

**DEL MACHINE LEARNING AL APRENDIZAJE DE TALLER LITERARIO.
HACIA UNA METODOLOGÍA DE LAS RELACIONES ENTRE
FICCIÓN LITERARIA Y GRANDES MODELOS DE LENGUAJE**

FROM MACHINE LEARNING TO LITERARY WORKSHOP LEARNING.
TOWARDS A METHODOLOGY OF THE RELATIONSHIPS BETWEEN
LITERARY FICTION AND LARGE LANGUAGE MODELS

Miguel AMORES

Universidad de Burgos

mamores@ubu.es

Resumen: Este artículo propone una reflexión metodológica y epistemológica sobre el modo de concebir y evaluar la capacidad de escribir textos literarios ficcionales de los grandes modelos de lenguaje, o *large language models*, en inglés (LLM). Comenzamos describiendo algunas de las dificultades estructurales de esta relación, que se resumen en el antagonismo entre la naturaleza autorregresiva de los LLM —que los lleva a elegir unas palabras sobre otras por una cuestión de probabilidad estadística— y el carácter asistemático de la escritura literaria de ficción. Posteriormente justificamos que la mejor forma de comprobar el potencial de los LLM para esta tarea es una metodología de *prompt* múltiple expresada en términos literarios —es decir, subjetivos y asistemáticos—. Por último, se analizan los resultados de un experimento que siguió esta premisa, y en el que se apreció una clara mejoría entre la primera versión y la quinta del texto, entre las que mediaron cuatro rondas de *prompts* con veintinueve correcciones concretas.

Palabras clave: Grandes modelos de lenguaje. Aprendizaje automático. Inteligencia Artificial Generativa. Literatura.

Abstract: This article proposes a methodological and epistemological reflection on the conceptualization and evaluation of the capacity of Large Language Models (LLMs) to generate fictional literary texts. It, first, describes some structural difficulties of this relationship, summarized in the antagonism between the autoregressive nature of LLMs —which leads them to choose certain words over others due to a matter of statistical probability— and the unsystematic nature of fictional literary writing. Secondly, the article proposes that the best way to test the potential of LLMs for this task is a multiple-prompt methodology expressed in literary terms —that is, subjective and unsystematic. In an experiment following this premise, the results show a clear improvement between the first and the fifth version of the text, with four rounds of prompts and twenty-nine specific corrections.

Keywords: Large Language Models. Machine Learning. Generative Artificial Intelligence. Literature.

1. INTRODUCCIÓN

El auge de las diversas herramientas de IA generativa está suponiendo una revolución social integral a la que los estudios literarios no son inmunes; todo lo contrario. Y la razón profunda es que esta tecnología, sobre todo a través de sus manifestaciones más populares, como los *chatbots* tipo ChatGPT, amenaza seriamente una de las premisas básicas sobre las que la crítica, con independencia de su orientación exacta, se había basado hasta hace muy poco: que todo texto literario era una creación exclusivamente humana¹. Esto explica que en los últimos años se hayan publicado numerosas investigaciones que, desde distintas perspectivas, tratan de evaluar las verdaderas capacidades de la inteligencia artificial para con la escritura literaria².

El problema, sin embargo, como en toda cuestión realmente transversal y trascendental, no es solo que cualquier respuesta a este tema esté lejos de ser sencilla o inequívoca; el verdadero problema es que hay tantas formas de abordar este asunto que simplemente dar con un planteamiento que sea del todo consciente de la complejidad del objeto de estudio, con independencia de lo acertado o no de su desarrollo posterior, es ya una tarea titánica. Por ello, el objetivo de este artículo no es tanto ofrecer respuestas definitivas sobre la cuestión mencionada como proponer un enfoque analítico que pueda asumir de la forma más rigurosa posible los términos exactos en los que debe concebirse y evaluarse el potencial de la IA generativa en relación con la ficción literaria narrativa³. En este sentido, consideramos adecuado considerar este análisis como una suerte de exploración metodológica y epistemológica.

¹ En realidad, desde hace ya algunos años —como mínimo desde 2018, cuando apareció la primera versión de ChatGPT— podemos afirmar que la literatura ha dejado de ser un proceso exclusivamente humano. Se trata de algo sobre lo que no existe duda si tenemos en cuenta el uso auxiliar de la IA generativa por parte de algunos escritores, en procesos como la generación de texto para superar bloqueos creativos (Zhu *et al.*, 2024: s. p.). En lo que respecta a la elaboración de textos por parte de programas de IA generativa y con una participación humana marginal, o a lo más secundaria (lo que Ledesma [2024: 204] denomina “creación asistida inversa” y otros autores “machine-in-the-loop” [Yupeng *et al.*, 2024: s. p.]), podemos hablar de la existencia de textos completos de nuevo al menos desde 2018, con la aparición de *I the Road*, considerada una de las primeras novelas creadas casi íntegramente con IA generativa (Merchant, 2018). Sin embargo, hasta el momento la atención de la crítica académica a los textos literarios generados exclusivamente por IA ha sido escasa.

² Nos referimos a trabajos que abarcan de manera específica el potencial literario de los LLM, como los de Mikros (2025); Marco *et al.* (2024 y 2025); Bellemare-Pepin *et al.* (2024); Guo *et al.* (2024), Garrido-Merchán *et al.* (2023), Chakrabarty *et al.* (2023 y 2024) o Gómez-Rodríguez y Williams (2023). Para ver una revisión actualizada de las investigaciones sobre las capacidades de los LLM tanto para el lenguaje literario como para el lenguaje humano en general ver Mikros, 2025: 588-590.

³ Las conclusiones de este experimento se limitan al potencial de los LLM con respecto a los textos literarios narrativos de tipo ficcional, que en virtud de tal ficcionalidad serían aún más asistemáticos y ofrecerían especiales dificultades a la IA generativa (ver secciones 2 y 3) que los textos literarios narrativos factuales, como ciertos ensayos o biografías. En este artículo siempre que citemos el término *literatura* nos estaremos refiriendo a *texto literario ficcional de tipo narrativo*.

2. SOBRE LA NATURALEZA GENERATIVA DE LOS LLM Y DE LA FICCIÓN LITERARIA

Dentro del vasto universo de la IA generativa, en este artículo nos limitaremos a sus manifestaciones más comunes, los *chatbots* basados en grandes modelos de lenguaje (*Large Language Models*, LLM) que se articulan, a su vez, a través de redes neuronales que procesan textos de Internet sin etiquetar a través de técnicas de *machine learning*, autosupervisado o semisupervisado (Radford *et al.*, 2018: 1-2). Y, dentro de estos programas, basaremos nuestro análisis en el que se considera más avanzado, ChatGPT, creado por la empresa Open AI. En concreto, usaremos su versión más reciente en el momento de escribir estas líneas, ChatGPT-4o⁴.

Sin embargo, antes de proceder a explicar nuestros presupuestos metodológicos y epistemológicos y de describir también los resultados de nuestro experimento, creemos necesario hacer una serie de precisiones sobre la naturaleza básica del funcionamiento de este tipo de programas. Y lo creemos importante porque, a nuestro juicio, en demasiadas ocasiones los estudios que evalúan las capacidades de los LLM para con la literatura no prestan atención a una serie de elementos que apuntarían a ciertos antagonismos esenciales entre la IA generativa actual y el texto literario. Y con ello no nos referimos solo al hecho de que tales sistemas, al no ser humanos, carezcan evidentemente de algunos de los atributos —subjetividad, sentimientos, experiencias vitales directas, identidades dinámicas, evoluciones subjetivas, vivir inmersos en una serie de convenciones socioculturales concretas, etc.— (Zhu *et al.*, 2024: s. p.), que tradicionalmente se han relacionado con la esencia última de la literatura. Hablamos de una serie de grandes antagonismos de tipo estrictamente técnico que afectan a la capacidad de los LLM de escribir buenas ficciones literarias; unos LLM que, a pesar de que aparentan tener una capacidad casi mágica para elaborar todo tipo de textos en segundos y con una calidad generalmente aceptable⁵, están diseñados para llevar a cabo una serie de tareas textuales en detrimento de otras.

⁴ A pesar de que existen otros muchos LLM comerciales distintos de ChatGPT —Claude, Gemini, Copilot, etc.—, y que incluso hay investigaciones que se fundamentan en comparar sus diferentes capacidades para con la escritura literaria (Bellemare-Pepin *et al.*, 2023; Gómez-Rodríguez & Williams, 2023), aquí nos limitaremos al modelo de Open AI. Más allá de que se trata del LLM actual más relevante, la razón de limitarse a él es que, a día de hoy, todos estos programas poseen la misma tecnología básica —el tipo de arquitectura neuronal conocida como *transformador* (“A Comprehensive Comparison of All LLMs”, 2024)—, al punto de que la única diferencia estructural entre ellos es la cantidad y el tipo de entrenamiento que incorporan —un dato, por otra parte, que, dada la naturaleza de “caja negra” del funcionamiento de los LLM, desconoce el público en general, entre los que por supuesto se encuentran los investigadores (Kosinski, 2024)—.

