda producir y comprender lenguaje natural. El objeto de estudio de la lingüística computacional es entonces la competencia lingüística y el conjunto de factores que determinan la actuación lingüística. (Morales, 1997, p. 39)

Este objetivo de emitir lenguaje natural no lleva, por sí solo, a desarrollar competencias lingüísticas. En este caso particular, que un ordenador *comprenda* el lenguaje significa que ha aprendido algunas de las reglas de composición gramatical, sobre todo las sintácticas, pero no asegura que comprenda a un nivel semántico profundo lo que se está formulando. Más aún, es evidente que una buena

1 El término *texto* aquí debería concebirse más abiertamente: cuentos, novelas, imágenes, etc. Pero el enfoque del artículo hace referencia a textos literarios-narrativos.

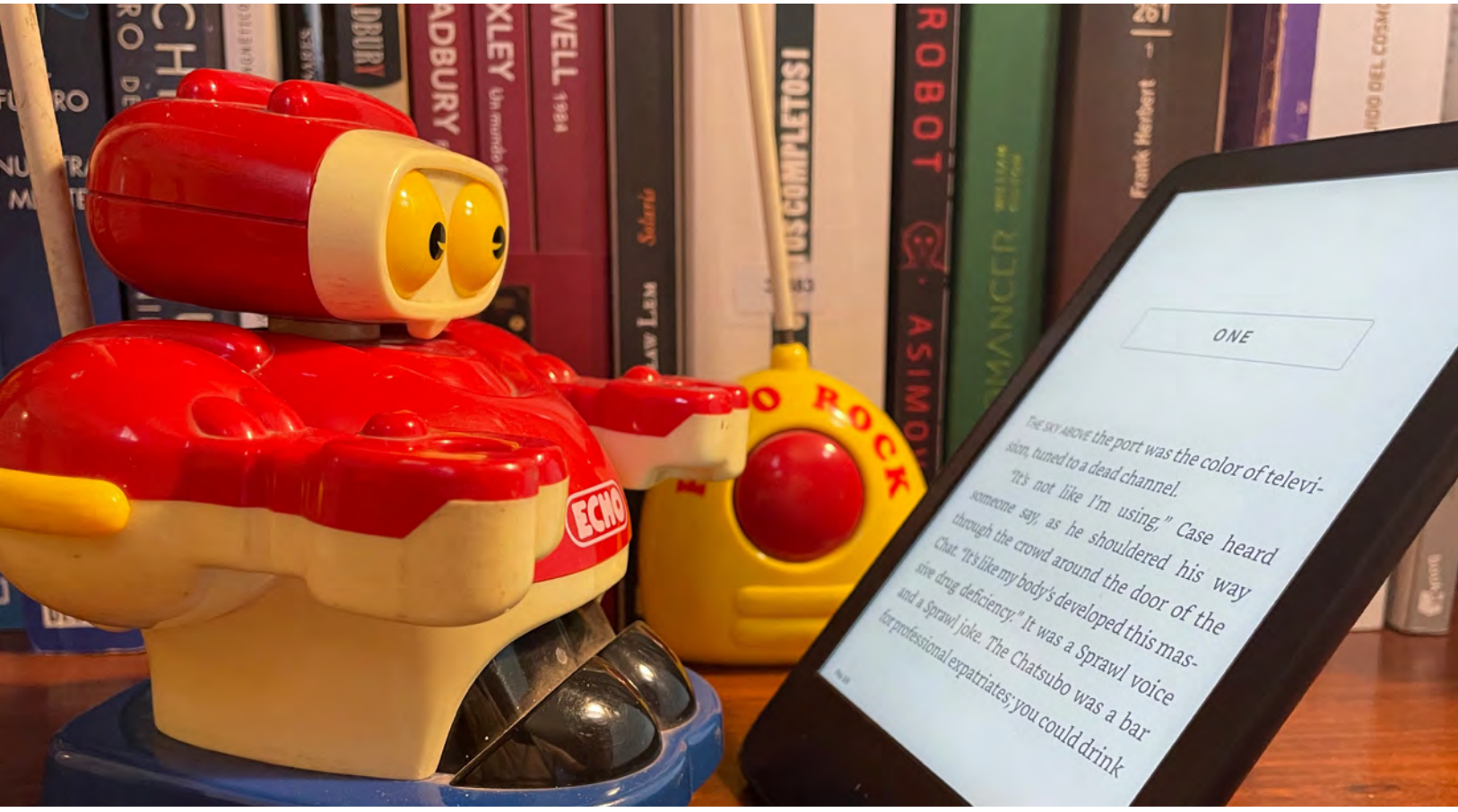
No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

parte de lo que propone la programación del lenguaje natural es llevar a cabo una interacción adecuada y pertinente, una comunicación efectiva con el usuario —esto en el caso de *chatbots* como ChatGPT-4, por ejemplo—, lo que no excluiría, por cierto, una porción importante de las máximas de Grice. Es decir, que dentro de este proceso de formación de competencias lingüísticas deben tenerse en cuenta también ciertas competencias pragmáticas. El mismo artículo especifica la importancia de esto:

> Si los conocimientos gramaticales son el fundamento de la competencia lingüística, los conocimientos pragmáticos son el fundamento de la competencia comunicativa, es decir, capacitan a los participantes para desarrollar deducciones adecuadas sobre el contexto y así, por ejemplo, poder mantener una conversación eficiente. (p. 42)

Esta falta de pertinencia semántico-pragmática, pese a los esfuerzos elaborados desde el campo de estudio de las IA, ya ha sido señalada desde áreas próximas a la literatura —aunque, creemos, no se ha problematizado lo suficiente—. Por ejemplo, el artículo titulado "The Art of Narration and Artificial Narrative Intelligence: Implications for Interdisciplinary Research" discute las posibilidades de un campo emergente y puntual, definido como "narrativa computacional", o "inteligencia narrativa computacional". Esta última es una competencia dentro de dicha área, la cual comprende la narratividad como una capacidad cognitiva inherente al ser humano, y desde ese punto focal explora las posibilidades de trasladar dicha inteligencia narrativa a ordenadores (Adamivna, 2019, p. 315). Pero no todo ordenador es una IAG que analiza, descompone y genera textos, y este campo más especializado no equivale a la ciencia multidisciplinaria de la IA. El artículo destaca, pues, las ventajas en la competencia narrativa de esta naciente disciplina frente a las desventajas de las IA. Lo que nos concierne es que menciona el origen de tales desventajas: reducciones en la comprensión de los textos por parte de las IA, de índole semántica.