⁵ La evaluación de las capacidades de los LLM para los textos creativos es un ámbito en construcción en el que aún faltan consensos claros (Yupeng *et al.*, 2024; Chakrabarty *et al.*, 2023 y 2024; Lamb *et al.*, 2018). Sin embargo, de los estudios que comparan la capacidad para la escritura creativa de LLM y seres humanos, estos últimos aún conservan su superioridad —al menos, con respecto a la mayoría de LLM— (Gómez-Rodríguez & Williams, 2023; Marco *et al.*, 2024 y 2025).

La reflexión de McCoy *et al.* (2024), a pesar de que no está enfocada a evaluar las habilidades literarias de los LLM, resulta clarificadora al respecto. Estos autores recuerdan que ChatGPT, lo mismo que otros grandes modelos de lenguaje comerciales similares, son ante todo programas basados en una lógica autorregresiva que se proyecta a partir del conocimiento que le proporcionan textos de Internet, a los que toman como referencia y descomponen en términos de patrones (Radford *et al.*, 2018). Esto, en esencia, quiere decir que se trata de sistemas de generación textual cuyo núcleo de funcionamiento es la probabilidad estadística de generar una palabra que tenga sentido con respecto a la anterior (“statistical next-word prediction systems” [McCoy *et al.*, 2024: s.p.]). Lo dicho implica no solo que estos programas no *razonen* en un sentido humano, o que realmente no entiendan los términos que ponen en juego —pues, en puridad, lo único que hacen es manejar posibilidades que identifican como correctas o incorrectas; se trata de una carencia que los autores denominan “difficulty on tasks that depend on meaning” (McCoy *et al.*, 2024: s.p.)—. Tampoco nos referimos solo a otras dificultades como la incapacidad de cambiar el texto generado si no es a través de un nuevo *prompt*, o los problemas que acusan estos programas en lo relativo a la memorización idiosincrática —y que los lleva en ocasiones a repetir frases y expresiones—. La clave estaría en que, dado que lo que se busca es que estos programas tengan el mejor rendimiento posible, pero teniendo en cuenta también que la posibilidad de entrenamiento supervisado y de optimización de tareas siempre es limitada, lo que hacen los diseñadores es concentrar esfuerzos en aquellas tareas que estadísticamente es más probable que se le pidan a los LLM. Esto implica que, a pesar de la apariencia de poder omnímodo, los grandes modelos de lenguaje están mucho mejor entrenados para llevar a cabo ciertas tareas textuales que otras. Y a ello hay que añadir, además, que ciertos tipos de textos y procedimientos de Internet —que, como se ha dicho, estos programas descomponen en patrones de generación textual y les sirven como referencia— son más habituales que otros, lo que generaría todo tipo de sesgos en función de la sobrerrepresentación e infrarrepresentación de tales contenidos en la Red.

Todo lo anterior lleva a los autores a sostener que, de manera general, los LLM tendrán un mejor desempeño en tareas de generación textual que estén sujetas a una alta probabilidad —y que paralelamente estén respaldadas por una gran cantidad de contenidos en Internet que les puedan servir de referencia—, y baja en aquellas en las que ocurra lo contrario. Para demostrar esta hipótesis McCoy *et al.* ponen como ejemplo una serie de 13 tareas sencillas pero anómalas —en el sentido de *estadísticamente improbables*— de tipo matemático, lógico o sintáctico. Una de ellas evaluaba el rendimiento de ChatGPT en relación con un método criptológico tan rudimentario como el cifrado por sustitución. Según los autores (McCoy *et al.*: 9), en Internet, el valor de rotación 13 —es decir, que cada letra sufre un desplazamiento alfabético de 13 caracteres, de forma que la *a* sería la *n* y una palabra como *casa* se escribiría *onfn*— es hasta 100 veces más habitual que el de valor 12. De este modo, cuando los investigadores le pidieron a ChatGPT que tradujera una serie de fragmentos con una rotación de valor 12,

descubrieron que se equivocaba en mucha mayor medida... sin que en términos reales la dificultad de la tarea hubiera variado.

Este sencillo experimento respaldaría su tesis de hasta qué punto el acierto de este tipo de programas, incluso en tareas sistemáticas y relativamente sencillas, depende no tanto de lo que podríamos llamar *dificultad intrínseca de una tarea*, sino del hecho de que el LLM haya recibido entrenamiento supervisado a ese respecto y que en Internet exista un número suficiente de textos —con sus correspondientes patrones generativos— análogos al de aquel que se pretende obtener. Ahora bien, si se da esto por bueno, ¿en qué situación estaría la capacidad de elaboración de textos de la IA generativa con respecto a la ficción literaria?

En nuestra opinión, el problema fundamental, el antagonismo básico entre IA generativa y literatura, vendría del hecho de que la primera, como se ha dicho, sigue una lógica autorregresiva, es decir, que crea sus textos en función de la probabilidad de la siguiente palabra, mientras que la literatura se muestra refractaria —o directamente incompatible— con respecto a esta lógica meramente probabilística y muchas veces *asemántica*. Y es que, sea cual sea el prisma desde el que observemos los textos literarios, y con independencia de a qué tipología concreta nos refiramos, lo menos que se puede decir es que su generación no se rige por esquemas probabilísticos estables. Y para ello no sería necesario estar de acuerdo con la frase comúnmente atribuida a Valle Inclán, según la cual la literatura es poner juntas palabras que normalmente comparecen por separado (Bosque, 2001: 31). Bastaría con mostrarse de acuerdo con un principio literario tan conservador como el del *extrañamiento*, formulado por Shklovski (1991), según el cual la literatura supone un tipo de comunicación que, en virtud del marcado carácter autotélico de sus signos, supone una comunicación alterada y extraña que “desautomatiza” los cauces habituales del entendimiento verbal. Si se acepta este principio, habría que concluir necesariamente que la esencia profunda e inevitable del texto literario es aspirar a ser un tipo de comunicación singular, extrañante, lo que necesariamente pasa por proponer cadenas de palabras poco probables desde un punto de vista estadístico.

Pero dado que no se puede afirmar que en la Red no exista un número suficiente de textos literarios de todo tipo que les puede servir de referencia, las dificultades de la IA generativa para con la ficción literaria también se deben explicar por el hecho de que el texto literario, dependiendo de cuál sea su tipo, es entre *dudosamente compatible* y *radicalmente contrario* a ser reducido a patrones de generación realmente sistemáticos.

Antes de seguir consideramos que es necesario aclarar este último aserto, pues de él va a depender en buena medida nuestra argumentación posterior. A primera vista podría decirse que es como mínimo discutible, pues a pesar de que las obras literarias están sujetas a una lógica sistemática mucho menos férrea que otros tipos de textos, siempre es posible encontrar en ellas una serie de patrones reconocibles y hasta cierto punto cuantificables —y, por tanto, susceptibles de ser asimilados por parte de los LLM—. A un nivel intuitivo, el hecho mismo de que, con independencia de las características

concretas de un determinado texto, seamos capaces de encuadrarlo dentro de categorías como *novela*, *relato*, *elegía* o *comedia de enredo* ya presupone que los lectores competentes manejan cierta estructura arquetípica —lo que Van Dijk (1980) denomina “superestructura”— en relación con los distintos tipos de obras literarias. Además, a un nivel teórico, disciplinas como la narratología —y especialmente la de raigambre estructuralista (VV.AA., 1982)— o la estilística computacional⁶, cada una en su ámbito correspondiente, también han determinado una serie de patrones que se repiten en las obras pertenecientes a la misma tipología literaria.

Sin embargo, consideramos que esta posibilidad de capturar la creación literaria —o siquiera la variante que aquí nos ocupa: la ficción literaria narrativa— en términos de patrones estructurales será siempre relativa si atendemos a la perspectiva de la generación textual práctica. Y la razón fundamental es que, a diferencia de otro tipo de discursos, en el literario no hay una conexión directa o sistemática entre la determinación de los elementos típicos —o incluso inherentes— de una determinada variedad discursiva y su ejecución textual práctica. Por ejemplo, el hecho de que según la teoría clásica de Propp (2014 [1928]), uno de los puntales de la narratología estructuralista, todas las variantes argumentales de los relatos del folclore ruso se pudieran capturar en 31 funciones narrativas no añade demasiada luz al hecho de cómo podría elaborarse un buen texto de ese tipo. Así, determinar que en muchos cuentos tradicionales puede distinguirse la función *prohibición* no ilumina de forma natural un procedimiento predeterminado o sistemático sobre cómo puede funcionar e interaccionar entre sí el conjunto de elementos que la hacen posible —personajes, trama, tono del texto, implicaciones morales implícitas, etc.—.