> Mientras que la narratología computacional se esfuerza por formular concepciones logísticas, formalistas y objetivistas del lenguaje, la cognición y la computación, las teorías narrativas existentes en la IA reflejan una comprensión de la representación basada únicamente en



> construcciones textuales, lógicas o matemáticas. Por lo tanto, se requiere la motivación para una nueva interpretación de los conceptos tradicionales de significado, agencia, comprensión y lenguaje de la IA a la luz de un enfoque multimodal de la naturaleza narrativa.[2] (p. 315)

La clave radica en que estas teorías narrativas, que se basan en construcciones textuales, lógicas y matemáticas, carecen de enfoques más amplios en la semántica de los textos que están analizando y produciendo. Su valor puede destacarse más bien en el plano de la sintaxis, para el que las IA deben contar con fórmulas de composición mucho mejor estructuradas y codificadas que en lo semántico y aun en lo pragmático. De ahí que se sugiera ese cambio de enfoque centrado en la naturaleza narrativa y en una interpretación más rica de varios conceptos clave.

Este juego de aparente fortaleza sintáctica es un factor presente en el enclave de las IA, y lejos de soslayarse, debe considerarse. Sucede esencialmente lo mismo que en el experimento mental de la *Chinese Room* propuesto por John R. Searle (2009). En este experimento, que busca dilucidar la capacidad de ejecución de las IA a la manera de la prueba de Turing, un individuo que no habla chino es colocado en un cuarto oscuro y provisto de toda la base de datos de logogramas chinos, las instrucciones para construir palabras con los símbolos de esa lengua, un lápiz y una hoja. Este individuo podría estar capacitado para responder preguntas que se le formulen desde fuera del cuarto en dicho idioma, pero eso no implicaría que entienda nada sobre él, salvo las reglas de su sintaxis. No tiene acceso al significado de los símbolos, sino solo a su valor oposicional, y podría construir oraciones gramaticalmente correctas sin conocer el significado de sus expresiones ni los efectos que estas podrían tener a gran escala en su interlocutor como actos de habla.

> Si la persona que está en la sala no entiende el significado de los símbolos a partir de la implementación del programa, entonces tampoco lo entenderá ningún otro ordenador

2 En el original: "While computational narratology strives at formulating logistic, formalist and objectivist conceptions of language, cognition, and computation, existing theories of narrative in AI reflect understanding of representation based only on textual, logical, or mathematical constructs. Thus, the motivation for a new interpretation of AI's traditional concepts of meaning, agency, comprehension and language in the light of a multimodal approach to narrative nature is required". (todas las traducciones de este artículo son propias).

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

> a partir de esa única base, porque ningún ordenador, por el mero hecho de sus propiedades computacionales, tiene algo que el hombre no tenga. Además, la sala entera no tiene forma de llegar desde los símbolos a sus significados. El sistema entero no tiene forma de asignar una semántica a la sintaxis de los símbolos del ordenador.[3] (p. 144)

Como sugiere Searle, quizá esta incompetencia semántica sea un factor inherente a una inteligencia artificial que no está dotada de ciertas competencias humanas, como el sentido común (Kaku, 2019, p. 144) o ciertos códigos semióticos que permiten puntualizar contextos y actualizarlos para llevar a cabo procesos recursivos que doten de significado a la experiencia, entre otras capacidades cognitivas propias del ser humano, y que son parte del proceso de elaboración de un texto.[4]

De modo que el ejemplo de Searle, aplicado al contexto de la inteligencia de las máquinas, es conceptual, pero también fundamental. No obstante, apenas roza el ámbito de la literatura porque no se refiere directamente a cuestiones del lenguaje. La propuesta que más propiamente se instala en este campo viene de la mano de Emily Bender *et al.* (2021), en su artículo "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?". Su aporte representa, dentro de la crítica semántica, el argumento más fuerte e incisivo, e implica un proceder cuidadoso ante los textos generados por estas tecnologías, con enfoque en el lenguaje verbal escrito.

Los autores son claros al respecto: el desarrollo de *modelos de lenguaje* —término que agrupa a las IA de corte generativo—, que desde 2018 no ha hecho más que incrementarse, conlleva riesgos, y se hacen necesarias estrategias para mitigarlos (p. 610). Discute, en primer lugar, riesgos ambientales (en particular, emisiones de CO_2) y luego los riesgos derivados de la absorción de una cantidad insondable de datos de Internet para entrenar a las IAG y sus redes neuronales. En este punto surgen los problemas de sesgo o *bias*. La presencia de estos no indica sino la falta de una verdadera competencia gramatical en cuanto a la semántica y una ausencia de previsión pragmática real. En el caso investigado por Bender y su equipo, las causas de estos problemas derivan del aprendizaje no moderado a partir de datos disponibles en la red, pero también de la incapacidad inherente de las IA —o de los grandes modelos de lenguaje (LLM, por su sigla en inglés)— para comprender la información y su uso. En la sección en que utiliza la oportuna metáfora de los "loros estocásticos", expone un conjunto de problemas que terminan de delinear lo que hemos venido discutiendo hasta el momento:

> La tendencia de los datos de entrenamiento ingeridos desde Internet a codificar visiones de mundo hegemónicas, la tendencia de los ML[5] a amplificar sesgos y otros problemas en los datos de entrenamiento, y la tendencia de los investigadores y otras personas a confundir las ganancias de rendimiento impulsadas por ML con la comprensión real del lenguaje natural, presentan riesgos reales de daño, a medida que se implementan estas tecnologías. Después de explorar algunas razones por las que los humanos confunden el resultado de ML con texto significativo, pasamos a los riesgos y daños de implementar un modelo de este tipo a escala. Descubrimos que la combinación de sesgos humanos y lenguaje aparentemente coherente aumenta el potencial de sesgo de automatización, mal uso deliberado y amplificación de una visión de mundo hegemónica.[6] (p. 616)

Lo dicho hasta aquí pretende constituirse como un criterio utilizable desde la teoría y la crítica literarias, con el cual atender el hecho de que, si para las IAG no hay una comprensión semántica ni pragmática, existe el riesgo de la reproducción de sesgos presentes en los materiales con que fueron entrenadas. La expansión y sobreproducción de

3 En el original: "If the person in the room does not understand the meanings of the symbols on the basis of implementing the program, then neither does any other computer solely on that basis, because no computer, just in virtue of its computational properties, has anything that the man does not have. Furthermore, the whole room has no way of getting from the symbols to their meanings. The whole system has no way to attach a semantics to the syntax of the computer symbols".

4 Sin embargo, véase el concepto de *frames* propuesto por Marvin Minsky, el cual otorga a la IA ciertas posibilidades de comprensión contextual. No obstante, esta comprensión sigue dependiendo de una competencia meramente performativa, similar a la que se plantea en el juego del cuarto chino de Searle.

5 Modelos de lenguaje.

6 En el original: "The tendency of training data ingested from the Internet to encode hegemonic worldviews, the tendency of LMs to amplify biases and other issues in the training data, and the tendency of researchers and other people to mistake LM-driven performance gains for actual natural language understanding — present real-world risks of harm, as these technologies are deployed. After exploring some reasons why humans mistake LM output for meaningful text, we turn to the risks and harms from deploying such a model at scale. We find that the mix of human biases and seemingly coherent language heightens the potential for automation bias, deliberate misuse, and amplification of a hegemonic worldview".

textos que han codificado estos sesgos en su entramado es, por tanto, problemática. Así, ante el texto literario producido por una inteligencia artificial, debe existir una mirada crítica, de sospecha, que más allá de enfocarse en hacer un rastreo puntual de sesgos en el texto, permita emitir juicios sobre el acto literario en su conjunto, y considere que la utilización de una IAG como herramienta para la creación literaria muy probablemente acarrea sesgos importantes, agravados por la ausencia de competencias semánticas reales. Este juicio involucraría no solo a la IAG en sí, sino también al coautor humano.

Además, es evidente que incluso si el foco de la crítica semántica parte de lo textual, no permanece en ese plano: amenaza con salirse de él y emprender un camino que explore los posibles daños provocados por los sesgos y, en algunos casos, incluso sus orígenes. De modo que parecería que los criterios textuales no harían sino abrir paso a criterios extratextuales. Por el momento, tanto el experimento del cuarto chino como los riesgos de sesgos descritos por Bender proporcionan al investigador literario dos herramientas críticas con las que abordar mejor la lectura y comprensión de una obra escrita con inteligencia artificial. Todo acto creativo de esta índole presentaría, así, sesgos importantes que sería necesario examinar, y que podrían revelar un problema epistemológico subyacente aún más digno de análisis.

Los criterios de la semiótica textual

El libro de Eco (1987), del cual tomamos las observaciones sobre la semiótica textual, tiene como objetivo definir las particularidades del proceso de lectura e interpretación, con el fin de establecer cómo una obra se abre a su lector. De toda la metodología, nos interesa en concreto lo que en el proceso de interpretación se denomina *elección de isotopías*, así como otros conceptos colindantes.

La *isotopía* es un concepto de origen greimasiano, pero Eco, para el contexto de su obra —y del que participaremos, pues se enfoca en el texto narrativo—, la define como un recorrido de lectura a través de una unidad de coherencia semántica; es decir, como un concepto que "se refiere a la constancia de un trayecto de sentido que un texto exhibe cuando se le somete a ciertas reglas de coherencia interpretativa" (p. 132). Debe saberse, a su vez, que el hecho de trazar un recorrido de coherencia en una narrativa tiene mucho que ver con plantear una hipótesis sobre el tema, o *topic*, de dicha narrativa. Efectuar este acto es ya comenzar una interpretación: "El reconocimiento del *topic* es un movimiento cooperativo (pragmático) que guía al lector hacia el reconocimiento de las isotopías como propiedades semánticas de un texto" (p. 144).

El reconocimiento de estos criterios permitiría emitir juicios interesantes no tanto sobre la creatividad, sino sobre el criterio estético —según la crítica especializada que opte por ello—, que oriente la mirada, de entrada, hacia las disyunciones en las narrativas y la elección de los temas. La idea es establecer un modo de determinar si, en cuanto a su trama y sus disyunciones, un cuento —digamos— generado mediante una IAG resulta interesante o incluso novedoso.

Debido a que operan privilegiando el plano sintáctico combinatorio del lenguaje, las IAG tienen altas probabilidades de incursionar en combinaciones no antes establecidas o poco codificadas en la literatura. Así lo sugiere el artículo de Margaret A. Boden (1998), "Creativity and Artificial Intelligence", aunque en un plano más general sobre la creatividad. Si bien hay que tomar en cuenta el año de publicación del artículo —que, en términos de inteligencias artificiales se percibe como lejano—, sus planteamientos conservan una vigencia semejante a la de algunos del ensayo de Turing —con sus afirmaciones sobre la capacidad performativa de las máquinas[7]—. Boden sostiene que las máquinas tienen capacidad para la creatividad, al menos en el nivel que ella llama creatividad combinatoria, es decir, "la combinación novedosa (improbable) de ideas familiares"[8] (p. 348; véase también la conclusión, p. 355), un tipo o nivel de creatividad inferior a aquellos transformadores de estructuras.[9] En lo que respecta a lo literario, estas afirmaciones resultan aplicables: no es improbable que, con todas las bases de datos proporcionadas como entrenamiento, las IAG —a través de redes neuronales o algoritmos— obtengan modelos con los que puedan proceder analíticamente y, tras una descomposición, generar recomposiciones en patrones poco recurrentes o novedosos que, finalmente, puedan considerarse creativos.

7 Lo que incomoda respecto de esta afirmación es la fecha en que se emitieron estas conclusiones. Artículos como "Using Artificial Intelligence for Enhancing Human Creativity" (2023) llegan a conclusiones similares veinticinco años después (Elfar y Dawood, 2023, pp. 118-119).

8 En el original: *novel (improbable) combinations of familiar ideas.*

9 Esta discusión sobre la creatividad basada en combinaciones poco probables remite a lo expuesto en el subapartado anterior, donde se corroboró que, aunque pueden existir competencias sintácticas sólidas para comprender y elaborar un sistema, las competencias semánticas y pragmáticas suelen ser mucho más limitadas.

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

Para tal efecto, la semiótica textual se presenta como una herramienta crítica idónea para leer textos producidos por inteligencias artificiales. La conjunción de los conceptos de *topic*, *isotopía* y *frame* permitiría determinar cuán creativas pueden ser tales disyunciones y cuán inusuales los patrones generados. El procedimiento sería, más o menos, el siguiente: el investigador establecería un *topic* del texto narrativo generado por la IA, lo cual remitiría a una hipótesis sobre qué trata ese cuento o novela. El recorrido isotópico abriría entonces un abanico de posibilidades: un mismo *topic* puede derivar en diversas historias, es decir, en distintas isotopías. ¿Cuáles de estas resultan en la disyunción más *creativa* o novedosa posible?

El concepto de *marco* o *frame* determinaría la calidad de dicha creatividad. "Un marco es una estructura conceptual que representa el conocimiento convencional de los usuarios de una lengua. Los marcos definen lo que esperaríamos que fueran los transcursos de eventos posibles, normales o necesarios" (Dijk, 2005, p. 34). En un contexto literario, un *frame* permite la correcta comprensión de un texto. Volviendo a *Lector in fabula*, Eco afirma: "Consideramos que la comprensión textual se encuentra ampliamente dominada por la aplicación de cuadros pertinentes" (p. 116). Eco distingue entre cuadros comunes y cuadros de tipo intertextual. El investigador literario avisado e intertextualmente competente podría establecer el cuadro en cuestión que enmarca el *topic* y la isotopía, y reconocer si el marco seleccionado por la IA corresponde, en efecto, a un ejemplo ya dado en la literatura, a un motivo bien codificado, y de qué forma. Esta acción exige, por supuesto, una competencia filológica básica que permita, al menos, establecer conexiones genéricas entre el texto creado por la IAG y otro u otros textos de la literatura.