En cuanto a la estilometría computacional, otra de las disciplinas basadas en premisas que problematizan el citado aserto, también habría dificultades serias para sistematizar e instrumentalizar los posibles patrones del texto literario, a pesar de que en este caso sí que abordan esta escritura de un modo más directo y práctico. Esto sería así, en primer lugar, porque muchos de estos análisis se limitan a evaluar la capacidad de los LLM de imitar los estilos de ciertos autores. Se trata, por tanto, de una vía investigadora distinta a la de este trabajo, que parte de la premisa de que lo que realmente debe concentrar el interés investigador, lo que realmente puede cambiarlo todo en nuestro ámbito hasta dejarlo irreconocible, es que los LLM puedan llegar a escribir textos literarios de calidad de forma mínimamente autónoma, y no que lleguen a imitar ciertos textos al punto de hacerlos indistinguibles del original humano. A ello habría que añadir, además, que las investigaciones publicadas hasta ahora ofrecen resultados muy diversos acerca de la capacidad de los LLM de descomponer el texto literario en patrones sistematizables para replicar en toda su calidad y complejidad ciertos estilos literarios (Garrido-Merchán *et al.*, 2024; Mikros, 2025).

⁶ Véase también: Chakrabarty *et al.* (2024), Riemer & Peter (2024) o Hicke & Mimno (2023).

Consideramos que la clave acerca de la asistemacidad literaria que defendemos reside en algo que afirma John Gardner al comienzo de su manual de escritura creativa *El arte de la ficción*, al que luego nos referiremos. Allí sostiene que la experiencia le ha enseñado que, en lo relativo a la escritura de ficciones literarias, es posible fijar “unos cuantos principios generales” traducibles en “algunas advertencias” más o menos concretas (Gardner, 2001: 23). El problema, afirma el autor, es que el modo de existencia de estas pautas las hace poco compatibles con cualquier lógica sistemática. Y es que dichas pautas —como por ejemplo no defraudar las expectativas argumentales creadas— en efecto existen, pero lo hacen “a un nivel de abstracción tan elevado que no aportan ninguna pauta útil para un escritor” (Gardner, 2001: 23). Sostiene Gardner, además, que aun en el caso de que pudiéramos determinar una norma de escritura literaria de validez universal, siempre cabría la posibilidad de que se creara una obra que la contradijera punto por punto... y que sin embargo fuera juzgada como buena. O dicho con otras palabras: la normatividad literaria tiene una naturaleza tal que, aunque no sea lo habitual, potencialmente, el ejemplo siempre estará al mismo nivel que el contraejemplo. Es esto, de hecho, lo que explica el surgimiento de categorías como la llamada *unnatural narratology* (Biwu, 2015), destinada a estudiar ciertas posibilidades narrativas que la narratología tradicional no contemplaba —o lo hacía con grandes dificultades—: narradores imposibles, cronologías alteradas, mundos ficcionales contradictorios, etc.

Sobre esta naturaleza refractaria a las normas de la ficción literaria narrativa, en otros trabajos (Amores, 2018: 47-151) hemos propuesto una serie de reflexiones que en este punto creemos resultan pertinentes. Hemos sostenido que la ficción literaria se distingue del resto de discursos posibles, incluso de las ficciones no literarias, por el hecho de que, al no tener ninguna obligación de correspondencia rígida o sistemática con los contenidos o normas del mundo real, no posee tampoco ningún tipo de restricción enunciativa *a priori*, ni en cuanto a forma ni en cuanto a contenido. Dicho de forma más sencilla: al contrario que el resto de discursos imaginables, la ficción literaria puede expresar potencialmente lo que quiera y como quiera, mientras que los otros discursos, más o menos obligados a guardar una relación de correspondencia con las normas o los estados de cosas del mundo real, siempre encontrarán algún tipo de restricción relativa a la forma o al contenido en sus manifestaciones. Y es que hay muchas posibilidades enunciativas que, de forma apriorística, anularían de raíz la validez de textos como un informe policial o un relato histórico, como por ejemplo dar crédito a la existencia de hipogrifos, o expresarse mediante monólogos interiores. Pero no hay ninguna posibilidad formal o conceptual que, de antemano, esté prohibida en una novela; que la inhabilite en tanto que novela. Esto no quiere decir, según el autor, que literalmente *cualquier cosa* pueda considerarse ficción literaria; pero sí que debe admitirse que potencialmente la ficción literaria puede acoger cualquier posibilidad expresiva imaginable, o formularse a través de ella, con entera naturalidad, sin que ninguna le arrebate de forma previa su condición de ficción literaria, lo que ya de por sí supone una diferencia estructural con el resto de los discursos humanos.

Esta hipótesis nos lleva a formular una serie de tesis asociadas, y que resultan relevantes a la hora de explicar las dificultades de los LLM para la elaboración de buenos textos literarios. Porque lo anterior obliga a concebir la literatura no solo como un conjunto de textos que poseen una serie de características compartidas, sino adicionalmente como “una dinámica de generación de prototipos” (2018: 98-115). Y eso se debería a que el carácter potencialmente infinito de la enunciación literaria ficcional, unido a la exigencia estructural de originalidad, muy acusada en nuestra época contemporánea, convierten cada uno de sus textos en una creación que no se ajustaría de forma exacta a unas normas y lógicas predeterminadas, sino en algo que, de alguna manera, pugnaría por dotarse de sus propias normas constitutivas. De esta forma, un texto literario sería *prototípico* en el sentido de que supondría la última creación de una serie ya existente en la que en principio se encuadraría, pero con respecto a la cual pretendería simultáneamente encarnar una cierta ruptura; pretendería incorporar una serie de innovaciones que convertirían a dicho texto literario en singular y difícilmente reproducible, y que por tanto, en un plano normativo y ontológico, lo que buscaría sería no tanto —y no solo— constituirse como un caso novedoso dentro de la serie dada, sino ampliar con el hecho mismo de su existencia los límites de lo tenido por posible dentro de dicha serie. Ejemplificamos este concepto de prototipo con esas creaciones por lo común extravagantes conocidas como *concept cars*, que a menudo se exponen en salones de automóviles internacionales. El objetivo principal de los fabricantes no sería que estos modelos, en ocasiones estrafalarios o con características técnicas imposibles de producir en masa, sirvieran de adelanto a lo que podría ser un modelo de venta al público; el verdadero objetivo sería exhibir el potencial de su imaginación con respecto al concepto *coche*, ampliarlo en términos ontológicos, sin las tradicionales ataduras que impone la lógica comercial e industrial.

Por todo ello, tanto el carácter profundamente subjetivo de los consejos —que no reglas fijas— como la concepción prototípica de los textos literarios —con su obligación estructural de innovación o incluso de disrupción— no hacen sino subrayar su inadecuación con respecto a la lógica probabilística de la IA generativa —por más que puedan existir ciertos patrones narratológicos o estilométricos en relación a elementos literarios concretos—; no dejan de destacar la dificultad de diseñar protocolos de entrenamiento para LLM para ejecutar la tarea, en cierto modo paradójica, de elaborar patrones de generación textual para un tipo de escritura basada en una exigencia estructural de singularidad.

3. METODOLOGÍA DE TALLER DE ESCRITURA CREATIVA PARA LA GENERACIÓN ARTIFICIAL DE FICCIÓN LITERARIA

Dado que muchas de las investigaciones que tratan de evaluar la capacidad de la IA generativa para la creación literaria no tienen en cuenta alguno —o incluso ninguno— de los aspectos descritos, aquí quisimos proponer una metodología que nos parecía más

ajustada a la naturaleza de los elementos en juego. Para ello nos decantamos por un enfoque (hiper)cualitativo que, a pesar de contar con una serie de limitaciones⁷, pudiera asumir de la forma lo menos imperfecta posible tanto las singularidades epistemológicas de la literatura como las enormes dificultades que los LLM poseen para con esta debido a su consustancial lógica autorregresiva.

Nuestra primera decisión fue que, aunque nos mantendríamos dentro de la lógica de la creación asistida inversa o “machine-in-the-loop” (ver nota 1), sería necesaria una amplia interacción con ChatGPT en forma de numerosos *prompts*⁸. Con esta fórmula (“numerosos prompts”) nos referimos a algo que no termina de encajar en las tipologías habituales de técnicas de *prompting* basadas en arquitecturas de *prompt* múltiple, desde el *prompt chaining* hasta el *multimodal CoT prompting*⁹. El motivo es que estas técnicas —que en absoluto fueron concebidas para la generación de textos literarios— están centradas casi exclusivamente en el contenido de los mensajes —o, a lo sumo, en características de su exposición—; se basan por tanto en un enfoque que resulta insuficiente para la literatura, cuyos textos jamás son puramente instrumentales, pues su naturaleza se aproxima mucho más a lo autotélico, al predominio de lo que Jakobson denominó “función poética” (1960). Lo que nosotros pretendíamos era diseñar un proceso de introducción de instrucciones que se ajustara a las singulares características de la literatura, y no solo en el sentido de que se le prestara una especial atención a la forma. Por estos motivos, lo primero que había que hacer era corregir una tendencia habitual en este campo investigador, la de juzgar los textos literarios escritos por los LLM que son fruto de un solo *prompt*, por sofisticado o complejo que este pueda ser¹⁰. Partíamos de la base de que evaluar la capacidad literaria de los LLM de este modo es engañoso, porque en ese caso el texto resultante suele ser de muy baja calidad. Dichos textos literarios tienen tendencia a que cada contenido esté reflejado de forma explícita o sobreexplicada; a estar llenos de clichés, tanto expresivos como narratológicos; a presentar una caracterización superficial de los personajes...