En términos narrativos, si se solicita a una IAG que escriba un relato de misterio dentro del género policiaco, la semiótica textual —idealmente a través de los tres conceptos explicados, aunque sin limitarse a ellos— podría establecer si el resultado de la trama y el tema propuesto por la máquina son realmente creativos o novedosos, o realizar una verificación intertextual que remita a historias previas y desmonte la noción de creatividad, ya sea en un plano estructural (siguiendo a Boden) o meramente sintáctico-combinatorio. Una forma habitual de lograr

un efecto interesante en una obra es, precisamente, jugar con los marcos intertextuales. De modo que un autor intertextualmente competente, conocedor de los códigos del género en que incursiona y de los cuadros que le son propios, "puede decidir *ex profeso* prescindir de ellos precisamente para sorprender, para engañar o para deleitar al lector" (Eco, 1987, p. 119). Lo mismo podría aplicarse a una IAG, en tanto la creatividad consiste también en probar combinaciones improbables.

La semiótica textual, a través del *topic*, la *isotopía* y el *frame*, como criterios, permite conceptualizar esa creatividad, esas combinaciones improbables, y señalarlas en el texto, proporcionando así pruebas claras y contundentes que avalen —o desmonten— la creatividad y lo inusitado en los patrones elegidos por las IAG en su labor creativa.

Las críticas extratextuales

Los criterios extratextuales presentan un alcance más amplio en cuanto a las implicaciones que traen consigo. La crítica cultural, por ejemplo, es un campo vasto, y a través de la observación de determinados fenómenos —como en este caso proponemos que sea la creación literaria de las IAG— se interroga por problemáticas vigentes cuyos contornos exceden los de un área específica. Es decir, la crítica cultural que delineamos a continuación conduce a reflexiones sobre la relación de los seres humanos con las máquinas. Sin embargo, nuestro objetivo es retornar, en determinado momento, a las observaciones sobre lo literario, si lo que deseamos es delinear criterios de utilidad para valorar las creaciones de las IAG.

El criterio de la razón instrumental es el primero dentro de estas críticas extratextuales. Lo tomamos como guía, como un recorrido para comprender una implicación más profunda que acompaña a la existencia misma de máquinas que ejecutan tareas humanas, y que desemboca en la problemática de las máquinas como fin. A la razón instrumental la articulamos con una reflexión en torno a la reproducción y la sobreproducción del arte,

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

donde hallaremos otros criterios viables para los fines de este apartado. Finalmente, el criterio ético propone que las creaciones de las IAG deberían ser atribuidas primordialmente a los humanos, como forma de asumir responsabilidad sobre toda la gama de implicaciones que puedan surgir de estas tecnologías.

La crítica cultura a partir de la razón instrumental

Por crítica cultural aludimos, por supuesto, a un planteamiento crítico de uno de los integrantes de la Escuela de Frankfurt: Max Horkheimer (1973), en su obra *Crítica de la razón instrumental*. Utilizar el concepto de *razón instrumental* como criterio para observar las creaciones de las IAG remite a la idea de una tecnología asumida como fin y no como medio: un fin despreocupado, poco atento a la calidad de sus medios. A partir de esta justificación, podremos articular otros conceptos que acompañen esta crítica.

No podemos proceder sin considerar el cambio que, según Horkheimer, ha habido respecto al concepto de *razón* tal como se concebía con todo su poder durante la Ilustración. No nos detendremos aquí en su crítica al pragmatismo —aunque constituye la base de lo que explicaremos más adelante— ni en las soluciones que propone el autor o la Escuela de Frankfurt. Lo que queremos observar es cómo las IAG, como agentes creativos, pueden comprenderse desde el concepto de una razón que ha sido instrumentalizada. Este revestimiento opera como la entrada a otros criterios más especializados, de los que hablaremos en las secciones siguientes.

Cuando Horkheimer afirma que "la razón aparece totalmente sujeta al proceso social", señala un fenómeno: la simplificación de la razón.

> Su valor operativo, el papel que desempeña en el dominio sobre los hombres y la naturaleza, ha sido convertido en criterio exclusivo. Las nociones se redujeron a síntesis de síntomas comunes a varios ejemplares […]. Vemos en ellas meras abreviaturas de los objetos particulares a los que se refieren. Todo uso que va más allá de la sintetización técnica de datos fácticos, que sirve de ayuda, se ve extirpado como una huella última de la superstición. Las nociones se han convertido en medios racionalizados, que no ofrecen resistencia, que ahorran trabajo. Es como si el pensar mismo se hubiese reducido al nivel de los procesos industriales sometiéndose a un plan exacto; dicho brevemente, como si se hubiese convertido en un componente fijo de la producción. (p. 32)

Esta simplificación, sin embargo, no es trivial: revela un mundo dominado por el cumplimiento de objetivos, una sistematización del pensamiento que corre el riesgo de enajenar cualquier forma de pensar distinta de aquella que privilegia la metáfora del proceso industrial.

El uso de herramientas analíticas y generadoras de texto tan potentes como las IAG promueve un adelanto de la razón como instrumento que exige del pensamiento un actuar sistemático y reducido. Las propias inteligencias artificiales parecen privilegiar este actuar, pues en su funcionamiento codificado radican formas abstractas —imitaciones del pensamiento— fundamentadas en la matemática y la lógica: las redes neuronales, los *embeddings* vectoriales y la recolección masiva de datos se presentan como paralelismos de la inteligencia humana, del cerebro orgánico, pero no son sino, en última instancia, lo que Horkheimer señala: una reducción de la razón, un intento de replicar el órgano del pensamiento.

Lejos de producir entusiasmo ante las posibilidades creativas de las IAG, el criterio de la razón instrumental las posiciona como herramientas propias de su época, con toda la carga que ello conlleva: su existencia problemática como fin; el embelesamiento por las máquinas que desemboca en fenómenos como el *AI Washing*[10] o la farsa; el reemplazo laboral de seres humanos por máquinas inteligentes; e incluso la consideración del miedo y la incertidumbre que despiertan estas posibilidades —aquí podría también retomarse el viejo sueño modernista del que beben los ideales poshumanistas y transhumanistas, revisiones que no abordaremos por cuestiones de extensión—. Para una teoría literaria que emplee como criterio de análisis la teoría de Horkheimer, las creaciones de las IAG adquieren un matiz distinto: ya no interesa tanto la cuestión de la creatividad, sino su instrumentalización. Esta idea se antepone a cualquier criterio estético.