⁷ Nos referimos a elementos como que nuestra investigación se limita a comprobar las capacidades literarias de Chat GPT 4o, sin atender a otros LLM, tanto comerciales como especializados; o que analizamos un único texto (un relato breve), por lo que no se tienen en cuenta ni las capacidades de ChatGPT con otros textos del mismo tipo ni de tipología diferente. Se trata de carencias de las que somos conscientes y que trataremos de paliar en futuras investigaciones, pero que en buena medida se justifican por el carácter hipercualitativo de nuestro análisis (ver Conclusiones).

⁸ En términos de *prompt engineering*, la metodología que se ha aplicado es algo entre medias entre el *zero shot* y el *few-shot prompting*.

⁹ Una metodología que en los últimos años ha ofrecido prometedores avances en las investigaciones acerca de las capacidades literarias de los LLM ha sido la *Retrieval Augmented Generation* (RAG), enfocada a que estos dispositivos dispongan de mejores corpus a la hora de elaborar sus textos (Neelakanteswara *et al.*, 2024). En esta investigación es algo que no se ha aplicado debido a que, como se argumentó en la sección anterior, parece algo más enfocado a meramente replicar los estilos de un determinado autor que a evaluar la capacidad de los LLM de crear literatura de la forma más autónoma posible, que es lo que concentra nuestro foco de interés.

¹⁰ A este respecto, resulta modélica la investigación de Garrido-Merchán *et al.* (2023), donde se elaboran bajo minuciosos criterios literario-filológicos unos *prompts* realmente extensos que cumplen el objetivo de reproducir unos textos casi indistinguibles con respecto a los originales de H. P. Lovecraft.

Antes de continuar con nuestro razonamiento y con la eventual descripción del experimento y sus resultados, creemos que se hace imprescindible una aclaración. El motivo es que antes se ha señalado que una de las mayores dificultades de escritura de las ficciones literarias, tanto para los humanos como —por motivos diferentes— para los LLM, es la ausencia de normas y criterios claros acerca de cómo se establece la calidad de un determinado texto. En nuestro experimento, sin embargo, actuamos como si dichas normas y criterios sí existieran; el hecho mismo de su planteamiento, el pretender *corregir* un primer texto generado por ChatGPT a través de sucesivas rondas de *prompts* con indicaciones concretas, implica necesariamente que sí creemos en al menos ciertas pautas de qué es un buen texto literario y qué normas sigue. Justificamos, sin embargo, esta evidente contradicción de la forma siguiente.

Cuando afirmamos que la ficción literaria es refractaria a los esquemas estables, al punto de ser prototípica, y que por ello resulta difícil establecer para ella valores estables de calidad que sean asequibles a la lógica probabilística de los LLM, nos estamos refiriendo a la literatura como sistema y como conjunto heterogéneo formado por millones de obras escritas a lo largo de miles de años. Que obras tan distintas como *La Iliada*, *El Quijote* o *Las flores del mal* sean consideradas obras maestras sería prueba de la extrema dificultad de encontrar unos valores de calidad objetivos y estables. Sin embargo, en el análisis práctico, y conforme vamos aumentando el grado de concreción genérica e histórica, sí que se pueden apreciar patrones compartidos entre los textos; y, con ello, de criterios de calidad —y de ausencia de ella— más o menos ciertos.

Con esto en mente, y teniendo en cuenta que lo que le pedimos a Chat GPT es la creación de un relato breve, las pautas de calidad literaria que han guiado nuestras correcciones las hemos tomado del manual de escritura creativa de John Gardner *El arte de la ficción*. Somos plenamente conscientes de que ni este manual —a pesar de ser uno de los más extendidos— ni ningún otro de este tipo posee las claves absolutas de la calidad literaria. Pero creemos que puede resultar válido, al menos en este punto de nuestras investigaciones, por dos motivos. El primero, que propone una serie de principios más o menos específicos que se pueden traducir en una serie de recomendaciones —y *prompts*— concretos. Así, las posibles dudas que este tipo de obras pudieran despertar se verían compensadas —al menos en parte— por el modo en que favorece la operatividad de las investigaciones. Y, en segundo lugar, porque este tipo de obras, al tener una finalidad didáctica, están fuertemente enfocadas a evitar que el escritor principiante cometa una serie de errores considerados como básicos, lo cual encaja bien con la actual posición de los LLM con respecto al desempeño literario. Somos conscientes de que estos motivos —operatividad y adecuación a una supuesta circunstancia de los LLM— no parecen las bases metodológicas más sólidas. Pero dado que de lo que hablamos no es de *análisis* de textos literarios, sino de *las claves de su generación práctica*, dudamos que haya otras bases más válidas. Y, en cualquier caso, reiteramos que, en nuestra opinión, y contra lo que quizá cabría pensar, esto es mucho menos imperfecto —metodológica y

epistemológicamente— que *prompts* únicos que no dan indicaciones en términos de escritura creativa.

Una vez establecido esto, los grandes principios que han guiado nuestra corrección, extraídos de la obra de Gardner (2001: 127-200), han sido los siguientes:

- Evitar los tópicos, tanto formales como narrativos.
- Evitar la tendencia a la sobreexplicación de los contenidos.
- Evitar las meras abstracciones explicativas, por entender que en la ficción narrativa es mucho más eficaz permitir que las cosas se manifiesten por sí solas antes que explicarlas en términos referenciales. Esto implica desde privilegiar el diálogo y la acción frente al relato descriptivo o abstracto de hechos o ideas hasta la relevancia del detalle como “sangre que riega la ficción y le da vida” (Gardner, 2001: 56), y que por tanto mantiene la credibilidad del universo ficcional.
- Que los personajes den la sensación de tener consistencia y vida propia —por ejemplo, mostrando sus contradicciones, aspiraciones y pensamientos—, sin que parezcan meros instrumentos unidimensionales al servicio de las tesis del autor.
- Evitar errores como cambios injustificados en el tono del relato, falta de variedad en la construcción de las frases o rimas internas.

De este modo, con nuestras correcciones no pretendíamos que Chat GPT creara una obra maestra, sino que evitara una serie de errores de escritor principiante en función de un criterio seguramente imperfecto, pero desde luego concreto, operativo y —a nuestro entender— razonable. Opinamos que esta subjetividad, convenientemente traducida a *prompts* y desplegada en rondas sucesivas —ver a continuación—, era preferible a las instrucciones aparentemente más objetivas —y generalmente limitadas a un solo *prompt*— de las que se valen otras investigaciones.

Como decíamos antes de este inciso, estimamos que uno de los principales motivos de la discreta calidad de los textos de ficción literaria creados por los LLM es la lógica autorregresiva que los articula. Con esto nos referimos no tanto —y no solo— a la posibilidad de que, en ausencia de ajustes, es muy probable¹¹ que los LLM se inspiren en muestras textuales donde lo que podríamos considerar *mala literatura* sea más numerosa que la *buena*. Estimamos que la dificultad es más profunda, pues si el problema se limitara a ofrecerle muestras de alta calidad literaria a los LLM, estos dispositivos, a través de técnicas de RAG, hace tiempo que dominarían por completo la escritura literaria. Creemos que el problema es más bien de método; del método de generación textual de los LLM con respecto a los elementos que —siquiera siguiendo planteamientos como los de Gardner— le otorgan verdadero valor al texto literario. Así, por ejemplo, Chat GPT es

¹¹ Nos limitamos a decir que es *muy probable* porque, dada la naturaleza de caja negra del funcionamiento de los LLM comerciales, los investigadores no sabemos qué tipo de muestras textuales manejan estos programas, ni tampoco los criterios exactos de su composición algorítmica de patrones de generación (Kocinski, 2024).

bastante bueno reformulando el texto para evitar clichés cuando estos le son identificados de forma previa. Algo similar ocurre con las sobreexplicaciones, o con la importancia de los detalles para la credibilidad narrativa, cuando se dan las instrucciones adecuadas. Pero ante cuestiones más complejas, asistemáticas, subjetivas y sobre todo sujetas a convenciones poco articuladas y variables, como por ejemplo crear personajes multidimensionales, o mantener de forma uniforme el tono del relato, o crear conexiones simbólicas entre, por ejemplo, ciertas aspiraciones de los personajes y ciertos objetos o situaciones... ahí Chat GPT muestra notables dificultades.

En nuestra opinión, eso podría deberse a que en tales casos hablamos solo de elementos *profundamente humanos*, sino sobre todo de elementos *profundamente representativos de un tipo de entendimiento puramente humano*. Mitchell & Krakauer (2023), haciéndose eco de un debate presente entre los investigadores en IA generativa, contraponen dos tipos de entendimiento. Por un lado estaría el de los LLM, que a pesar de su inmensísima complejidad y los millones de parámetros que maneja, es en último término estadístico; y por otro el de los seres humanos, basado en “conceptos”, y que se caracterizaría por “requiring few data, minimal or parsimonious models, clear causal dependencies, and strong mechanistic intuition” (Mitchell & Krakauer, 2023: s.p.). Afirman Mitchell & Krakauer que existen investigadores que sostienen que es cuestión de tiempo que el tipo de pensamiento estadístico de los LLM logre crear mejores patrones al aumentar su capacidad para integrar muchos más datos que en la actualidad, y que de este modo alcance o incluso supere al pensamiento conceptual humano —se hacen eco del mantra “scale is all you need” (Mitchell & Krakauer, 2023: 2)—. Sea o no cierto esto último, lo que la experiencia práctica nos demuestra es que, en su configuración actual, la lógica —estadística— de los LLM encuentra particulares dificultades a la hora de plasmar textualmente este tipo de elementos.

Puntualizado esto, nuestra metodología consistió en que a partir de un primer texto generado por un *prompt* muy breve y genérico, contestáramos a ChatGPT con un nuevo *prompt* que tratara de enmendar los errores. En este sentido, la metodología no era distinta a la que se sigue en muchos talleres de creación literaria, en los que un profesor se dedica a corregir los textos que los alumnos le entregan. Cada ronda de correcciones realizadas por el profesor genera un nuevo texto por parte del alumno, y de este modo gradual se termina por llegar a versiones definitivas de los textos; unos textos que, con independencia de que sigan teniendo carencias, son desde luego mejores que el primer resultado entregado.

Esta metodología implicaba, además de numerosos *prompts*, un tipo concreto de instrucciones. Con ello nos referimos a que cuando afirmamos que los textos generados por IA tenían una baja calidad no se debía a fallos que podrían haberse corregido con instrucciones más o menos objetivas, como las que corresponden a errores gramaticales. Las carencias de los textos generados por ChatGPT se debían a aspectos poco compatibles con una normatividad rígida —explicitud excesiva, mala calidad de los conflictos narrativos, etc.— que exigían, a su vez, instrucciones subjetivas, abstractas, o incluso

equivocas, del tipo: “El simbolismo de la brújula cuya flecha cambia constantemente de dirección es poco refinado. Hay que mejorar ese elemento simbólico o prescindir de él”.

Muy relacionado con esto cabe destacar una circunstancia que creemos resulta increíblemente relevante, y que también suele ser ignorada. Nos referimos al hecho de que la literatura, rodeada de todo tipo de incertidumbres epistemológicas, teleológicas e incluso existenciales, es especialmente refractaria a su ejecución a través de unas instrucciones textuales sencillas que además —y en esto queremos incidir— puedan darse de una sola vez. La literatura como categoría discursiva ofrece respuestas muy poco concretas —al menos *a priori*— a preguntas como *qué quiero escribir, cómo quiero escribirlo, para qué quiero escribirlo o para quién quiero hacerlo*. Las respuestas a estas preguntas, o al menos una precisión suficiente con respecto a ellas, a menudo solo se consigue a mitad del proceso de escritura, tras descartar numerosos caminos que se han revelado erróneos —lo que en el escritor humano se traduce en bocetos, reescrituras, descartes textuales, etc.—.

La clave, sin embargo, está en que este modo de proceder no se debería solo a que la creación literaria humana sea imperfecta. La clave residiría más bien en que la escritura literaria se parece más a un itinerario exploratorio que se va enriqueciendo a través de la necesaria superación de errores que a la ejecución precisa de un plan textual predeterminado. Todo ello hizo que para medir el potencial literario de los LLM nos basáramos en una metodología de *prompt* múltiple.

4. UN EXPERIMENTO LITERARIO DE *PROMPT* MÚLTIPLE

Una vez explicado nuestro proceder analítico, presentamos un experimento cuya prioridad fue ser consecuente con el antagonismo entre la naturaleza autorregresiva de los LLM y la asistematicidad literaria.

Nuestro primer paso fue pedirle a ChatGPT que redactara un relato breve, de 800 palabras, cuyo protagonista emprendiera “un viaje que le cambiara para siempre”. Optamos por dicha extensión por suponer un punto intermedio entre una duración demasiado corta, que le impidiera desarrollar mínimamente la trama y los personajes, y otra demasiado larga, que dificultara una evaluación lo bastante rápida por nuestra parte. Por otro lado, nos decantamos por la temática del viaje porque, dentro de su generalidad, daba una pauta narrativa relativamente clara.

Nuestra intención era responder a este texto con una serie de correcciones —necesariamente subjetivas, como hemos explicado— en aquellos aspectos que consideráramos mejorables; unas correcciones que transformaríamos en un nuevo *prompt*, que a su vez daría lugar a un nuevo texto creado por ChatGPT. Repetiríamos este proceso hasta alcanzar el límite de las posibilidades literarias del programa —entendiendo por esto último tanto que ciertos aspectos esenciales del texto estuvieran ejecutados de forma mínimamente solvente (según las pautas del manual de Gardner) como el hecho de que, tras detectar un error y pedirle a ChatGPT que lo corrigiera, este no consiguiera

hacerlo con éxito, tras varias reiteraciones—. Y todo ello desde la hipótesis de trabajo de que entre el primer texto generado y el último habría una mejora en favor de este último.

Por una cuestión de espacio no podemos reproducir aquí todos los textos y *prompts* de que constó nuestro experimento, pero pueden consultarse en una carpeta creada para este artículo en un servidor remoto¹². En cambio, reproduciremos el primer texto generado y el último, obtenido tras las sucesivas correcciones, y resumiremos las directrices más destacadas que introdujimos en forma de *prompt*.

El experimento arrancó con la siguiente instrucción: “Buenos días, ChatGPT. Escríbeme por favor un relato literario que tenga un título, que posea una extensión máxima de 800 palabras y cuya trama gire en torno a un viaje que cambie a su protagonista para siempre. Muchas gracias”. El texto ofrecido ante esta petición fue el siguiente:

El tren de los sueños perdidos

El silbato del tren resonó en el aire frío del amanecer, un llamado metálico que parecía desgarrar la niebla que cubría la pequeña estación. Clara estaba allí, con una maleta vieja y un cuaderno en el bolsillo de su abrigo. Había comprado el boleto sin un destino claro, eligiendo simplemente el tren que partía más pronto. No sabía hacia dónde iba, pero sabía que debía irse. Su vida, un engranaje oxidado de rutinas y silencios, la había convertido en una sombra de sí misma.

Cuando el tren se puso en marcha, Clara se instaló junto a una ventana. El paisaje se deslizaba como un cuadro inacabado: campos cubiertos de escarcha, árboles desnudos que parecían alzar sus ramas hacia el cielo gris, y pequeñas casas cuyos techos parecían susurrar historias de otros tiempos. Pronto, el traqueteo rítmico del tren se convirtió en una melodía hipnótica, y Clara abrió su cuaderno para escribir. Pero las palabras no llegaban. Las páginas seguían en blanco, como lo había estado su mente durante meses.

El tren hizo una breve parada en un pueblo casi desierto. Subió un anciano con una bolsa de cuero que llevaba a la espalda. Sus ojos eran oscuros, pero había una chispa juguetona en ellos, como si guardaran secretos. Se sentó frente a Clara y le dedicó una sonrisa que era más un desafío que una cortesía.

—¿Hacia dónde viajas? —preguntó el hombre.

Clara dudó antes de responder:

—No lo sé exactamente.

El anciano arqueó una ceja, intrigado:

—Interesante. Pocas personas se atreven a viajar sin rumbo. Pero dime, ¿qué buscas?

La pregunta era sencilla, pero Clara sintió que se le atascaba en la garganta. Buscaba algo, sí, pero no sabía qué. Después de un momento de silencio, murmuró:

—Tal vez... una forma de empezar de nuevo.

El anciano asintió como si comprendiera algo que Clara no podía articular. Luego sacó un objeto de su bolsa: una brújula antigua. Era un artefacto peculiar, con una aguja que parecía moverse caprichosamente, ignorando cualquier noción de norte o sur.

—Llévala contigo —dijo mientras se la tendía—. Esta brújula no te guiará a un lugar físico, sino al lugar que necesitas encontrar.

¹² Véase: <https://github.com/MiguelAmores9/Anexo-de-prompts-y-textos-del-articulo-Del-deep-learning-al-aprendizaje-de-taller-literario-.git> [17/10/2025].

Clara aceptó la brújula con escepticismo, pero algo en la mirada del anciano la convenció de que debía guardarla. Poco después, el hombre se levantó y se dirigió al siguiente vagón, dejándola sola con sus pensamientos y la extraña brújula

Durante las horas siguientes, Clara observó la aguja de la brújula moverse de forma errática. Finalmente, el tren se detuvo en una estación rodeada de montañas y un lago que reflejaba el cielo como un espejo roto. La aguja señalaba con insistencia hacia el lago. Impulsada por una curiosidad que no podía explicar, Clara bajó del tren y caminó hacia la orilla.

El aire era puro, y el silencio envolvía el lugar como un manto. Mientras se acercaba al agua, un destello llamó su atención. Era un pequeño bote amarrado al muelle. Sin pensarlo demasiado, Clara subió al bote y comenzó a remar. La brújula seguía apuntando hacia adelante, como si quisiera guiarla a través del lago.

Cuando llegó al centro, dejó los remos y permitió que el bote se meciera suavemente. Fue entonces cuando lo sintió: una calma profunda, como si todo el peso que llevaba sobre sus hombros se disolviera en el agua. Cerró los ojos y, por primera vez en mucho tiempo, dejó que sus pensamientos fluyeran libremente.