Los criterios textuales, por ejemplo, enmarcados en el de la razón instrumental, adquieren una mayor pertinencia al interpretarse desde la perspectiva de que el uso de

10 El *AI washing* o 'lavado de imagen' con inteligencia artificial es una estrategia de *marketing* engañosa que consiste en exagerar la integración de la IA en un producto o servicio para hacerlo parecer más avanzado o innovador de lo que realmente es (Lagos, 21 de agosto de 2024).

las IAG evidencia cómo la razón instrumental que domina nuestra época permite transformar en procesos sistematizados ciertas áreas y facultades humanas que no habían sido concebidas así. Esta idea se enlaza con la continuación del comentario de Horkheimer: "Cuanto más automáticas y cuanto más instrumentalizadas se vuelven las ideas, tanto menos descubre uno en ellas la subsistencia de pensamientos con sentido propio. Se las tiene por cosas, por máquinas" (1973, p. 33). De hecho, un ejemplo estrechamente paralelo al de las IAG como máquinas diseñadas para sistematizar el pensamiento o la creatividad es el del lenguaje mismo, en su caída bajo el dominio de la razón instrumental:

> El lenguaje, en el gigantesco aparato de producción de la sociedad moderna, se redujo a un instrumento entre otros. Toda frase que no constituye el equivalente de una operación dentro de ese aparato, se presenta ante el profano tan desprovista de significado […]. La significación aparece desplazada por la función o el efecto que tienen en el mundo las cosas y los sucesos. Las palabras, en la medida en que no se utilizan de un modo evidente con el fin de valorar probabilidades técnicamente relevantes o al servicio de otros fines prácticos, entre los que debe incluirse hasta el recreo, corren el peligro de hacerse sospechosas de ser pura cháchara, pues la verdad no es un fin en sí misma. (p. 33)

Para actualizar esta perspectiva, reconocemos que el concepto de *razón instrumental*, aplicado a las inteligencias artificiales, ha sido trabajado por Éric Sadin (2020) en su libro *Inteligencia artificial o el desafío del siglo*. Su mención de este concepto resalta la hegemonía de la razón instrumental en detrimento de una racionalidad abierta y crítica, sin caer en posiciones de *tecnofilia* ni de *tecnofobia*:

> Ya no asistiríamos a la manifestación de desacuerdos que tratan respecto de las herramientas, sino respecto de un espíritu de la técnica, en este caso aquel que está predominando y que alimenta un tipo específico de racionalidad: una razón instrumental extrema. Esta razón no genera controversias a la altura de los dilemas que plantea mientras que hoy en día se impuso de modo masivo y se convirtió en hegemónica.
>
> En la medida en que nos oponemos a sus fundamentos, y que esta razón llegó a neutralizar todas las demás modalidades posibles, es nuestro deber afirmar que su posición dominante es ilegítima y defender una racionalidad abierta, crítica, inventiva. y que respete el axioma intangible de la pluralidad humana. (p. 281)

Sin embargo, es sobre todo el siguiente pasaje, más tajante y directo, el que más llama la atención por su fuerza crítica frente a la racionalidad instrumental propia de las IA. Se trata, más que de un criterio, de un manifiesto, que resulta pertinente recuperar cuando se compara la creatividad humana con la artificial:

> En oposición a una racionalidad que se aplica a hacer de cada hecho y cada gesto el objeto de una transacción mercantil, y a desplazar indefinidamente los límites del mercado, pretendemos ubicar nuestras existencias al abrigo de estas ambiciones totalizadoras y, más todavía, pretendemos desplazar el acto de consumo desde el centro hacia la periferia, y no recurrir a él a menos que sea necesario […]. Sostenemos que son las imperfecciones de la vida (que nunca se resuelven de una vez por toda) las que estimulan nuestro deseo de realizarnos y trabajar sin descanso para la construcción de un mundo común que se base en el axioma cardinal de no dañar a nadie. En oposición a una racionalidad que concibe lo humano como henchido de defectos y cuenta con paliarlos en favor de máquinas infalibles y todavía más productivas, que hacen de nosotros seres "superfluos" (Hannah Arendt), celebramos los poderes virtualmente infinitos de cada ser humano, y queremos obrar para que todos podamos beneficiamos de las mejores condiciones que presiden su eclosión y expresión. En oposición a una racionalidad que genera una furia "innovadora" que converge en la extensión de su imperio y contribuye a la instauración de un utilitarismo generalizado, nos negamos a contar permanentemente con una ganancia en nuestras relaciones con lo real y los demás, ya que cultivamos los poderes de nuestra inventiva en vistas a experimentar múltiples modos de existencia que participen

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

> de nuestra plenitud individual y colectiva. (pp. 281-282)

No hemos querido desaprovechar la oportunidad de partir desde Horkheimer y llegar hasta Sadin, con el objetivo de actualizar y focalizar el concepto de razón instrumental: Sadin concreta la crítica del teórico de Frankfurt aplicándola a la tecnología más sugerente para el caso: la inteligencia artificial. Sin embargo, como señalamos antes, este concepto no hace sino abrir la puerta a otros criterios, más propios de una crítica especializada, que permiten acercarse con mayor precisión al campo de la literatura. Este puente, proponemos, puede encontrarse en las teorías de Jean Baudrillard y Walter Benjamin.

Como hicimos con Horkheimer, no nos extenderemos aquí en la teoría de ambos autores —lo suficientemente vasta como para abrir un nuevo artículo—, pero sí mencionaremos, como ya hicimos en la introducción, un pasaje de Baudrillard que entronca con los postulados frankfurtianos, especialmente en lo relativo al pensamiento rebajado a un proceso industrial. Escribe Baudrillard (1985) en "El Xérox y el infinito": "Si los hombres sueñan con máquinas originales y geniales, es porque desesperan de su originalidad, o porque prefieren desasirse de ella y gozarla por máquina interpuesta. Pues lo que ofrecen esas máquinas es el espectáculo del pensamiento" (p. 6). Esta es, también, una forma de pensar la razón instrumental.

El artículo en cuestión señala dos razones por las que las máquinas presentan una desventaja frente a los seres humanos: la carencia de un cuerpo y la imposibilidad de sentir placer. Estos factores conducen a implicaciones más profundas. Pero dejando de lado ese desarrollo, quizá los conceptos más útiles de Baudrillard —más conectados con el peso teórico de la razón instrumental— sean los de *hiperrealidad* y *simulacro*, expuestos en *Cultura y simulacro* (1977).