Recordó los sueños que había abandonado: ser escritora, viajar por el mundo, amar sin reservas. Recordó también los miedos que la habían paralizado y las decisiones que la habían llevado a ese punto de su vida. Allí, en medio del lago, entendió que el viaje que debía emprender no era hacia un lugar, sino hacia dentro de sí misma.

Cuando finalmente abrió los ojos, el sol comenzaba a ponerse, y el cielo estaba teñido de naranja y púrpura. Regresó al muelle con una sensación renovada de propósito. No sabía exactamente cómo, pero algo dentro de ella había cambiado. Subió al tren de regreso con la brújula todavía en su mano, pero ya no la miraba con escepticismo. Sabía que tenía el poder de decidir hacia dónde dirigirse.

De vuelta en su asiento, Clara abrió su cuaderno y comenzó a escribir. Las palabras fluían como un río desbordado, llenando las páginas con historias, sueños y promesas. Mientras el tren avanzaba por la noche, Clara entendió que no se trataba de escapar de su vida, sino de construirla de nuevo, una página a la vez.

Y así, con cada kilómetro que dejaba atrás, Clara renacía.

Nuestra opinión inicial de este cuento fue desfavorable. De forma similar a los malos relatos escritos por personas, había una sensación general de falta de pericia en el manejo de los elementos estilísticos y narrativos que daba lugar a un relato impostado, donde nada parecía funcionar. Para intentar corregirlo, le pedimos a ChatGPT que introdujera cambios como los siguientes:

- No ser tan explícito a la hora de expresar los pensamientos y sentimientos de los personajes, y especialmente los de la protagonista, Clara. Que se reflejaran de forma más sutil e implícita, a través de acciones, por ejemplo.
- Reformular el personaje del anciano que se sube al tren. Su influencia sobre Clara debía ser más sutil e indirecta.
- Perfeccionar el elemento simbólico de la brújula cuya flecha cambia constantemente de dirección, que resultaba demasiado explícito y burdo.
- Darle más importancia al hecho de que Clara sufría por el hecho de no haber podido cumplir su aspiración de ser escritora, que poseía una importancia vital en la evolución del personaje.

Tras introducir esas correcciones en forma de *prompt* y pedir a ChatGPT que elaborara un nuevo texto, podría decirse que seguimos sin obtener un buen relato, pero que logramos una sensible mejora. El motivo principal fue que se corrigió el que en nuestra opinión era el problema básico de la primera versión, la explicitud y sobreexplicación de los contenidos del texto, y en particular de los sentimientos de Clara. Más allá de que se hubieran eliminado ciertas menciones explícitas, el verdadero cambio, lo que suponía moverse ya dentro de unos códigos puramente literarios, era haber establecido desde el principio —aunque de modo indirecto— un vínculo entre el desencanto vital de la protagonista y el hecho de que hubiera traicionado su vocación de escritora. Eso otorgó cierta solidez y coherencia estructurales al texto y permitió que la evolución psicológica de la protagonista se expresase de forma indirecta, aunque lo bastante precisa.

También se mejoró el episodio del anciano que propicia el cambio en Clara. Ya no era una versión torpe de los ayudantes de los cuentos de hadas que se valía del simbolismo pueril de la brújula sin dirección, sino una mediación mucho más sutil, aunque igualmente orientada a dar respuesta a su desorientación vital. Sin embargo, en esta segunda versión del texto persistían otros errores graves, como los siguientes:

- Faltaba una caracterización de Clara más profunda, desde una mínima descripción física a alguna pista sobre las circunstancias personales que le habían llevado a esa situación de zozobra espiritual.
- Era necesaria una mayor claridad en lo referente al mapa que le brinda el desconocido a Clara, que se explicara con mayor precisión por qué le provocan ese cambio.
- Había una serie de problemas de verosimilitud, en particular el hecho de que la protagonista usara su propia mano como “página” en la que escribir.
- Era conveniente un título menos melodramático.

Tras introducir estas correcciones en forma de *prompt*, el resultado obtenido fue un texto mejorado y razonablemente correcto. La evolución psicológica de Clara mostraba ahora un equilibrio más depurado entre explicitud y claridad, y los problemas de verosimilitud más llamativos se habían suavizado. Por ello, en la siguiente ronda de correcciones, las cuestiones de contenido fueron cediendo espacio a otras de tipo formal:

- Describir con mayor detalle y claridad por qué el enigmático mapa dejado por el desconocido tenía tanto impacto en Clara.
- Explicar mejor por qué Clara decide bajarse en esa parada solitaria y emprender el camino hacia el lago.
- Eliminar los clichés que aún quedaban: “mechones rebeldes”, “ideas cobrando vida”, etc.

La cuarta versión del relato no distaba mucho de esta anterior, y, dentro de la limitación general, la mayoría de los errores graves volvían a ser relativos a la forma. Por ello, a continuación se reproduce la quinta y última versión del relato, con la que dimos por terminado el experimento:

Deriva

El tren avanzaba con un ritmo pausado, como si pesara más de lo habitual. Clara, mediada la treintena, miraba por la ventana, pero no parecía ver el paisaje. Su rostro, de facciones suaves aunque tensas, mostraba un cansancio que no era solo físico. Vestía un abrigo oscuro que, aunque cuidado, mostraba signos de desgaste, y una bufanda que se había aflojado ligeramente durante el trayecto. En el bolsillo interior del abrigo llevaba un cuaderno cerrado, cuyas páginas permanecían intactas desde hacía demasiado tiempo.

Había dejado su trabajo en una editorial hacía semanas, incapaz de seguir revisando las palabras de otros mientras las suyas se acumulaban en su mente como escombros. Los días se habían convertido en una sucesión de silencios, llamadas que evitaba y correos que no respondía. Había tomado el tren casi por impulso, huyendo de una sensación que no podía nombrar pero que la comprimía por dentro. Sabía que algo debía cambiar, aunque no estaba segura de cómo o cuándo.

En una de las paradas, un hombre subió al vagón con movimientos tranquilos. Llevaba consigo un libro cuya portada gastada capturó de inmediato la atención de Clara: “Los mapas de la deriva”. Sin decir nada, el hombre se sentó frente a ella y lo colocó sobre la mesa. Durante varios minutos no lo abrió, como si su presencia bastara para llenar el espacio. Finalmente, sin mirarla, pasó una página con un gesto casi ceremonial. El mapa que quedó a la vista no se parecía a nada que Clara hubiera visto antes.

No era un mapa tradicional. Las líneas serpenteaban y se bifurcaban sin lógica aparente, como si cada ruta contuviera sus propios caminos secretos. Había nombres que apenas se sostenían sobre el papel, borrándose o mutando bajo la luz tenue del vagón. Los puntos de referencia eran vagos, a veces meros símbolos que evocaban más emociones que lugares. Clara no pudo evitar inclinarse para verlo con mayor detalle, sintiendo un cosquilleo extraño, como si esas rutas imposibles resonaran con algo que había olvidado. El hombre, sin pronunciar una palabra, dejó una pluma sobre la mesa antes de bajar en la siguiente parada. Clara la miró un momento antes de tomarla, preguntándose si debía hacerlo. El metal era frío al tacto y más liviano de lo que esperaba, pero había algo en su diseño simple que la inquietaba.

Cuando el tren llegó a una estación rodeada de montañas, Clara decidió bajar. No fue un impulso inconsciente, sino una certeza silenciosa, como si la quietud del lugar le ofreciera algo que necesitaba. Al pisar el andén, sintió el aire fresco contra su piel y, sin pensarlo demasiado, comenzó a caminar por un sendero que se adentraba en el paisaje. Había algo en el sonido de sus pasos sobre la tierra que la hacía sentirse menos extraviada. No sabía adónde la llevaba, pero tampoco tenía prisa por descubrirlo.

El sendero desembocó en un lago que reflejaba los últimos tonos cálidos del atardecer mezclados con las sombras de las montañas. Clara se detuvo junto a la orilla, observando cómo las ondulaciones rompían el reflejo del cielo. Sacó el cuaderno del bolsillo y la pluma que había tomado, ambas ahora más presentes, como si esperaran ser utilizadas. En silencio, se sentó en una roca plana y abrió el cuaderno. Al principio, las palabras se resistieron, pero pronto empezó a escribir. Las líneas que aparecían en el papel no eran

respuestas claras, sino preguntas transformadas en imágenes, fragmentos de historias que había dejado atrás. Lo que escribía no era perfecto, pero tampoco lo necesitaba. Por primera vez en mucho tiempo, el acto mismo era suficiente.

Cuando la noche se hizo más fría, regresó a la estación. Encontró un banco de madera bajo un poste de luz tenue y se sentó allí, envolviéndose en su abrigo. El silencio de la noche parecía darle espacio para pensar sin apremios. No durmió, pero escribió durante horas, hasta que el cuaderno comenzó a llenarse de palabras que no sabía que aún podía encontrar.

El amanecer llegó con un resplandor suave, y el tren apareció en el horizonte, avanzando con su característico sonido rítmico. Clara subió al vagón con el cuaderno aún entre las manos. Mientras se acomodaba junto a la ventana, pasó la punta de los dedos sobre las palabras recién escritas, sintiendo que cada página, aunque imperfecta, era un paso. Miró por la ventana, y aunque sus ojos aún guardaban un eco de duda, había en ellos una claridad nueva.

El tren reanudó su marcha, y con cada kilómetro que dejaba atrás, Clara comprendía que estaba empezando a avanzar.

5. CONCLUSIONES

En su artículo arriba citado, McCoy *et al.* destacan que, por más que los LLM se usen para producir todo tipo de textos, en ningún caso se puede olvidar su “autoregressive substrate, from which all of their abilities emerge” (2024: s.p.). Así, sostienen que, aunque un LLM pueda usarse para, por ejemplo, resolver un problema matemático, es importante no confundirlo con un dispositivo diseñado para tal fin, pues de lo que hablamos en realidad es de un sistema de predicción de la próxima palabra utilizado para resolver un problema matemático.

Algo similar podría decirse de los usos que se le están dando para generar textos literarios. Instrumentos como ChatGPT no son *generadores de textos literarios*, sino *sistemas de predicción estadística de la próxima palabra que, en ocasiones, pueden utilizarse para elaborar textos literarios*. La capacidad casi mágica de estos programas para elaborar todo tipo de textos en segundos hace que, en muchos casos, los análisis pasen a una fase posterior, la de evaluación de los resultados obtenidos, y olviden hasta qué punto un LLM está realmente dotado desde un punto de vista estructural para acometer con éxito la tarea de la creación de textos literarios de ficción.

En la primera parte de este artículo hemos repasado algunos de los antagonismos esenciales entre la lógica generativa de los LLM y la naturaleza fundamental de la literatura —adicionales los derivados de la naturaleza no humana de la IA generativa—. El principal de ellos es que, mientras que los LLM se basan en una lógica probabilística centrada en la elaboración de patrones textuales, el texto literario es por naturaleza singular y asistemático —o incluso *prototípico*, en el sentido explicado— lo que lo convierte en esencialmente refractario a cualquier cálculo estadístico, por complejo que este pueda resultar. Y aunque es cierto que existen disciplinas capaces de encontrar ciertos patrones estructurales dentro de lo literario, como la narratología estructural o la estilística computacional, hemos expuesto una serie de motivos por los que dichos patrones no

podrían arrojar una luz procedimental decisiva para la elaboración práctica de un texto —realmente— original, que es lo que en nuestra opinión realmente marcaría un cambio verdaderamente decisivo. A este antagonismo fundamental habría que añadir otros impedimentos, como por ejemplo el carácter profundamente variable de la calidad —o incluso la finalidad— del texto literario, lo que dificulta la asimilación algorítmica de sus códigos; o la incompatibilidad de la literatura con una planificación previa, detallada y sistemática de su ejecución textual, lo que hace muy difícil la obtención de un buen texto a partir de un único *prompt*.

Teniendo todo ello en cuenta, en la segunda parte del artículo quisimos hacer un experimento para comprobar las capacidades de un LLM concreto, Chat GPT 4o, para la generación de un texto literario determinado, un relato breve. Para ello, y en coherencia epistemológica y metodológica con lo apuntado, nos decantamos por un análisis hipercualitativo basado en la aplicación de rondas sucesivas de *prompts* destinadas a corregir las carencias de un único texto generado por ChatGPT a partir de unas instrucciones poco detalladas. Esta metodología poseía una serie de carencias inherentes en lo relativo a su alcance y representatividad (ver nota 7), pero solo de esta forma era operativo no caer en los errores de concepto descritos, que en nuestra opinión perjudican en mayor medida la luz que puedan aportar este tipo de investigaciones sobre el potencial literario de los LLM.

Nuestro experimento consistió en pedirle a Chat GPT que, a partir de un *prompt* poco detallado, elaborara un cuento ficcional de 800 palabras en el que la trama estuviera centrada en cómo un viaje cambia al protagonista para siempre. Hicimos hasta cinco versiones de ese texto, con cuatro rondas de correcciones, y nuestra referencia a la hora de aplicar las correcciones fueron las pautas fundamentales que da Gardner en su manual *El arte de la ficción*.

En los primeros textos se centraron en carencias muy básicas, como la explicitud de todos los enunciados, y especialmente de los pensamientos y sentimientos de Clara, la protagonista, que parecían más propias de un cuento infantil. También se pidió a ChatGPT que eliminara el personaje del anciano, por considerarlo un puro cliché; que refinara el simbolismo de la brújula cuya aguja giraba sin parar, o que corrigiera ciertos elementos inverosímiles. Por otro lado, se le pidió que estableciera una identificación más clara entre la desorientación vital de Clara y su vocación frustrada de ser escritora. Se trataba de indicaciones concretas, pero su ejecución estaba lejos de ser sistemática u objetiva. Y, sin embargo, en su segunda versión, el texto había mejorado sustancialmente en estos aspectos, por lo que entonces las correcciones se limitaron a aspectos relacionados con mejoras formales —evitar clichés expresivos, fundamentalmente— y a pulir aspectos de verosimilitud. Esa fue también la dinámica básica de la tercera y cuarta rondas de correcciones, hasta que en la quinta versión del texto dejamos de corregir debido a que, o bien los errores detectados se habían corregido de forma mínimamente satisfactoria, o bien Chat GPT parecía incapaz de hacerlo tras sucesivos intentos —fue el caso de la explicación del profundo impacto que el mapa tiene en Clara—.

El resultado fundamental del experimento fue que entre la primera versión del texto y la quinta y última —entre las que mediaron cuatro rondas de correcciones con 29 anotaciones concretas— hubo una sustancial mejora. Esto no implica que el relato, en nuestra opinión, tuviera una gran calidad literaria, pero tampoco hablamos en términos meramente subjetivos; hablamos de *mejora* sobre todo en el sentido de que ChatGPT, a través de las sucesivas rondas de correcciones, supo ir ajustando su texto a las pautas subjetivas, asistemáticas y variables que Gardner atribuye a las narraciones literarias de calidad.

Como se ha afirmado, nuestras decisiones metodológicas pueden limitar el alcance de las conclusiones de este trabajo. Y, sin embargo, creemos que esta investigación apunta en la dirección de apostar por análisis (hiper)cualitativos a la hora de evaluar el verdadero potencial literario de los LLM. Al igual que cualquier otro producto textual que se quiera obtener de estos dispositivos, creemos que es un error juzgar la capacidad literaria de los LLM en función de un único *prompt* y no de una larga cadena de ellos. Y, por otro lado, la especial naturaleza de la escritura literaria —asistemática, prototípica y no instrumental— desaconseja ciertas metodologías de encadenamiento de instrucciones como el *prompt chaining* —o también el RAG, si nuestro interés es evaluar la capacidad de los LLM de crear textos mínimamente originales, que es donde en nuestra opinión se encuentra el verdadero núcleo de interés de esta cuestión—. Como hemos afirmado, en nuestra opinión la clave de todas las dificultades de la escritura literaria no está solo en que esta sea un producto *profundamente humano*, sino que supone *un producto profundamente representativo de un tipo de entendimiento puramente humano*. Y esto resulta antagónico con respecto a la lógica autorregresiva —es decir, en último término *estadística*— de los LLM, cuya inmensa potencia generativa se ve inevitablemente menguada.

Nuestro experimento apunta a que, sin corrección ulterior, y al menos en lo que respecta a la escritura de un relato breve y a Chat GPT 4o —aunque su tecnología básica es la misma que la del resto de LLM generalistas—, este dispositivo cae en errores de escritor principiante como son las sobreexplicaciones, los errores de verosimilitud y los tópicos temáticos y formales. Y, además, no es hasta la segunda versión del texto cuando se engrana la clave simbólica fundamental del relato —que la zozobra vital de Clara se identifica con su renuncia a seguir su vocación vital de ser escritora—. De este modo, nuestra propuesta de someter a la poderosísima capacidad de generación de ChatGPT, basada en el *machine learning*, a un método de aprendizaje *texto-corrección-nuevo texto*, propio de los talleres de escritura tradicionales, no se explica por una supuesta intención —sesgada de antropocentrismo— de aplicarle a las máquinas el —imperfecto— modo de elaboración literaria propio de los seres humanos. Se trata más bien de la metodología que, en vista del antagonismo entre la lógica generativa de los LLM y la naturaleza de la literatura, mejor se adapta a la tarea de evaluar el potencial de estas herramientas para con los textos literarios de ficción.