Artificialidad y *reproducción* son conceptos que permean esa obra. Una crítica literaria especializada podría considerar que la reducción del concepto de *razón* abre paso a la búsqueda de lo *hiperreal* y del *simulacro*. En términos literarios, el fenómeno debe abordarse incluso desde un enfoque semiótico: las máquinas reproducen signos de signos, lo cual no es en sí mismo nuevo en la naturaleza ni en la literatura, pero la teoría de Baudrillard sugiere que la hiperproducción de la realidad a través de ciertos artificios conlleva un desgaste de la sustancia, como si se tratara de una pérdida: un simulacro, signo a su vez de una crisis cultural. Y si bien no es nuestro interés señalar someramente una decadencia de toda una cultura vislumbrada a través de unos pocos artificios, es cierto que parte del funcionamiento de las IA consiste precisamente en reproducir: obras, estilos, motivos, etc. La masificación y diseminación de estas producciones remiten a lo que Baudrillard expone como un mundo contemporáneo de simulacro, donde "la simulación no corresponde a un territorio, a una referencia, a una sustancia, sino que es la generación por los modelos de algo real sin origen ni realidad: lo hiperreal" (p. 5). De manera que el criterio al que podemos llegar con lo hiperreal tendría que preguntarse por la calidad —no precisamente estética— de un artificio que reproduce y copia exhaustivamente —y *sin una claridad semántica ni pragmática*— todo aquello que aprende de sus bases de datos de entrenamiento.

> De ahí la histeria característica de nuestro tiempo: la de la producción y reproducción de lo real […]. Aquello que toda una sociedad busca al continuar produciendo, y superproduciendo, es resucitar lo real que se le escapa. Por eso, tal producción "material" se convierte hoy en hiperreal. Retiene todos los rasgos y discursos de la producción tradicional, pero no es más que una metáfora. De este modo, los hiperrealistas fijan con un parecido alucinante una realidad de la que se ha esfumado todo el sentido y toda la profundidad y la energía de la representación. Y así, el hiperrealismo de la simulación se traduce por doquier en el alucinante parecido de lo real consigo mismo. (p. 49)

Especialmente el señalamiento de cómo esta producción material se convierte en hiperreal es lo que nos permite argumentar que el concepto puede transformarse en un criterio cultural para observar las creaciones literarias de las IAG. Plantea la cuestión de por qué existe la necesidad de esta reproducción de las reproducciones de las reproducciones, y permite seguir elaborando el argumento sobre cómo este fenómeno concierne a las obras de arte en general, en la medida en que se inscriben en una lógica del simulacro.

Sin embargo, si por algo el concepto de Baudrillard no termina de instalarse plenamente en la literatura —o en el arte mismo—, es porque Walter Benjamin (2003) desemboca directamente en dicho territorio. También él trabaja

desde la idea de reproducción, pero lo que atiende en particular es la historicidad y la espacialidad de la obra de arte, elementos que añaden valor más allá del contenido en sí. El concepto de *autenticidad* —que designa el "aquí y ahora" de la obra (p. 42)— constituye, entonces, otro criterio que se articula con los de simulacro, hiperrealidad y razón instrumental. Las IAG no son sujetos empíricos en el sentido en que lo son los humanos: su tiempo es estático y su experiencia no es histórica. Por eso, el concepto adyacente de destrucción del aura representa también un criterio válido de valoración crítica de las creaciones literarias de las IAG: "Día a día se hace vigente, de manera cada vez más irresistible, la necesidad de apoderarse del objeto en su más próxima cercanía, pero en imagen, y más aún en copia, en reproducción" (p. 48). Este desgaste o pérdida del aura es lo que habría que considerar ante la posibilidad de que se generalice un gusto por la literatura producida por inteligencia artificial. Y entonces la duda surge, casi como una curiosidad incómoda: ¿Por qué nos atrae este fenómeno? ¿Por qué preferir una literatura artificial, procesada por una máquina? Con esta última pregunta tocamos un umbral que este artículo no busca sobrepasar.

Solo hemos de añadir una serie de cuestionamientos: ¿Qué consecuencias entraña la necesidad de poner a escribir —a crear— a las máquinas? ¿Qué implica que un agente humano racional decida colocarse *detrás* del ordenador y emitir obras literarias que simulen estilos de autorías canónicas, que contengan registros de sesgos o que, en general, aborden la creación literaria necesariamente a través de la asistencia de una máquina?

Estas son apenas algunas de las cuestiones que permite plantear la crítica cultural aquí expuesta. En otro de sus ensayos, Max Horkheimer (1970) dedica un pasaje especialmente revelador a la relación entre las máquinas y los seres humanos, que puede servir de punto de partida para abordar estas preguntas. Allí reflexiona sobre el trabajo humano y la idea de liberar a las personas de sus tareas a través de las máquinas. La misma lógica puede aplicarse al trabajo creativo:

> Si es cierto que hoy se ha tornado real el sueño de que las máquinas realicen las tareas humanas, los humanos obran cada vez más y más como máquinas … Pese a toda su actividad, los hombres se tornan más pasivos; pese a todo su dominio de la naturaleza, se vuelven más impotentes frente a la sociedad y a sí mismo. (p. 30)

Parecería, por un momento, que todos estos conceptos enlazados —*razón instrumental*, *simulacro*, *hiperrealidad*, *pérdida del aura* y el sueño cumplido del ser humano de ver a las máquinas tomando control de ciertos sectores de su dominio— convergen hacia ideas aún más profundas, que revelan, como ya se sugirió, una crítica cultural que debe ser más amplia y que podría llevarnos a detectar problemáticas más tajantes.

Byung-Chul Han (2022) lo expresa de un modo semejante, al trasladar la problemática de las IA a campos más amplios y, si cabe, más preocupantes, como el del poder ejercido a través del dominio de la información:

> Llamamos *régimen de la información* a la forma de dominio en que la información y su procesamiento mediante algoritmos e inteligencia artificial determinan de modo decisivo los procesos sociales, económicos y políticos […]. El factor decisivo para obtener el poder no es ahora la posesión de medios de producción, sino el acceso a la información, que se utiliza para la vigilancia psicopolítica y el control y pronóstico del comportamiento. El régimen de la información está acoplado al del capitalismo de la información que hoy deviene en un capitalismo de vigilancia y que degrada a las personas a la condición de datos y ganado consumidor. (p. 9)

Dicho esto, parece entonces que una crítica cultural debería ir aún más allá del texto que la crítica ética, y observar el fenómeno de las IA con un lente mucho más amplio, incluso global. En ello radicaría el verdadero valor de una crítica cultural y filosófica. No obstante, los textos generados mediante las IAG son precisamente el detonante que activa tan importantes reflexiones.