REFERENCIAS BIBLIOGRÁFICAS

- “A Comprehensive Comparison of All LLM” (2024). *AI-PRO*, 15 de octubre. Disponible en línea: <https://ai-pro.org/learn-ai/articles/a-comprehensive-comparison-of-all-llms> [02/09/2025].
- AMORES, M. (2018). *Ficción. Anatomía de un potencial infinito*. Madrid: Visor Libros.
- BELLEMARE-PEPIN, A.; LESPINASSE, F.; THÖLKE, P.; HAREL, Y.; MATHEWSON, K.; OLSON, J. A.; BENGIO, Y. & JERBI, K. (2024). “Divergent Creativity in Humans and Large Language Models”. *ArXiv*, 13 May (revised 1 July 2025), s. p. Disponible en línea: <https://doi.org/10.48550/arXiv.2405.13012> [02/09/2025].
- BIWU, S. (2015). “Unnatural Narratology: Core Issues and Critical Debates”. *Journal of Literary Semantics* 44.2, 169-194.
- BOSQUE, I. (2001). “Sobre el concepto de ‘lugar común’ desde un punto de vista gramatical”. *Pandora. Revue d’Études Hispaniques* 1, 31-46.
- CHAKRABARTY, T.; PADMAKUMAR, V.; HE, H. & PENG, N. (2023). “Creative Natural Language Generation”. En *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, Q. Zhang & H. Sajjad (eds.), 34-40. Stroudsburg, PA: Association for Computational Linguistics Disponible en línea: <https://doi.org/10.18653/v1/2023.emnlp-tutorial.6> [12/10/2025].
- CHAKRABARTY, T.; LABAN, P.; AGARWAL, D.; MURESAN, S. & WU, C-S. (2024). “Art or Artifice? Large Language Models and the False Promise of Creativity”. En *CHI ’24: Proceedings of the CHI Conference on Human Factors in Computing Systems*, F. F. Mueller & P. Kyburz (eds.), s.p. New York: Association for Computing Machinery. <https://dl.acm.org/doi/10.1145/3613904.3642731> [16/10/2025].
- GARDNER, J. (2001). *El arte de la ficción*. M. Martínez-Lage (ed. y trad.). Madrid: Fuentetaja.
- GARRIDO-MERCHÁN, E. C.; ARROYO-BARRIGÜETE, J. L. & GOZALO-BRIHUELA, R. (2023). “Simulating H.P. Lovecraft Horror Literature with the ChatGPT Large Language Model”. *ArXiv*, 5 May, s. p. <https://doi.org/10.48550/arXiv.2305.03429> [02/09/2025].
- GÓMEZ-RODRÍGUEZ, C. & WILLIAMS, P. (2023). “A Confederacy of Models. A Comprehensive Evaluation of LLMs on Creative Writing”. *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino & K. Bali (eds.), 14504-14528. Stroudsburg, PA: Association for Computational Linguistics. Disponible en línea: <https://doi.org/10.18653/v1/2023.findings-emnlp.966> [17/10/2025].
- GUO, R.; XU, W. & RITTER, A. (2024). “Meta-Tuning LLMs to Leverage Lexical Knowledge for Generalizable Language Style Understanding”. En *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*

- (*Volume 1: Long Papers*), L.-W. Ku, A. Martins & V. Srikumar (eds), 13708-13731. Bangkok: Association for Computational Linguistics. Disponible en línea: <https://doi.org/10.18653/v1/2024.acl-long.740> [02/09/2025].
- HICKE, R. M. M. & MIMNO, D. (2023). “T5 Meets Tybalt: Author Attribution in Early Modern English Drama Using Large Language Models”. En *Computational Humanities Research 2023 (CHR 2023). Proceedings of the Computational Humanities Research Conference 2023, Paris, France, December 6-8, 2023*, A. ŠeĽa, F. Jannidis & I. Romanowska (eds.), 274-302. Paris: CEUR Workshop Proceedings. Disponible en línea: <https://ceur-ws.org/Vol-3558/paper2757.pdf> [02/09/2025].
- JAKOBSON, R. (1960). “Linguistics and Poetics”. En *Style in Language*, T. A. Sebeok (ed.), 350-377. Cambridge, MA: MIT Press.
- KOCINSKI, M. (2024). “What is Black Box Artificial Intelligence (AI)?”. *IBM Think*, s. f. <https://www.ibm.com/think/topics/black-box-ai> [02/09/2025].
- LAMB, C.; BROWN, D. G. & CLARKE, C. L. A. (2018). “Evaluating Computational Creativity: An Interdisciplinary Tutorial”. *ACM Computing Surveys* 51.2, s. p.
- LEDESMA, V. (2024). “¿Sueña el ChatGPT con ovejas eléctricas? El problema de la creatividad literaria en los textos generados por IA”. En *Travesía por los límites de la literatura*, S. Adelaida, J. L. Ortega y J. A. Nisa (coords.), 203-220. Madrid: Dykinson.
- MARCO, G.; GONZALO, J.; DEL CASTILLO, R. y MATEO-GIRONA, M. T. (2024). “Pron vs Prompt: Can Large Language Models Already Challenge a World-Class Fiction Author at Creative Text Writing?”. *ArXiv*, 1 July (revised 28 October), s. p. Disponible en línea: <https://doi.org/10.48550/arXiv.2407.01119> [02/09/2025].
- MARCO, G.; GONZALO, J. & RELLO, L. (2025). “Small Language Models can Outperform Humans in Short Creative Writing: A Study Comparing SLMs with Humans and LLMs”. En *Proceedings of the 31st International Conference on Computational Linguistics, Abu Dhabi, January 19-24*, O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. Di Eugenio & S. Schockaert (eds.), 6552-6570. Stroudsburg, PA: Association for Computational Linguistics. Disponible en línea: <https://aclanthology.org/2025.coling-main.437/> [17/10/2025].
- MCCOY, T.; YAO, S.; FRIEDMAN, D. & GRIFFITHS, T. L. (2024). “Embers of Autoregression Show How Large Language Models Are Shaped by the Problem They Are Trained to Solve”. *PNAS* 121.41, s. p. Disponible en línea: <https://doi.org/10.1073/pnas.2322420121> [17/10/2025].
- MERCHANT, B. (2018). “When an AI Goes Full Jack Kerouac”. *The Atlantic*, 1 Oct, s. p. <https://www.theatlantic.com/technology/archive/2018/10/automated-on-the-road/571345/> [02/09/2025].
- MIKROS, G. (2025). “Beyond the Surface: Stylometric Analysis of GPT-4o’s Capacity for Literary Style Imitation”. *Digital Scholarship in the Humanities* 40.2, 587-600. Disponible en línea: <https://doi.org/10.1093/llc/fqaf035> [02/09/2025].

- MITCHELL, M. & KRAKAUER, D. (2023). "The Debate over Understanding in AI's Large Language Models". *PNAS* 120.13, s. p. Disponible en línea: <https://doi.org/10.1073/pnas.2215907120> [17/10/2025].
- NEELAKANTESWARA, A.; CHAUDHARI, S. & ZAMANI, H. (2024). "RAGs to Style: Personalizing LLMs with Style Embeddings". En *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*, A. Deshpande et al. (eds.), 119-123. Stroudsburg, PA: Association for Computational Linguistics. Disponible en línea: <https://aclanthology.org/2024.personalize-1.11/> [02/09/2025].
- PROPP, V. (2014 [1928]). *Morfología del cuento*, F. Díez del Corral (trad.). Madrid: Akal.
- RADFORD, A.; NARASIMHAN, K.; SALIMANS, T. & SUTSKEVER, I. (2018). "Improving Language Understanding by Generative Pre-Training". *OpenAI*, s. f., s. p. Disponible en línea: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf [17/10/2025].
- RIEMER, K. & PETER, S. (2024). "Conceptualizing Generative AI as Style Engines: Application Archetypes and Implications". *International Journal of Information Management* 79, s. p. <https://doi.org/10.1016/j.ijinfomgt.2024.102824> [02/09/2025].
- SHKLOVSKI, V. (1991). "El arte como artificio". En *Teoría de la literatura de los formalistas rusos*, T. Todorov (ed.), A. M. Nethol (trad.), 55-70. México D. F.: Siglo XXI.
- VAN DIJK, T. (1980). *Macrostructures. An Interdisciplinary Study of Global Structures in Discourse, Interaction and Cognition*. Hillsdale, NJ: Erlbaum.
- VV.AA. (1982). *Análisis estructural del relato*, B. Dorriots (trad.). Buenos Aires: Tiempo Contemporáneo.
- YUPENG, C.; XU, W.; JINDONG, W.; YUAN, W.; LINYI Y.; KAIJIE Z.; HAO C.; XIAOYUAN Y.; CUNXIANG W.; YIDONG W.; WEI Y.; YUE Z.; YI C.; PHILIP S. Y.; QUIANG Y. & XING, X. (2024). "A Survey on Evaluation of Large Language Models". *ACM Transactions on Intelligent Systems and Technology* 15.3, s. p. Disponible en línea: <https://doi.org/10.1145/3641289> [17/10/2025].
- ZHU, S.; WANG, Z.; ZHUANG, Y.; JIANG, Y.; GUO, M.; ZHANG X. & GAO Z. (2024). "Exploring the Impact of ChatGPT on Art Creation and Collaboration: Benefits, Challenges and Ethical Implications", *Telematics and Informatics Reports* 14, s. p. Disponible en línea: <https://doi.org/10.1016/j.teler.2024.100138> [17/10/2025].



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND).

Fecha de recepción: 14/01/2025

Fecha de aceptación: 22/10/2025