El criterio ético

Establecer un criterio ético permite abordar otros riesgos y desencuentros entre las disciplinas de las ciencias y las humanidades. Incluso cuando pudiera creerse que las leyes de regulación serían lo más adecuado para atender problemas presentes y futuros, conviene recordar que las regulaciones tienden a caer en la obsolescencia. En este caso en particular, creemos que, según lo que se evalúa

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

de los problemas planteados —más allá de las demandas y protestas, que tienen más relación con los derechos de autor y los plagios—, habría que considerar lo expuesto ya en los puntos anteriores, de manera que dichas cuestiones desemboquen en esta crítica desde la ética: los problemas de los sesgos, la percepción de que existe una posibilidad de superación creativa de los autores humanos —fenómeno que toca varios puntos relacionados con el *AI Washing*— y la utilización de las IA como un fin antes que como un medio, como se ha observado tras la reflexión sobre el concepto de razón instrumental como criterio. De este modo, establecer un marco ético nos ofrece una dirección sobre cómo deberíamos relacionarnos con las inteligencias artificiales ante esta contingencia tecnológica.

El criterio ético por el que se opta se encuentra en la propuesta del libro *Human-centered* AI. Ben Shneiderman, su autor, toma como base la idea del control humano sobre los sistemas de IA, junto con otros postulados que incluyen una base teórica denominada *estructuras de gobernanza*, un modelo de gestión de proyectos gubernamentales.

Pero antes de esta propuesta de Shneiderman, habría que considerar que ha existido también algo llamado *Friendly AI*, que bien ha podido derivar en varias reflexiones dentro de un campo más amplio conocido como *ética de máquinas*, del que James Moor es uno de sus mayores exponentes. La IA *amigable* conceptualiza que la inteligencia artificial debería incluir en su diseño y funcionamiento el deseo de no causar ningún daño a los seres humanos, sea del modo que fuere, incluso manteniendo conciencia de que esta suerte de decreto debe evolucionar a la par con la propia IA:

> La amabilidad (el deseo de no dañar a los humanos) debería estar incorporada en el diseño desde el principio, pero los diseñadores deberían reconocer que sus propios diseños pueden fallar y que el robot aprenderá y evolucionará con el tiempo. Por lo tanto, el desafío es de diseño de mecanismos: definir un mecanismo para que los sistemas de IA evolucionen bajo un sistema de controles y contrapesos, y darles funciones de utilidad que sigan siendo amigables ante tales cambios.
>
> No podemos simplemente darle a un programa una función de utilidad estática, porque las circunstancias y nuestras respuestas deseadas a las circunstancias cambian con el tiempo.[11] (Russel y Norvig, 2010, p. 1039)

Como dijimos, esta primera pauta es conceptual, pero útil: reconoce la necesidad de que las IA mantengan una verificación de este concepto de *amabilidad* hacia el ser humano mientras evolucionan. Toda esta cadena de

11 En el original: "Friendliness (a desire not to harm humans) should be designed in from the start, but that the designers should recognize both that their own designs may be flawed, and that the robot will learn and evolve over time. Thus, the challenge is one of mechanism design—to define a mechanism for evolving AI systems under a system of checks and balances, and to give the systems utility functions that will remain friendly in the face of such changes.
We can't just give a program a static utility function, because circumstances, and our desired responses to circumstances, change over time".

marcos éticos —que va desde la ética de máquinas, que reconoce la necesidad de contar con una ética para salvaguardar la integridad de los seres humanos, y pasa por la idea de la IA *amigable*, que por momentos se manifiesta demasiado abstracta— desemboca en la HCAI (inteligencia artificial centrada en el ser humano), que emite propuestas todavía más puntuales.

En su libro, Shneiderman (2022) no pasa por alto los señalamientos de la IA *amigable*: también propone soluciones específicas para el diseño de algunos sistemas inteligentes mediante el método que denomina *design metaphors* (p. 10). No obstante, más allá de la fuerza de esta idea emparentada con la IA *amigable*, lo que nos interesa directamente es el principio central de la HCAI: que los seres humanos deben mantener el control principal sobre los sistemas inteligentes. Explica el autor:

> Este libro propone una nueva síntesis en la que los algoritmos inteligentes basados en IA se combinan con el pensamiento centrado en el ser humano para crear [la] HCAI. Este enfoque aumentará la posibilidad de que la tecnología empodere a las personas en lugar de reemplazarlas. En el pasado, los investigadores y desarrolladores se centraban en crear algoritmos y sistemas de IA, haciendo hincapié en la autonomía de las máquinas y midiendo el rendimiento de los algoritmos. La nueva síntesis presta la misma atención a los usuarios humanos y a otras partes interesadas al aumentar el valor del diseño de la experiencia del usuario y al medir el rendimiento humano. Los investigadores y desarrolladores de sistemas de IA en el ámbito de la salud valoran el control humano significativo, poniendo a las personas en primer lugar al servir a valores humanos como los derechos, la justicia y la dignidad, y apoyando objetivos como la autoeficacia, la creatividad, la responsabilidad y las conexiones sociales.
>
> Esta nueva síntesis refleja el creciente movimiento para expandir el pensamiento centrado en la tecnología a fin de incluir aspiraciones centradas en el ser humano que destaquen el beneficio social. El interés en la HCAI se ha fortalecido desde la Declaración de Montreal de 2017 para el Desarrollo Responsable de la IA.[12] (p. 7)

Pero no se trata solo de actuar en beneficio de la experiencia del usuario —y de favorecerlo—, sino también de una cuestión de responsabilidad: que un usuario o una compañía desarrolladora que cuente con

12 En el original: "This book proposes a new synthesis in which AI-based intelligent algorithms are combined with human-centered thinking to make HCAI. This approach will increase the chance that technology will empower rather than replace people. In the past, researchers and developers focused on building AI algorithms and systems, stressing machine autonomy and measuring algorithm performance. The new synthesis gives equal attention to human users and other stakeholders by raising the value of user experience design and by measuring human performance. Researchers and developers for HCAI systems value meaningful human control, putting people first by serving human values such as rights, justice, and dignity, and supporting goals such as self-efficacy, creativity, responsibility, and social connections.
This new synthesis reflects the growing movement to expand from technology-centered thinking to include human-centered aspirations that highlight societal benefit. The interest in HCAI has grown stronger since the 2017 Montreal Declaration for Responsible Development of AI".

No. 34, julio-diciembre de 2025. e22825
ISSN: 2011-804X · e-ISSN: 2462-8441

el control de un sistema inteligente tenga también una responsabilidad proporcional sobre este.

De aplicarse este marco ético como criterio, se buscaría una verificación puntual: las obras creativas de las IA deberían atribuirse a sus autores humanos o, cuando menos, aclararse la coautoría, aunque siempre atribuyendo un mayor peso autoral al ser humano, incluso cuando la IA haya tenido un papel predominante en el proceso de creación del producto final. Esto permitiría facilitar el abordaje de las problemáticas jurídicas, por ejemplo. Así, lo que hoy deriva en problemas de derechos de autor y plagio —por mencionar solo dos de los conflictos relacionados con las creaciones de las IAG— podría abordarse con las compañías desarrolladoras de las tecnologías en cuestión. Y quizá estas, en un actuar igualmente ético, puedan amplificar las nociones implícitas en la IA *amigable*, junto con el enfoque de la HCAI, y ajustar el diseño de sus tecnologías para reducir futuros resultados negativos.

Es evidente que existen otros tipos de riesgos que un marco ético como el de la HCAI quizá no pueda atender. En cuanto a los *bias* y al criterio de la razón instrumental extrema —ambos aunados al problema de la (hiper)reproductibilidad técnica y la masificación de la obra de arte—, estaríamos hablando de retos de índole, quizá, epistemológica, para lo cual habría que pensar en un futuro próximo las posibilidades de atender dicho vacío.

De momento, la HCAI debería entenderse también dentro de marcos más amplios, como el de una ética del cuidado, por ejemplo, que contemple el cuidado tanto de uno mismo como de los demás en el actuar y en la toma de decisiones; y el de una ética de la responsabilidad que considere la obligación que las generaciones presentes tienen hacia las futuras.

De este modo, la acción ética de otorgar el control y la responsabilidad a los seres humanos sobre las acciones de las máquinas —de las IAG—, aplicada como criterio para las creaciones literarias de estas tecnologías, contribuiría a proporcionar una nueva perspectiva para valorar tales producciones. Todos los posibles sesgos, pero también las hazañas creativas —o la ausencia de estas— implícitas en los textos, tendrían una conexión directa con los usuarios humanos que operan las inteligencias artificiales o con sus desarrolladores. Eso abriría nuevos derroteros a las críticas especializadas, permitiendo valorar —y, por ende, también cuestionar— la presencia de las inteligencias artificiales en la literatura, y más ampliamente en las humanidades. Con ello, la cuestión quizá se desplace aún más hacia lo filosófico, pero tal giro sería un resultado positivo, en tanto que nos permitiría plantear con mayor precisión preguntas como: "¿Cuál debería ser nuestra relación con estas tecnologías?".

Conclusiones

La búsqueda de criterios que hemos llevado a cabo no emite sus juicios de manera separada. Así, aunque hayamos dividido la labor en cuatro segmentos, no es difícil constatar cómo, de un criterio, se puede aludir a otro. De cierta manera, el criterio que engloba a los otros tres es el ético. Los criterios textuales abren la posibilidad de reflexión desde fuera del texto: permiten observar las problemáticas que plantean los criterios de la crítica cultural, y todas las problemáticas que evidencien los criterios precedentes encuentran un fin en el planteamiento ético. Así pues, estas dos categorías de crítica bien podrían conjuntarse en un modelo general de observación crítica del fenómeno de las IA en las humanidades. El método implicaría observar el texto bajo los criterios establecidos, y salir y volver a él constantemente, según lo remitan las críticas extratextuales.

De este modo se cumple la finalidad que hemos establecido al inicio: postular un aparato crítico especial que delibere sobre una problemática, relevar a una crítica literaria general de la labor falible de abordar, con los mismos criterios con que se abordan fenómenos humanos, las creaciones de las IAG. Creemos que es la única manera de mantener a la vanguardia a la literatura, a las humanidades, ante un acontecimiento tan relevante y explosivo como lo es la participación de estas tecnologías en el interior de su esfera. Pero ello no representa el cierre de las posibilidades de acceso: no es sino la toma de una postura. Todavía es posible discernir la manera de utilizar estas tecnologías a favor de la literatura, para el análisis o la creación. Eso, sin dudas. Pero, insistimos, el primer paso es siempre tomar una postura; luego, formular un aparato crítico sobre la base de esta postura. Es la tarea que creemos haber cumplido en este trabajo.

Referencias

Adamivna, I. (2019). The Art of Narration and Artificial Narrative Intelligence: Implications for Interdisciplinary Research. *Journal of Narrative and Language Studies*, *7*(13), 309-318. https://www.nalans.com/index.php/nalans/article/view/197?utm

Baudrillard, J. (1977). *Cultura y simulacro*. Kairós.

Baudrillard, J. (1985). El Xérox y el infinito. *Revista de Occidente*, (113), 5-14.

Bender, E., Gebru, T., McMillan-Major, A. y Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be too Big? En M. Elish, W. Isaac y R. Zemel (eds.), *Proceedings of the 2021* ACM *Conference on Fairness, Accountability, and Transparency* (pp. 610-623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Benjamin, W. (2003). *La obra de arte en la época de su reproductibilidad técnica*. Ítaca.

Boden, M. (1998). Creativity and Artificial Intelligence. *Artificial Intelligence*, *103*(1-2), 347-356. https://doi.org/10.1016/S0004-3702(98)00055-1

Dijk, T. van. (2005). *Estructuras y funciones del discurso*. Siglo XXI.

Eco, U. (1987). *Lector in fabula*. Lumen.

Elfar, M. y Dawood, M. (2023). Using Artificial Intelligence for Enhancing Human Creativity. *Journal Of Art, Design And Music*, *2*(2), 106-120. https://doi.org/10.55554/2785-9649.1017

Han, B-C. (2022). *Infocracia*. Taurus.

Horkheimer, M. (1970). *Sobre el concepto del hombre y otros ensayos*. Editorial Sur.

Horkheimer, M. (1973). *Crítica de la razón instrumental*. Editorial Sur.

Kaku, M. (2019). *El futuro de la humanidad*. Debate.

Lagos, A. (21 de agosto de 2024). AI *Washing*, el engaño detrás del boom de la inteligencia artificial. *Wired en Español*. https://es.wired.com/articulos/ai-washing-el-engano-detras-del-boom-de-la-inteligencia-artificial

Minsky, M. (1975). A Framework for Representing Knowledge. En P. Winston (ed.), *The Psychology of Computer Vision* (pp. 211-276). McGraw-Hill.

Morales, B. (1997). La lingüística en el contexto de la inteligencia artificial. *Forma y Función*, (10), 25-50. https://revistas.unal.edu.co/index.php/formayfuncion/article/view/17072

Russel, S., Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Pearson.

Sadin, É. (2020). *Inteligencia artificial o el desafío del siglo*. Caja Negra.

Searle, J. (2009). The Turing Test: 55 Years Later. En R. Epstein, G. Roberts y G. Beber (eds.), *Parsing the Turing Test: Philosophical and methodological Issues in the Quest for the Thinking Computer* (pp. 139-150). Springer.

Shneiderman, B. (2022). *Human centered* AI. Oxford University Press.